ISOLATED WORD RECOGNITION USING LPC & VECTOR QUANTIZATION

M. K. Linga Murthy[1], G.L.N. Murthy[2]

[1,2]*Asst. Professor, ECE, Lakireddy Balireddy College of Engineering, Andhra Pradesh, India,*
lingamurthy413@gmail.com, murthyfromtenali@gmail.com

Abstract

Speech recognition is always looked upon as a fascinating field in human computer interaction. It is one of the fundamental steps towards understanding human recognition and their behavior. This paper explicates the theory and implementation of Speech recognition. This is a speaker-dependent real time isolated word recognizer. The major logic used was to first obtain the feature vectors using LPC which was followed by vector quantization. The quantized vectors were then recognized by measuring the Minimum average distortion.

All Speech Recognition systems contain Two Main Phases, namely Training Phase and Testing Phase. In the Training Phase, the Features of the words are extracted and during the recognition phase feature matching Takes place. The feature or the template thus extracted is stored in the data base, during the recognition phase the extracted features are compared with the template in the database. The features of the words are extracted by using LPC analysis. Vector Quantization is used for generating the code books. Finally the recognition decision is made based on the matching score. MATLAB will be used to implement this concept to achieve further understanding.

Index Terms: *Speech Recognition, LPC, Vector Quantization, and Code Book.*

----------***----------

1. INTRODUCTION

Speech is a natural mode of communication for people. We learn all the relevant skills during early childhood, without instruction, and we continue to rely on speech communication throughout our lives. It comes so naturally to us that we don't realize how complex a phenomenon speech.

Speech recognition, or more commonly known as automatic speech recognition (ASR), is the process of interpreting human speech in a computer. A more technical definition is given by Jurafsky, where he defines ASR as the building of system for mapping acoustic signals to a string of words. He continues by defining automatic speech understanding (ASU) as extending the goal to producing some sort of understanding of the sentence.

1.1 Challenges

The general problem of automatic transcription of speech by any speaker in any environment is still far from solved. But recent years have seen ASR technology mature to the point where it is viable in certain limited domains.

1.2 Difficulties

One dimension of variation in speech recognition tasks is the vocabulary size.

A second dimension of variation is how fluent, natural or conversational the speech is isolated word recognition, in which each word is surrounded by some sort of pause, is much easier than recognizing continuous speech

A third dimension of variation is channel and noise. Commercial dictation systems, and much laboratory research in speech recognition, is done with high quality, head mounted microphones

A final dimension of variation is accent or speaker-class characteristics. The objective of this paper is to recognize the isolated words spoken by the speaker. These results are very useful for implementing the recognition systems. The words or utterances are recorded by Microphone and are stored in work space, then processed using MATLAB signal processing toolbox. It involves pre-emphasis, frame blocking autocorrelation analysis, LPC Analysis and Vector quantization.

2. LPC FOR SPEECH RECOGNITION:

There are three basic steps in Speech Recognition, they are

- Parameter estimation. (in which the test pattern is created)
- Parameter Comparison.
- Decision making.

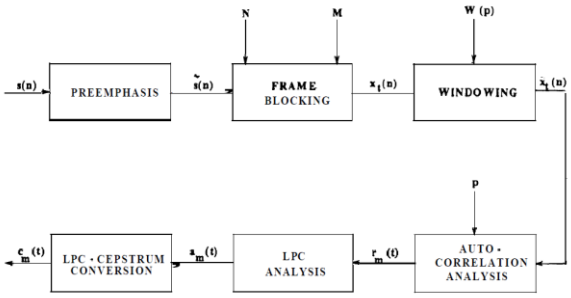

Fig2.1 Block diagram for LPC Processor for Speech Recognition.

2.1 Pre – emphasis:

From the speech production model it is known that the speech undergoes a spectral tilt of -6dB/oct. To counteract this fact a pre-emphasis filter is used. The main goal of the pre-emphasis filter is to boost the higher frequencies in order to flatten the spectrum. Pre emphasis follows a 6 dB per octave rate. This means that as the frequency doubles, the amplitude increases 6 dB. This is usually done between 300 - 3000 cycles. Pre emphasis is needed in FM to maintain good signal to noise ratio. Perhaps the most widely used pre emphasis network is the fixed first-order system:

$$H(z) = 1 - az^{-1}, \qquad 0.9 \leq a \leq .0 \qquad (2.1)$$

2.2 Frame – Blocking:

In this step the pre-emphasized speech signal is blocked into frames of N samples, with adjacent frames being separated by M samples .Thus frame blocking is done to reduce the mean squared predication error over a short segment of the speech wave form. In this step the pre emphasized speech signal, $S(n)$ is blocked into frames of *N* samples, with adjacent frames being separated by *M* samples.

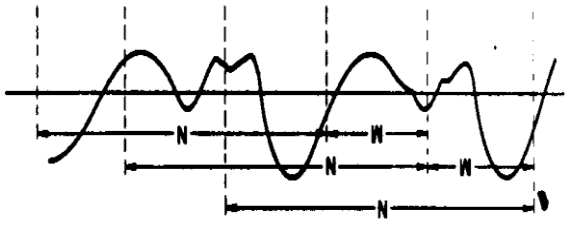

Fig2.2 blocking of speech into overlapping frames.

Typical values for N and M are 256 and 128 when the sampling rate of the speech is 6.67 kHz. These correspond to 45-msec frames, separated by 15-msec, or a 66.7-Hz frame rate.

2.3 Windowing:

Here we want to extract spectral features of entire utterance or conversation, but the spectrum changes very quickly. Technically, we say that speech is a non-stationary signal, meaning that its statistical properties are not constant across time. Instead, we want or extract spectral features from a small window of speech that characterizes a particular sub-phone and for which we can make the assumption that the signal is stationary. this is done by using a window which is non-zero inside some region and zero elsewhere, running this window across the speech signal, and extracting the waveform inside this window. A more common window used in feature extraction is the Hamming window, which shrinks the values of the signal toward zero at the window boundaries, avoiding discontinuities.

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \qquad 0 \leq n \leq N-1 \quad (2.2)$$
$$0 \qquad , \qquad Else\ where$$

2.4 Autocorrelation analysis:

Each frame of windowing signal is next auto correlated to give

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}_l(n)\tilde{x}_l(n+m), m = 0,1,2, \ldots\ldots\ldots\ldots, P, \quad (2.3)$$

Where the highest autocorrelation value, p, is the order of the LPC analysis. Typically, values of *p* from 8 to 16 have been used, with *p* = 10 being the value used for this systems. A side benefit of the autocorrelation analysis is that the zeroth autocorrelation, $R_l(0)$, is the energy of the l^{th} frame.

2.5 LPC Analysis:

The next processing step is the LPC analysis, which converts each frame of P+1 autocorrelations into an LPC parameter set in which the set might be the LPC coefficients, the reflection coefficients(PARCOR-co-efficient), the log area ratio coefficients, or any desired information of the above sets. The formal method for converting from autocorrelation coefficients to an LPC parameter set (for the Autocorrelation method) is known as Durbin's method

2.6 LPC parameter conversion to Cepstral coefficients:

A very important LPC parameter set, which can be derived directly from the LPC coefficients set, is the LPC Cepstral coefficients, $c(m)$. The recursion method is used. The Cepstral coefficients, which are the coefficients of the Fourier transform representation of the log magnitude spectrum, have been shown to be a more robust, reliable feature set for speech

recognition than the LPC coefficients, the PARCOR coefficients, or the log area ratio coefficients.

3. VECTOR QUANTIZATION:

Vector quantization is one very efficient source-coding technique. Vector quantization is a procedure that encodes a vector of input (e.g., a segment of waveform or a parameter vector that represents the segment spectrum) into an integer (index) that is associated with an entry of a collection (codebook) of reproduction vectors.

Using the basic pattern-recognition approach, the input (unknown class or word) Speech pattern is next compared with each class (or word) reference pattern and a measure (score) of similarity between the unknown pattern and each reference pattern is calculated.

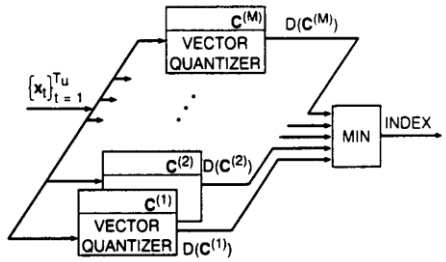

Fig 4.1 A Vector Quantized based speech recognition system

It is the pattern recognition based approach to speech recognition, we see that the *M* codebooks are analogous to *M* (sets of) reference patterns (or templates) and the dissimilarity measure is defined according to Eq. (4.1) and Eq. (4.2), where no explicit time alignment is required.

$$D\left(C^{(i)}\right) = \frac{1}{T_u} \sum_{t=1}^{T_u} d\left(X_t, \tilde{X}_t^i\right) \qquad (4.1)$$

The utterance is recognized as class K if

$$D\left(C^{(k)}\right) = \min_i D\left(C^{(i)}\right) \qquad (4.2)$$

Where $D\left(C^{(k)}\right)$ is a minimum average distortion.

And $D\left(C^{(i)}\right)$ is an M average distortion score of all code books.

4. SYSTEM IMPLEMENTATION:

The training set for the vector quantizer was obtained by recording the utterances of set isolated words .The words are recorded for a two different speakers. The recognition vocabulary consisted of the names (Forward, Back, Left, Right, and Stop). Here each word is applied for 10 times .The results obtained are shown in the table below.

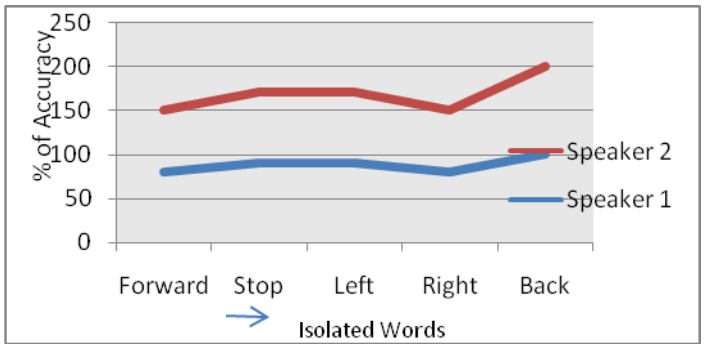

Fig 4.1 Comparisons for isolated word recognition system for two speakers

CONCLUSIONS

Isolated Word Recognition using Linear Predictive Coding and Vector Quantization provides basic idea for implementing for Speech Recognition for Isolated Words. The Vector Quantized based speech recognition is very simple method. In this method by using LPC analysis extracts the features of given words & vector quantization is used for feature matching in Speech Recognition.

In the successful implementation the results were found to be satisfactory considering less number of training data. The accuracy of the real time system can be increased significantly by using an improved speech detection/noise elimination algorithm.

REFERENCES

[1]. L.R.Rabiner and B.H.Juang, "Fundamentals of speech recognition", Prentice Hall (Signal Processing series) 1993.

[2]. Richard O.Duda, Peter E.Hart, David G.Stork,"Pattern Classification", John Wiley & Sons (ASIA) Pte Ltd.

[3]. Y.Linde ,A.Buzo and R.M.Gray, "An algorithm for vector quantizer design" ,IEEE Trans .COM-28,January 1980 .

[4]. Mayukh Bhaowal and kunal chawla , "Isolated word recognition for English language using LPC,VQ & HMM" **,** students of IIT, Allahabad, India.

[5]. Poonam Bansal , Amita dev and Shail Bala Jain "Automatic speaker Identification using VQ ", Medwell journals,6(9) : 938 – 942 ,2007.

[6]. Lawrence R. Rabiner "Applications of speech recognition in the area of Telecommunications"**,** AT&T Labs Florham Park, New Jersey 07932, 0-7803-3698-4/97/$10.00 *0* 1997 IEEE

[7]. L. R. Rabiner, "Applications of Voice Processing to Telecommunications", Proc. IEEE, Vol. 82, No. 4, pp. 199-228, Feb. 1994.

[8]. L. R. Rabiner and R.W. Schafer "Digital processing of Speech signals", Prentice Hall (Signal Processing series).

[9]. J.E.Shore and D.K.Burton "Discrete Utterance Speech recognition without Time alignment", IEEE Transactions on IT, IT – 29(4), July 1983, pp. 473 – 491.

[10]. D. K. Burton, J. T. Buck, and F. Shore, "Parameter Selection for Isolated Word Recognition Using Vector Quantization," Proc. ICASSP 84, San Diego, CA, pp. 9.4.1- 9.4.4, March 1984.

[11]. Douglas o'shaughnesy "Speech Communication", Universities press Electrical engineering Series, 2/e ,2001

BIOGRAPHIES:

M. K. Linga Murthy is currently working as Asst. Professor in ECE in LBRCE, Mylavaram. He completed his B.Tech in the year 2001 at SJCET, Yemmiganur. He completed his M.Tech in the year 2008 at MITS, Madanapalle. He has over 06 years of teaching experience & his research an area is signal processing.

G.L.N Murthy is currently working as Asso. Professor in ECE in LBRCE, Mylavaram. He completed his B.Tech in JNTU, Anantapu. He completed his M.Tech in JNTU, Anantapur. He pursuing his PhD in SVU Tirupathi, he has over 12 years of teaching experience & his research an area is signal processing