Scientific Journal of Impact Factor (SJIF): 5.71
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 02, February -2018

Modified MFCC Algorithm for Speech Recognition

Boda Aruna[1], D Srinivas[2] Prakash Nakirekanti[3] Ravi Boda[4]

[1] *ECE department & MIETW,Hyderabad, India*
[2] *ECE department & VJIT,Hyderabad, India*
[3] *CSE department & SIST,Hyderabad, India*
[4]*ECE department & UCE Osmania university,Hyderabad, India*

Abstract — *Automatic Speech Recognition has been an active research topic for more than four decades. With the advent of digital computing and signal processing, the problem of speech recognition was clearly posed and thoroughly studied .These developments were complemented with an increased awareness of the advantages of conversational systems. The goal of Automatic Speech Recognition is to develop techniques and systems that enable computers to accept speech input. The speech recognition problem may be interpreted as speech to text conversion problem. Since the early 80's,compact implementations of accurate, real-time speech recognizers have found wide spread applications, which includes voice activated transcription, simplified man machine communication, aids for hearing impaired individuals and the physically disabled telephone assistance and other man-machine interface tasks. The aim of this work is to develop a speech recognition (SR) system based on Vector Quantization (VQ) approach. This system receives speech inputs from users, analyzes the speech inputs, searches and matches the input speech with the pre-recorded and stored speeches in the trained database or codebook. First the feature extraction from the speech signal is done by a parameterization of the wave formed signal into relevant feature vectors by Mel Frequency Cepstral Coefficients (MFCC) algorithm and Modified MFCC .This parametric form is then used by the recognition system both in training the models and testing.*

Keywords- *Speech Recognition, Vector Quantization, Mel Frequency Cepstral Coefficients (MFCC), Feature Extraction, voice activated transcription.*

I. INTRODUCTION

In this technological era, information technology continues making more impact on many aspects of our daily lives, however, the problems of communication between human beings and information processing machines become increasingly important. So far, such communication has been done almost entirely by means of keyboards and screens, but there are substantial disadvantages of this method for many applications. Speech is considered as the most widely used and natural means of communication between humans, and it is an obvious substitute for such means of keyboards and screens in the communication process. However, this deceptively simple means of exchanging information is, in fact, extremely complicated. Although the application of speech in the man-machine interface [1] is growing rapidly, in the present forms machine capabilities for generating and interpreting speech are still incomplete and imperfect.

Here it is concerned with speech recognition technology, which is part of speech and signal processing, as well as human computer interaction. Speech recognition is highly demanded and has many useful applications Speech recognition was also defined as the technology by which sounds, words or phrases spoken by humans are converted into electrical signals, and these signals are transformed into coding patterns that can be identified by a computer, Based on this identification, the computer usually takes some actions. Speech recognition also refers to the ability of a machine or computer program to receive and interpret spoken commands and act upon those commands. What clinches the case in favor of speech recognition is the experimental evidence, that voice communication is critical to the single and multi-modality communication links.

Various speech recognition techniques have been proposed and applied to English language, for example the Dynamic Time Warping (DTW) [3], the Hidden Markov Model[4], Vector Quantization[5], Continuous Density Hidden Markov Modeling and the Artificial Neural Networks

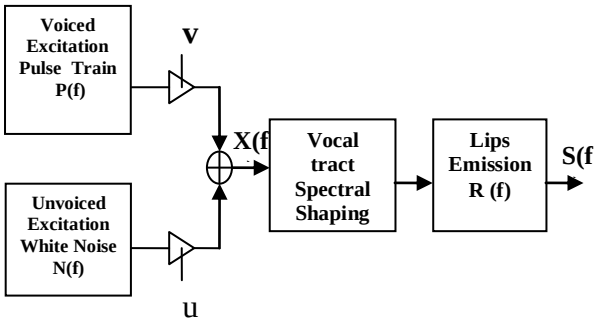


Figure:1.1 A simple model of speech production

$$S(f) = \big(V.P(f) + u.N(f)\big).H(f).R(f)$$
$$= X(f).H(f).R(f) \qquad (1)$$

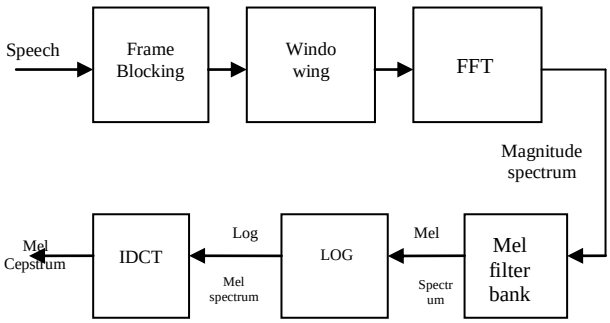
To influence the speech sound, we have the following parameters in our speech production model

1. the mixture between voiced and unvoiced excitation (determined by v and u)
2. the fundamental frequency (determined by P(f))
3. the spectral shaping (determined by H(f))
4. the signal amplitude (depending on v and u)

These are the technical parameters describing a speech signal. To perform speech recognition, the parameters given above have to be computed from the time signal (this is called speech signal analysis or "acoustic preprocessing") and then forwarded to the speech recognizer. For the speech recognizer, the most valuable information is contained in the way the spectral shape of the speech signal changes in time. To reflect these dynamic changes, the spectral shape is determined in short intervals of time, e.g., every 20ms. By directly computing the spectrum of the speech signal, the fundamental frequency would be implicitly contained in the measured spectrum (resulting in unwanted "ripples" in the spectrum).

II. PROPOSED METHOD

The main steps include the following preprocessing, frame blocking, windowing using hamming window, performing Discrete Fourier Transform (DFT) [6], applying the Mel-scale filter bank in order to find the spectrum as it might be perceived by the human auditory system, performing the Logarithm[7], and finally taking the inverse DCT of the logarithm of the magnitude spectrum



Steps For Feature Extraction Using Modified MFCC:

Modified MFCCS extraction also involves a frame-based analysis of a speech signal where the speech signal is broken down into a sequence of frames.[8,11] Each frame undergoes a sinusoidal transform (Fast Fourier Transform) in order to obtain certain parameters which then undergo Mel-scale perceptual weighting and dc-correlation. The result is a sequence of feature vectors describing useful logarithmically compressed amplitude and simplified frequency information. Modified MFCC features are obtained by first performing a standard Fourier analysis, and then converting the power-spectrum to a mel-frequency spectrum. Therefore, Modified MFCC will be obtained by taking the logarithm of that spectrum and by computing its inverse Discrete Cosine Transform the main steps required for the Modified MFCC computations

The basic concept of vector quantization as applied to speech recognition is schematically depicted in Figure 4.1 [2, 12]. A training speech sequence is first used to generate the codebook. As we described, the speech signal is segmented (windowed) into successive short frames and each frame of speech is represented by a vector of finite dimensionality.

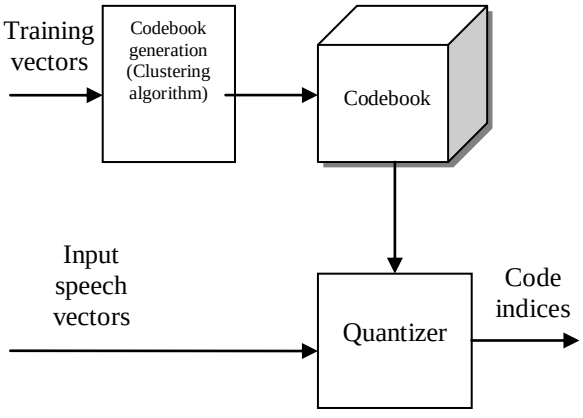


Figure 4.1.Block diagram for a Vector Quantizer

In our case, the vector is the result of either the filter bank analysis which captures the time, various spectral characteristics of the speech signal. If we compare the information rate of the vector representation to that of the raw (uncoded) speech waveform, we see that the spectral analysis has significantly reduced the required information rate. Consider, for example, 10-kHz sampled speech with 16-bit speech amplitudes. A raw signal information rate of 160 kbps is required to store the speech samples in PCM format [9]. For the spectral analysis, consider vectors of dimension p = 10 using a 10 ms frame rate. If we again represent each spectral component to 16-bit precision, the required storage is about 16 kbps. It is a 10-to-1 reduction over the uncompressed signal. Such compressions in storage rate are very impressive. Based on the concept of ultimately needing only a single spectral representation[10] for each basic speech unit, it may be possible to further reduce the raw spectral representation of speech to those drawn from a small, finite number of unique spectral vectors, each corresponding to one of the basic speech units (that is the phonemes).

This ideal representation was, of course, impractical, because there is so much variability in the spectral properties of each of the basic speech units. However, the concept of building a codebook of distinct analysis vectors, although with significantly more code words than the basic set of phonemes, remains an attractive idea and is the basis behind a set of techniques commonly called vector quantization methods. Assume that we require a codebook with about 1024 unique spectral vectors (that is about 25 variants for each of the 40 basic speech phonetic units in American English). Then to present an arbitrary spectral vector all we need is a 10 bit number which is the index of the codebook vector that best matches the input vector. A total bit rate of about 1 kbps is required to present the spectral vectors of a speech signal. This rate is about 1/16 the rate required by the continuous spectral vectors. Hence the VQ representation is potentially an extremely efficient presentation of the spectral information in the speech signal. This is one of the main reasons for the interest in VQ methods.

III. RESULT AND DISCUSSION

Test	zero	one	two	three	four	five	six	seven	eight	nine
sample1	0.761	0.926	0.484	0.516	0.517	0.524	0.724	0.625	0.714	0.725
sample2	0.659	0.813	0.719	0.625	0.615	0.618	0.751	0.652	0.763	0.713
sample3	0.634	0.826	0.547	0.531	0.527	0.812	0.725	0.651	0.682	0.711
sample4	0.682	0.719	0.687	0.453	0.651	0.571	0.684	0.649	0.678	0.672
sample5	0.715	0.657	0.437	0.718	0.542	0.496	0.71	0.592	0.626	0.633
sample6	0.791	0.734	0.515	0.609	0.624	0.613	0.623	0.712	0.721	0.588
sample7	0.615	0.672	0.672	0.718	0.712	0.582	0.636	0.704	0.707	0.59
sample8	0.636	0.75	0.75	0.656	0.615	0.582	0.628	0.68	0.712	0.621
sample9	0.628	0.703	0.703	0.532	0.519	0.622	0.761	0.612	0.687	0.629
sample10	0.673	0.516	0.531	0.631	0.596	0.525	0.715	0.663	0.644	0.675
Average	0.679	0.732	0.605	0.599	0.592	0.595	0.696	0.654	0.693	0.656

Table 3.1: MFCC and Vector Quantization based recognition

Gives the CPU execution time for each utterance using Modified MFCC and VQ .For each digit average time is calculated in the above table 5.3 the sample implies the utterances taken by the different persons. Thus the Average time for the Modified MFCC and Vector Quantization algorithm is 0.54 seconds. The program developed took 0.54 seconds on a average for a speech signal of an average duration of 1 second. The main advantage with the use of a vector quantization in speech recognition is that it can be observed in the time required for the execution of such a program.

zero	one	two	three	four	five	six	seven	eight	nine
3.9	4.069	3.027	3.888	4.801	5.056	5.476	5.145	4.579	5.388
3.532	6.872	6.348	5.782	5.189	5.458	5.147	5.082	5.328	4.435
3.836	4.297	5.404	4.887	5.992	5.681	4.517	5.420	4.547	5.343
3.089	5.147	5.083	4.767	5.716	5.903	5.402	4.973	3.5502	6.447
3.189	5.227	3.459	3.960	5.976	5.493	5.487	5.094	4.5804	5.655
4.266	5.273	4.354	5.920	4.055	5.829	3.663	5.149	5.6756	6.478
3.015	5.712	4.815	5.665	5.641	5.440	4.787	5.269	5.2138	5.927
2.216	4.478	5.343	4.369	4.972	5.872	5.732	5.243	5.0581	6.747
3.120	5.599	5.619	3.471	5.659	5.452	5.313	4.292	6.0048	5.781
3.391	5.284	5.682	5.640	5.796	5.502	5.173	4.209	5.9427	5.923
3.855	5.785	5.155	4.678	5.672	5.453	5.234	4.923	5.3421	6.102

Table 3.2: Modified MFCC and Vector Quantization based recognition

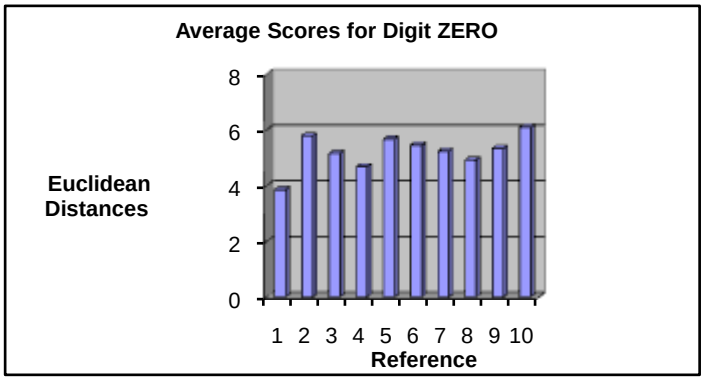


Fig 3.1: Average scores for Digit Zero

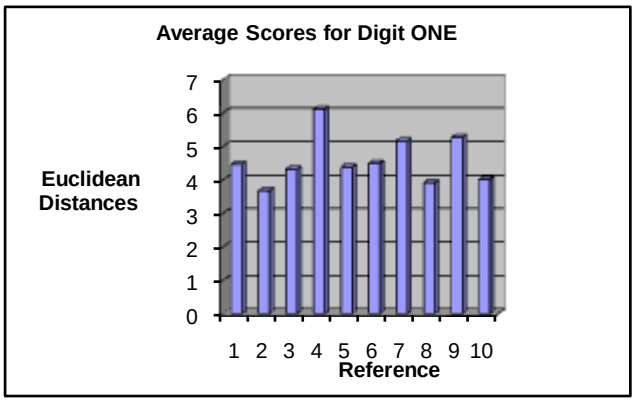


Fig 3.2: Average scores for Digit ONE

From the figures 3.1 and 3.2 can observe that the average value of the individual scores are decreased for digits Zero to Nine using the combination modified MFCC and VQ than MFCC and VQ that is the Euclidean distances are decreased. This implies that there is perfect match between the trained word and the tested word using Modified MFCC and VQ. From the tables 5.2 and 5.4 we can observe that the CPU execution time for the utterance of the digits from Zero to Nine by different persons for the combination of Modified MFCC and Vector Quantization algorithm is less when compared to the MFCC and VQ, which is advantageous. Thus Modified MFCC and VQ is used in Speech Recognition for reducing the execution time

IV. CONCLUSION

The theory for extracting Mel Frequency Cepstrum Coefficient and Modified Mel Frequency Cepstrum Coefficient from a speech signal, Vector Quantization Models for testing utterances against the trained models has been successfully implemented. MFCC and Modified MFCC are compared by means of CPU execution time. The computational time using MFCC and VQ is 0.65 sec and using Modified MFCC and VQ is 0.54sec.thus modified MFCC and VQ is used for speech recognition improves the performance of the system by reducing the execution time. The strengths of speech recognition includes that , in this technological era, users want to achieve their targets in a very easy and fast manner. Therefore, using speech interaction would achieve the user's needs, whereby this speech recognition system will ease and fasten the process of communication.

REFERENCES

[1] J.N.Holmes Wokingham, Van Nostrand Reinhold, "Speech Synthesis and Recognition" 2nd Edition, Prentice Hall 2001.
[2] Rabiner, Lawrence, Juang, Biing Hwang, "Fundamentals of Speech Recognition", Prentice Hall,NJ,USA
[3] BenGold, Nelson Morgan, "Speech and Audio Signal Processing", John Wiley and Sons Pte.Ltd.Singapore 2002.
[4] Shaik.R.A. and Yousaf-Zai,F.Q.(2004), " A Novel Approach to Noisy Speech Recognition Using DTW Algorithm with Mel-Frequency Cepstral Coefficients", Journal of Engineering and Technology(JET_IUT),pp.21to 24.
[5] Soong, F.Rosenberg, A.Rabiner, L.Juang, B., "A Vector Quantization Approach to Speaker Recognition", Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP April 1985, Vol.10, page no 387-390.
[6] Wai,C.C, "Speech Coding Algorithms Foundation and Evolution of Standardized Coders",John Wiley and Sons Inc.,NJ,USA.
[7] Hermansky.H., "Perceptual Linear Predictive (PLP) Analysis of Speech", the Journal of the American Society of America, Vol.87, no.4, pp.1738-1752
[8] Poonam Bansal ,Amita Dev,Shail Bala Jain , "Enhanced Feature Vector Set for VQ Recognizer in Isolated Word Recognition", International Conference information Research and Applications-I.Tech 2007.
[9] SunYoungPark and Hyung Soon Kim, "Modified MFCC Features for Speech Recognition", Proceedings of ICSP-2001, Vol.2, pp 659-662, August 2001.
[10] Linde.Y,Buzo.A,Gray.R.M, "An Algorithm For Vector Quantizer Design",IEEE Transactions on Communications,Vol.28,No.1,pp.84-95.
[11] Rabiner,L.R(1989),"ATutorial On Hidden Markov Models and Selected Applications In Speech Recognition", Proceedings of the IEEE,Vol.77,No.2.pp.257-286.
[12] Xing He Scordilis,M.Gongiun Li, "A Novel VQ-Based Speech Recognition approach for mobile terminals" , Southeastcon,2005,IEEE Proceedings, 10 April 2005,pp.433-436.

