International Journal of Recent Trends in Engineering & Research
ISSN(ONLINE) : 2455-1457
IMPACT FACTOR : 3.344

Voice Based Shapes Recognition using Mel Frequency Cepstrum Coefficents

Banumathi.P[1], RajanaveenS.V[2] ,Saberi R.S[3] , Thilak Vasu Deva K[4]

[1]Associate professor,Computer Science And Engineering, Kathir College of Engineering,
[2,3,4] Computer Science And Engineering ,Kathir College of Engineering

Abstract: Signal processing is computing an extracting the feature set is an important stage in any speech recognition system. The feasible feature set is still not yet decided though the vast efforts of researchers. There are many types of features, which are derived distinct and have good impact on the recognition rate. This project presents one of the techniques to take out the feature set from a speech signal, which can be used in speech recognition systems. The key is to adjust the speech wave form to parametric representation. To achieve this, we have first made a comparative study of the Mel Frequency Cepstrum Coefficients approach. The voice based biometric system is based on isolated or single word recognition. A particular microphone pronounce the words once in the training session so as to train and store the accentuate of the way in word. Later in the testing session the user pronounce the word again in order to achieve recognition if there is a same content. The quality vectors unique to that speaker are obtained in the training phase and this is made use of later on enter the parameters to the same microphone who once again pronounce the same word in the testing phase. At this stage with parameter value can also test the system.

Keywords: Voice based techniques to draw efficiently using Hamming window and Mel frequency Cepstrum.

I. INTRODUCTION

Speech is the most natural way to interface for humans. While this has been true since the dawn of civilization , the invention and global spread use of the telephone ,audio phonic storage media , radio , and television has given even further importance to speech transmit and speech processing[2]. The move nearer in digital signal processing technology has led the use of speech processing in many distinct application areas like speech compression ,enhancement ,synthesis ,and recognition[4] .In this premise ,the issue of speech recognition is studied and a speech recognition system is developed for Isolated word using Vector quantization model.

From a technological perspective it is possible to distinguish between two broad types are Direct Voice Input (DVI) and Large Vocabulary Continuous Speech Recognition (LVCSR) . DVI devices are primarily aimed at voice command and audio control, whereas LVCSR systems are used for form filling or voice- based document creation. In both cases the unrevealed technology is more or less the same. DVI systems are typically configured for small to mid-sized vocabularies (up to several thousand words) and might employ word or phrase spotting techniques . Also, DVI systems are usually required to acknowledge immediately to a voice command. LVCSR systems involve lexiconof perhaps hundreds of thousands of words, and are generally configured to transcribe continuous speech. Also,LVCSR need not be completed in real-time-for example , at least one vendor has offered a telephone-based dictation service in which the transcribed document is mail information back to the user. From an application view point, the benefits of using ASR derive from providing an extra interface channel in hands-

busy,eyes-busy, human-machine interaction (HMI),or simply from the fact that talking can be faster than typing.

II. RELATED WORK

In Speaker Adaptive TrainingIt reducing the speaker-specific variation in the speech signal allowing the compact model to represent more accurately the phonetically relevant context dependent variation.The first involves training a multi-style,or noise *condition independent (CI) system on speech data collected* in a wide range of diverse noise environments.This exploits the implicit modelling ability of the underlying statistical models to achieve a good generalization to concealed noise conditions.The second category is based on uncertainty decoding(UD).The uncertainty that varies with the noise delivered by, for example, a conditional distribution of the distorted speech, is propagated into the recognizer.

III. EXISTING SYSTEM

These approaches ignore the fact that different weight vectors makes mistakes.The main downside of these approaches is that they fail to decidedly model relationships between different voice gives a weight age based on voice low or high.More complex in case of large number of views/expressions. In Existing approach is only coefficents based The application that is being implemented is voice input as equations plotter.Initially a set of numbers in the form of words will be given as input by voice. It makes through Hidden Markov model which gives more complex and trained details are stored in database and while collecting input recognises very difficult and ittakes time.

IV. PROPOSED SYSTEM

Obtaining the acoustic characteristics of the speech signal is considered as Feature Extraction. Feature Extraction is used in both training and testing phases.
It contain following steps:

1. Frame Blocking
2. Windowing
3. Fast Fourier Transform.
4. Mel-Frequency Wrapping
5. Cepstrum

Feature Extraction

The main goal of Feature Extraction is to simplify recognition by reporting the vast amount of speech data without disaster the acoustic properties that defines the speech [12]. The simplified diagram of the step is depicted in Figure 1.1.

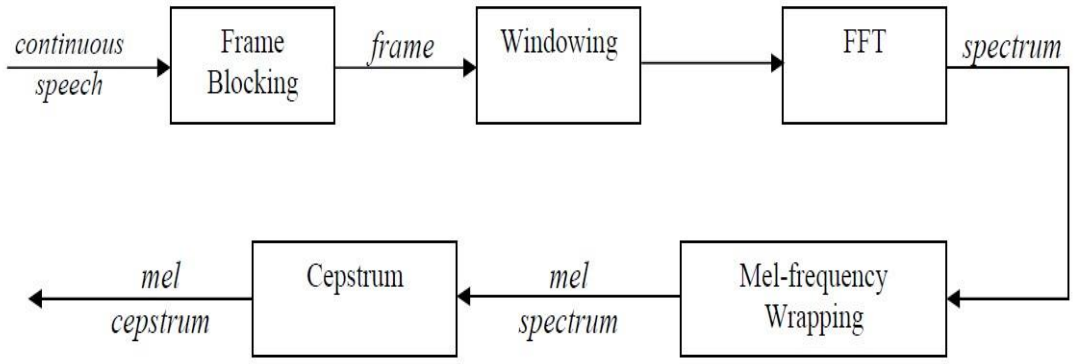


Fig. 1.1. Feature Extraction Steps

Frame Blocking

The speech signal characteristics stays stationary in a deficiently short period of time interval (It is called quasi-stationary). For this senses, speech signals are processed in short intervals. It is divided into frames with size between 30 to 100 milliseconds. Each frame overrun its previous frame by a predefined size. The aim of the replicate scheme is to smooth the transition from one frame to another frame.

Windowing

The second step is to window all frames. This is done in order to remove discontinuity at the edges of the frames. If the windowing function is defined as w(n), [0 < n < N-1] where N is the number of samples in each successive frame, the resulting signal will be; y(n) = x(n)*w(n). Generally hamming windows are used.

Fast Fourier Transform

Then each frame in FFT. This renewal is a fast way of Discrete Fourier Transform and it changes the domain from time to frequency.

Mel Frequency Warping

The human ear realizes the frequencies non-linearly. Researches show that the rising is linear up to 1 kHz and logarithmic above that. The Mel-Scale (Melody Scale) filter bank which concide with the human ear preciseness of frequency.The signals for each frame is passed through Mel-Scale band pass filter to impersonator the human ear. As mentioned above, psychophysical studies have shown that human perception of the frequency details of sounds for speech signals does not follow a linear rise. Thus for each tone with an absolute frequency, *f*, measured in Hz, a approximate pitch is measured on a scale called the mel scale. The *mel-frequency* scale is a direct frequency spacing below 1000 Hz and a algebraic spacing above 1000 Hz. As a point, the pitch of a 1 kHz tone, 40 dB above the visceral hearing threshold, is defined as 1000 mels. Therefore we can use the following proper formula to summate the mels for a given frequency *f* in Hz:

$$mel(f) = 2595 * \log_{10}(1 + f / 700)$$

One approach to simulating the optional spectrum is to use a filter bank, one filter for each animus mel-frequency component. That filter bank has a triangular bandpass frequency acknowledge, and the spacing as well as the bandwidth is determined by a constant mel frequency interval. The modified spectrum of S(□) thus consists of the product power of these filters whenS(□) is the input. The number of melcepstral coefficients, K, is generally chosen as 20.

Note that this filter bank is applied in the frequency domain; therefore it simply aggregate to taking those triangle shape windows in the Fig 1.2 on the spectrum. A useful way of thinking in mel-warped filter bank is to view each filter as a scattered diagram bin (where bins have overlap) in the frequency domain. A useful and able way of implementing this is to consider these triangular filters in the Mel rise where they would in effect be equally spaced filters.

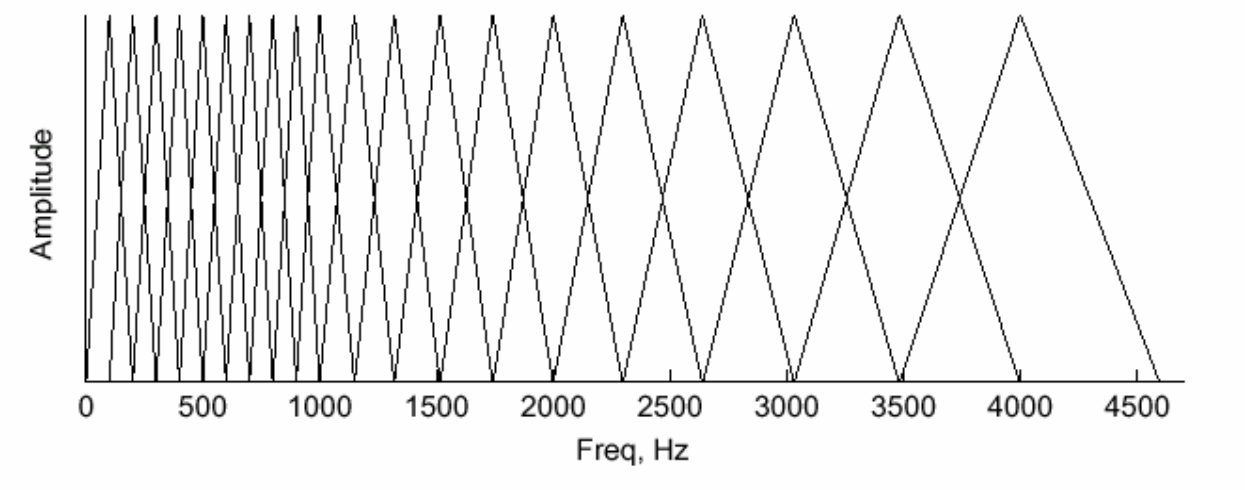


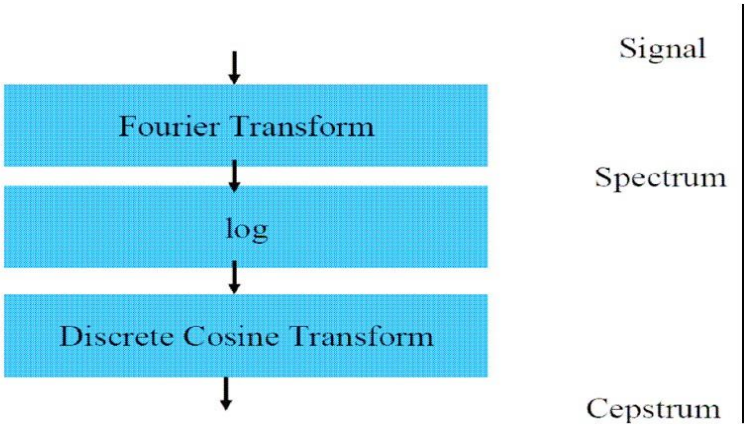
Fig. 1.2. Filter Bank in Mel frequency scale

Cepstrum

Cepstrum name was derived from the spectrum by recalling the first four letters of spectrum. We can say cepstrum is the Fourier Transformer of the log with uncrated phase of the Fourier Transformer.

- Mathematically we can say Cepstrum of signal = FT(log(FT(thesignal))+j2IIm)
  Where m is the interger required to properly uncrated the angle or imaginarypart of the complex log function.
- Algorithmically we can say – Signal - FT - log - phase uncrating - FT -Cepstrum.

For defining the real values real cepstrum uses the logarithm functiont. While for defining the recondite values whereas the complex cepstrum uses the complex algebraic function. The real cepstrum uses the information of the magnitude of the spectrum. where as complex cepstrum holds proposition about both amount and phase of the initial spectrum, which allows the modification of the signal. We can compute the cepstrum by distinct ways. A few of them need a phase-warping algorithm, rest do not. Figure below shows the pipeline from signal to



V. CONCLUSION

Speech synthesis has been ripened regularly over the last decades and it has been incorporated into several new applications. For most applications, the cognizability and palability of synthetic speech have reached the passable level. However, in presido, text preprocessing, and pronunciation fields there is still

much work and renovation to be done to achieve more legitimate sounding speech. Legitimate speech has so many dynamic changes that perfect naturalness may be impossible to reach. However, since the markets of speech synthesis related applications are raising regularly, the interest for giving more attempt and funds into this research area is also raising. Present speech synthesis systems are so hard that one researcher can not handle the entire system. With good mortality it is possible to divide the system into several indivisible modules whose developing process can be done separately if the communication between the modules is made precisely. However testing is done and speech is recognized finally extracted in future to implement and develop in real time.

REFERENCES

1. Rongfeng Su ,Xunying Liu, Member IEE and Lan Wang, Member IEE.
2. T. Anastasakos, J. McDonough, R. Schwartz, and J. Makhoul, "Acompact model for speaker-adaptive training," in *Proc. ICSLP'96*, Philadelphia, PA, USA, 1996, pp.

