An Enhanced Dc Motor Control Using the Speech Recognition System

Radhakrishna Karne, Chakradhar.A

Abstract: In this paper we present a new enhanced methodology for increasing the accuracy among the speech recorded in order to control the DC motor. The novel method takes in to the consideration the MFCC and VQ algorithm in order to calculate the coeefiecients . The speech moves DC motor with the different speeds and controls it's direction also with the only on command which can be issued inorder to stop the motor.

Keywords- DC, MFCC, VQ

I. INTRODUCTION

The speech is a natural communication mechanism between two persons but it is very complex from automatic analysis viewpoint. The human speech is, technically speaking, air pressure variations caused by movements of muscles and other tissue within the speaker. The speech organ system is a complicated mechanism but in short one can say that the lungs pump air through the windpipe to the surrounding environment via mouth and nose and speech sounds are formed during this process. The brain drives the muscle system that controls the lungs and vocal cords, the shape and volume of the windpipe, size and shape of the oral cavity, nasal passage controlling airflow through the nose, and finally, the lips. The signal can be though of arising from the speaker articulating the message that he wants to express. The audio signal is sampled and quantized. This digital speech signal is the input of speaker profile management and automatic speaker recognition systems.

II.PRINCIPLES OF SPEECH RECOGNITION

Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance. Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker. All speaker recognition systems have to serve two distinguishes phases. The first one is referred to the enrollment sessions or training phase while the second one is referred to as the operation sessions or testing phase. In the training phase, each registered speaker has to provide samples of their speech so that the system can build or train a reference model for that speaker. In case of speaker verification systems, in addition, a speaker-specific threshold is also computed from the training samples. During the testing (Operational) phase, the input speech is matched with stored reference models and recognition decision is made. Speaker recognition methods can also be divided into textindependent and text-dependent methods. In a textindependent system, speaker models capture the characteristics of somebody's speech which show up irrespective of what one is saying. In a textdependent system, on the other hand, the recognition of the speaker's identity is based on his or her speaking one or more specific phrases, like passwords, card numbers, PIN codes, etc. At the highest level, all speaker recognition systems contain two main modules: Feature Extraction and Feature Matching.

III. FEATURE EXTRACTION

The main objectives of feature extraction are to extract characteristics from the speech signal that are unique to each individual which will be used to differentiate speakers. The purpose of this module is to convert the speech waveform to some type of parametric representation for further analysis and processing. This is often referred as the signalprocessing front end. The speech signal is a slowly timed varying signal (It is called quasi-stationary). When examined over a sufficiently short period of time (Between 5 and 100 ms), its characteristics are fairly stationary. However, over long periods of time
(On the order of 1/5 seconds or more) the signal characteristic change to reflect the different speech sounds being spoken. Therefore, short-time spectral analysis is the most common way to characterize the
speech signal. A wide range of possibilities exist for parametrically representing the speech signal for the speaker recognition task, such as Linear Prediction Coding (LPC), Mel-Frequency Cepstrum Coefficients (MFCC), and others. MFCC is perhaps the best known and most popular, and these will be
used here.

3.1 Preprocessing

The first step of speech signal processing involves the conversion of analog speech signal into digital speech signal. This is a crucial step in order to enable further processing. Here the continuous time signal (Speech) is sampled at a discrete time points to form a sample data signal representing the
continuous time signal. The method of obtaining a discrete time representation of a continuous time signal through periodic sampling, where a sequence of samples, x[n] is obtained from a continuous signal s (t). It is apparent that more signal data will be obtained if the samples are taken closer together[4].

3.2 Frame Blocking

Framing is the process of segmenting the speech samples obtained from the analog to digital conversion into small frames with time length in the range of 20ms to 40 ms. In this step the continuous speech signal is blocked into frames of N

Manuscript received on September, 2013.

Radhakrishna Karne, M.Tech (Embedded Systems), Jayamukhi Institute Of Technological Sciences, Narsampet,Waranagl.

Chakradhar.A, Assistant Professor, Jayamukhi Institute Of Technological Sciences, Narsampet,Waranagl.

samples, with adjacent frames being separated by M (M < N). The human speech production is known to exhibit quasi-stationary behavior over a short period of time (20ms to 40 ms)[1][2].

3.3 Windowing

It is necessary to work with short term or frames of the signal. This is to select a portion of the signal that can reasonably be assumed stationary. Windowing is performed on each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame.

3.4 Mel-frequency Wrapping

Psychophysical studies have shown that human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency f, measured in Hz, a subjective pitch is measured on a scale called the 'mel' scale. The mel-frequency scale has linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 mels[1][6]. Therefore the following approximate formula is used to compute the mels for a given frequency f in Hz:

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

3.5 Mel Filterbank

One approach to simulating the subjective spectrum is to use a filter bank. That filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant mel-frequency interval. The modified spectrum of S (w) thus consists of the output power of these filters when S (w) is the input. The number of mel spectrum coefficients K, is typically chosen as 20.

3.6 Cepstrum

In this final step, the conversion of the log mel spectrum back to time. The result is called the mel frequency cepstrum coefficients (MFCC). Discrete Cosine Transform (DCT) is used to convert them back to the time domain.

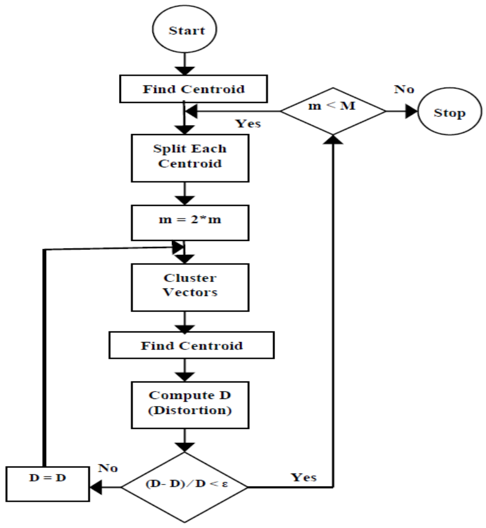

Figure 1 Flowchart of VQ-LBG Algorithm

IV. CONCLUSION

The unique characteristics of human speech, Mel Frequency Cepstrum Coefficients (MFCC) are used for feature extraction and Vector Quantization is used for feature matching technique. Clustering algorithm such as VQ-LBG algorithm is taken to implement the vector quantization for this purpose.

It is very important to investigate feature parameters that are stable over time, insensitive to the variation of speaking manner, including the speaking rate and level and robust against variations in voice quality

due to causes such as voice disguise or colds. It is also important to develop a method to cope with the problem of distortion due to telephone sets and channels, and background and channel noises. From the human-interface point of view, it is important to consider how the users should be prompted, and

how recognition errors should be handled. Furthermore various other analyses could be carried out on Voiceprints, Age and Emotion Identification, Combination of Features, Spoken Language, Natural Language etc. from the voice pattern of human speech.

REFERENCES

[1] An Efficient MFCC Extraction Method in Speech Recognition. Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong-Pang PUN, Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, IEEE 2006.

[2] Differential MFCC and Vector Quantization used for Real-Time Speaker Recognition System, 2008 IEEE Congress on Image and Signal Processing, Wang Chen□Miao Zhenjiang, Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

[3] "Speaker Identification Using MEL Frequency Cepstral Coefficient", Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman. Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology. 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh.

[4] Lawrence Rabiner and Biing-Hwang Juang, Fundamental of Speech Recognition", Prentice-Hall, Englewood Cliffs, N.J., 1993.

[5] Zhong-Xuan, Yuan & Bo-Ling, Xu & Chong-Zhi, Yu. (1999). "Binary Quantization of Feature Vectors for Robust Text-Independent Speaker Identification" in IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 1, January 1999. IEEE, New York, NY, U.S.A.

[6] F. Soong, E. Rosenberg, B. Juang, and L. Rabiner, "A Vector Quantization Appr