VOICE INTEGRATED SPEED AND DIRECTION CONTROL OF DC MOTOR

[1]Chandini K,[2]Priyanka M K,[3]Shalini S,[4]Sushma P
[1]Student,[2]Student,[3]Student,[4]Student
[1]EEE,
[1]GSSSIETW, Mysuru, India.

Abstract: It is very difficult to work in hazardous environment in many of the industries. Human can survive only certain amount of humidity, temperature, pressure, etc. Working in environments like this will cause threat to human life, so precautions should be taken against this. To overcome this huge loss voice control was developed. Due to the advancement of wireless technology, there are several connections are introduced such as GSM, Wi-Fi and Bluetooth. Each of the connection has their own unique specifications and applications. The speed control was implemented using Bluetooth technology to provide communication access from smart phone. Communication plays a major role in day today's life and can be used as a better tool in control system. It deals with wireless communication and voice recognition and is used to control the motor speed. There are numerous techniques for speed control. Using voice as input control will reduce the manual operation. Voice recognition applications can be interfaced and speed control of DC motors can be done using the Arduino UNO microcontroller or Raspberry Pi. In addition to this IR sensor is used to sense the motor speed and in turn speed of the motor can be received via Bluetooth to the android mobile. If we control the Industrial Devices system using Speech then we can save enough time to do other sophisticated work. The voice control is highly reliable and fast. Thus the IOT makes the industrial automation much more easier.

I. INTRODUCTION

Speech is one of the natural forms of communication. Speech recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech wave. Speech recognition can be classified into identification and verification. Speech identification is the process of determining which registered speech provides a given utterance. Speech verification, on the other hand, is the process of accepting or rejecting the identity claim of a speech and it is the method of automatically identify who is speaking on the basis of individual information integrated in speech waves. The main aim of the project is speech identification, which consists of comparing a speech signal from an unknown speech to a database of known speaker. The system can recognize the speaker, which has been trained with a number of speakers and the motor is controlled as according to the voice input by the user.

II. PRINCIPLES OF SPEECH RECOGNITION

Speech recognition methods can be divided into text-independent and text-dependent methods. In text dependent method the speech has to say key words or sentences having the same text for both training and recognition trials. Whereas in the text independent does not rely on a specific text being speaks. Formerly text dependent methods were widely in application.

Every technology of speech recognition, identification and verification, whether text-independent and text- dependent, each has its own advantages and disadvantages and may require different treatments and techniques. The choice of which technology to use is application-specific. At the highest level, all speech recognition systems contain two main modules feature extraction and feature matching. Text- independent system the system recognize the speech without having a certain word in a database, the system extracts the singular characteristics of the speaker´s voice, making possible recognition without saying a precise word. The Text-dependent system recognize the speech based on some words or phrases that were previously recorded and stored in a database, for example the speech say a PIN number or his name to activate a device

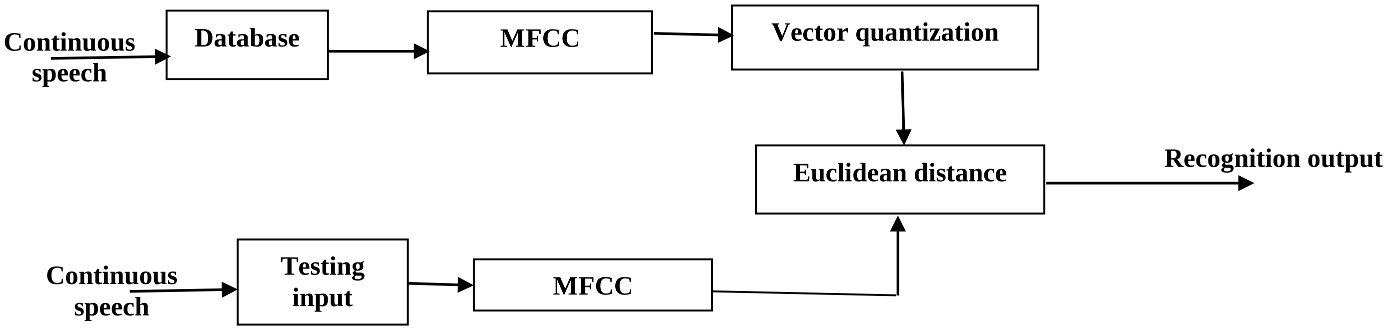
III. SOFTWARE REQUIREMENT

MATLAB is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, you can analyze data, develop algorithms, and create models and applications. The language, tools, and built-in math functions enable you to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages such as C/C++ or java.

Compared to other numerically oriented languages like C++ and FORTRAN, MATLAB is much easier to use and comes with a huge standard library. The unfavorable comparison here is a gap in execution speed. This gap is not always as dramatics popular lore has it, and it can often be narrowed or closed with good MATLAB programming.

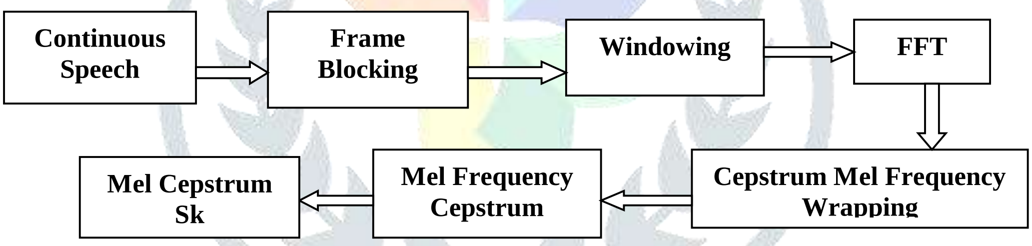
IV. BLOCK DIAGRAM OF SPEECH RECOGNITION SYSTEM

Speech recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech wave. Here using Mel frequency Cepstral Coefficient (MFCC) and Vector quantization technique. Mel frequency Cepstral Coefficient (MFCC) to extract the features from voice and vector quantization to identify the speaker, this technique is usually used in data compression, it allows to model a probability functions by the distribution of different vector, the results that we achieve were hundred percent of precision with a database of speakers.



A. **Speech Feature Extraction:** The purpose of this module is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate). The speech signal is a slowly time 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 0.2s or more) the signal characteristics 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 speech recognition task, such as Linear Prediction Coding (LPC), Mel Frequency Cepstrum Coefficients (MFCC), and others The LPC features were very popular in the early speaker-identification and speaker-verification systems. However, comparison of two LPC feature vectors requires the use of computationally expensive similarity and hence LPC features are unsuitable for use in real-time systems. MFCC is perhaps the best known and most popular, and this feature has been used here. It was originally developed for speech recognition.

B. **MFCC Technique:** MFCCs are based on the known variation of the human ear's critical bandwidths with frequency; filters spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the phonetically important characteristics of speech. This is expressed in the Mel-frequency scale, which is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Here the Mel scale is being used which translates regular frequencies to a scale that is more appropriate for speech, since the human ear perceives sound in a nonlinear manner. This is useful since our whole understanding of speech is through our ears, and so the computer should know about this, too. A block diagram of the structure of an MFCC processor is given in Figure below. The speech input is typically recorded at a sampling rate above 12500 Hz. This sampling frequency was chosen to minimize the effects of aliasing in the analog to-digital conversion is shown below.



1. **Frame Blocking:** The first step is framing. Is the segmentation of the speech samples in boxes within the range of 20ms to 40ms. The speech signal is split up into frames typically with the length of 20 to 40 milliseconds. The frame length is important due to the trade off between time and frequency resolution. If it is too long it will not be able to capture local spectral properties and if too short the frequency resolution would degrade and if too short the frequency resolution would degrade. The frames overlap each other typically by 25% to 70% of their own length.
2. **Windowing:** In signal processing, a window is used when a signal we are interested has a limited length. Indeed, a real signal has to be finite in time; in addition, a calculation is only possible from a finite number of points. To observe a signal in a finite time, we multiply it by a window function. After the signal is split up into frames each frame is multiplied by a window function. A good Window function has a narrow main lobe and a low side lobe. A smooth tapering at the edges is desired to minimize discontinuities.
3. **Fast Fourier Transform (FFT):** The next processing step is the Fast Fourier Transform, which converts each frame of *N* samples from the time domain into the frequency domain. These algorithms are based on decomposing and breaking the transform into smaller transforms and combining them to give the total transform. FFT reduces the computation time required to compute a discrete Fourier transform and improves the performance by a factor of 100 or more over direct evaluation of the DFT. FFT reduces the number of complex multiplications from N2 to (N/2) log2N and its speed improvement factor is N2 /(N/2) Log2N). It typically uses a window of N samples, N is a number that is a power of 2, it is because the FFT algorithm is much faster for these numbers.
4. **Mel Frequency Wrapping:** The frequency range in the FFT spectrum is very wide, so much data to process. So, we must use a filter bank in the Mel scale. The power spectrum of the speech that is obtained is to be integrated within overlapping critical band filter responses. Human perception of the frequency contents of sounds for speech signal does not follow a linear scale. Hence we use a frequency scale called Mel scale which is based on pitch perception and is used in the filter bank for the Mel cepstral approach. This scale has rough linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz. The speech consists of tones with different frequencies. For each tone with an actual Frequency, f, measured in Hz, a subjective pitch is measured

on the 'Mel' scale. As a reference point, the pitch of a 1 kHz tone, 40dB above the perceptual hearing threshold, is defined as 1000 Mel's.

Therefore we can use the following formula to compute the Mels for a given frequency *f* in Hz.

$$Mel(f) = 2595*\log10(1+f/700)$$

Where f is the actual frequency and Mel(f) is the perceived one.

The Mel frequency warping when calculating MFCCs is accomplished by the use of a triangular Mel filter bank. It consists of several triangular shaped and Mel spaced filters, and their outputs are described by

$$y(i) = \sum_{j=1}^{N} S_j H_{i\,j}$$

Where *S j* is the N-point magnitude spectrum and *Hi j* the sampled magnitude response of an M-channel filter Bank. The Mel frequency bank filter is applied in frequency domain.

5. **Cepsrum:** In the final step, the log Mel spectrum has to be converted back to time. The result is called the Mel frequency Cepstrum coefficients (MFCCs). The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the Mel spectrum coefficients are real numbers (and so are their logarithms), they may be converted to the time domain using the Discrete Cosine Transform (DCT). Doing this we got the Mel Frequency Cepstrum Coefficients, we called the set of coefficients acoustic vectors, consequently each input expression is transformed into a sequence of acoustic vectors. Discrete Cosine Transform (DCT) removes the correlation between the output values of the filter bank and collects features of parameters. Since we have performed FFT, DCT transforms the frequency domain into a time-like domain called quefrency domain. The obtained features are similar to cepstrum, thus it is referred to as the mel-scale cepstral coefficients, or MFCC.

C. **Vector Quantization:** The technique of VQ consists of extracting a small number of representative feature vectors as an efficient means of characterizing the speech specific features. By means of VQ, storing every single vector that we generate from the training is impossible. By using these training data features are clustered to form a codebook for each speaker. In the recognition stage, the data from the tested speech is compared to the codebook of each speech and measure the difference. These differences are then use to make the recognition decision. The VQ design algorithm requires an initial codebook. The initial codebook is obtained by the splitting method. In this method, an initial code vectors is set as the average of the entire training sequence. This code vector is then split into two. The iterative algorithm is run with these two vectors as the initial codebook. The final two code vectors are split into four and the process is repeated until the desired number of code vectors is obtained.

D. **Euclidean Distance:** In the recognition stage, the data from the tested speech is compared to the codebook of each speech and measure the difference. These differences are then use to make the recognition decision.

$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i-q_i)^2}$$

CONCLUSION

The proposed system can be used in hazardous environments where humans cannot survive. With the speed feedback mechanism, the system is much more reliable than any other open loop control mechanism. In this project android mobile phone acts as a microphone and the voice command is given to the mobile and speed is varied, and thus the system provides a new technology for industrial and home automation.

REFERENCES

[1] L. Rabiner, B.H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.

[2] Md. R. Hasan, M. Jamil, Md. G. Rabbani, Md. S. Rahman, "Speech Identification using Mel Frequency Cepstral Coefficients," Third International Conference on Electrical & Computer Engineering ICECE, Dhaka, 2004.

[3] W. Yutai, J. Xiaoqing, L. Feng,"Speech Recognition Based on Dynamic MFCC Parameters, "School of Information Science and Engineering, University of Jina , 2002. [4] A. Zulfiqar, T. Enriquez, "A Speech Identification System Using MFCC Features with VQ Technique," Third International Symposium on Intelligent Information Technology Application, vol.3, pp.115 – 118, Mar. 2009.

[5] W.Han, C.F. Chan, C.S. Choy and K.P. Pun, "An Efficient MFCC Extraction Method in Speech Recognition," Department of Electronic Engineering, The Chinese University of Hong Kong, Hong, IEEE – ISCAS, 2006.

[6] L. D. Alsteris and K. Paliwal, "ASR on Speech Reconstructed from short- time Fourier Phase Spectra', School of Microelectronic Engineering Griffth University, Brisbane, Australia, ICLSP – 2004.

[7] P. Kumar and P. Rao, "A Study of Frequency-Scale Warping for Speech Recognition", Dept of Electrical Engineering, IIT-Bombay, National Conference on Communications, NCC 2004, IISC Bangalore, Jan 30 -Feb 1, 2004.

[8] V. Tiwari, " MFCC and its applications in speech recognition", International Journal on Emerging Technologies",vol.1, pp.19-22,Feb.2010.

[9] K. Samudravijaya, R. Madan," A novel approach to speech verification", http://speech.tifr.res.in

[10] S.H. Chen and Y.R. Luo, "Speech Verification Using MFCC and Support Vector Machine", Proceedings of the

International Multi Conference of Engineers and Computer Scientists Vol.1 IMECS 2009, March 18 - 20, 2009, Hong Kong.
[11] H. Lei, E.L. Gonzalo-" Mel, Linear, and AntiMel Frequency Cepstral Coefficients in Broad Phonetic Regions for Telephone Speech Recognition" The International Computer Science Institute, Berkeley, CA.
[12] Y. Linde, A. Buzo, R. M. Gray, "An algorithm for Vector Quantizer Design". IEEE Transaction on Communications, 28: 1980, pp 84-95.
[13] Kinnunen T.and Kärkkäinen I., "Class-Discriminative Weighted Distortion Measure for VQ-Based Speech Identification". Joint IAPR Int. Workshop on Statistical Pattern Recognition (SPR'2002), Windsor, Canada, 681-688, August 2002.
[14] Masahisa M Suzuki "speech recognition using MFCC and vector quantization tech, the university of Eletro-communications (UEC).2012.IEEE.
[15] Kavitha K J, "Automatic speech recognition system using MATHLAB", Sr.lecturer, ECE Dip, PESITM, Shivamoga, Karnataka, India.
[16] S. M. Kamruzzaman1, A. N. M. Rezaul Karim2, Md. Saiful Islam3 and Md. Emdadul Haque, "speech identification using MFCC –Domain Support Vector Machine" Department of information and communication engineering. University of Rajshahi, rajshahi-605.Bangladesh.
[17] Atahan Tolunay, "Text-dependent speech verification implemented in Matlab using MFCC and DTW" Department of Electrical Engineering.