A Systematic Approach for Taking Academic Attendance Using Voice Identification Technique

Ajinkya Rajhans
Department of Design and Automation, SMEC
School of Mechanical Engineering
VIT University, Vellore, Tamil Nadu
India
ajinkya.rajhans@outlook.com

Jay Kesarkar
Department of Design and Automation, SMEC
School of Mechanical Engineering
VIT University, Vellore, Tamil Nadu
India
jaykesarkar22@gmail.com

Akshay Ashok Ruikar
Department of Design and Automation, SMEC
School of Mechanical Engineering
VIT University, Vellore, Tamil Nadu
India
akshayruikar91@yahoo.com

Abstract: Automatic voice acknowledgment has made extraordinary steps with the improvement of advanced sign handling equipment and programming. Voice is essential method of Communication among of individual. The correspondence among human computer collaboration is called human computer interface. Voice has capability of being vital method of cooperation with computer The configuration of voice acknowledgment framework requires watchful considerations to the accompanying issues: Meaning of different sorts of voice classes, voice representation, systems, voice classifiers, database and execution assessment. This project present the suitability of (MFCC) Mel Frequency Cepstral coefficient Algorithm and vector quantization to concentrate highlights and neural network System model for highlight choice, by lessening the dimensionality of the separated elements. There is an expanding requirement for another Feature choice strategy, to build the preparing rate and acknowledgment exactness of the classifier, by selecting the discriminative components. Henceforth a neural network system and back propagation framework model is utilized selecting the idealcomponents from Voice vectors which are separated utilizing MFCC. The work has been done on MATLAB 13a and trial results demonstrates that framework is capable perceive voices of various understudy for participation precision.

*****

I. INTRODUCTION

Voice Recognition is otherwise called Automatic Voice Recognition (AVR), or Voice acknowledgment which is the procedure of changing over a Voice sign to a succession of voice by method for a calculation executed as a computer program. It has the capability of being a critical method of communication amongst people and computers [1]. By and large, machine acknowledgment of voice is done by coordinating the given Voice signal against the succession of words which best matches the given Voice test [2]. The principle objective of Voice acknowledgment range is to create methods and frameworks for Voice info to machine. Voice is the essential method for correspondence between people. For reasons running from innovative interest about the components for mechanical acknowledgment of human Voice abilities to yearning to computerize straight forward undertakings require human machine cooperations. The examination in AVR by machines has pulled in a lot of consideration for around sixty years [3] and AVR today finds boundless application in undertakings that require human machine interface, for example, programmed call preparing [4].

1.1 AVR System Classification:

Voice Recognition is an exceptional instance of pattern recognition. There are two stages in administered design acknowledgment, viz., Training and Testing. The procedure of extraction of elements pertinent for order is normal in both stages. During the preparation stage, the parameters of characterization model are evaluated by utilizing an extensive number of class cases (Training Data). During the testing or acknowledgment stage, the component of test example (Test voice Data) is coordinated with the prepared model of every last class. The test example is pronounced to have a place with that whose model matches the test design best

1.2 Continuous Voice

Consistent VOICE recognizers permit clients to talk actually, while the computer decides the substance. Fundamentally, it is computer transcription [8]. Recognizers with nonstop VOICE capacities are the absolute most troublesome occupation to make since they use uncommon techniques to decide articulation limits. As vocabulary develops bigger, confusability between various word groupings develops.

1.3 Spontaneous voice

This kind of voice is characteristic and not practiced. An AVR framework with spontaneous voice ought to have the capacity to handle an assortment of common voice elements, for example, words being run together, "ums" and "ahs" and even slight stammers . Spontaneous or unrehearsed voice may incorporate errors, false-begins, and non-words

II. TYPES OF SPEAKER MODEL:

All speakers have their exceptional voices, because of their special physical body and identity. Voice acknowledgment framework is extensively ordered into primary classes in view of speaker models, to be specific, speaker ward and speaker autonomous.

2.1 Speaker subordinate models:

Speaker subordinate frameworks are intended for a particular speaker. They are for the most part more exact for the specific speaker, yet a great deal less precise for others speakers. This frameworks are typically less demanding to create, less expensive and more exact, yet not as adaptable as speaker versatile or speaker free frameworks.

2.2.Speaker autonomous models

Speaker autonomous framework are intended for assortment of speakers. It perceives the discourse examples of an extensive gathering of individuals. This framework is most hard to grow, most costly and offers less exactness than speaker subordinate frameworks. Be that as it may, they are more adaptable.

III. FEATURE EXTRACTION

The voice feature extraction in an order issue is about lessening the dimensionality of the info vector while keeping up the separating force of the sign. As we probably am aware from major development of speaker distinguishing proof and check the voice feature extraction in an order issue is about lessening the dimensionality of the data vector while keeping up the segregating force of the sign . As we probably am aware from key arrangement of speaker distinguishing proof and check framework, that the quantity of preparing and test vector required for the grouping issue develops with the measurement of the given information so we require feature extraction of voice signal. Taking after some component extraction strategies:

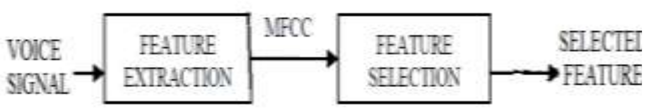


Fig.1 Basic Model Of Voice Recognition

Linear Predictive Coding (LCOMPUTER)
Perceptually Based Linear Predictive analysis(PLP)
Cepstrum method

Mel-Frequency Cepstrumcoefficient (MFCC)

3.1.Lenear Prediction Coding

Equation for predicting can be written as:

$$X^\wedge(N+1) = \sum_{L=0}^{P-1} w(l)x(n-1)$$

whereX^(N + 1) is the predicted value of X(N + 1)and w(0), w(1)….w(p-1) are the weights to be multiplied for a ‘p’ length filter to obtain X^(N + 1). It is called linear predicting because, here we are trying to predict the next sample as linear combination of ‘p’ previous samples

3.2 Error In Prediction

$e(n)=x(n=1)- x^\wedge(n+1)$

mean square error = $E[fe(n)^2] = E[e(n)\, e^*(n)]$

$\mathcal{E} = E[e(n)\{x^*(n=1)- \sum_{l=0}^{p-1} w*(l)x*(n-l)\}]$

To get more accurate prediction, we should minimize the mean square error.
Since Ɛ is a real function of complex variable it is enough of we differentiate with respect to $w^*(x)$ or w(k) and equate it to zero.

Therefore $\frac{d\mathcal{E}}{d*w*(k)} = 0$ k=0, ……p-1

$E[\frac{d}{dw*(k)}\{e(n)x*(n+1)\} - \frac{d}{dw*(k)}\{e(n)\ \sum_{l=0}^{p-1} w*(l)x*(n-l)]= 0$

$= E[-e(n)x^*(n-k)]=0$

$= E[e(n)x^*(n-k)]=0$ k=0…… p-1.

Substituting for e(n)

$E\{x(n+1)\ x^*(n-k)\}- \sum_{l=0}^{p-1} w(l)E\{x(n-1)\, x*(n-k)\}=0$ …… (1)

We are assuming the process to be wide sense stationary i.e

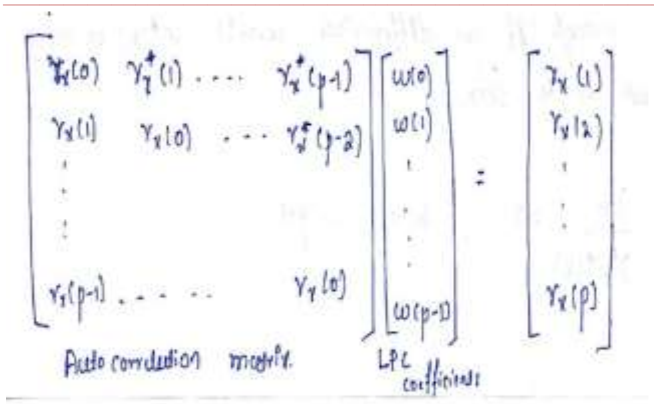
$E\{x(k)x^*(l)\}=r_x(k-l)$

Equation (1) reduces to

$= r_x(k+1) - \sum_{l=0}^{p-1} w(l)\, rx(k-1) = 0$; k=0…..p-1.

Writing it in the matrix form.

 ISSN: 2321-8169
210 - 213



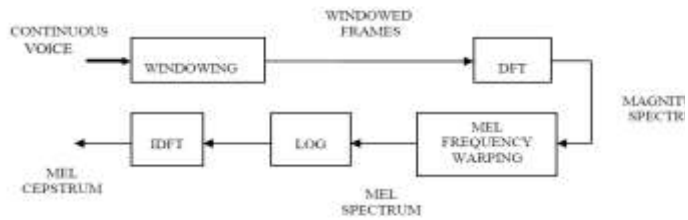


Fig. 2 Feature Extraction Using MFCC

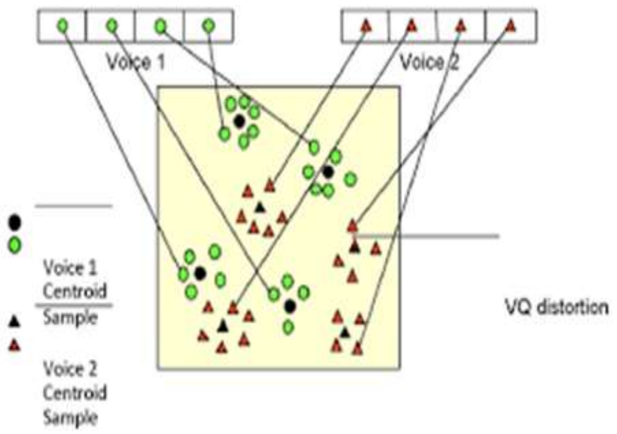


Fig. 3 Vector quantization code-book formation

3.3 Neural Network

3.3.1 Training The Neural Network:

1 Feed Forward:

Each feature is multiplied with the weights of the error sum will go to the next node. The weighted sum is mapped to the the range 0 to 1 using an activation function. The error is used to updates the weights in each iteration

2. Back Propagation:

The error calculated in the feed forward is propagated back to updatw the weights in order to match the obtained output to the desired value. This is called back propagation algorithm.

IMPLIMENTATION:

1. Record the voice samples for 1 second with the sampling rate of 8000.So we will get 8000 samples
2. Noise removal
3. After removing the noise by keeping the thresholdnumber of samples will be reduced greatly.

It is passed through high pass filter to boost the high frequency components,since the high frequency components are more susceptible for noise.

4.Frame Blocking:

Array of voice samples are divided into many number of overlapping blocks.This is done because over the small length of the block we can assume the process as WSS process.
We have taken the block length as 240 samples and the overlapping as 80 samples.So the first block will be from 0 to 239 and the second block will start from the index 160 i.e (240-80).

5.Hamming Window:

Each block is multiplied by the hamming window.This is done because if we consider the block directly for the analysis there will be sudden discontinuity at the ends.So to remove the sudden discontinuity the ends are tapered using the hamming window.
So the ends will be attenuated in the process.That is why we consider the overlapping windows in order to accommodate all the information correctly.

6.Auto correlation Matrix Formation:

The matrix is formed as explained in the theory.
MFCC coefficients are obtained for the each block. These coefficients obtained are used to train the neural network.To get accurate result many samples are required to train it. In our examples we tell the numbers from 0 to 9 and the system should recognize it and display the number.
We trained 15 samples of each 0 to 9 in a constant acoustic and could achieve an accuracy of 90 % for a speaker.

IV. RESULT AND DISCUSSION:

We initially tested with a group of people consisting of 10 person it gave results accurately for 6 of them the we further improved the algorithm with few more filters and tested it again with 10 people and this time it showed better results with recognising 8 of them correctly then we performed with 30 people with noise proof environment this time it recognizes 27 people correctly. With more research and industrial grade instruments it can be more accurate and could be more viable to use in the real world.

V. CONCLUSION:

In this project we have made a graphical user interface for taking student attendance for academic purpose. After recording of students voices, we apply pattern recognition algorithm to find out MFCC's of each voice sample. After storing all these MFCC's into database, we apply vector quantization algorithm to generate codebook which are useful in recognizing voice of particular student and mark his attendance.

VI. ACKNOWLEDGMENT:

We are very grateful to our guide **PROF.S KRISHNA**(Asst. Prof. VIT UNIVERSITY, VELLORE) School of mechanical engineering VIT UNIVERSITY, vellore, for his provision of expertise, and technical support in the project. We would also like to thank for his patience and belief in us as mentor.We would also like to appreciate the contribution of each member of this group towards the project.

REFERENCES:

[1] Santosh K.Gaikwad, Bharti W.Gawali. A Review on Speech Recognition Technique, International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010

[2] Samudravijaya K. Speech and Speaker recognition tutorial TIFR Mumbai 400005

[3] Catherine J Nereveettil1, M.Kalamani2, Dr.S.Valarmathy3“ Feature Selection Algorithm for Automatic Speech Recognition Based On Fuzzy Logic”, International Journal of Advanced Research in Electrical,Electronics and Instrumentation Engineering ISSN (Print) : 2320 – 3765

[4] M.A.Anusuya , S.K.Katti “Speech Recognition by Machine: A Review” International journal of computer science and Information Security 2009

[5] Samudravijay K “Speech and Speaker recognition report” source: http://cs.jounsuu.fi/pages/tkinnu/reaserch/index.html Viewed on 23 Feb. 2010

[6] Sanjib Das,” Speech Recognition Technique: A Review”, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622

[7] IBM (2010) online IBM Research

[8] slide share.com

[9] intellectual interchange.com

[10] Dannis Norris, “Merging Information in Speech Recognition” feedback is never Necessary workshop.1995