ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

A Novel Method of Digitization & Noise Elimination of Digital Signals Using Image Processing Concepts

[1] M.R. Prasad [2] Pavithra G [3] Dr. T.C.Manjunath
[1] B.E., M.Tech. [Ph.D.-VTU-Pursuing]
Research Scholar, VTU – RRC, Belagavi, Karnataka &
Assistant Professor, CSE Dept., JSSATE, Bangalore
[2] B.E., M.Tech., [Ph.D.-VTU-Pursuing]
Research Scholar, VTU – RRC, Belagavi, Karnataka
[3] Ph.D. (IIT Bombay), FIETE, FIE, CE, SMIEEE
Professor & HOD, ECE Dept., Dayananda Sagar College of Engg., Bangalore

***Abstract:* -- The primary purpose of this research work is to develop a speech recognition system, which will be capable of recognizing and interpreting an individual's voice. This will thus obviate the use of keyboards and other such input devices to output the data. The scope of this work is limited to the recognition of digits. Matlab, a versatile software tool is made use of for the processing of the signals. This processing involves filtering, shaping, correlating the speech signal which will enable the computer to interpret the commands given by the user. The signature of every spoken digit is unique from which we can derive the parameters which differ from digit to digit. These parameters serve as a concrete base for the implementation of a speaker independent voice recognition system. Using such parameters simultaneously and using probability concept the digits are sorted and arranged in the descending order of their probability. The digit with the highest probability is displayed.**

I. INTRODUCTION

1. Scope of the research work

The subject of our work is an isolated word, speaker independent 0 to 9 digit recognition system that can achieve a higher level of accuracy. We have chosen speaker independent approach because for a small vocabulary systems (e.g., digit recognizers) with large amount of users it is not feasible to store training data for every individual user.

Furthermore, most systems cannot train themselves to a new speaker very rapidly. Hence, we aim to make research in the field of speaker independent isolated work 0 to 9 digit recognition. Digitization of the analog speech signal and then elimination of the unwanted noise signal. The purpose of digitization is to produce sampled data representation of the speech signal with high Signal to Noise Ration (SNR).

The primary purpose of this research work is to develop a speech recognition system, which will be capable of recognizing and interpreting an individual's voice. This will thus obviate the use of keyboards and other such input devices to output the data. The scope of this work is limited to the recognition of digits.

II. DIGITIZATION AND NOISE ELIMINATION

Digitization is a process of converting the speech signal from sound pressure wave to a digital signal and emphasizing the important frequency components in the signal. It involves digitization of the analog speech signal and then elimination of the unwanted noise signal. The purpose of digitization is to produced sampled data representation of the speech signal with high Signal to Noise Ration (SNR).

The block diagram of the system is shown in Fig. 1. We know that speech is a continuously varying analog signal. The first step in any system is to capture the speech and a general purpose high sensitivity low

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

noise microphone is well suited for such an application. Audio is inherently an analog phenomenon.

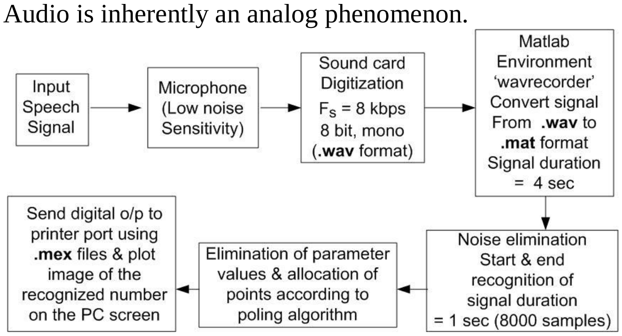

Fig.1 Block diagram of the implement speech

Since the analog data to be digitized is human speech, the signal content is upto 3 KHz. Hence, according to nyquist criterion of sampling frequency (Fs > 2 Fm), we select sampling frequency as 8 KHz. Selection of higher sampling rate will increase the accuracy of detection but at the cost of processing time.

III. MATLAB ENVIRONMENT

In order to recognize the digit spoken we need to process the sampled speech signal. Hence, we require a versatile software tool for high performance numerical computation and visualization. "Matlab" is used as the tool. Mex files are written for the software design.

IV. PARAMETERS

Basically, we aim to develop methods and techniques that enable computer systems to accept speech input and to transcribe the recognized utterances into normal orthographic writing. 4 basic approaches to attain this goal have been implemented.

a). Template based approaches, where the incoming speech is compared with stored units in an effort to find the best match.

b). Knowledge based approaches that attempts to emulate the human expert ability to recognize the speech.

c). Stochastic approaches, which exploit the inherent statistical properties of the occurrence and co-occurrence of individual speech sounds.

d). Connectionist approaches which use networks of a large number of simple, interconnected nodes which are trained to recognize speech.

In the search for a sound feature that would be relevant in respect to our classification purpose, we developed signal processing algorithms to extract several physical parameters. Parameters are those features which are unique for a particular digit. Thus, we have to derive some parameters out of the input speech signal which should have the following characteristics :

- ♣ Values of parameters should vary widely from digit to digit.
- ♣ For a particular digit, they should be insensitive to the change in speaker.
- ♣ Parameters should be stable over a period of time.
- ♣ Such parameters should be easy to derive.

4.1 Energy related parameters

The energy of a signal is a crucial time domain parameter used in separation of digits. It is a measure of intensity of the signal. The energy of a spectrum is computed as a sum of the squares of the amplitude of each sample as . The entire signal in TD in divided into windows of size 50. The energy of all such windows is computed and plotted. Some digits show peculiar characteristic in some windows, for e.g., the number of lobes produced by every digit. Hence, these characteristics can be used to derive parameters like number of lobes and lobe size.

4.2 Number of lobes

According to observations made the number of lobes generated for the energy signal is different for different digits. It basically depends on the utterance of digits. On basis of utterances, the digits can be classified as single sounding (1,2,5,6,8,9) and multi sounding (0,3,7). The digit 4 (four) cannot be classified under a specific group because for some dummy inputs it was single sounding while for others it was multi sounding. The single sounding digits like one, six produce a single lobe in the selected window. Lobe size : The size of the extracted lobe is calculated using the mathematical formula, lobe size = count 2 – count, where count 2 = the energy window where the signal falls below max / 5. Count 1 = the energy window where the signal rises

above max / 3. The lobe size is a striking feature which enables us to differentiate between various digits.
E.g. 1 : Dummy inputs for various single sounding digits were studied and it was found that for the digit 6 (six), the lobe size was < 20 samples. This differentiated the digit 6 from other single sounding digits.

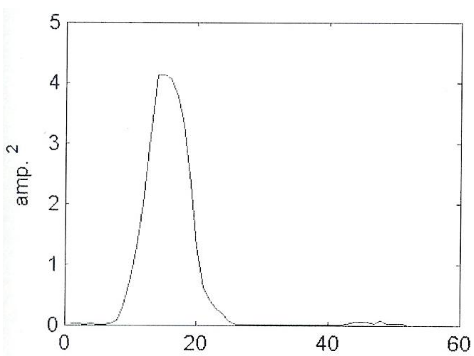

Fig. 2 : Digit 6 energy window

E.g. 2 : For double sounding digits, the size of the 1st lobe was considered. Dummy inputs for various multi sounding digits were studied and it was found that for the digit 3 (three), the lobe size was < 20 samples. This differentiated the digit 3 from other multi sounding digits. Hence, the above mentioned energy related parameters can be determined using a suitable Matlab program designed by us.

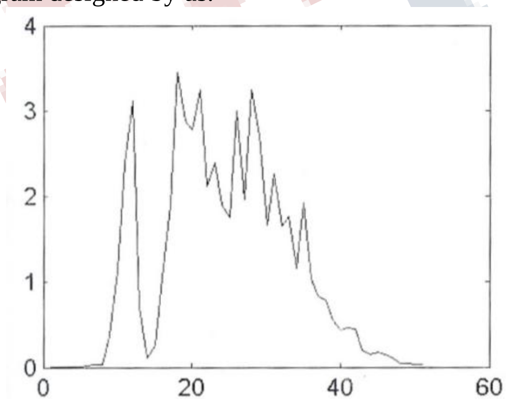

Fig. 3 : Digit 3 energy window

Correlation related parameters : Correlation is simply defined as a mathematical operation between two sequences which produce another sequence called either 'cross correlation sequence', when the 2 sequences are different or 'auto correlation', when the 2 sequences are identical. The correlated sequence is represented in TD as follows

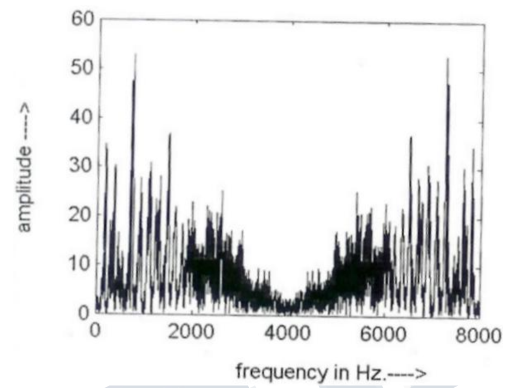

Fig. 4 Digit 9 frequency spectrum

$$acf_{xx}(k) = \int_{-\infty}^{\infty} x(t)\, x(t-k)\, dt$$

and in DT as follows :

$$acf_{xx}(\tau) = \sum_{n=1}^{N-1} x[n]\ x[t-\tau]\ .$$

4.3 Formatting of speech signal

A slot for microphone input is available on every PC with a sound card. Every windows operating system provides a sound recorder to record the analog speech and digitize it. The user can provide the sampling frequency and the number of quantization levels.

Since the analog data to be digitized is human speech, the signal content is upto 3 KHz. Hence, according to nyquist criterion of sampling frequency (Fs > 2 Fm), we select sampling frequency as 8 KHz. Selection of higher sampling rate will increase the accuracy of detection but at the cost of processing time. We select quantization levels of 8, i.e., 8 bits / sample.

4.4 Spectral analysis parameters

In spectral analysis, the frequency spectrum of the speech signal is used for finding out the common signatures of the input signal. The FFT is used for converting the N samples of input signal from TD to FD. The FFT is a fast algorithm to implement the DFT which is defined on the set of N samples {xn}, as follows :

Note that we use j here to denote the imaginary unit, i.e., . In general, Xn's are complex numbers. The result after this step yields complex sequence hence the

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

absolute value is taken. The frequency spectrum of every digit is also observed.

a). Pitch : The frequency at which maximum amplitude of the input signal is obtained is defined as 'pitch' of the respective signal. It is the fundamental frequency of the speech signal. It is the frequency with which vocal chord vibrate. This pitch may vary from person to person. In order to compensate for this variation, we have provided a frequency range with the pitch frequency as the central frequency.

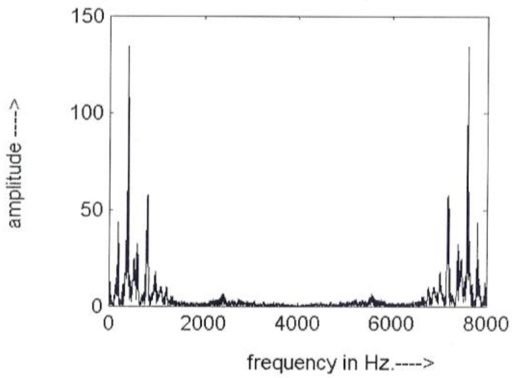

Fig. 5 : Digit 4 frequency spectrum

This frequency range will be used as a parameter to differentiate between various digits. E.g., dummy inputs of various people for the digit 4 (four) were studied. The frequency range selected for the inputs was from 325 to 450 Hz. The pitch is determined using a Matlab program. After determining the pitch we have decided the frequency range with the precaution of providing a guard band. The guard band is nearly 50 Hz on either side.

b). Threshold crossing for all frequencies : Every signal is corrupted with noise, however most of the times, such noise does not cross a particular threshold. Hence, the signal above such threshold is actual signal. The pitch of the input signal was found out and the signal was normalized, by dividing every sample's amplitude by the maximum amplitude, which occurs at pitch frequency. According to observations made the different digits have different number of samples crossing a threshold of 0.2, hence, 0.2 was fixed as the threshold level. E.g., Dummy inputs of various people for the digit 8 (eight) were studied. All inputs had a threshold crossing value <100. This parameter is also determined by a Matlab program.

c). Threshold crossing at high frequencies : The signal crossing a particular threshold was extracted again. According to observations made, for certain digits the signal crossed a particular threshold for high frequencies. The threshold used for this parameter was calculated by computing the average of the extracted signal over the period where the extracted signal existed. Such high frequency components were unique for that particular digit and hence proved to be a powerful aid in separating the digits.

d). Deviation : is the variation of the 2nd maximum amplitude from the 1st maximum amplitude w.r.t. to the 1st maximum amplitude. It can be mathematically expressed as

deviation = (max 1 – max 2) / max 1,

where max 1 is the maximum amplitude at pitch frequency and max 2 is the second maximum amplitude. E.g., Dummy inputs of various people for digit 2 were studied. All the inputs studied had a deviation ranging from 0.7 to 0.85, which could be used to separate the digit 2 from others. The parameter is determined by a Matlab program. However, this parameter proved to be an extra parameter and increased the computational time, the parameter could be included in the program if necessary.

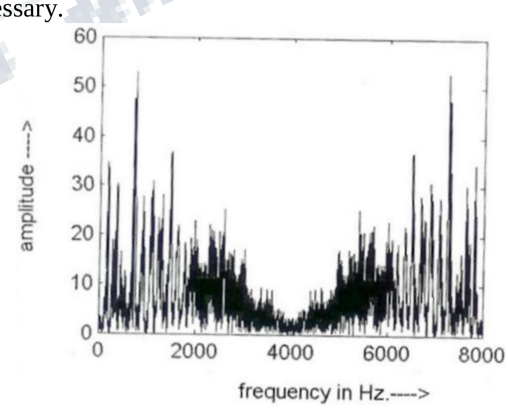

Fig. 6 : Digit 9 frequency spectrum

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

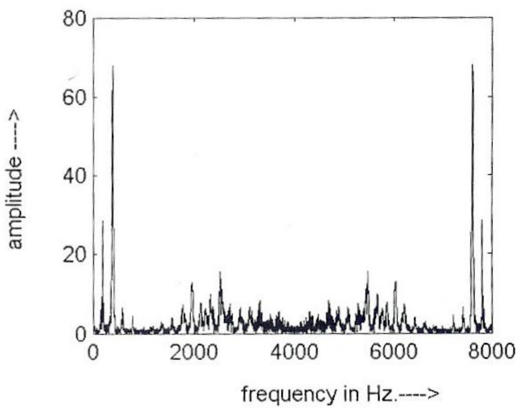

Fig. 7 : Digit 6 frequency spectrum

e). Zero crossing rate : is the number of times the speech signal crosses the zero level. The rate at which zero crossings occur is a simple measure of the frequency content and the repetition in the speech signal. Hence, the major information is contained in the zero crossing rate of a signal. To eliminate noise in the signal, we have chosen a threshold level rather than zero level.

f). Maximum amplitude after pitch frequency amplitude : Another important parameter derived by observation of the frequency spectrum used the maximum amplitude after pitch frequency amplitude. The frequency spectrum showed a similarity for all the inputs of a particular digit.

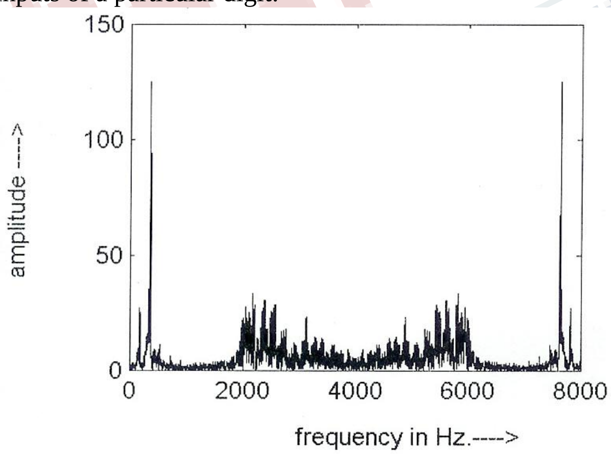

Fig. 8 : Digit 8 frequency spectrum

Few other parameters were also tried and tested. However, they have not been used. These parameters were the size of main lobe occurring in the frequency spectrum and the frequency spectrum decimation in amplitude. However, their usage just increases the processing time and these parameters do not separate any digits but give a general division of digits into 2 broad classes, hence ewe have excluded the usage of these parameters. They can be included if any when necessary. These parameters were determined by a Matlab program.

Compatibility of data types :

The recognized digit in Matlab needs to be passed to this mex file to give digital output at the printer port. To achieve this it is necessary to maintain parallelism between data types of C and Matlab. The mxArray is a special structure or a data type that contains Matlab data. It is the C representation of a Matlab array. All types of Matlab arrays (scalars, vectors, matrices, strings, cell arrays, etc.,.) are mxArrays in C language.

Obtaining output

The basic aim behind the recognition of digits is to control a process or a machine or a robot using speech. Also, many software applications can be built using the digit recognizer.

V CONCLUSIONS

A speaker independent isolated (0-9) recognition system has been successfully implemented. The system under best conditions showed accuracy more than 90 %. In implementation of the system, a low noise high sensitivity microphone is used as the transducer element for receiving the digit input. This analog signal is sampled, quantized and then converted into digital signal by the soundcard present in every PC. Matlab, a versatile software tool is used for further signal processing. After recording the sound, the proper content of the signal is determined by endpoint and start point algorithm. This algorithm, in addition with end and start algorithm, helps in eliminating the noise to a considerable extent.

The signal is then further processed to drive certain common parameters in same digit spoken by

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

different parameters. Such parameters are consistent for the same digit spoken by different people and vary for other digits. The parameters found by us can be broadly classified on the basis of extraction techniques as temporal analysis parameters and spectral analysis parameters. The temporal analysis parameters basically derived from time domain analysis of signal consists of very important parameters like energy related parameters. The spectral analysis parameters derived from frequency domain analysis of signal consists of parameters like pitch, threshold crossings, zero crossing rate, etc.,.

Due to the inconsistency of parameters for males and females, separate programs for males and females are developed. As a result, the accuracy is increased. Using these parameters, the digits are sorted according to their probability of their occurrence. We make use of the polling system to sort the digits. The digit with the highest probability is displayed onto the computer screen. An interface between Matlab and MS-paint has been established. In our program, we have given an option of re-recording the digit if digit is wrongly recognized. All the interfaces made with the user are either visual or audio and hence require no expert knowledge of the system.

Another way of displaying the recognized digit is 7-segment displays. The recognized digit is sent to the printer port by using mex-files in C language. The binary output at printer port is converted into decimal using IC-7447. The IC 7447 drives the 7-segment displays. Our system is basically a prototype and the scope of it can be enhanced beyond imagination. We hope that our humble effort can serve as the foundation in research in speaker independent speech recognition.

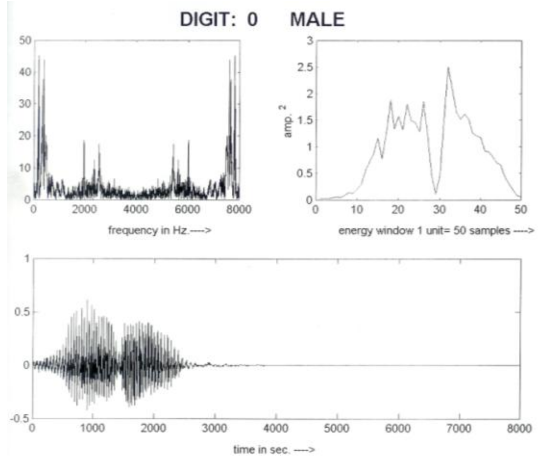

Fig. 9 : Response of digit 0 of Male

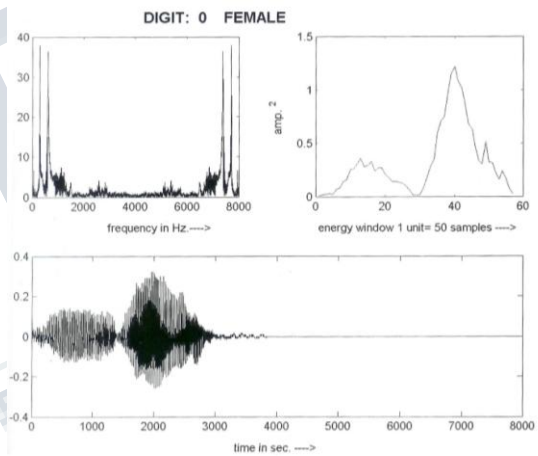

Fig. 10 : Response of digit 0 of Female

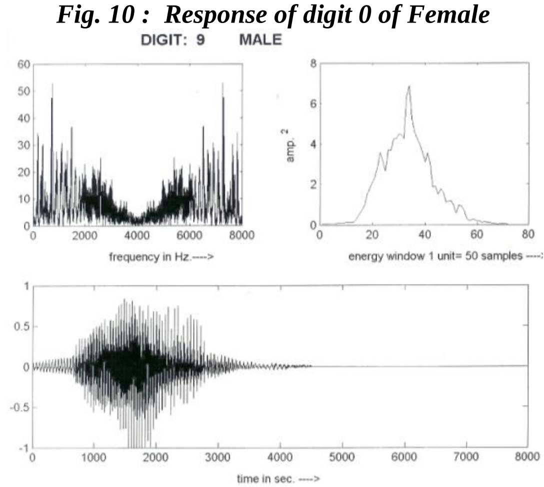

Fig. 11 : Response of digit 9 of Male

The responses of digits 1 to 8 were observed similarly for both males and females. The responses of digits 1 to 8 were observed similarly for both males and females.

Software codes designed and implemented :

- Code for speech recognition for males
- Code for speech recognition for females

Different functions used are

- Energy consumption
- Computing the threshold ratio
- Computing the pitch of the signal
- Calculation of 2nd max. frequency after pitch frequency and deviation
- Computing threshold crossing in frequency domain
- Computing the maximum probable digit
- To check whether the recognized digit is correct or not
- To plot the image of the recognized
- To generate the output at the printer port
- Different functions used are
- Energy consumption.
- Computing the threshold ratio.
- Computing the pitch of the signal.
- Calculation of 2nd max. frequency after pitch frequency and deviation.
- Computing threshold crossing in frequency domain.
- Computing the maximum probable digit.
- To check whether the recognized digit is correct or not.
- To plot the image of the recognized.
- To generate the output at the printer port.

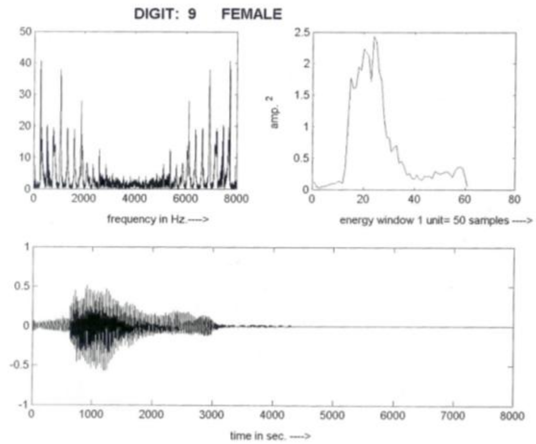

Fig. 12 : Response of digit 9 of Female

REFERENCES

[1] Proakis, Manolakis, Digital Signal Processing - Principles, Algorithms and Applications, PHI, New Delhi, India.

[2] Rudra Pratap, Getting started with Matalb, TMH.

[3] Bernard Grob, Theory and applications of DSP.