International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Estimating the Closeness of Known and Unknown Speech Signal using MATLAB

Shivayogi G K[1], Asha K S[2]

[1]M.Tech Student (SP & VLSI) Department of Electronics and communication Engineering,
Jain University, Karnataka, India

[2]Assistant Professor, Department of Electronics and Communication Engineering,
Jain University, Karnataka, India

Abstract: ***In today's world implementation of any expert system with maximum data and networking security becomes a real necessity in academic organizations as well as industrial communities. In this paper, an expert security system is developed using speech recognition technology. A human can easily recognize a familiar voice however; getting a computer to distinguish a particular voice among others is a more difficult task. This paper describes the process behind implementing a speaker recognition algorithm in MATLAB. The algorithm utilizes the Discrete Fourier Transform in order to compare the frequency spectra of two voices. A plot is then generated depicting how the normalized frequency spectra in your voice and compare to the average normal vector of voice data. The algorithm makes a comparison and displays in the command window 'YOU ARE NOT!!!!' if you do not fall within 2 standard deviation of the normal average voice. If you do happen to fall within 2 standard deviation, then the command window displays 'HELLO!!!'***

Keywords: Speech Recognition Technology, Discrete Fourier Transform, spectra, secure system

1. Introduction

Since the rapid development of the information technology in the past years, human dependence on 3C products is higher and higher. The 3C products must have attractive functions and good services. The interface between product and user is then quite important. For example handwritten input and touch screen monitor [2,3,4] are favored by users.

Recently, the topic on the process of audio signal attracts much attention. There are many researches about speech recognition [5,6,7] because speech recognition will be a standard interface in the future. Speech is a complicated signal produced as a result of several transformations occurring at several different levels: semantic, linguistic, articulator and acoustic. Differences in these transformations are reflected in the differences in the acoustic properties of the speech signal. Besides there are speaker related differences which are a result of a combination of anatomical differences inherent in the vocal tract and the learned speaking habits different individuals. In speaker verification, all these differences are taken into account and used to discriminate between speakers.

2. Methodology

In this system there are two phases
2.1. Train phase
2.2. Test phase

2.1 Train Phase

2.1.1 Signal Acquisition

This is also called creating data files

- Record the word by recorder which is in audio format.
- Convert the audio format to wave format by use the converter say wave file.
- Read the wave file in MATLAB.
- This is the data size of wave file, used for analysis.
- Spectra plus also used for analysis on wave file.

2.1.2 Signal input

In this project record time of voice samples is two second and that is fixed, every second 44100 samples will record so that total 88200 samples recorded. This voice are converted to wave format and stored in same directory, these voice samples are taken to compare with the external input voice sample. Then the recordings are cropped and placed in a 88200*20 matrix.

2.1.3 Resizing of Signals

Resizing of a signal to a desired size is often required in many applications. Sometimes, we need to reduce or increase the length or size of a signal. Most often we use to insert zeros to increase the length, and delete some portion of the signal at the end or beginning to reduce the length. This is not the proper way to manipulate size. Inserting zeros unnecessarily introduces silent period in case of sound wave. Deleting again removes the portion of the signal which is not desirable in many cases. S simpler improved way to reduce or increase length of a signal twice, thrice etc without losing much information in the signal is done in the following way.

For reduction, select the signal at position multiple of 2 or 3 etc to reduce twice or thrice the length of the original signal. To increase the length twice o thrice, replicate each element twice o thrice. Such process of length or size manipulation is often followed by an interpolation process to smoothen the signal.

2.1.4 Spectrum Analysis

Any voice analysis in time domain would be extremely impractical. Instead, an analysis of the frequency spectra in a voice (which remains predominately unchanged as speech is slightly varied) turned out to be a more viable option. Converting all recordings into frequency domain (by applying the Discrete Fourier Transform) greatly simplified the process of comparing two recordings. That being said, working in frequency domain also provided a new set of issues that required attention.

The frequency spectrum of a time-domain signal is a representation of that signal in the frequency domain. The frequency spectrum can be generated via a Fourier transform of the signal, and the resulting values are usually presented as amplitude and phase, both plotted versus frequency.

Due to nature of human speech, all data pertaining to frequencies above 600Hz can safely be discarded. Therefore, once a recording is converted into frequency domain, it could then be simply regarded as a vector in 600-dimensional Euclidean space. At this point, a comparison between two vectors could easily be carried out by normalizing the vectors (giving them length1) then computing the norm of the difference between the two (of course, the difference between two vectors in R600 is performed by subtracting component wise) [8].

This completes the train phase and next we have to analyze test phase to compare with train phase.

2.2 Test Phase

A test set is a set of data used in various areas of information science to assess the strength and utility of a predictive relationship. Test sets are used in artificial intelligence, machine learning, genetic programming and statistics. In all these fields, a test set has much the same role.

In train phase we are taking 10 sets of voice samples from same person in same environment condition and each voice file containing 88200 samples then it is resizing to 600 samples, so that when we applied FFT on that samples it creates real and imaginary parts.

In train phase we are taking 10 voice samples, that is resized to 600 samples. After that when we applied FFT it will create 600*10*2(real & imaginary) array. But same in test phase we are verifying for single voice file through external input to the stored samples in the same directory, so it will become 600*2 array. After subtracting both i.e. (600*10*2-600*2) it will create 600*2 array that array is passed to standard deviation.

2.2.1 Standard deviation

In statistics and probability theory, the standard deviation (SD) (represented by the Greek letter sigma, σ) measures the amount of variation or dispersion from the average. A low standard deviation indicated that the data points tend to be very close to the mean (also called expected value); a high standard deviation indicates that the data points are spread out over a large range of values. The standard deviation of a random variable, statistical population, data set or probability distribution is the square root of its variance. It is algebraically simpler though in practice less robust than the average absolute deviation. A useful property of the standard deviation is that, unlike the variance, it is expressed in the same units as the data. Note, however, that for measurements with percentage as the unit, the standard deviation will have percentage points as the unit.

Equation

$$\sigma = \sqrt{\sum(x-\bar{x})/(n-1)}$$

Where:

x=each sample

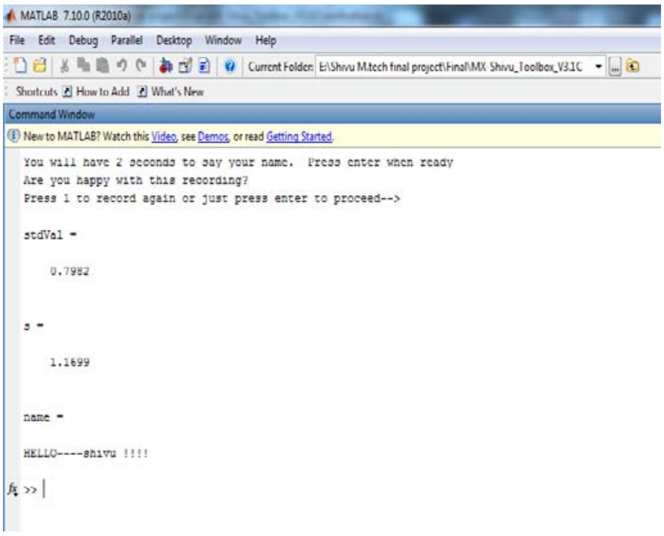
N=the number of average

∑=mean of average

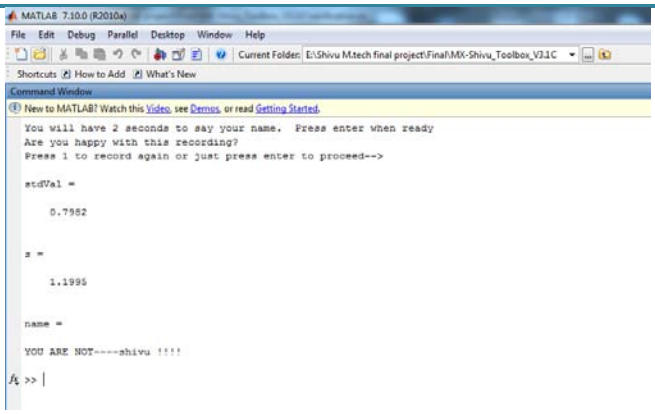
After finding standard deviation individually, if the standard deviation of data set is less than external input then it will show HELLO!!!. If the standard deviation of the external input is more than the data set than it will show YOU ARE NOT!!!.

3. Results

If the standard deviation of external voice input is less than the standard deviation of stored voice files then it will display HELLO----name!!!!.



If the standard deviation of external voice input is greater than the standard deviation of stored voice files then it will display YOU ARE NOT----name!!!!.



4. Application

4.1 Speech verification in forensic department

One of the important area of Automatic Speaker Verification application is in forensics. Usually in the case where a crime has been committed and the voice of criminal needs to be verified from a recorded message. Traditionally this was done by training a specialist who can able to identify the speaker's voice by comparing the visual speech features (spectrograms voice prints) of the speakers. But the accuracy in these methods were found not reliable and not effective. To prove that the suspect is the criminal, it needs to be verified beyond reasonable doubt that the voice of the criminal and the voice of the suspect are the same. So to overcome this problem a Automatic and reliable Speaker Verification system is desired.

4.2 Speech Verification in Organizations

Speech verification is widely used in many organization for the purpose of attendance system, which controls the employee timekeeping.

4.3 Speech Verification in Home Security

Speech verification provides strong security at entrances to homeowners. In today's world implementation of any expert system with maximum data and networking security becomes a real necessity in academic organizations as well as in industries communities. A smart microphone is situated in front of the door, will receive the voice samples and allows the hardware sensor to open/close the door of the system.

4.4 Speech Verification in Computers

Speech verification is also devolved for security access to computers and providing single logon facilities. So by keeping security with the documents only owner can access the system and it provides high security and safeguard the documents by other persons.

5. Conclusion

In this work, 10 voice files are collected and analyzed on them. Each Sample spectrum parameter values are obtained using MATLAB. FFT spectrum methodology is used to find the parameter. Some words are same but they still have some different parameters which tell us about the word. E.g. in speech FILL and KILL words are similar and have some parameters which are same, but even though parameters like Maximum peak, THD, THD+N are different which can be differentiate between them. So to recognition of any word, first of all values of the selected parameters are obtained. When these parameter values lie in between the bounded values then we find the correct word. The accuracy of this system is 75%. However this can be enhanced by using AI techniques to train them.

5. Future Recommendation

Now a day's biometric technology such as finger prints [9], voice prints, iris scan [10], face detection [11], signatures or the geometry of the hand are becoming increasingly popular due to the use of unique physical traits. In voice identification technology is still slow to take off in many markets because it is not as accurate as other biometric technologies due to the tendency to have a high flash reject because of background noise and other variables. However with the advancement of signal processing technology better vocal synthesis, analysis and measurements using sophisticated algorithms can be taken and converted into voice print, a unique digital representation of an individual's voice. This technology helps businesses and governments to fight identity threat and fraud, secure transactions, project confidential information, reduce costs and enhance level of service.

References

[1] R. Dutta1, S. Dutta2, K. Mitra3 Speaker Verification for Security System using Spartan6, International Journal of Electronics and Communications (IJEC), Volume -1, Issue -1, August 2012.

[2] X. Huang, Y. Abe and I. Echizen, Capacity daptive synchronized acoustic steganography scheme, Journal of information Hinding and Multimedia Signal Processing, vol. 1, no. 2, pp. 72-90, 2010.

[3] J. McAuley, J. Ming =, D. Stewart, and P. Hanna, Subband correlation and robust speech recognition, IEEE Trans. Speech and Audio Processing, vol. 13, no. 5, pp. 956-964, 2005.

[4] K. Yamamoto and M. Iwakiri, Real-time audio watermarking based on characteristics of PCM in digital instrument, Journal of Information Hiding and Multimedia Signal Processing, vol. 1, no. 2, pp.59-71, 2010.

[5] C. Wan and L. Liu, Research and improvement on embedded system application of DTW-based speech recognition, Proc. Of the 2nd International Conference on Anti-counterfeiting, Security and Identi_cation, Guiyang, China, pp.401-404, 2008.

[6] T. Kinjo and K. Funaki, On HMM speech recognition based on complex speech analysis, Proc. Of the 32nd Annual Conference of the IEEE Industrial Electronics, Paris, France, pp.3477-3480, 2006.

[7] H. Sayoud and S. Ouamour, Proposal of a new confidence paramenter estimating the number of speakers-An experimental investigation, Journal of Information Hiding and Multimedia Signal Processing, vol. 1, no.2, pp.101-109, 2010.
[8] Voice Recognition Using MATLAB David Roberts.
[9] F. A. Afsar, M. Arif and M. Hussain, "Fringerprint Identification and verification system using minute matching", proceedings of National Conference on Emerging Technologies 2004, pp. 141-146.
[10] Weicheng Shen Khanna R., "Prolog to Iris Recognition : An Emerging Biometric Technology", Proceedings of IEEE, vol. 85, no. 9, September 1997.
[11] Bowyer K. W., "Face Recognition technology: security versus Privacy", Technology and Society Magazine, IEEE, vol. 23, no. 1, pp. 09-19, June 2004.

Author Profile

Shivayogi G k received the Bachelor of Engineering in Electrical and Electronics Engineering from Visvesvaraya Technological University. Currently perusing Master of Technology in Signal Processing and VLSI from Jain University. His area of interest is Signal Processing, VLSI analog design.

Asha K S obtained the Bachelor of Engineering from Siddaganga institute of Technology, Visvesvaraya Technological University, Master of Technology from Sri Siddhartha institute of Technology, Visvesvaraya Technological University. Her area of interest is wireless communication.