Volume 16 Issue 1
Originally issued as G1



Global Journal of human-social science: G
Linguistics & Education
Volume 16 Issue 1 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-460X & Print ISSN: 0975-587X

Development of Voice Recognition for Student Attendance

By Md. Nasir Uddin, MM Rashid, MG Mostafa, Belayet H, SM Salam, NA Nithe, MW Rahman & S Halder

Islamic University Malaysia, Malaysia

Abstract- Development of voice recognition for student attendance system is beneficial in many ways. It helps the lecturer in administrative the attendance of their student with efficiency. This is because students always cheat with their attendancy by signing on behalf of their friend who did not attend class. With this project, voice biometric is used as a medium for student to mark their attendance. Cheating among students will be prevented because like fingerprints, each voice is different. The objectives of this project are to study and understand the properties understand the properties of speaker recognition and to analyze the effectiveness of using Euclidean distance feature for speaker recognition. Databases of 26 volunteers were collected consisting of only male. The report result is tabulated. Three types of analysis were done, first same train is used as test data reported 100% correct. The remaining two analyses used different test data recording. Volunteers use the same sentence as test data reported 76.92% correct. Lastly volunteers used their name and the correct percentage is 46.15%.

GJHSS-G Classification : FOR Code: 139999



Strictly as per the compliance and regulations of:



© 2016. Md. Nasir Uddin, MM Rashid, MG Mostafa, Belayet H, SM Salam, NA Nithe, MW Rahman & S Halder. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License http:// creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Development of Voice Recognition for Student Attendance

Md. Nasir Uddin [α], MM Rashid [σ], MG Mostafa [ρ], Belayet H [ω], SM Salam [¥], NA Nithe [§], MW Rahman [χ] & S Halder [ν]

Abstract- Development of voice recognition for student attendance system is beneficial in many ways. It helps the lecturer in administrative the attendance of their student with efficiency. This is because students always cheat with their attendancy by signing on behalf of their friend who did not attend class. With this project, voice biometric is used as a medium for student to mark their attendance. Cheating among students will be prevented because like fingerprints, each voice is different. The objectives of this project are to study and understand the properties understand the properties of speaker recognition and to analyze the effectiveness of using Euclidean distance feature for speaker recognition. Databases of 26 volunteers were collected consisting of only male. The report result is tabulated. Three types of analysis were done, first same train is used as test data reported 100% correct. The remaining two analyses used different test data recording. Volunteers use the same sentence as test data reported 76.92% correct. Lastly volunteers used their name and the correct percentage is 46.15%.

Chapter 1

I. Introduction

a) Overview

Initially, lecture attendance systems in IIUM University were based on a piece of attendance paper that is passed around by students, which requires student to sign on the date column under their name. Sometimes, lecturer calls the person name one by one to mark as attendance but this method are time consuming. As an alternative solution, biometrics technologies can be introduced to construct a more powerful version of attendance system. Biometric is an authentications technique that recognizes unique features in each human being. In this case, voice recognition is used as biometric because it is a natural signal to produce. Each person has their unique characteristic in speech and voice that can be captured and analyse to make this new class attendance more efficient and effective.

Voice recognition can be divided into two, which are speech recognition and speaker recognition. Both are using voice biometric differently. Speech recognition is the ability to recognize what have been said while speaker recognition is the ability to recognize who is speaking. In brief, speech recognition covers the ability to match a voice pattern against an acquired or provided vocabulary. Normally, the vocabulary given is small and the user needs to record a new word to expand the vocabulary.

Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech signals. It can be divided into two tasks, which are identification and verification. Speaker identification is uded to decide which unknown voice belongs to from amonst a set of known speakers. Speaker verification accepts or rejects the identity claim of a speaker. For this project speaker identification in speaker recognition is used. The unknown speaker is later labeled as test data and the set of known speaker is labeled as train data. This is possible because different speakers have different spectra for similar sound. Spectra are the location and magnitude of peaks in spectrum.

b) Problem Statement

As mentioned earlier, a piece of paper is used as an attendance in our lecture. This method comes with many problems. If a student attendance is less than 80% or missed six classes per semester, they will be barred from the final examination. Three classes miss, warning letter will be issued by the lecturer and sent to their parent.

To avoid all of the above, if a student did not come on that particular lecture, he or she will ask their friend to sign on their behalf. Sometimes, the class is empty and when the lecturer checks the attendance, it is fully sign. When ask who is responsible for this action, nobody will admit it. This is not fair to those students who come regularly, as their attendance is the same with those who seldom comes. There is a case where the student only comes a few times per semester that is only during quiz and exams day. When the result turn out bad, that student will blame the lecturer.

Furthermore, when a student has a habit of missing classes, they tend to bring this bad habit in their working world. This will badly affect their performance and reputation. To overcome this problem voice recognition can be used in class attendance system. The advantages of the system are as follow:

- Only student who comes to the class will be mark present in the attendance.
- Warning letter and barred letter will be issued without any mistake.

Author α σ ρ χ ν: Department of Mechatronics Engineering, International Islamic University Malaysia, Kuala-Lumpur, Malaysia.
e-mail: nasir.u@live.iium.edu.my
Author ω ¥ §: Department of Electrical & Electronics Engineering, ADUST, BUET, I&E, Dhaka, Bangladesh.

© 2016 Global Journals Inc. (US)

- Make the student attend the lecture and think twice before missing the class.
- The class will be full house attendance and encourage the lecturer to teach passionately.

c) Objectives

The main objective of this project is to design and develop a voice recognition system for class attendance. The objectives of this project are:

- To study and understand the properties of speaker recognition.
- To study and understand Euclidean distance feature.
- To collect voice to be used as database.
- To study and analyze the output of Matlab code.

d) Project Methodology

To achieve the objectives stated above, the following approach will be taken in this project:

- Literature review
- Voice data collection.
- Coding using Matalab software.
- Testing and analyze the coding

e) Scope of Work

In this project the scope can be divided into two that is data collection and coding. In data collection part, the participant need to pronounce loud and clear a sentence repeatedly for five time as this will be the train data for this project. For the test data, the same participant needs to pronounce the same sentence and their own name to be used for the analysis later.

The second part consists of the coding. The coding is written in Matlab software using all the data collected before. After that, the data is analyze to know how accurate the feature being used.

f) Report Content

This report consists of five chapters. The first chapter, which is the introduction consists of back-ground of the project, problem statement, objectives, project methodology and scope of work that need to be done in accomplishing this project. The second chapter focuses on the theoretical background and literature reviews of speaker recognition. In addition, a review of past method and features of speaker recognition is also included.

Chapter three discusses the methodology used in this project. Euclidean distance is used as the feature of this project and all the equation is stated here. Next chapter 4, is the result and analysis. All the preliminary results obtain are presented. These include table of analysis. Finally, chapter 5 is the conclusion of this report and future work for Final Year Project 2.

Chapter 2

II. Literature Review

a) Overview

Chapter 2 consists of all theoretical background and literature reviews of speaker recognition. In addition, a review of past method and features of speaker recognition is also included.

b) Voice Recognition

Voice recognition suggests that the computer do not understand it but only can take command and perform it. Comprehending human languages falls under a different field of computer science called natural language processing. Nowadays, a lot of voice recognition systems are available. The most powerful one can recognize thousands of words.

Traditionally, voice recognition system only used in a few specialized situations because of their limitations and high cost. As the time goes on, the cost decreases and performance improves, speech recognition systems are entering the mainstream. For example, voice recognition is used as an alternative to keyboards. These systems are useful in instances when the user us unable to use a keyboard to enter data because his or her hands are occupied or disabled. Instead of typing commands, the user can simply speak into a headest.

Speaker recognition is a biometric modality that uses an individual's voice for recognition purposes. It is a different technology than speech recognition, which recognizes words as they are articulated from. Speech recognition is not a biometric. The speaker recognition process relies on features influenced by both the physical structure of an individual's vocal and the individual's behavioral characteristics. This project will concentrate on speaker recognition.

c) Speaker Recgnition

Speaker recognition is divided into two which are verification and identification. Speaker verification is use to validate a person's claimed identity from his voice. Many terms which has the same meaning with speaker verification is usually used. For example, voice verification, speaker authentication, voice authentication, talker authentication and talker verification. A person can makes an identity claim with the help of other source. For example, by entering an employee number or presenting his smart card. (Reynolds, 2008)

In the other hand, speaker identification means there is no prior identity claim and the system decides who the person is, what group the person is a member of or that the person is unknown. In a simple word, speaker verification is defined as deciding if a speaker is who he claims to be. This is different from the speaker identification problem. (Peacocke & Graf, 1990)

© 2016 Global Journals Inc. (US)

Speaker verification can be divided further into text dependent and non-independent text. In text-dependent recognition, the phrase is known to the system and can be fixed. While in non-independent, the speaker can use any phrase and then analyze by the sytem. The typical speaker recognition setup is further explained. (Reynolds, 2008)

i. *Speaker Recognition Setup*

The person speaks the phrase into a microphone. This signal is analyzed by a verification system that makes the binary decision to accept or reject the user's identity claim or possibly to report insufficient confidence and request additional input before making the decision. The person, who has previously enrolled in the system, presents an encrypted smart card containing his identification information. He then attempts to be authenticated by speaking a prompted phrase (s) into the microphone. (Campbell, 1997).

Before verification session, the person voice must be recorded earlier in the system. Usually it is under a supervised conditions and environment. During this time, voice models are generated and stored on asmart card for use in later verification sessions. There is generally a difference between accuracy and the duration and number of enrolment sessions.

ii. *Speaker Recognition Errors*

Table 2.1 : Sources Verification and Identification Error

Misspoken or misread prompted phrase
Extreme emotional states (stress)
Time varying microphone placement
Poor or inconsistent room acoustics
Channel mismatch (using different microphone for enrolment and verification)
Sickness (head colds can alter the vocal tract
Aging (vocal tract can drift away from models with ages)

There are many factors of verification and identification errors. Some of the human and environmental factors that contribute to these errors are list in Table 2.1. These factors generally are outside the scope of algorithms or are better corrected by means other than algorithms. For example the use of a better set of microphones. These factors are very important. No matter how good a speaker recognition algorithm is, human error ultimately limits its performance. As an example human may misreading or misspeaking the phrase provided (Campbell, 1997).

d) Previous Speaker Recognition Work

For the past years, there is a lot of speaker – recognition activity. Among those who have researched and designed about speaker-recognition systems are AT&T., the Dalle Molle Institute for Perceptual Artificial Intelligence Switzerland and many more. Table 2.2 shows a sampling of the chronological advancement in speaker verification. The following terms are used to define the columns in Table 2.

Source refers to a citation in the references, org is the company or school where the work was done and features are the signal measurements. Input is the type of input speech. For example laboratory, office quality or telephone. Text indicates whether a text-dependent or text-independent mode of operation is used. Moreover, method is the heart of the pattern-matching process and pop is the population size of the test or known as number of people. Finally error is the equal error percentage for speaker-verification systems or speaker identification systems given the specified duration of test speech in seconds. (Campbell, 1997)

e) Speech Production

The important physical distinguishing factor of speech is the vocal tract. The vocal tract is generally considered as the speech production organs above the vocal folds. Adult human vocal tract systems consist of this five which are laryngeal pharynx, oral pharynx, oral cavity, nasal pharynx and nasal cavity.

As the acoustic wave passes through the vocal tract, its frequency content (spectrum) is altered by the resonances of the vocal tract. Vocal tract resonances are called formants. Thus the vocal tract shape can be estimated from the spectral shape of the voice signal.

Typically, voice verification systems features derived only from the vocal tract. The human vocal mechanism is driven by an excitation source, which also contains speaker-dependent information. The excitation is generated by airflow from the lungs, carried by the trachea through the vocal folds. The excitation can be characterized as phonation, whispering, frication, compression, vibration or a combination of these (Cambell, 1997)

Chapter 3

III. Methodology

a) Overview

This chapter will discuss all the methodology used to achieve the objective of this project. This project can be divided into five stages. Those are planning, feature selection, collecting data, programming and design specification.

b) Planning

The first stage of Methodology is to plan the project properly. This is highly required so that it is easy to estimate the timing and duration of each activity for this project to be done efficiently. The important tasks and activities related to the project are displayed in the Gantt chart in table 3.1.

c) Feature Selection

As been mentioned before voice recognition biometrics is different from one human being to another.

© 2016 Global Journals Inc. (US)

It is suitable to choose voice as a medium for the class attendance system. More specifically this project will concentrate on the speaker recognition. Speaker recognition can be divided into two, which are independent speech and non-independent speech.

In non-independent speech, a specific text or phrase known by the system is speaks into the system microphone. Then it will analyze and validate or identify the owner of the voice. In the other hand, independent speech is a free text or phrase that can be used to identify the unknown voice belongs to whom. This can be achieved provided the data has been recorded earlier in the system. For the system to match the unknown voice to their respective name feature selection is needed.

Features can be defined as the signal measurement. As stated in chapter two, many features has been used before. Among those are Cepstrum, Normalized Cepstrum, Mel-Cepstrum and many more. For this particular project, Euclidean Distance is used as the features.

i. *Euclidean Distance*

In this project, the vector is in matrices. When the audio is read into Matlab it converts the signal into matrices. This is because computer only understands vector or number.

Euclidean distance is used because of its simplicity. In general it can be difine as the distance between two points. The equation is as follow

---

In the speaker recognition phase, the database or train data is compared with the unknown voice which is represented by a sequence of vector (Y1, Y2,…..Yn). In order to identify the unknown voice, Euclidean distance is used. This can be done by measuring the shortest distance of the two vector sets. The Euclidean distance is only and ordinary distance between the two points that can be measure with a ruler. This can be proven by repeated application of the Pythagorean Theorem. (Per-Erik Danielsson, 1980)

The following can be defined as the Euclidean distance formula:

The Euclidean distance between two points P= (p1, p2…..pn) and Q= (q1, q2……..qn)

The answer with the shortest distance is chosen to be identified as the unknown person voice.

d) Data Collection

Data collection is crucial part for this project. It must be done carefully so that the voice recorded can be used later in the system. For this project, 26 voices from 26 different volunteers have been recorded. It can be divided into two part which are train data and test data.

The voice data is collected using earphone that has a small microphone attach and connect it through a laptop. The volunteers need to pronounce loud and clear a sentence for a few times through the microphone. Then, the data is saved in the laptop. The environment must be constant for all the 26 voices to minimize noise and error. The data is recorded in mahallah room as a constant environment.

i. *Train Data*

Train data is used to train the system in identifies the correct answer. The train voice data must be recorded a few times to take the average and make the result more accurate. In this project, the volunteers must pronounce and repeated for five times this sentence “The quick brown fox jumps over the lazy dog”.

The sentence is known as pangram. A pangram is a sentence that comprises all the letters of the alphabet appeared at least once. This is the most famous English pangram because it is simple and easy to pronounce. Generally, an interesting pangrams are short ones and consist of a sentence that includes the fewest repeat letters possible. There are many other types of English language pangram. For example, “ The job requires extra pluck and zeal from every young wage earner’ and ‘A quart jar of oil mixed with zinc oxide makes a very bright paint.’

ii. *Test Data*

Test data is data which has been specifically identified for use in tests. It can be used as an unknown data, so that the system will come out with the correct answer. For this project, there are two test data. First, is the pangram itself. Full name of each of the 26 volunteers is used as the second test data.

In total, the volunteers need to pronounce six pangrams and one full name to be recorded. All of the above is record continually and then separate it accordingly.

iii. *Audacity Software*

Audacity is a free open source digital audio editor and recording computer software application. It is used to separate the continuous recording of voice audio, test and train data. The audio recorded is in wma format while Matlab can only read in wav format.

Audacity can convert wma format into wav format.

First, the audio is uploaded into the Audacity software. Then, it is trim for only the voice wanted. After that, the noise and unrelated sound is cut from the audio. Finally, the audio is saved in wav format. Each of the seven recorded audio per volunteer undergo this process.

e) Programming

The programming is constructed in Matlab software. The full coding is in the Appendix A. The coding consists of two functions. One is known as frame spectra function and the second, is Score function.

© 2016 Global Journals Inc. (US)

Frame Spectra function is the important part of this coding. All of the recorded audio data must go through this function to be process. When the audio wave read in Matlab software it s in matrices. The audio wave is usually amplitude over time. It needs to undergo Fast Fourier Transform (FFT) to become amplitude over frequency. Then, the frequency is sample at 44.1 kHz. After that it undergoes mormalization, to adjust the matrics so that it can be used in the next process.

In the Score function, the Euclidean distance principle is used. The unknown function or test data is compared one by one with all the train data to get it smallest difference in distance. The output is stored in D. The equation used in this function is as follow:

.....

There is also a loop that is used to calculate the mean distance of D obtain from the score function. With this, the average distance of the 26 person is obtained. The minimum mean obtained match with the unknown voice data. In other words, minimum mean distance is the shortest distance from the unknown data.

f) Design Specification

Design specification part list all the hardware needed to construct the prototype of this voice recognition for class attendance system. This part will be done in next semester for Final Year Project 2.

i. *Microcontroller*

After looking into various microcontrollers, the Arduino Uno board shown in figure below is choosen. Arduino is choose because it is cheap. Plug straight into computer USB port and simple to set up and used.

Table 3.2 shows list of specification of Arduino Uno obtained in the official Arduino website that is http://www.arduino.cc/en/Main/Arduino Board Uno.

Table 3.2 : Arduino Uno specification

Microcontroller	ATmega328
Operating Voltage	5V
Input Voltage (recommended)	7-12V
Input Voltage (limits)	6-20V
Digital I/O pins	14 (of whichprovide PWM output)
Analog Input Pins	6
DC Current per I/O pin	40 mA
DC Current for 3.3V Pin	50 mA
Flash Memory	32 KB (ATmega328) of which 0.5 KB used by bootloader
SRAM	2 KB (ATmega328)
EEPROM	1 KB (ATmega328)
Clock Speed	16 MHz
Length	68.6 mm
Width	53.4 mm
Weight	25 g

ii. *Microsd Card Module For Arduino*

This is a Micro SD (TF) module. It is tiniest card in the market available now and compatible with TF SD cad which is commonly used in Mobile phone. SD module has various applications such as data logger, audio, video and graphics. This module will greatly expand the capability of an Arduino board. Arduino cannot do many things because of their poor limited memory. This module has SPI interface and V power supply which is compatible with Arduino UNO/Mega. With this module it can store a big amount of information up until 2 GB depends on the micro SD used.

This project contains a big database to store the train data voice of a whole class. Each class has around twenty five students. Each student needs at least five voices to be data recorded and stored as train data. The 2GB memory included is more than enough to store this data.

iii. *Voice Recognition Module V3*

The voice recognition module V3 could recognize any voice command. It receives configuration commands or responds through serial port interface/software serial and voice recognition library for Arduino. With this module, the unknown voice can be recorded and then analyze by the Arduino. It also included one set of microphone.

iv. *12c Serial Lcs 1602 Module*

This module has 2.6 inch LCD display screen support. It allows only 2 input and output pins. This screen is used to display the output after the system has been program. If the output is correct it will be saved in the SD card. If it is wrong the student needs to run the programmed one more times.

Chapter 4

IV. Result and Analysis

a) Overview

In chapter 4 the result obtained in the Matlab software is put into a table to easily analyze. The table showed the mean distance of each of the 26 volunteers. As mentioned efore each volunteer has to repeat the pangrams for five times. The mean of these pangrams per volunteer is then calculated to get the average. There are two outputs from the programming which are minimum means and minimum distance.

Minimum mean can be explained as the nearest distance between unknown and mean that is calculated before. In the other hand, minimum distance is the shortest distance between unknown and each of the pangrams that is used as train data. If the answer is correct 1 is given and if the answer is wrong 0 is given. The percentage is then calculated by sum up all the correct identification (labeled 1) and divided by the total number of samples that is 26.

Three sets of test data is used to be analyze. First part, the train data is used back as the test data.

© 2016 Global Journals Inc. (US)

Secondly, another set of pangram that has been recorded earlier is used as test data. This act as a dependent text and labeled as fox test data. Lastly, name of each volunteer is used for the non-independent test data and labeled as name test data.

b) Summary

First part and second part, the percentage of the answer correct is within the range which can be satisfied. The third part answer is far away from correct. There are a few factor contributed to these problem. The main factor that has been identifies is the feature extraction used.

This project only usd Euclidean distance as their feature extraction. It is assumed to improve this project, more number of features are needed. There is a lot of feature extraction that is suitable to be combined with Euclidean distance. The examples are Cepstrum and filter –bank. The improvement will be done in Final Year Project 2.

Other than that, there is also noise in the voice recoded. It makes it hard for the system to match to the correct voice. To solve this, five pangrams per volunteers is recorded. Then, the average of the five pangrams is calculated and compares with the unknown voice. This should give more accurate and solid result.

Minimum mean and minimum distance in this three part analysis gives exactly the same percentage as each other. Minimum mean is the shortest distance between unknown voice and average of each volunteer train data. While, minimum dstance is the shortest distance between unknown voice and each of the pangram used as the database. In simple terms, minimum distance is the nearest distance from the train data in the database. This shows that, there is no difference between minimum mean and minimum distance.

Chapter 5

V. Conclusion and Future Work

a) Conclusion

The main objective of this Final Year Project was to develop voice recognition for class attendance system. It is divided into two, which are Project 1 in this current semester and Project 2 in the upcoming semester. In project 1, the theoretical background of this project was studied in details. This includes the speaker recognition, Euclidean distance and the hardware that will be used in project 2. Furthermore, past research and application related to voice recognition is also studied in detail. Finally, this project is beneficial in many ways especially towards the lecturer in IIUM University.

In short, Final Year Project 1 will give a great advantage to the future development in constructing the hardware part and further improve the programming of the device in Final Year project 2. Hence, the idea needed for Final Year Project 2 is fully gather.

b) Future Work

Future work in the upcoming semester is to build the hard ware part for this system. Moreover, the programming part of this project can be improved. Feature selection also can be added to get a more accurate result. Table 5.1 display the proposed activities and task that can be performed in project 2 with their estimated time to completion.

Bibliography

1. Joseph p. Campbell (1997). Speaker Recognition: A Tutorial. Proceedings of the IEEE, vol.85, no. 9.
2. Duglas A. Reynolds (2008). Overview of Automatic Speaker Recognition. JHU 2008 Workshop Summer School M. I.T. Lincoln Laboratory.
3. Per-Erik Danielsson (1980). Euclidean Distance Mapping. Department of Electrical Engineering, Linkoping University, Linkoping (pp 581-583), Sweden.
4. Richard D. Peacocke and Daryl H. Graf (1990). An Introduction to speech and speaker recognition.
5. Dongdong Li, Yingchun Yang, Zhenyu Shan, Gang Pan and Zhaohui Wu (2006). AVAS: An Audio-Visual Attendance System. Department of computer Science and Technology, In LCNS 4261 (pp. 667-675), Springer-Verlag Berlin Hidelberg 2006.
6. What is the difference between speech Recognition and speaker recognition. Retrived from https://-answers.yahoo.com/question/index?qid=20080620080053AA 02 UHQ
7. Fun with word. RETRIVED FROM HTTP://-RINKWORKS. COM/WORDS/PANGRAMS. SHTML
8. Intro To Arduino By Randofo. Retrived from HTTP://WWW.INSTRUCTABLES.COM/ID/ITRO-TO-ARDUINO/STEP2/ARDUINO-UNO-FEATURES

© 2016 Global Journals Inc. (US)