

1 Development of Voice Recognition for Student Attendance

2 MD. Nasir Uddin¹, Nasir Uddin², Nasir Uddin³ and MD. Nasir Uddin⁴

3 ¹ International Islamic University Malaysia

4 Received: 14 December 2015 Accepted: 5 January 2016 Published: 15 January 2016

5

6 **Abstract**

7 Development of voice recognition for student attendance system is beneficial in many ways. It
8 helps the lecturer in administrative the attendance of their student with efficiency. This is
9 because students always cheat with their attendancy by signing on behalf of their friend who
10 did not attend class. With this project, voice biometric is used as a medium for student to
11 mark their attendance. Cheating among students will be prevented because like fingerprints,
12 each voice is different. The objectives of this project are to study and understand the
13 properties understand the properties of speaker recognition and to analyze the effectiveness of
14 using Euclidean distance feature for speaker recognition. Databases of 26 volunteers were
15 collected consisting of only male. The report result is tabulated. Three types of analysis were
16 done, first same train is used as test data reported 100

17

18 **Index terms**— speaker recognition and to analyze the effectiveness.

19 **1 Introduction a) Overview**

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

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

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

40 **2 b) Problem Statement**

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

8 B) VOICE RECOGNITION

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

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

- 53 ? Only student who comes to the class will be mark present in the attendance.
- 54 ? Make the student attend the lecture and think twice before missing the class.
- 55 ? The class will be full house attendance and encourage the lecturer to teach passionately.

56 3 c) Objectives

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

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

63 4 d) Project Methodology

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

- 65 ? Literature review
- 66 ? Voice data collection.
- 67 ? Coding using Matlab software.

68 5 ? Testing and analyze the coding e) Scope of Work

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

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

75 6 f) Report Content

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

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

84 7 Literature Review a) Overview

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

87 8 b) Voice Recognition

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

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

97 Speaker recognition is a biometric modality that uses an individual's voice for recognition purposes. It is
98 a different technology than speech recognition, which recognizes words as they are articulated from. Speech

99 recognition is not a biometric. The speaker recognition process relies on features influenced by both the physical
100 structure of an individual's vocal and the individual's behavioral characteristics. This project will concentrate
101 on speaker recognition.

102 **9 c) Speaker Recognition**

103 Speaker recognition is divided into two which are verification and identification. Speaker verification is used to
104 validate a person's claimed identity from his voice. Many terms which has the same meaning with speaker
105 verification is usually used. For example, voice verification, speaker authentication, voice authentication, talker
106 authentication and talker verification. A person can make an identity claim with the help of other source. For
107 example, by entering an employee number or presenting his smart card. (Reynolds, 2008) In the other hand,
108 speaker identification means there is no prior identity claim and the system decides who the person is, what group
109 the person is a member of or that the person is unknown. In a simple word, speaker verification is defined as
110 deciding if a speaker is Speaker verification can be divided further into text dependent and non-independent text.
111 In textdependent recognition, the phrase is known to the system and can be fixed. While in non-independent,
112 the speaker can use any phrase and then analyze by the system. The typical speaker recognition setup is further
113 explained. (Reynolds, 2008) i. Speaker Recognition Setup The person speaks the phrase into a microphone. This
114 signal is analyzed by a verification system that makes the binary decision to accept or reject the user's identity
115 claim or possibly to report insufficient confidence and request additional input before making the decision. The
116 person, who has previously enrolled in the system, presents an encrypted smart card containing his identification
117 information. He then attempts to be authenticated by speaking a prompted phrase (s) into the microphone.
118 (Campbell, 1997).

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

123 ii. Speaker Recognition Errors There are many factors of verification and identification errors. Some of the
124 human and environmental factors that contribute to these errors are listed in Table ??-1. These factors generally are
125 outside the scope of algorithms or are better corrected by means other than algorithms. For example the use of a
126 better set of microphones. These factors are very important. No matter how good a speaker recognition algorithm
127 is, human error ultimately limits its performance. As an example human may misreading or misspeaking the
128 phrase provided (Campbell, 1997).

129 **10 d) Previous Speaker Recognition Work**

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

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

140 **11 e) Speech Production**

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

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

147 Typically, voice verification systems features derived only from the vocal tract. The human vocal mechanism
148 is driven by an excitation source, which also contains speaker-dependent information. The excitation is generated
149 by airflow from the lungs, carried by the trachea through the vocal folds. The excitation can be characterized as
150 phonation, whispering, frication, compression, vibration or a combination of these (Campbell, 1997) Chapter 3 III.

151 **12 Methodology a) Overview**

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

155 **13 b) Planning**

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

159 **14 Volume XVI Issue I Version I**

160 **15 c) Feature Selection**

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

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

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

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

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

175 Euclidean distance is used because of its simplicity. In general it can be define as the distance between two
176 points. The equation is as follow - ??—————In the speaker recognition phase, the database or train
177 data is compared with the unknown voice which is represented by a sequence of vector (Y1, Y2,?..Yn). In order
178 to identify the unknown voice, Euclidean distance is used. This can be done by measuring the shortest distance
179 of the two vector sets. The Euclidean distance is only and ordinary distance between the two points that can
180 be measure with a ruler. This can be proven by repeated application of the Pythagorean Theorem. (Per-Erik
181 Danielsson, 1980) The Euclidean distance between two points P= (p1, p2?..pn) and Q= (q1, q2?..qn)

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

183 **16 d) Data Collection**

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

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

191 **17 i. Train Data**

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

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

200 **18 ii. Test Data**

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

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

206 **19 iii. Audacity Software**

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

210 Audacity can convert wma format into wav format.

211 First, the audio is uploaded into the Audacity software. Then, it is trim for only the voice wanted. After that, 212 the noise and unrelated sound is cut from the audio. Finally, the audio is saved in wav format. Each of the seven 213 recorded audio per volunteer undergo this process.

214 Frame Spectra function is the important part of this coding. All of the recorded audio data must go through 215 this function to be process. When the audio wave read in Matlab software it s in matrices. The audio wave 216 is usually amplitude over time. It needs to undergo Fast Fourier Transform (FFT) to become amplitude over 217 frequency. Then, the frequency is sample at 44.1 kHz. After that it undergoes mormalization, to adjust the 218 matrices so that it can be used in the next process.

219 In the Score function, the Euclidean distance principle is used. The unknown function or test data is compared 220 one by one with all the train data to get it smallest difference in distance. The output is stored in D. The equation 221 used in this function is as follow: ???????.

222 There is also a loop that is used to calculate the mean distance of D obtain from the score function. With 223 this, the average distance of the 26 person is obtained. The minimum mean obtained match with the unknown 224 voice data. In other words, minimum mean distance is the shortest distance from the unknown data.

225 **20 f) Design Specification**

226 Design specification part list all the hardware needed to construct the prototype of this voice recognition for class 227 attendance system. This part will be done in next semester for Final Year Project 2.

228 **21 i. Microcontroller**

229 After looking into various microcontrollers, the Arduino Uno board shown in figure below is choosen. Arduino 230 is choose because it is cheap. Plug straight into computer USB port and simple to set up and used. ii. Microsd 231 Card Module For Arduino This is a Micro SD (TF) module. It is tiniest card in the market available now and 232 compatible with TF SD card which is commonly used in Mobile phone. SD module has various applications such 233 as data logger, audio, video and graphics. This module will greatly expand the capability of an Arduino board. 234 Arduino cannot do many things because of their poor limited memory. This module has SPI interface and V 235 power supply which is compatible with Arduino UNO/Mega. With this module it can store a big amount of 236 information up until 2 GB depends on the micro SD used.

237 This project contains a big database to store the train data voice of a whole class. Each class has around 238 twenty five students. Each student needs at least five voices to be data recorded and stored as train data. The 239 2GB memory included is more than enough to store this data.

240 iii. Voice Recognition Module V3

241 The voice recognition module V3 could recognize any voice command. It receives configuration commands or 242 responds through serial port interface/software serial and voice recognition library for Arduino. With this module, 243 the unknown voice can be recorded and then analyze by the Arduino. It also included one set of microphone.

244 iv. 12c Serial Lcs 1602 Module This module has 2.6 inch LCD display screen support. It allows only 2 input 245 and output pins. This screen is used to display the output after the system has been program. If the output is 246 correct it will be saved in the SD card. If it is wrong the student needs to run the programmed one more times.

247 **22 Chapter 4**

248 IV.

249 **23 Result and Analysis a) Overview**

250 In chapter 4 the result obtained in the Matlab software is put into a table to easily analyze. The table showed 251 the mean distance of each of the 26 volunteers. As mentioned efore each volunteer has to repeat the pangrams 252 for five times. The mean of these pangrams per volunteer is then calculated to get the average. There are two 253 outputs from the programming which are minimum means and minimum distance.

254 Minimum mean can be explained as the nearest distance between unknown and mean that is calculated before. 255 In the other hand, minimum distance is the shortest distance between unknown and each of the pangrams that 256 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 257 then calculated by sum up all the Three sets of test data is used to be analyze. First part, the train data is used 258 back as the test data.

259 Secondly, another set of pangram that has been recorded earlier is used as test data. This act as a dependent 260 text and labeled as fox test data. Lastly, name of each volunteer is used for the non-independent test data and 261 labeled as name test data.

262 **24 b) Summary**

263 First part and second part, the percentage of the answer correct is within the range which can be satisfied. The 264 third part answer is far away from correct. There are a few factor contributed to these problem. The main factor 265 that has been identifies is the feature extraction used.

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

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

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

278 Chapter 5 V.

279 25 Conclusion and Future Work a) Conclusion

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

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

289 26 b) Future Work

290 Future work in the upcoming semester is to build the hardware part for this system. Moreover, the programming
291 part of this project can be improved. Feature selection also can be added to get a more accurate result. Table
292 ???.1 display the proposed activities and task that can be performed in project 2 with their estimated time to
293 completion.

294 27 Bibliography

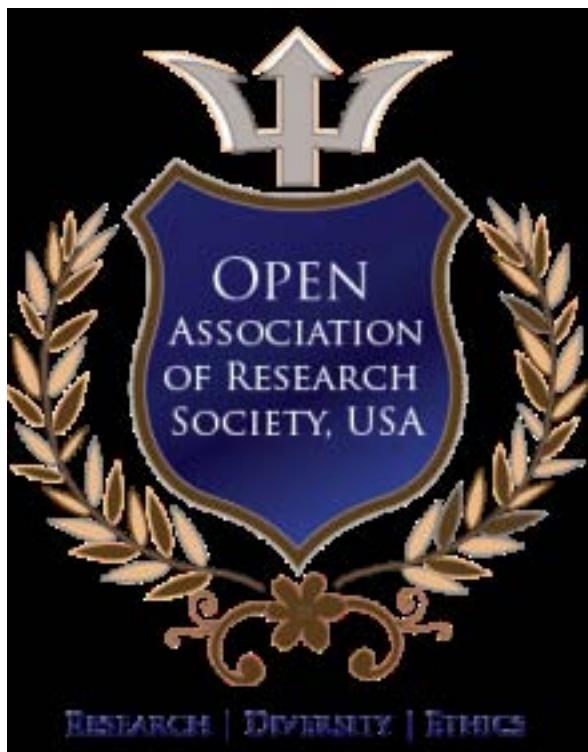


Figure 1:

21

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

mismatches different

Figure 2: Table 2 . 1 :

3

Figure 3: Table 3 .

32

Microcontroller	ATmega328
Operating Voltage	5V
Input Voltage (recommended)	7-12V
Input Voltage (limits)	6-20V
Digital I/O pins	14 (of which provide 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

Figure 4: Table 3 . 2 :

296 [Richard et al. ()] *An Introduction to speech and speaker recognition*, D Richard , Daryl H Peacocke , Graf . 1990.

297 [Li et al. ()] ‘AVAS: An Audio-Visual Attendance System. Department of computer Science and Technology’.

298 Dongdong Li , Yingchun Yang , Zhenyu Shan , Gang Pan , Zhaohui Wu . *LCNS 4261*, 2006. 2006. Springer-

299 Verlag Berlin Hidelberg. p. .

300 [Per-Erik Danielsson ()] *Euclidean Distance Mapping*, Per-Erik Danielsson . 1980. Linkoping; Sweden. p. .

301 Department of Electrical Engineering, Linkoping University

302 [Duglas and Reynolds ()] *Overview of Automatic Speaker Recognition*, A Duglas , Reynolds . 2008. 2008.

303 Workshop Summer School M. I.T. Lincoln Laboratory

304 [Campbell ()] ‘Speaker Recognition: A Tutorial’. Joseph P Campbell . *Proceedings of the IEEE* 1997. 85 (9) .

305 [What is the difference between speech Recognition and speaker recognition] *What is the difference between*

306 *speech Recognition and speaker recognition*, <https://-answers.yahoo.com/question/index?qid=20080620080053AA02UHQ7.Funwithword.RETRIVEDFROMHTTP://-RINKWORKS.COM/WORDS/PANGRAMS.SHTML>

307

308