DEVELOPMENT OF ACADEMIC ATTENDANCE SYSTEM USING VOICE VERIFICATION

Nor Azlinah Md Lazam[1*], Azilah Saparon[2]

[1]University Malaysia of Computer Science and Engineering,Cyberjaya 63000 Selangor, Malaysia

[2]Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Email: norazlinah@unimy.edu.my[1*] (corresponding author), azilah574@uitm.edu.my[2]

DOI: https://doi.org/10.22452/mjcs.sp2021no1.10

ABSTRACT

Based on the fact that all Human Voices are different, where every individual's voice contains unique characteristics that can be distinguished from others by using special analysis, voice verification can also be used for biometric recognition of individual. The aim of this project is to develop an automated attendance process captivating and student databases record maintaining using C++ programming and voice verification technique. The Input speech will undergo a series of voice recognition processes. The process starts with Feature extraction to obtain the attribute of voice. The features or specific characteristics for each voice are extracted by linear predictive coding technique (LPC) into feature vectors which later be used in speaker modeling. In speaker modeling stage the feature vectors of each speaker will be then processed by Gaussian Mixture Modeling (GMM) technique to generate speaker models and will be stored in a database. Verification process is done through template matching technique. Based on log-likelihood logic decision (template matching technique) with respect to the FAR (False Acceptance Rate) and FRR (False Rejection Rate), the identity of that speaker is accepted if the match is above the threshold. Once the verification result is accepted, the attendance in MySQL database for that speaker will be auto-updated.

Keywords: Speech analysis, feature extraction, speaker modeling, voice recognition, attendance database

1.0 INTRODUCTION

Taking attendance of students and maitaining the records manually are a major task. The project developed an automated attendance monitoring system to replace the manual attendance system.

This project adopts the idea of Development of Academic Attendance Monitoring System Using Fingerprint Identification [1][2]. Fingerprint Identification System that has been developed has a certain limitation especially to a handicapped person with amputated finger due to certain diseases and fingers (thumb) injuries. Instead, the idea of attendance captivating systems using facial recognition techniques was developed however it requires expensive gadget and yet still not getting the required accuracy [3-6]

The ultimate goal of this project is to provide proof of concept solution of auto-updating student attendance list in student database using his/her owns voice. The development of Academic Attendance using Voice Recognition System is an alternative solution and a complementary approach to other biometric system.

In general, voice recognition can be clasified as in figure 1.

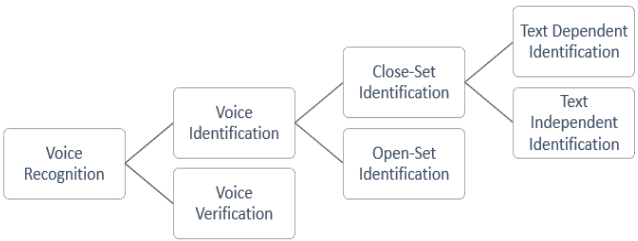


Fig. 1: Voice Recognition Clasification

Voice verification is mainly used for security purposes or individual authentication. Verification refer to a process in identifying if a query biometric sample that is claimed belongs to the owner is identical or not. The recognition is against his/hers own voice (1:1 matching) that has been recorded earlier and registered in the database [7-8].

On the other hand, Voice identification is a process to recognize the person among registered voices (1: N matching). Identification on the other hands is to find if the query biometric sample is correspondant to those the samples that have undergone the registration process and stored in the database.

Closed-set identification is a process in identifying the speaker who is known to be a member of the set of N enrolled speakers. While, Open-set identification is to identify a speaker who is not known to be a member of the set of N enrolled speakers. Text-dependent identification- requires the speaker to speak of key words or password of the same text being used during both training and recognition process.

Text -independent identification is perfomed with no specific passwords. It does not rely on the same text being spoken during training and verification process.

Voice recognition is the ability of a machine or program to receive and interpret dictation or to understand and carry out spoken commands [9-10]. The project developed is using these approach;

- Voice recognition
- Voice identification
- Close set identification
- Text dependent identification

2.0 VOICE RECOGNITION

Generally, all speaker recognition systems contain four main modules: Speech Analysis, Feature Extraction, Speaker Model and Decision (Template Matching) [11].In this project voice recognition is working in 3 stages: Feature Extraction, Speaker Modeling and Verification.

Feature extraction is used to obtain the attribute of voice. The features or specific characteristics for each voice are extracted by linear predictive coding technique (LPC) into feature vectors which later be used in speaker modeling. In speaker modeling stage the feature vectors of each speaker will be then processed by Gaussian Mixture Modeling (GMM) technique to generate speaker models and will be stored in a database [12] Verification process is done through template matching technique. Based on log-likelihood logic decision, the identity of that speaker is accepted if the match is above a threshold. Once the verification result is accepted, the attendance in database for that speaker will be updated [13]

2.1 Speech Analysis Technique

The speech analysis works with selecting an appropriate frame size and shift to segmentize the speech signal that later be analyzed in feature extraction process . The speech analysis processes include DC Offset Removal, Silence Removal, Pre-emphasis, Frame & Windowing and Auto-Correlation [14]

This project utilizes freely available audio digital editor software; Audacity 1.2.6 to perform the first 3 processes in speech analysis; DC Offset Removal, Silence Removal & Pre-Emphasis.

2.1.1 Frame and Windowing

Short -time spectral analysis is the commonly used technique in analysing speaker identification due to periodic and quasi-stationary for small interval frames.

In making the speech signal amendable for analysis, speech signal is divided into small fixed-length frame, where each frame is considered as a stand-alone signal. The signal are slow moving and amendable in small frame.

To smoothen the signal, a windowing function is applied. In this project Hamming Window techniques is used. Output from the windowed signal ,is a coeeficient of the most dominant features can be obtained. The frame can be cnsidered as quasi-stationary over a short period of time usually around 10-30ms.

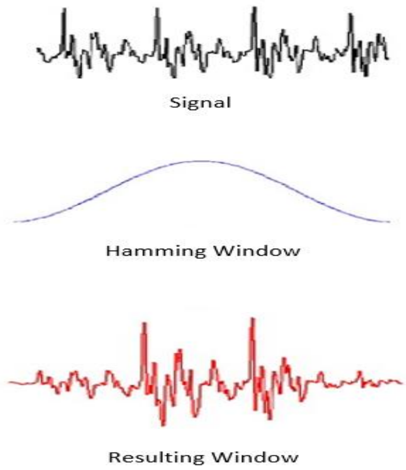


Fig. 2: Hamming Window Technique

2.1.2 Autocorrelation

A signal matches its own time-shifted version can be measured by using Autocorrelation techniques. In auto correlation signals, there are two important information;

- The periodicity of a periodic signal is preserved in autocorrelation.
- At all lags, a random signal is zero except for a lag zero in autocorrelation.

The ability of feature extraction methods can be improved, and the amount of noise can be reduced when frames are taken over by autocorrelation.

2.2 Feature Extraction Technique

Parameterization of the speech signal through short-time analysis can be performed using feature extraction. To produce speaker-specific feature vectors, feature extraction technique is applied to frame speech.

The acoustic parameters for each frame are extracted by linear predictive coding (LPC) technique. Fundamental frequency (F0) and harmonic structure are among acoustic parameters that show voice quality [15-17]

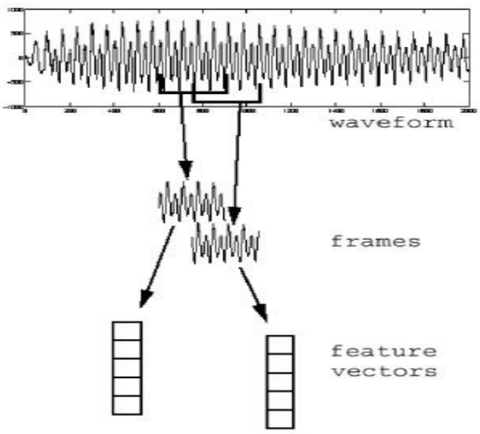


Fig. 3: Feature Vector

In an attempt to predict the speech samples such as linear combination of past samples, weights that contributes to best prediction, LP Coefficients (LPC) is formed. Generally, the past speech samples are analyzed using LPC to predict a given speech sample. The prediction signal s(n) can be represented as in equation 1. *P* is the number of coefficients, *a* is the predictor coefficients, *k* is the prediction of the nth value of the speech signal and *n* is the number of sample. To produce feature vectors, LPC are calculated over each speech frame.

$$\hat{s}(n) = a_1 s(n-1) + a_2 s(n-2) + a_p s(n-p) = \sum_{k=1}^{p} a_k s(n-k) \quad (1)$$

2.3 Speaker Modelling Technique

Speaker Modeling Technique is a classification process to identify the unknown speaker. The procedure is done by evaluating the extracted features from the original voice with the features that already registered in the set of known speakers.

The next process is to classify the speaker in order to identify the original author of the input speech signal. The identification process starts with enrolment of the speaker voice into the system using a modelling process. A matching technique is then used to identify the speaker. This is obtained once the speaker model have been created. The characterization of speaker specific patterns from the given speech requires training in speaker modelling.

In classification techniques, all individual's voice model will be created using speaker modeling technique. The construction of a speaker model is based upon the features extracted from their speech signal. The process is completed once the enrollment of the speaker model in speaker identification system database is registered [18].

2.3.1 Autocorrelation

There a several techniques to model speaker-specific features such as Vector Quantization (VQ) and Gaussian Mixture Modelling (GMM). This project used GMM (a statistical method) for its better estimating of data. (GMM) also known for its rapid response and simplicity in implementation. Furthermore, it is recommended for fast computing in training and identification process. A key advantage of GMM is a statistical method used to model speaker-specific features, which aims to provide better estimates of data. A GMM technique is said to have more accuracy in constructing the model speaker since the cluster are able to overlap. The speaker model was created using GMM with 16 mixtures. In addition, GMM Technique is also robust and able to form smooth estimation [19-20]

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation,

$$p(x|\lambda) = \sum_{i=1}^{M} \omega_i\, g(x|\mu_i \Sigma_i) \quad (2)$$

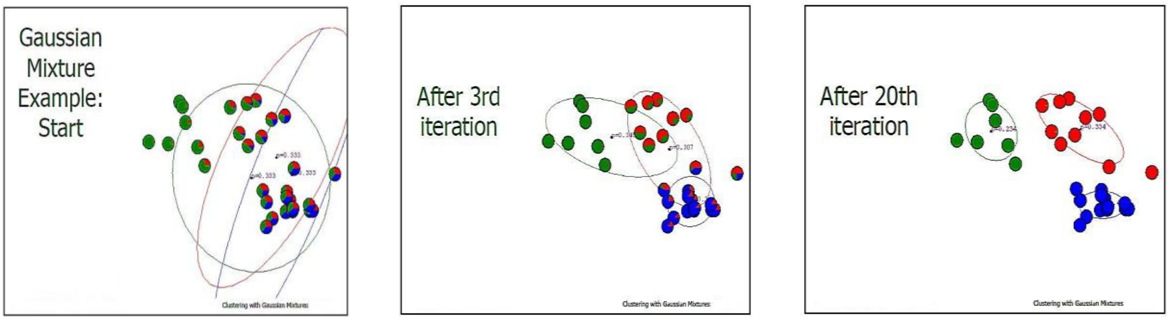


Fig. 4: GMM Iteration

2.4 Decision Technique Using Log Likelihood Ratio

The models in the speaker database are used to estimate the matching process in pattern matching. The similarity between each model in database with the features extracted from unknown speaker will influence the decision.

In this project, the system starts with identifying the speaker from the closed-set speaker identification system. Then, the system verifies the distance of this speaker with a threshold to come up with a decision. The system will generate an error message if the decision system fails to accept the speaker

2.5 Database Connectivity Module

Open Database Connectivity (ODBC) connection is used in this project. ODBC is a standard communication system for database access due to MySQL installation requirement. The module allows the connectivity between the verification results in C++ command line to the attendant list in MySQL Student Database. The connection is using DB Connector with ODBC Driver. The module is updated when Verification show Acceptance result, where the score is above threshold value. The module verifies speaker's authorization by computing match score between the unknown speaker's feature vector (sample voice) and the known speaker models (reference voice). The timestamp is printed every time the attendance list table is updated.

In general, the process on voice recognition can be visualized as in Figure 5.

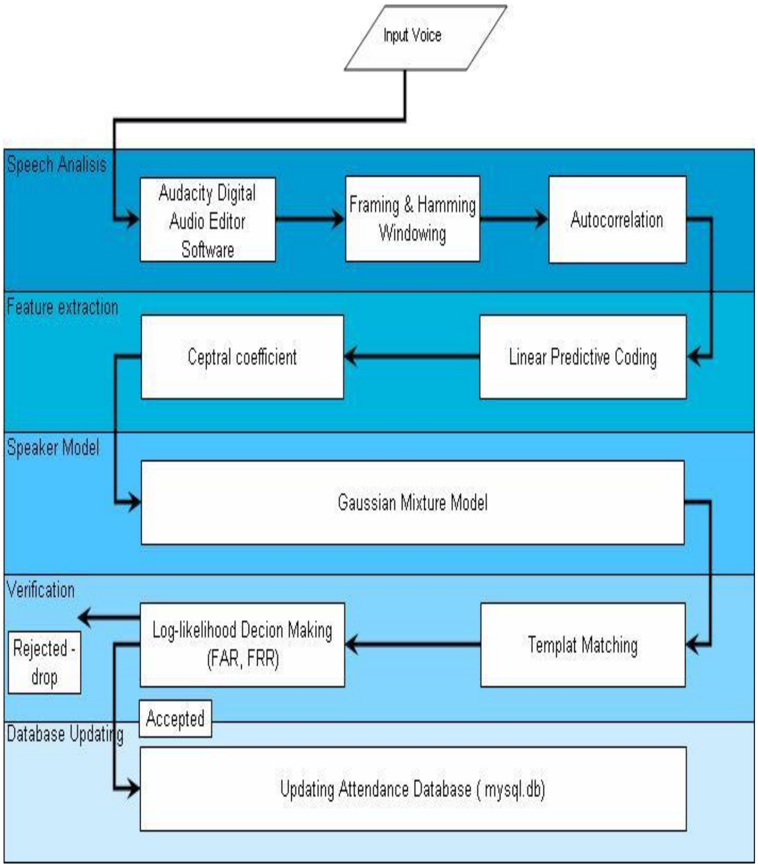


Fig. 5: Process of Voice Recognition

3.0 METHOD

Based on Voice Recognition Methodology in figure 5, proposed solution for this project involve 3 main Module namely Enrollment Module (Training Process), Verification Module (Testing Process) and Database Connectivity Module.

3.1 Enrollment Module

The module allows the user to record voice and makes a feature model of that voice. The module shall do end point detection on input wave to really separated speakers speech from and non-speech. The module shall translate the input voice record into wave form frequency. The module shall do Linear Predictive Coding (LPC) to determine the formants from the speech signal, estimate between the predictive signal and actual signal and represent the result as feature vector. The module shall train the speaker's feature vector using Gaussian Mixture Model (GMM) to generate density model of the corresponding speaker and attempt to isolate and identify the speaker.

3.3 Verification Module

The module shall verify speaker's authorization by computing match score between the unknown speaker's feature vector and the known speaker models (trained voice). The system shall be able to evaluate each density model to produce log-likelihood score for False Acceptance Rate (FAR) and False Rejection Rate (FRR). FAR occur when the value of the matched sample is above threshold value.

3.4 Database Connectivity Module

The module shall allow the connectivity between the verification result in C++ command line to the attendant list in MySQL Student Database. The connection is using DB Connector with ODBC Driver. The module shall be updated when Verification show Acceptance result, where the score is above threshold value. The module verify the speaker's authorization by computing match score between the unknown speaker's feature vector and the known speaker models (trained voice). The module shall print the timestamp.

4.0 IMPLEMENTATION

The deployment of this project is described in figure 6

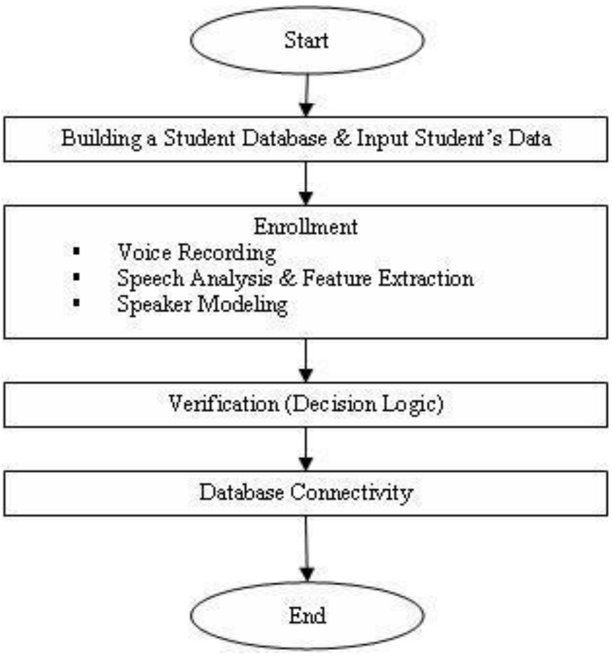


Fig. 6: Event Chronology

4.1 Student Database

Student database is created using open platform database, MySQL as shown in Fig 7. In Student Database, 3 tables were created: Attlist table (Attendance list), Info table and Result Table. Attlist table is the main table where the attendance will be updated in this table once the verification result is accepted. Student Info table contains the information of student: Student ID, Student Name, Sex, Address, Student IC and Contact No. All process in voice recognition starting from Enrollment toward the Verification will be referring to the Student ID.

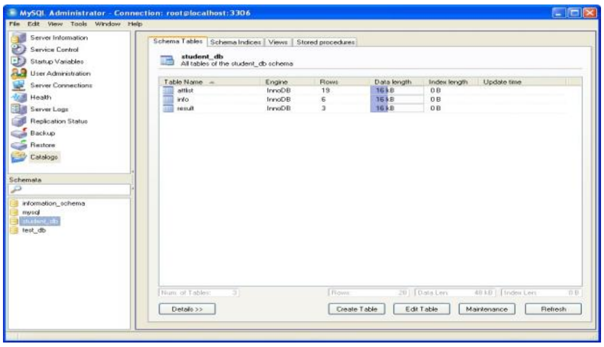


Fig. 7: Student Database

In Result table, result for each subject can be entered for record purposes.

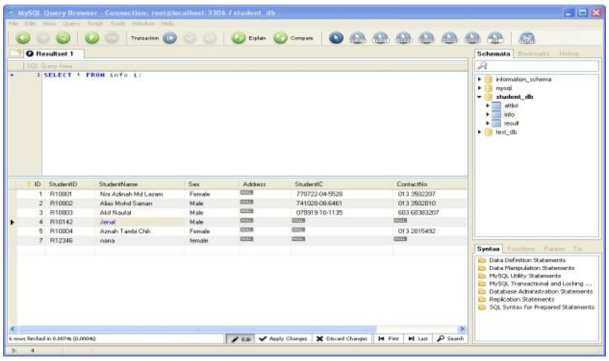


Fig. 8: Info Table

4.2 Enrolment Module

4.2.1 Voice Recording

Enrollment Process starts with voice recording. User must input their student ID in Student ID Box. Refer to figure 9. After clicking the Record button, Student must speak the pre-determined word “SATU”.

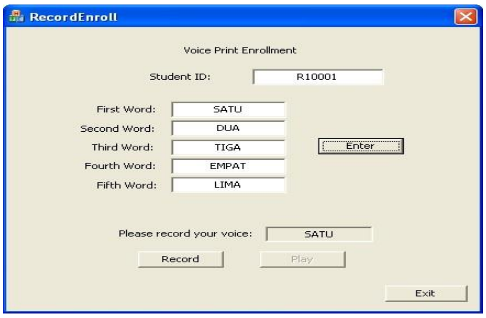


Fig. 9: Record-Enroll

The processes continue with saving the recorded voice in .wav format into System Folder. Next continue the record until the fifth words is saved.

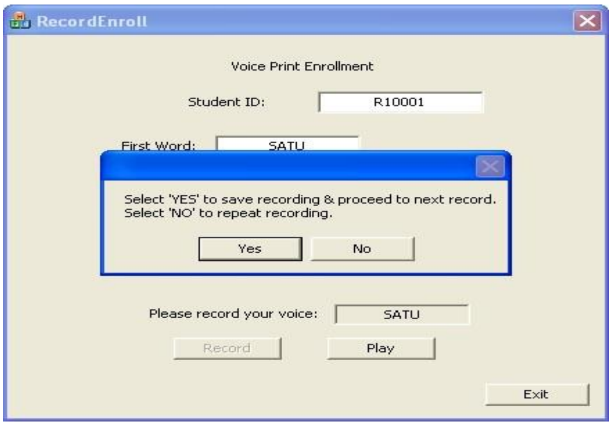


Fig. 10: Saving Recorded Voice

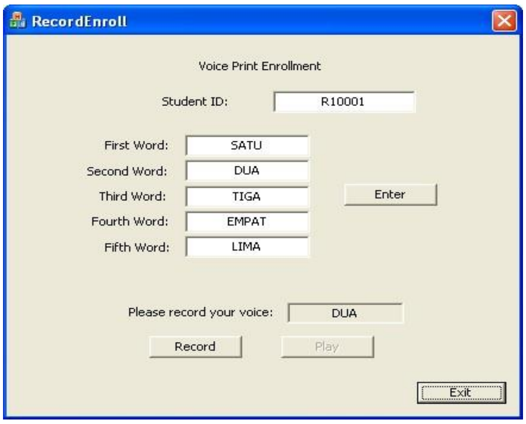


Fig. 11: Record 2nd – 5th Word

The sample voices for this Enrollment Process will be taken 5 times. The file of .wav for all samples will be stored as R10001_T01_W01.wav and so on.

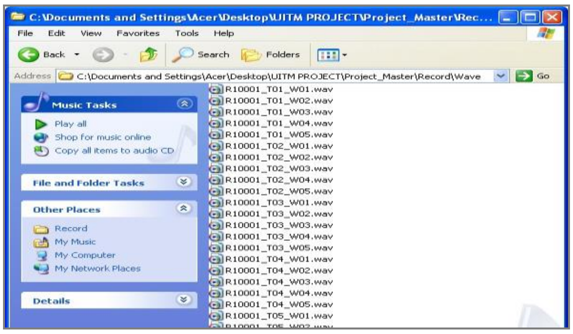


Fig. 12: Voice Sample in *.wav* format

4.2.2 Speech Analysis Feature Extraction

This sample will undergo the speech analysis & feature extraction process that will change the file to. vec in System Folder.

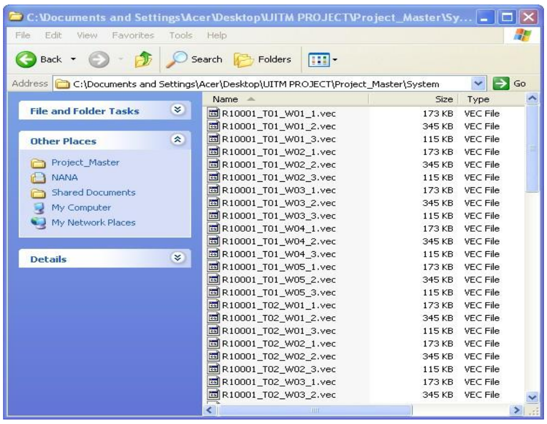


Fig. 13: Voice Sample in *.vec* format

4.2.3 Speaker Modelling

In speaker modeling, the sample voice will later be changed into *.mod* file, with 1 file *.thr* that contains the threshold value for that speaker. This generates the Speaker Model and will be stored in the System folder.

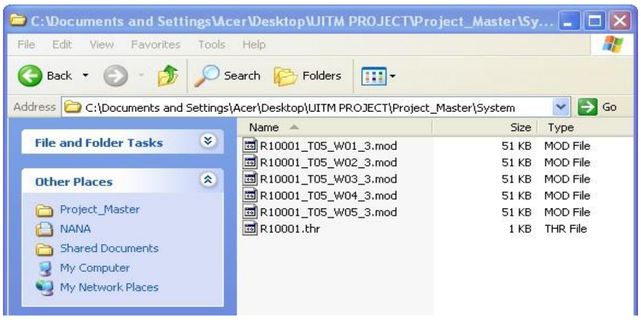


Fig. 14: Voice Sample in *.mod & .thr* format

4.3 Verification Process

In this Stage, the input voice will repeat the process of voice recording but this time the sample is stored in Auth Folder.

By using the Log-likelihood decision logic; the output of this sample voice will be compared with speaker model (reference voice data) in the System Folder. With respect to FAR (False Acceptance Rate) & FRR (False Rejection Rate), if the sample voice data score above the threshold value of the referenced voice data (trained data), the user is verified, and the attendance database will be time-stamped and updated accordingly.

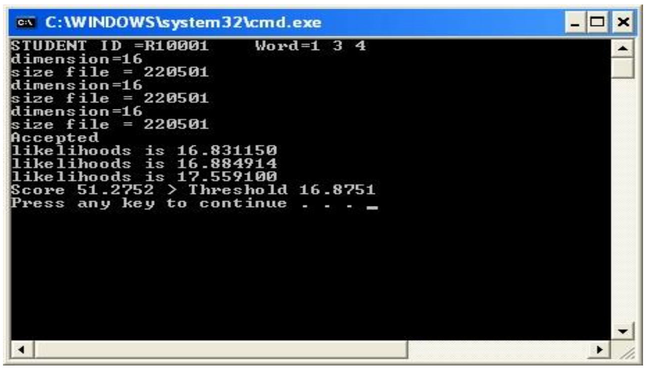
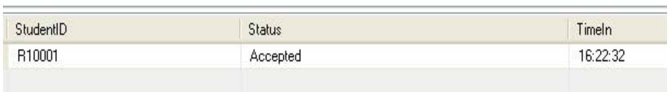


Fig. 15: Verification - Accepted

4.4 Updating Database

Once the Speaker is verified, then attendance list for that speaker will be updated automatically.



StudentID	Status	TimeIn
R10001	Accepted	16:22:32

Fig. 16: Attendance Status is updated with time-stamp

5.0 RESULT AND DISCUSSION

Total of 20 voice samples are executed to create 20 speaker models. Pre-recorded voices during enrollment session are stored in the system folder. Verification process will perform template matching technique between the sample speakers with the speaker model. All accepted result will be linked to the attendance database. A series of testing have been done to prove the program's accuracy. 6 test cases with different scenario are performed and all results are captured. Six test cases are as follows;

Test Case 1 : Verification against owns voice
Test Case 2 : Verification against same gender
Test Case 3 : Verification against opposite gender
Test Case 4 : Verification using own voice with variation during enrollment against same gender
Test Case 5 : Threshold value on vowels and consonants
Test Case 6 : Verification on consonants 'B' over 'P'

Test Case 1- 4 were recorded by using the same word during enrollment session that is "LIMA". The tests are carried out in the same environment however the level of noise is not measured. The type of microphone used is head phone speaker and the entire samples are taken with no variations in speaker voice. All sample voices undergone the freely available software; audacity version 1.2.6 before being modeled during the enrollment session. All verifications that are accepted bythe program will automatically be updated in the attendance list in student database. The program will print the timestamp once the verification is accepted.

5.1 Test Case 1 – Verification against owns voice

Objective of test case 1 is to test the program accuracy in identifying the spoken words and verifying the speaker model against owns voice (owns speaker model). Result is shown in table 1. Test done to 5 speakers' shows 100% accuracy. The project is capable of verifying the speaker against owns speaker model. By observation, male speaker has lower threshold value than female speaker.

Tabel 1 : Result for Test Case 1

Test	Test Scenario	Result
1	Speaker Model : R10001 Sample Speaker : R10001	Accepted Threshold Value : 16.8751
2	Speaker Model : R10002 Sample Speaker : R10002	Accepted Threshold Value : 17.1709

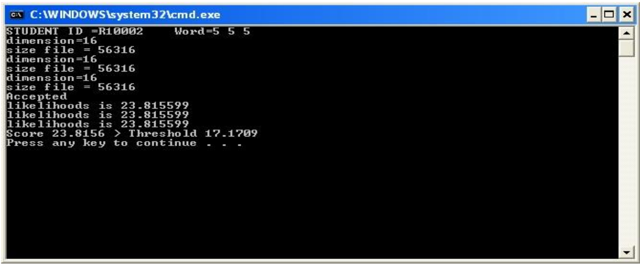


Fig. 17: Result using Speaker Model R10002 Vs Sample Speaker R10002

5.2 Test Case 2 – Verification against same gender

To test the program accuracy in identifying the spoken words and verifying the speaker model against same gender . Result is shaown in table 2. Test conducted to 2 female speakers' shows 100% accuracy. The project is capable of identifying the speaker against owns speaker model where same sample speaker with speaker model will be accepted while different sample speaker with speaker model will be rejected by the program.

Tabel 2 : Result for Test Case 2

Test Info	Result
Speaker Model (Female) : R10002 Sample Speaker (female) : R10017	Rejected Threshold Value : 16.5628 < 17.1709

Fig. 18: Speaker Model R10002 Vs Sample Speaker R100017

5.3 Test Case 3 – Verification against opposite gender

The objective of test case 3 is to test the program accuracy in identifying the spoken words and verifying the speaker model against different gender. Result is shown in table 3. Test done to match male and female speaker shows 100% accuracy. The system is capable in recognizing the speaker even for opposite gender. By observation, male speaker has lower threshold value than female speaker.

Tabel 3 : Result for Test Case 3

Test Info	Result
Speaker Model (female) : R10002 Sample Model (male): R10142	Rejected Threshold Value : - 47.0624 < 17.1709

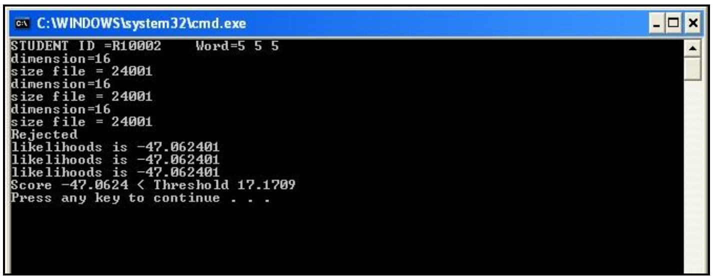


Fig. 19: Speaker Model R10002 Vs Sample Speaker R100142

5.4 Test Case 4 – Verification using own voice with variation during enrollment

The objective of test case 4 is to test the program accuracy in identifying the spoken words and verifying the sample speaker against speaker model with variation during enrollment (if speaker is sick or change in voice; husky, flu). Test done shows 100% accuracy. The extreme variation in speaker's model voice during training or sampling session and deployment session will directly affects the performance of the system. Variation in voice will give lower thresh hold value, so higher chances in accepting other sample speaker. In this case both samples from the same gender are accepted.

Tabel 4 : Result for Test Case 4

Test Info	Result
Speaker Model : R10001 Sample Speaker : R10020	Accepted Threshold Value : 16.8999 > 16.8751

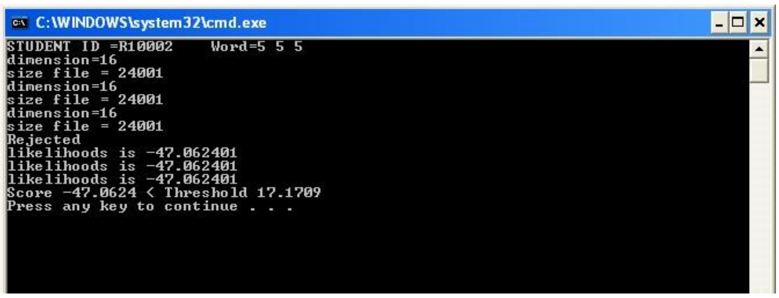


Fig. 20: Speaker Model R10001 Vs Sample Speaker R100020

5.5 Test Case 5 – Threshold value on vowels and consonant

The objective of test case 5 is to test the program accuracy in identifying the threshold values of spoken words between vowel sound letter and consonant letter. Test case 5 explains that vowel and consonant do affect the performance of the project. Sound of vowel ‘A’ and ‘U’ are having higher threshold value The consonant word ‘P’ & ‘B’ with lower threshold values get higher chances in accepting other sound-a-like sample speaker. This suggests the usages of word containing more sounds of vowel especially ‘A’ are recommended. Test case 5 helps to determine the best spoken word to be chosen during the enrollment session

Tabel 5 : Result for Test Case 5

Test Info	Result
Sound of Vowel A Speaker Model : R10005 Sample Speaker : R10005	Threshold Value : 17.3788
Sound of Vowel U Speaker Model : R10006 Sample Speaker : R10006	Threshold Value : 17.3451
Sound of Consonant P Speaker Model : R10007 Sample Speaker : R10007	Threshold Value : 16.9722
Sound of Consonant B Speaker Model : R10008 Sample Speaker : R10008	Threshold Value : 16.2346

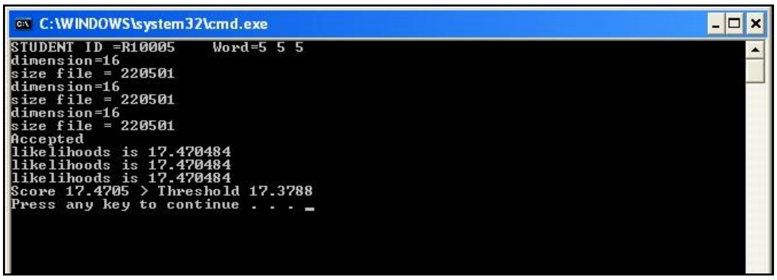


Fig. 21: Sound of Vowel A

5.6 Test Case 6 – Verification against consonant “B” over “P”

To test the program accuracy in identifying and verifying the sound-a-like consonants. Test 6 shows that the project has a high accuracy rate where the sound-a-like samples are tested and the project capable in differentiating them. In this case, consonant ‘P’ and ‘B’ are sound-a-like; however the word verification is rejected.

Tabel 6 : Result for Test Case 6

Test Info	Result
Speaker Model R10007 : Consonant ‘P’, Sample Speaker R10008 : Consonant ‘B’	Rejected Threshold Value : 16.2346 < 16.9722

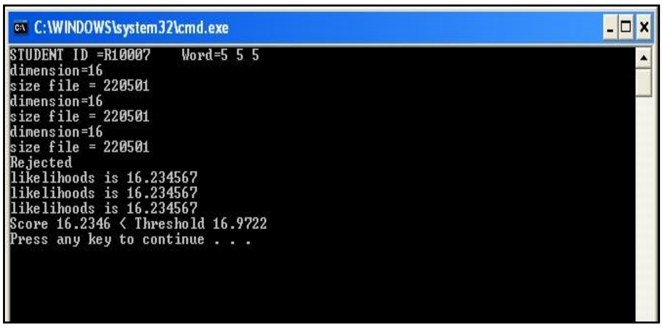


Fig. 22: Speaker model R10007 (consonant “P”) vs Sample Speaker R10008 (consonant “B”)

All accepted results will be linked to student database to auto-update the verified speaker attendance list as shown in figure 23.

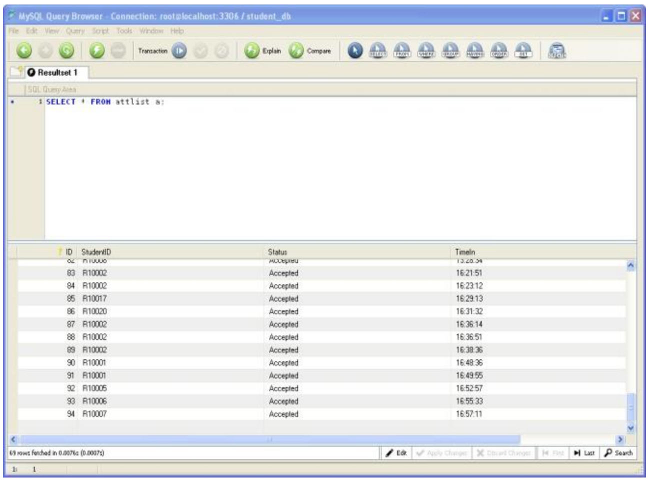


Fig. 23: Updated Staudent Attendance Database

6.0 CONCLUSION

Result from the series of test cases show that this project is proven for its high accuracy even for the sound-a-like consonant. Testing againts owns voice verification is 100% accurate. The project also suggested that the usage of sound with vowel especially ‘A’ is recommended in text-dependent verification (referring to Malay words)

The projects limitation includes noise 0f surrounding environment, type of microphone used and variation in voices will directly affects the performance of the program. Thus to prove this, it is recommended that more test to be done to further understand the program behavior such as:

- % of acceptance for person’s emotion
- % of acceptance for person’s health (husky, flu)
- % of acceptance for different microphone
- % of acceptance for different environment (noise)

Student database can be very informative and useful for lecturer as well as the registrar/admin staff. In this project, 3 tables have been created namely Attendance List Table, Student Information Table and Result Table. Since all tables in student database can be linked together, literally we'll have the idea whether the student performance are affected by student's attendance.

The findings of this project can be concluded that the main objective of this project that is to develop Smart Academic Attendance Database Using Voice Verification is achieved.

6.1 Future Recommendation

It can be further objective to develop a C++ program for 3 processes in speech analysis (DC Offset, Silent Removal and Pre-emphasis) that using freely available software as mention earlier in this document. This project is using text-dependent approach where the spoken word is determined earlier during the enrollment session and the same word will be using during verification. This will lead to some problem when speakers recorded their voice and use the pre-recorded to update the attendance while he/she is absent. Therefore it is recommended to develop the text-independent verification for non-specific word. Different word will be considered for verification accurateness.

It is yet another recommendation to update the student attendance list in remote server. To have a central server for student database will make an easy access to everybody.

REFERENCES

[1] Tabassam Nawaz, Saim Pervaiz, Arash Korrani, Azhar-Ud-Din, "Development of Academic Attendance Monitoring System Using Fingerprint Identification", IJCSNS International Journal of Computer Science and Network Security, 2009, Vol.9 No.5.

[2] Fahad-Bin-Mazhar, O. Ahamed and M. Rasedujjaman, "Biometric smart attendance kit with fingerprint scanner by using microcontroller," *2015 International Conference on Electrical & Electronic Engineering (ICEEE)*, 2015, pp. 13-16, doi: 10.1109/CEEE.2015.7428261.

[3] J. P. Jeong, M. Kim, Y. Lee and P. Lingga, "IAAS: IoT-Based Automatic Attendance System with Photo Face Recognition in Smart Campus," 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 363-366, doi: 10.1109/ICTC49870.2020.9289276.

[4] S. Bhattacharya, G. S. Nainala, P. Das and A. Routray, "Smart Attendance Monitoring System (SAMS): A Face Recognition Based Attendance System for Classroom Environment," *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 2018, pp. 358-360, doi: 10.1109/ICALT.2018.00090.

[5] U. Koppikar, S. Hiremath, A. Shiralkar, A. Rajoor and V. P. Baligar, "IoT based Smart Attendance Monitoring System using RFID," *2019 1st International Conference on Advances in Information Technology (ICAIT)*, 2019, pp. 193-197, doi: 10.1109/ICAIT47043.2019.8987434.

[6] M. S. Akbar, P. Sarker, A. T. Mansoor, A. M. Al Ashray and J. Uddin, "Face Recognition and RFID Verified Attendance System," *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, 2018, pp. 168-172, doi: 10.1109/iCCECOME.2018.8658705.

[7] R. A. Rashid, N. H. Mahalin, M. A. Sarijari and A. A. Abdul Aziz, "Security system using biometric technology: Design and implementation of Voice Recognition System (VRS)," *2008 International Conference on Computer and Communication Engineering*, 2008, pp. 898-902, doi: 10.1109/ICCCE.2008.4580735.

[8] N. M and A. S. Ponraj, "Speech Recognition with Gender Identification and Speaker Diarization," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298241.

[9] M. M. Toruk and R. Gokay, "Short Utterance Speaker Recognition Using Time-Delay Neural Network," *2019 16th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2019, pp. 383-386, doi: 10.1109/SSD.2019.8893188.

[10] B. Wang, J. Zhao, X. Peng and B. Li, "A Novel Speaker Clustering Algorithm in Speaker Recognition System," *2006 International Conference on Machine Learning and Cybernetics*, 2006, pp. 3298-3302, doi: 10.1109/ICMLC.2006.258463.

[11] S. A. Barlaskar, M. A. Laskar, N. Shome and R. H. Laskar, "Study on the varying degree of speaker identity information reflected across the different MFCCs," *2016 International Conference on Inventive Computation Technologies (ICICT)*, 2016, pp. 1-6, doi: 10.1109/INVENTIVE.2016.7823255.

[12] T. R. J. Kumari and H. S. Jayanna, "Comparison of LPCC and MFCC features and GMM and GMM-UBM modeling for limited data speaker verification," *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 2014, pp. 1-6, doi: 10.1109/ICCIC.2014.7238329.

[13] A. N. A. Kumar and S. A. Muthukumaraswamy, "Text dependent voice recognition system using MFCC and VQ for security applications," *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 2017, pp. 130-136, doi: 10.1109/ICECA.2017.8212779.

[14] R. Chakroun, L. Beltaïfa Zouari, M. Frikha and A. Ben Hamida, "Improving text-independent speaker recognition with GMM," *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2016, pp. 693-696, doi: 10.1109/ATSIP.2016.7523169.

[15] H. C. Bao and Z. C. Juan, "The research of speaker recognition based on GMM and SVM," *2012 International Conference on System Science and Engineering (ICSSE)*, 2012, pp. 373-375, doi: 10.1109/ICSSE.2012.6257210.

[16] D. Prabakaran and R. Shyamala, "A Review On Performance Of Voice Feature Extraction Techniques," *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*, 2019, pp. 221-231, doi: 10.1109/ICCCT2.2019.8824988.

[17] J. K. Hundal and S. T. Hamde, "Some feature extraction techniques for voice based authentication system," *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2017, pp. 419-421, doi: 10.1109/ICPCSI.2017.8392328.

[18] K. Wang, Z. Ren and X. Xiong, "Combination of Wavelet snd SIFT Features for Image Classification Using Trained Gaussion Mixture Model," *2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2008, pp. 79-82, doi: 10.1109/IIH-MSP.2008.76.

[19] X. Wu, M. Zhu, R. Wu and X. Zhu, "A Self-adapting GMM based Voice Activity Detection," *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, 2018, pp. 1-5, doi: 10.1109/ICDSP.2018.8631856.

[20] Danwen Peng, Xiongwei Zhang and Jian Sun, "Voice Conversion based on GMM and Artificial Neural Network," *2010 IEEE 12th International Conference on Communication Technology*, 2010, pp. 1121-1124, doi: 10.1109/ICCT.2010.5688637.

[21] P. Malathi, G. R. Sureshw and M. Moorthi, "Enhancement of electrolaryngeal speech using Frequency Auditory Masking and GMM based voice conversion," *2018 Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, 2018, pp. 1-4, doi: 10.1109/AEEICB.2018.8480968.