Speech-based Class Attendance

‘Umar Faizel Amri¹, Nik Nur Wahidah Nik Hashim² and Noor Hazrin Hany Mohamad Hanif³

¹,³ Department of Mechatronics Engineering, International Islamic University Malaysia

Abstract. In the department of engineering, students are required to fulfil at least 80 percent of class attendance. Conventional method requires student to sign his/her initial on the attendance sheet. However, this method is prone to cheating by having another student signing for their fellow classmate that is absent. We develop our hypothesis according to a verse in the Holy Qur’an (95:4), “We have created men in the best of mould”. Based on the verse, we believe each psychological characteristic of human being is unique and thus, their speech characteristic should be unique. In this paper we present the development of speech biometric-based attendance system. The system requires user’s voice to be installed in the system as trained data and it is saved in the system for registration of the user. The following voice of the user will be the test data in order to verify with the trained data stored in the system. The system uses PSD (Power Spectral Density) and Transition Parameter as the method for feature extraction of the voices. Euclidean and Mahalanobis distances are used in order to verified the user’s voice. For this research, ten subjects of five females and five males were chosen to be tested for the performance of the system. The system performance in term of recognition rate is found to be 60% correct identification of individuals.

1. Introduction
The attendance system is considered important and significant in every area. This is to make sure the participants attend the programmes related in order to achieve the KPI and objectives. In this study, the attendance system is projected in the area of class of lecture. As students, they need to attend the class because it is compulsory. Years ago, Kulliyyah of Engineering International Islamic University Malaysia (KOE IIUM) doesn’t enforce the obligation for engineering students to attend classes, provided obeying the following constraints: as long as the students scored the subjects of lectures, it matters. Paradoxically, not only the students chose to missed classes, but also don’t get the full quality of understandings regarding topics of subjects. Moreover, it is unethical for Muslim students to seek knowledge without having the way to meet the lecturers as the educators; the source of knowledge. This scenario showed how important the attendance system in the field of class of lectures.

The objectives of the research are to investigate the ability of speech features to recognize speaker as well as to analyse the effect of different types of speech in identifying unique individuals and to determine the factors affecting the distortion or increasing errors in speaker recognition. Based on these objectives, it then leads to the following questions, what is the best feature used in speaker recognition application? Is that possible to combine different features in a single set of data? These questions were expected to be answered throughout this research.

2. Previous works
In this section, other researcher’s papers were examined especially on the features that they used to extract the audio signal features and their results. According to [1], they used two types of features i.e. power spectral density (PSD) and also Mel-frequency Cepstral Coefficient (MFCC). For PSD, the results for male interview is 74.25% while female is 84.25% success rate. For the results for reading session, for male the score is 75.5% and female is 87%. For MFCC, they used 2-types which is 4-MFCC and 8-MFCC. For 4-MFCC, the success rate for interview for male is 70.3% and for female is 69.23%. For 8-MFCC, the success rate for male is 77.84% and female is 72.53%. For reading session, for 4-MFCC’s male, the score is 71.28% while female scored 67.82%. For 8-MFCC, the male scored 80.62% while female scored 73.3%. Here, the higher the MFCC used, the better the result.

As for the research conducted by [2], the methodology used is text-dependent or fixed-text category. This research is specifically use spectrograms and row mean. Features used to identify the speaker is depended on optimal spectrogram segmentations and Euclidean distance. Spectrogram is used to obtain the feature vectors while row mean is applied on the spectrograms to get feature vector 8, 16 or 32. Then, Euclidean distance is measured to identify the speaker. The result from this research is astonishing that the minimum success rate is 50% while the maximum hits 100% perfectly. Unfortunately, not all researchers get good results. For [3], using MFCC as method for speaker recognition is not the best in noisy environment. Not only MFCC, Cepstral Mean and Variance Normalization (CMVN), Relative Spectral (RS) and Feature Warping also being used. Voice Activity Detection (VAD) is also used to improve MFCC method. However, they managed to improves the correctness rate by increasing to 25.2%, 19.1% and 9.7% for every method respectively. Research by [4] also used MFCC for text-independent speaker recognition based on GMM-BUM. Accuracy of the speaker recognition in increased by 3.10%. The score-level fusion Teager Energy Operator (TEO) phase together with MFCC performs better compare to MFCC alone.

3. Methodology

In this section, several procedures are being made in order to achieved the objective of this study. The model used for this methods is the same as what Desai and Pujara [5]. Figure 3.1 shows the basic model for speaker recognition.

![Figure 3.1. Basic model for speaker recognition.](image)

In this research for the early stage, five females and five males’ students were chosen to test the system. The database that were collected are from students saying their names for five times. Four data per individual will be used as training data and one left for testing data. The pre-processing method was done using Audacity software. The feature extraction used for this research are PSD and Transition Parameter. The classification method used are the Euclidean and Mahalanobis distance equations.

3.1. Feature extraction

There are two method used for feature extraction i.e. PSD and Transition Parameter. Both of the features use are including in the system. The display of the system using MATLAB GUI function in order for users to choose which method of feature they prefer for.

3.1.1. Power spectral Density (PSD). Power Spectral Density (PSD) is also known as energy in frequency bands (2). PSD is used to characterized a stationary random process in frequency domain. The PSD used is to be limited from 0 Hz to 2 KHz because most of energy in speech signals is mostly present in this range. To minimize the number of feature extraction, the frequency of which 0 Hz to 2
KHz is being divided into four band equally. In band one, there are only 0 Hz to 500 Hz frequency. For band two, the frequency is from 500 Hz to 1000 Hz. For band three and band four the frequencies are 1000 Hz to 1500 Hz and 1500 Hz to 2000 Hz respectively. Even though there are now four PSD (which regarding to the four band), only the first three band is used in the feature extraction process (band one, band two and band three) for the fourth band can be predicted by the first three band. Figure 3.2 shows the PSD procedures.

3.1.2. Transition Parameter. Audio signal that consists voiced, unvoiced and silence was marked as 1, 2 and 3. The variation happens during the timing pattern of the speech. Here, the variations are captured in the form of transition from one state to another state. Interestingly, these states are interchangeable and may also not change; depending on the set of probabilities concerning the states. The probability is simply estimated using discrete-time Markov process [6]. According to Nik Hashim [6], “One of the output parameters given is the estimated transition matrix (T) which in this case is a three-by-three matrix where $t_{ij} = Pr (X_{k+1} = j \mid X_k = i)$ for $I = 1,2,3$ and $j = 1,2,3$”, p. 48. Here, for example, element $t_{23}$ is denoted as the conditional probability of unvoiced-to-silent. The transition matrix, $T$ elements were sequenced into one row vector $\{t_{11}, t_{12}, t_{13}, t_{21}, t_{22}, t_{23}, t_{31}, t_{32}, t_{33}\}$.

3.2. Speaker recognition
This is the final part of the speaker recognition method. Which is analysis of the features that were extracted. This project used one method in order to get the authentication of the speaker called Euclidean distance. Euclidean distance is also known as Euclidean metric is a popular method used to get the distance between two connecting points in a straight line in Euclidean space. The computation method used there is depending on the dimensions. Euclidean distance, $d_e$ is given in the equation,

$$d_e = ||x - \mu_i|| \tag{1}$$

where $x$ and $\mu_i$ are both present in the form of matrix (that denoted in bold forms). Here, $x$ is the test trained data and $\mu_i$ is the tested data. From (1), we know that feature vectors are assigned to classes which respected to their Euclidean distance from the mean points [8].

Then, Mahalanobis distance, $d_m$ equation is used as the continuation of this Euclidean distance given,

$$d_m = \sqrt{(x - \mu_i)^T \cdot (x - \mu_i)} \tag{2}$$

where (2) is defined as the square root of the transpose, $T$ of $(x - \mu_i)$ multiply by the same matrix $(x - \mu_i)$.
By using (2), we can now calculate the Euclidean distance of the tested and trained data. The idea is that, the least the distance between test and train data, the accurate it is; which indicates the authentication the speaker recognition.

To do this, train data from each subject must be calculated together using (2) with the test data from each subject. Here, we know that there are ten subjects. Which means, there are 10 x 10 possible combinations in order to analyse the data. Because this research used three PSDs, all possible combinations of three PSD will be used for testing and training data.

However, for transition parameter feature extraction, the possible combinations are Voiced-Voiced (t_{11}), Voiced-Unvoiced (t_{12}), Voiced-Silent (t_{13}), Unvoiced-Voiced (t_{21}), Unvoiced-Unvoiced (t_{22}), Unvoiced-Silent (t_{23}), Silent-Voiced (t_{31}), Silent-Unvoiced (t_{32}) and Silent-Silent (t_{33}).

3.3. Development of GUI-based Speech Attendance System

We develop a graphical user interface (GUI) based system for the extraction and analysis of features for speaker identification using MATLAB. User will be able to speak into the system and his/her name will be registered as present for that particular day. The flowchart shown in Figure 3.3 illustrates the system operated of this speech-based class attendance and Figure 3.4 shows displayed the GUI for this system.

![Figure 3.3. General flowchart of speech-based class attendance.](image-url)
4. Result and discussion
In this chapter, the results of the analysis are presented. Discussion on each segment of results is made especially upon the errors. The sets of data that are going to be presented in this chapter are the results for free-text category of recitation of own name. By verifying PSD method, the best result of minimum Euclidean distance is for PSD2 and PSD3 and the \( t_{11}, t_{13}, t_{31} \) and \( t_{33} \) for minimum Euclidean distance of 10 subjects verifying their voices by reciting their own names. Figure 4.1 shows the PSD result and Figure 4.2 shows the transition parameter result.

The result demonstrated 60 percent or less accurate identification. The low percentage of accuracy is mainly due to the component of speaker variability and speech features as mentioned by [7]. It is to be recommended that the first step of speaker recognition i.e. collection of data is being held in a more reliable and sensitive acts. For example, the session is held in a very closed room that free from acoustical noise. Also, the microphone used must be more quality compare to this research did. Not only that, the prime factor i.e. the speaker or subject itself must be remind to speak in a natural way. The condition of subject’s health also must be considered to get a better performance. With this, in shaa Allah the results will be much better. Not only by that, the subjects also must be increase quantitatively. The more subjects, the more data, the better the training.

5. Conclusion
In conclusion, the project went well although the results are not satisfying. This is the part and parcel of research. In addition, this topic is a long and continuous research which means, a set of time is really
needed to make this project a success. The PSD features and transition parameters features were not effective in identifying the speaker. The effect of different speech in identifying the speaker affected the performance of speaker recognition. Compared to others’ related work as mentioned in chapter 2, the result can be improved by using combinations of PSD and transition parameters and other features such as MFCC with different techniques of identification.

6. References
[1] Keskinpala HK, Yingthawornsuk T, Wilkes DM, Shiavi RG, Salomon RM. Screening for high risk suicidal states using mel-cepstral coefficients and energy in frequency bands. Eur Signal Process Conf. 2007;(Eusipco):2229–33.
[2] Kekre HB, Athawale A, Desai M. Speaker Identification using Row Mean Vector of Spectrogram. 2011;(Icwet):171–4.
[3] Wu D, Feng Z, Chen L, Xiao Z. Perception Auditory Factor For Speaker Recognition In Noisy Environment. 2016;1916–20.
[4] Agrawal P, Patil HA. Fusion of TEO Phase with MFCC Features for Speaker Verification. 2015;161–6.
[5] Desai KS, Pujara H. Speaker Recognition from the Mimicked Speech: A Review. 2016;2254–8.
[6] Nik Hashim NNW. Analysis of Speech Features as Potential Indicators for Depression and High Risk Suicide and Possible Predictors for The Hamilton Depression Rating (HAMD) And Beck Depression Inventory Scale (BDI-II). Vol. 53. Vanderbilt University; 2014.
[7] Singh N, Khan RA. Underlying Text Independent Speaker Recognition. 2016;6–10.
[8] Theodoridis S and Koutroumbas K 2009 Pattern Recognition (London: Academic Press).

Acknowledgments
This work is part of the ongoing project on the development of speech-based multi-level person verification system funded by the Department of Mechatronics Engineering, International Islamic University Malaysia.
Corrigendum: Speech-based Class Attendance

IOP Conf. Ser.: Mater. Sci. Eng. 260 (2017) 012008

Muhamad Fikri Zanil¹, Nik Nur Wahidah Nik Hashim² and Huda Azam³
¹ Department of Mechatronics Engineering, International Islamic University Malaysia
² Department of Mechatronics Engineering, International Islamic University Malaysia
³ Department of Mechatronics Engineering, International Islamic University Malaysia

Page 7:
In the Acknowledgements section, the following text appears:

“This work is part of the ongoing project on the development of speech-based multi-level
person verification system funded by the Department of Mechatronics Engineering,
International Islamic University Malaysia.”

This should read:

“This work is part of the ongoing project on the development of speech-based multi-level
person verification system funded by the Department of Mechatronics Engineering,
International Islamic University Malaysia. This research is supported by the Research
Acculturation Grant Scheme (RAGS 14-040-0103) by the Ministry of Higher Education
Malaysia.”