**Journal of Theoretical and Applied Information Technology**
15th October 2021. Vol.99. No 19
© 2021 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# DYNAMIC TIME WARPING AND HIDDEN MARKOV MODEL CLASSIFIER IN TEXT DEPENDENT SPEAKER VERIFICATION SYSTEM WITH DIFFERENT SPEAKING VARIANT

**[1]KEVIN KURNIAWAN, [2]RAYMOND SAMUEL, [3]BENFANO SOEWITO**

[123]Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480

E-mail: [1]kevin.kurniawan007@binus.ac.id, [2]raymond.samuel@binus.ac.id, [3]bsoewito@binus.edu

## ABSTRACT

One of the main focus on IoT is how to make the facility more secure in order to increase the safety and convenience of the facility. While traditional authentication approach for verifying identity such as Personal Identification Number (PIN) or password is widely used, traditional authentication approach have major flaws that can make the verified user lost access to the protected resource or even pose a major security breach. Speaker verification is one of the many biometric system solutions to solve that problem. In speaker verification, a different speaking pattern and loudness of the voice may affect the performance for the system to verify an individual identity, which a change of speaker speaking pattern will occur very likely since a speaking pattern is highly affected by the speaker's mental and physical condition, and loudness of the voice is affected by how far the speaker away from the microphone. In this paper, we compare two well research text-dependent speaker verification methods, dynamic time warping and hidden markov model, on verifying user's identity on different voice variant (loud, normal, quiet, shout, and soft) to replicate the condition when the speaker is experiencing mental or physical conditions and the distance between the speaker and the microphone that affects the loudness of the voice and speaker speaking pattern. This paper uses 330 total train data and 1,600 total test voice data where every test voice data will be retested for every registered user. Research done in this paper shows hidden markov model achieved better accuracy on normal, shout, and soft voice variant by 2,6%, 0,5%, and 0.74% respectively, whereas dynamic time warping achieves better accuracy on loud and quiet voice variant by 2,79% and 2,3% respectively.

**Keywords:** *Speaker Verification, Text Dependent Speaker Verification, Hidden Markov Model, Dynamic Time Warping*

## 1. INTRODUCTION

Along with the development of the digitalization era, human life is becoming more dependent to technologies such as digital application or software, which one of the reason why Internet of Things (also known as IoT) is becoming more popular over the year since IoT provides innovative solutions, such as smart home, to facilitate human life [1]. One of the main focus on IoT is how to make IoT more secure to authenticate identities in order to access something, such as limiting different individual access to smart home peripheral [2], in order to fulfill the users need for confidentiality, integrity and availability [3]. There are few traditional approaches for verifying identities like using a Personal Identification Number (also known as PIN) and passwords, but those approaches are very risky since it doesn't really verify an individual characteristic, meaning everyone can access the protected resource if the password or PIN is stolen or leaked [4][5]. Biometric system gives a solution to that problem by using individual characteristic in order to authenticate or verify the identity [6] since biometric system focuses on statistical analysis of biological characteristics [4][5] such as fingerprints, hand geometry, voice identification, and retina identification.

Speaker recognition is process of recognizing a person, who is speaking, by obtaining characteristics or speech wave parameters [7]. It enables access control from the speakers to verify their identity when accessing system or service. Most service such as voice call, confirming banking transaction, database access service, information service, reservation service are classified as speaker verification [8]. Speaker verification is part of speaker recognition [9]. Speaker verification is a process of which accept or reject speaker identity. There are two types of speaker verification: text-dependent speaker verification and text-independent speaker verification [7].

The difference between text-dependent speaker verification and text-independent speaker verification is the utterance the speaker. Text-dependent speaker verification uses same utterance for training and testing, whereas text-independent speaker verification uses different utterance for training and testing. There are two main steps for text-dependent speaker verification, which are training and decision making. Training process in text-dependent speaker verification requires the characteristic of the speaker's voice, which can be achieved by using an audio feature extraction algorithm (mel frequency cepstral coefficient (MFCC) being the most well-known and well researched feature extractor) and decision making process decides whether the input voice is indeed the verified user or an impostor trying to gain access by the similarity of the input voice characteristic and the trained voice characteristic. Two most well researched text-dependent speaker verification methods are dynamic time warping (DTW) and hidden markov model (HMM).

DTW represented its utterance by sequence of spectral features vectors, timing variation of the same text then normalize it by using DTW algorithm [10][11] while HMM efficiently model statistical variation in spectral features [8]. Speaker's speaking pattern may differ from their usual speaking pattern due to the cause of their mental and physical condition [12][13][14], which might affect speaker verification system performance for verifying the speaker identity. Other than different speaking pattern, loudness of the voice may also affect system performance for verifying the speaker identity, such as the input voice received from speaker speaking from 10 cm away from the microphone or 1 meter away from the microphone will have different input voice loudness. We compared both dynamic time warping and hidden markov model classifier with different speaker speaking variant or style such as loud, normal, quiet, shout, and soft with each variant having their own audio characteristic like decibels and speaking style to obtain which classifier yield better result on verifying individual identity with slightly different speaking pattern used in the training process to replicate the condition when the speaker is experiencing mental or physical conditions and the distance between the speaker and the microphone that affects the loudness of the voice and speaker speaking pattern.

In theory, DTW based speaker verification system should perform better on verifying quiet and shout voices due to the nature of DTW algorithm that compares the MFCC distance between train voices and test voices, whereas HMM based speaker verification system should perform better on verifying normal, loud, and soft voices due to the nature of HMM on classifying voices based on the MFCC characteristic of the voice. DTW based speaker verification system should also have a longer runtime compared to HMM based speaker verification system on verifying a speaker identity since DTW based speaker verification system compares all the distance in the train data to test data in order to get the audio data with least distance.

The rest of the paper is organized in the following manner. Review from previous work in Section 2, followed by methodology in Section 3. In Section 4, we discuss the experiment result of compared DTW and HMM classifier and finally Section 5 concludes the paper.

## 2. RELATED WORKS

Research by [10] shows that a speaker verification system with MFCC feature extraction and DTW classifier yields 96,2% accuracy on verifying speaker's voice, with *BM Millar* database which consists of utterance of "oh", "nought", and "one" to "ten". Research performed by [10] using a normal voice recording yields a high accuracy and research by [2] shows that a speaker verification system with MFCC feature extraction and DTW classifier yields 86,785% overall accuracy for loud voices. Similar research also performed by [15] shows that a speaker identification system with MFCC feature extraction and DTW classifier yields quite high accuracy for identifying speaker identity with noisy background condition with *NOIZEUS* database.

Research by [16] shows that a speaker verification with MFCC feature extraction and HMM classifier with *YOHO corpus* database yields high accuracy for verifying speakers identity. Research performed by [16] shows that the system produces 0% false rejection rate, 0% false acceptance rate for female speakers, and 0,09% false acceptance rate for male speakers. Similar research also done by [17] shows that a speaker identification with MFCC feature extraction and HMM classifier with noisy condition yields a high accuracy, where the system produces 99% to 100% accuracy on identifying individual voices for many background noises variant. Research by [18] proposes a more modern approach for speaker verification where the focus of the research is to implement automatic speaker verification with HMM where the systems performs high accuracy for verifying user where the system produces 99% overall accuracy. It is worth noting that every research with HMM classifier is done by normal speaking variant, which might be a contributing factor why the research in this area yields high accuracies.

Similar research has done by [12] which shows that a speaker identification system with MFCC feature extraction and DTW classifier yields rather low accuracy for similar voice variant used in this paper. Research performed by [12] shows that the system produces 67%, 84%, 80%, and 70% overall accuracies for shout, slow, loud, and soft voice variant respectively. The difference between research done by [12] and research done in this paper is research performed by [12] focuses on testing DTW based speaker identification system which the system try to guess the speaker identity from a voice and the focus of this paper is to test DTW based speaker verification system which the system is verifying the speaker identity from a voice.

Although there are many research has been done in this field, a research about different speaking style or variant in this field has been minimum. The closest research done with similar focus of this research has done by [12] where the research focuses on testing how accurate DTW classifier on identifying an individual on shout, slow, loud, and soft voice in a speaker identification system, and there are no similar research done with similar focus with HMM classifier.

## 3. METHODOLOGY

This section will explain the steps necessary to build a mel frequency cepstral coefficient feature extraction speaker verification system using hidden markov model and dynamic time warping classifier as shown in Figure 1.

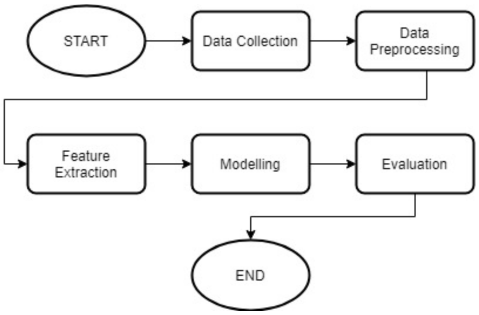


*Figure 1: Speaker Verification System Research Process*

### 3.1 Data Collection

Data used in this study consists of users and their voices. There are 10 users which are registered as legit users, and a total of 20 users which voices will be used to train and test the classifier's false acceptance and false rejection rate (which is called *impostor*). As for the different variants of voices, there are 1,930 total voices in which a total of 330 voices is used to train the model (10 registered user models and 1 impostor model) and the rest will be used for testing the accuracy for the classifiers (1,600 total voices that consists of 320 loud test voices, 320 normal test voices, 320 shout test voices, 320 soft test voices and 320 quiet test voices). Each of the variants differs one another by decibels measurement on how loud the voices are and the style of how the speaker talks. The voices are recorded in a quiet environment with no background noises. Loud variant are voices which have normal talking style and in range of 70 decibels up to 75 decibels, normal variant are voices which have a normal talking style and in range of 50 decibels up to 55 decibels, shout variant are voices which have a shout talking style and in range of 70 decibels up to 75 decibels, soft variant are voices which have a soft (whisper-like) talking style and in range of 35 decibels up to 40 decibels, and quiet variant are voices which have a normal talking style and in range of 35 decibels up to 40 decibels.

### 3.2 Data Preprocessing

After the voice data are collected, silence removal technique was applied on the gathered data to reduce system processing time and increase system performance by eliminating unvoiced segment of the recording. Silence removal technique will be achieved by using an audio editor and recording application software called *Audacity*.

### 3.3 Feature Extraction

After the gathered data preprocessed, feature extraction step begins. The goal of feature extraction step is to extract feature (such as the characteristic of the voice). Feature extraction step in this paper uses mel frequency cepstral coefficient algorithm, also known as MFCC, to obtain a vector of voice features for the recordings. MFCC algorithm first frames the audio signal into 20 milliseconds to 40 milliseconds frames then calculates the Discrete Fourier Transform, also known as DFT, on each frame using the formula as shown in Eq. (1):

$$S_i(k) = \sum_{n=1}^{N} S_i(n)h(n)e^{-j2\pi kn/N} \quad 1 \le k \le K \quad (1)$$

In Eq. (1), k represents the length of DFT, S(n) represents domain signal, which Si(n) represents the domain signal of each $i^{th}$ frame where n ranges from 1 to number of samples, and h(n) represents N sample long analysis window. After the DFT of each frame calculated, we then able to calculate Periodogram estimate of the power spectrum using the formula as shown in Eq. (2):

$$P_i(k) = \frac{1}{N}|S_i(k)|^2 \quad (2)$$

After power spectrum is calculated, we then able to apply mel-spaced filterbank to the power spectrum to obtain the sum of energy in each filter, which then we can obtain the logarithm of all filterbank energies to obtain the Discrete Cosine Transform of the respective log energies which results in features called mel frequency cepstral coefficients.

### 3.4 Modelling

In order to verify the speaker, two classifiers are used and compared, that are: dynamic time warping classifier and hidden markov model classifier. Two mentioned classifiers will use same preprocessed recording that was mentioned earlier. All models were run on a personal computer with MSI GTX 970 GPU.

#### 3.4.1 Dynamic Time Warping

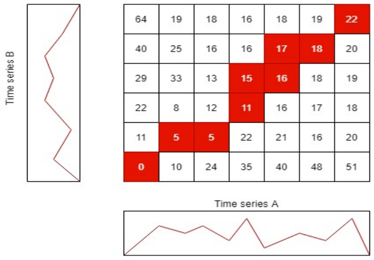
For dynamic time warping classifier, we first separate every training voice feature according to their respective speaker, which results us having two list of voice feature per verification process that consists of registered user voice feature and the impostor voice feature. Next, we extract the feature from the test voice data and then compare the test voice feature vector with each of the feature vector in the registered user voice feature and impostor voice feature list to obtain the voice recording with the smallest warping path. To obtain the warping path between two voice feature vectors, we need to first calculate the distance between the vector sequence for test voice feature vector and train voice feature vector using Euclidean distance formula as shown in Eq. (3):

$$d_{ij} = |i - j| = \left[(i_1 - j_1)^2 + (i_2 - j_2)^2 + \cdots + (i_n - j_n)^2\right]^{\frac{1}{2}} \,(3)$$

In Eq. (3), i and j represents points of train voice feature vector or test voice feature vector. After we obtained the distance between voice feature vectors, we need to calculate each grid of the cost matrix from the distance we obtained using Euclidean distance with a formula as shown in Eq. (4):

$$D_{ij} = \left|A_i - B_j\right| + min(D_{i-1\,j-1}, D_{i-1\,j}, D_{i\,j-1}) \quad (4)$$

In Eq. (4), A and B represents time series of test voice feature vector and train voice feature vectors that was obtained after calculating the Euclidean distance of the respective vector sequence. After we have a cost matrix, we can then compute the warping path and then calculate the total distance between two vectors by the sum of value in the cost matrix grid that is included in the warping path (like shown in an example in Figure 2). After each training voice feature vector distance towards the test voice feature vector is calculated, we then pick the most similar (in this case, the smallest distance) between all the train voice feature vectors to verify the user's identity as shown in Figure 3. If the most similar voice feature comes from the registered user voice feature list, then the system will label the test voice as the verified user. But if the most similar voice feature comes from the impostor voice feature list, then the system will label the test voice as the unverified user.



| | | | | | | |
|---|---|---|---|---|---|---|
| 64 | 19 | 18 | 16 | 18 | 19 | **22** |
| 40 | 25 | 16 | 16 | **17** | **18** | 20 |
| 29 | 33 | 13 | **15** | **16** | 18 | 19 |
| 22 | 8 | 12 | **11** | 16 | 17 | 18 |
| 11 | **5** | **5** | 22 | 21 | 16 | 20 |
| **0** | 10 | 24 | 35 | 40 | 48 | 51 |

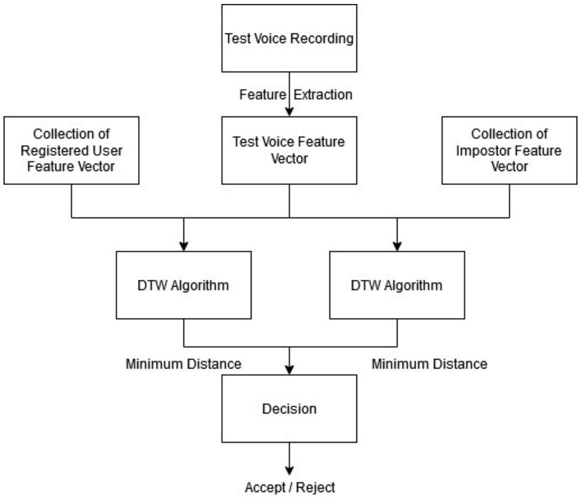


*Figure 3: DTW Speaker Verification Architecture*

### 3.4.2 Hidden Markov Model

Similar to dynamic time warping classifier approach, we use two speaker models in one verification process, which are registered user model and impostor model. Registered user model is trained by the voice feature vector from their respective speaker, while the impostor model is trained by the voice feature of impostor train data. This model was built by using a python library called *hmmlearn*. We use a total of 12 number of states in our hidden markov model with 200 iterations in order to train our model. In order to obtain the probability between the registered user model and the impostor model towards the test voice feature vector, we will use Viterbi algorithm to obtain a log-likelihood of the test voice feature vector for the registered user model and the impostor model as shown in Eq. (5):

$$log\ likelihood = \log P(O \mid \lambda) \qquad (5)$$

In Eq. (5), O represents observation sequence of the phrase and λ represents hidden markov model of the registered user or the impostor. The higher log-likelihood the model produces means more similar the test voice with the trained voice within the model. After we obtain the log-likelihood of each model towards the test voice feature vector, we check for the model with highest log-likelihood and decides if the test voice is a registered user or not as shown in Figure 4. If log-likelihood of the registered user model is higher than log-likelihood of the impostor model, the system will label the test voice as the verified user. But if log-likelihood of the impostor model, the system will label the test voice as verified user. But if log-likelihood of the impostor model is higher than the registered user model, the system will label the test voice as unverified user.

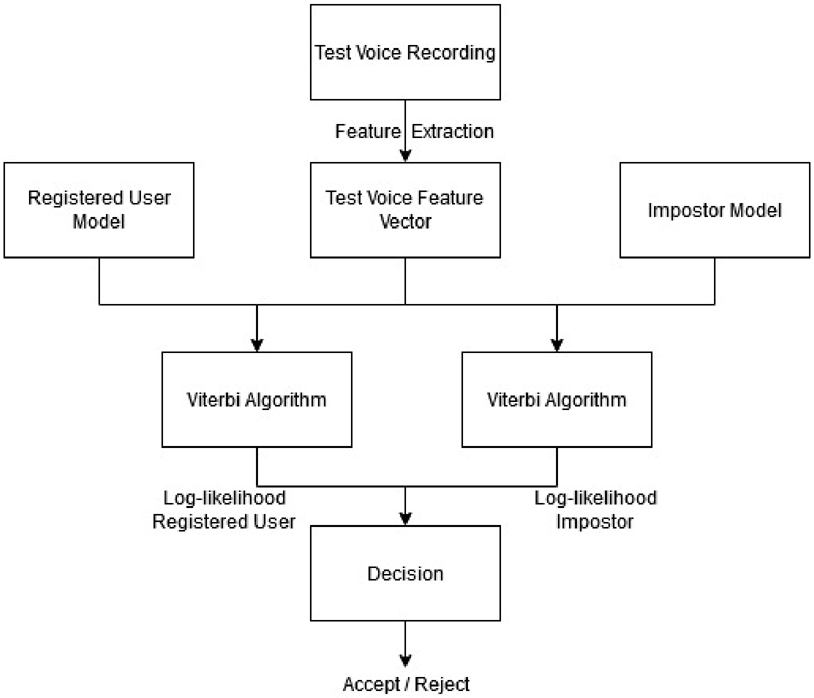


*Figure 4: HMM Speaker Verification Architecture*

### 3.5 Evaluation

Since the speaker verification system outputs a binary statement, which are accept or reject, we use confusion matrix (also known as error matrix) to analyze system's performance on accepting and rejecting voices. We use true positive (TP), false positive (FP), false negative (FN), true negative (TN), accuracy, and classification error for the calculation. The formula of each confusion matrix are shown in Eq. (6) to Eq. (11):

$$TP = \frac{number\ of\ correct\ accept}{total\ of\ verified\ voice} \quad (6)$$

$$FN = \frac{number\ of false\ reject}{total\ of\ verified\ voice} \quad (8)$$

$$TN = \frac{number\ of\ correct\ reject}{total\ of\ impostor\ voice} \quad (9)$$

$$Accuracy = \frac{TP+TN}{total\ of\ test\ voice} \quad (10)$$

$$Classification\ Error = \frac{FP+FN}{total\ of\ test\ voice} \quad (11)$$

Correct acceptance (CA), false acceptance (FA), false rejection (FR), and correct rejection (CR) can be illustrated with confusion matrix as shown in Table 1.

*Table 1: Confusion Matrix*

| | | Predicted | |
|---|---|---|---|
| | | Registered | Impostor |
| Actual | Registered | CA | FR |
| | Impostor | FR | CR |

## 4. RESULTS

Our goal in this experiment is to compare which classifier performs better on verifying different speaking variant or style. Table 2 shows the difference between each variant that we used in the experiment. Our experiment begins with training the classifier with train data that consists of 30 normal voice variant for registered users and the impostors, where the impostor normal voice training data are gathered from 15 different speakers. Table 3 shows the amount of data gathered and used for this experiment.

*Table 2: Sound Variant, Style and Loudness (dB) of the Dataset*

| Variant | Speaking Style | Loudness (dB) |
|---|---|---|
| Shout | Shout | 70-75 |
| Loud | Normal | 70-75 |
| Normal | Normal | 50-55 |
| Quiet | Normal | 35-40 |
| Soft | Whisper | 35-40 |

*Table 3: Voice Data Distribution*

| Variant | Speaker | # Train Voices | # Test Voices |
|---|---|---|---|
| Shout | Registered User | - | 300 |
|  | Impostor | - | 20 |
| Loud | Registered User | - | 300 |
|  | Impostor | - | 20 |
| Normal | Registered User | 300 | 300 |
|  | Impostor | 30 | 20 |
| Quiet | Registered User | - | 300 |
|  | Impostor | - | 20 |
| Soft | Registered User | - | 300 |
|  | Impostor | - | 20 |
|  | Total | 330 | 1600 |

Confusion matrix data of dynamic time warping classifier and hidden markov model classifier are shown Table 4 and performance percentage shown in Table 5 are calculated based on 1,600 total test voice data (1,500 test voice data from registered users and the other 100 test voice data are from speakers voice that is trained in impostor model and speakers voice that is not trained in any model) where every test voice data will be retested for every registered user (in this experiment there are a total of 10 registered users) which bring us to a total of 16,000 tests performed. For dynamic time warping classifier, we tested every voice data in the registered user list and impostor list to get the voice feature with the least distance from the test data. An example of warping path comparison of 2 voice features from an impostor voice feature vector and a registered user voice feature vector to a registered user test voice feature vector shown in Figure 5 shows the more similar two voice features, the more diagonally aligned the warping path is from the bottom left of the matrix to the top right of the matrix.

*Table 4: DTW and HMM Confusion Matrix Data*

| #Verified voice | #Impostor | Variant | Dynamic Time Warping |  |  |  | Hidden Markov Model |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|
|  |  |  | #True Accept | #False Accept | #False Reject | #True Reject | #True Accept | #False Accept | #False Reject | #True Reject |
| 300 | 2900 | Loud | 172 | 90 | 128 | 2,810 | 0 | 1 | 300 | 2,899 |
|  |  | Normal | 300 | 88 | 0 | 2,812 | 300 | 3 | 0 | 2,897 |
|  |  | Quiet | 223 | 177 | 77 | 2,723 | 45 | 8 | 255 | 2,892 |
|  |  | Shout | 68 | 76 | 232 | 2,824 | 11 | 4 | 289 | 2,896 |
|  |  | Soft | 75 | 132 | 225 | 2,768 | 65 | 101 | 235 | 2,799 |

*Table 5: DTW and HMM Performance Percentage*

| Variant | Dynamic Time Warping | | | | | | Hidden Markov Model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | %TP | %FP | %FN | %TN | %Acc | %Err | %TP | %FP | %FN | %TN | %Acc | %Err |
| Loud | 57,33 | 3,10 | 42,66 | 96,90 | 93,19 | 6,81 | 0,00 | 0,03 | 100,00 | 99,97 | 90,59 | 9,41 |
| Normal | 100,00 | 3,03 | 0,00 | 96,97 | 97,25 | 2,75 | 100,00 | 0,10 | 0,00 | 99,90 | 99,91 | 0,09 |
| Quiet | 74,33 | 6,10 | 25,67 | 96,00 | 92,06 | 7,93 | 15,00 | 0,28 | 85,00 | 99,72 | 91,78 | 8,22 |
| Shout | 74,67 | 8,00 | 25,33 | 92,00 | 90,38 | 9,63 | 3,66 | 0,14 | 96,33 | 99,86 | 90,84 | 9,16 |
| Soft | 25,00 | 4,55 | 75,00 | 95,44 | 88,84 | 11,16 | 21,67 | 3,48 | 78,33 | 96,52 | 89,50 | 10,50 |

Table 5 shows HMM classifier is slightly better at verifying speaker's identity for it produces better overall accuracy and classification error rate for normal, shout, and soft voice variant but have a high false rejection rate compared to DTW classifier, which HMM classifier's false negative percentage ranges from 85% to 100% whereas DTW classifier's false negative percentage ranges only from 25% to 75%. It is also worth noting that HMM classifier performs better at verifying user's identity with different speaking style (shout and soft) from the train data speaking style than DTW classifier which performs better at verifying user's identity with same speaking style as the one used in train data speaking style. As for computation time, DTW classifier shows a significant amount of time needed to finish the whole process. DTW classifier needed around 3 minutes to train the data and around 2 hours to finish whole test data, whereas HMM classifier only needs around 2 minutes to train the data and around 5 minutes to finish whole test data.

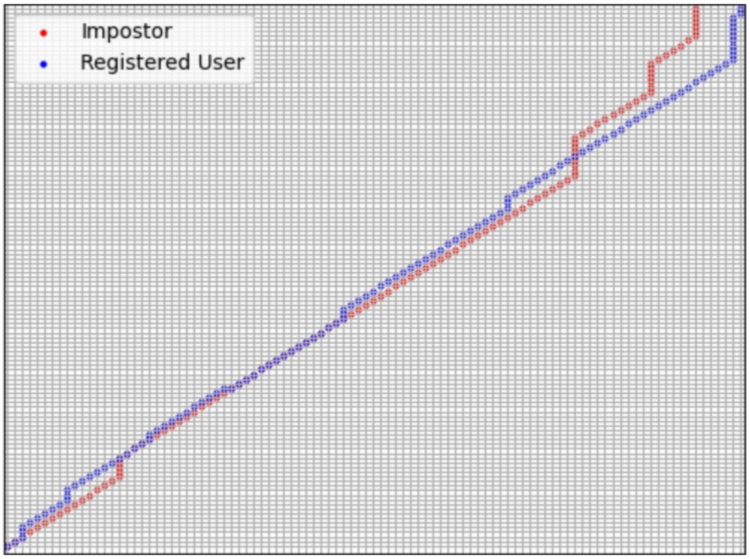


*Figure 5: An Example of Warping Path Comparison of 2 Voice Features to One of the Registered User Test Data*

## 5. Conclusion

Our study focuses on comparing both DTW classifier and HMM classifier on such different speaking variant or style. From the results discussed in Section 4, it can be concluded that HMM classifier produces better performance than DTW in terms of overall accuracy for normal, shout, and soft voice variant by 2,6%, 0,5%, and 0.74% respectively, whereas DTW achieves better accuracy on loud and quiet voice variant by 2,79% and 2,3% respectively. From the performance, we can conclude HMM classifier performs better on verifying voices with different speaking style as the one used for training, whereas DTW classifier performs better on verifying voices with same speaking style, but different loudness of the voice. HMM classifier also perform better in computation time compared to DTW classifier which need a significant amount of time. Although HMM classifier has better overall accuracy for verifying speaker identity, it is also worth noting that HMM classifier have significant false rejection accuracy compared to DTW. The accuracy can definitely be improved with a sufficient train audio data for each sound variant.

## ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for their support in making this research possible. This research was supported by Awarded Research Grant scheme 2021

## REFERENCES

[1] S. Kumar, P. Tiwari, M. Zymbler, Internet of things is a revolutionary approach for future technology enhancement: a review, *Journal of Big Data 6*, 2019

[2] B. H. Prasetio, D. Syauqy, Design of speaker verification using dynamic time warping (dtw) on graphical programming for authentication process, *Journal of Information Technology and Computer Science 2*, 2017.

[3] B. Soewito, Y. Marcellinus, Iot security system with modified zero knowledge proof algorithm for authentication, *Egyptian Informatics Journal*, 2020.

[4] M. Faundez-Zanuy, Biometric security technology, *Aerospace and Electronic Systems Magazine, IEEE*, vol. 21, no. 1 2006.

[5] K. Dharavath, F. A. Talukdar, R. H. Laskar, Study on biometric authentication systems, challenges and future trends: A review, *2013 IEEE International Conference on Computational Intelligence and Computing Research*, 2013.

[6] F. Orság, Speaker recognition in the biometric security systems, *Computing and Informatics*, vol. 25, 2006, pp. 369–391.

[7] S. Furui, 40 years of progress in automatic speaker recognition, *Advances in Biometrics Lecture Notes in Computer Science*, 2009, pp. 1050–1059

[8] S. Furui, An overview of speaker recognition technology, *The Kluwer International Series in Engineering and Computer Science Automatic Speech and Speaker Recognition*, 1996, pp. 31–56.

[9] S. Tranter, D. Reynolds, An overview of automatic speaker diarization systems, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, num. 5, 2006, pp. 1557–1565.

[10] K. Yu, J. Mason, J. Oglesby, Speaker recognition models, *EUROSPEECH*, 1995.

[11] Y. Permanasari, E. H. Harahap, E. P. Ali, Speech recognition usingdynamic time warping (dtw), *Journal of Physics: Conference Series*, vol. 1366, 2019.

[12] I. Shahin, N. Botros, Speaker identification using dynamic time warping with stress compensation technique, *Proceedings IEEE Southeastcon 98 Engineering for a New Era*, 1998.

[13] Y. Chen, Cepstral domain talker stress compensation for robust speech recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, num. 4, 1988, pp. 433–439.

[14] B. D. Womack, J. H. Hansen, Classification of speech under stress using target driven features, *Speech Communication*, vol. 20, num. 1-2, 1996, pp. 131–150.

[15] S. Malini, R. Koulsaya, Speaker identification using mfcc and dtw technique on the enhanced speech signal in a noisy environment, *International Journal of Engineering Research and Technology (IJERT)*, vol. 4, 2016.

[16] C. Che, Q. Lin, D.-S. Yuk, An hmm approach to text-prompted speaker

verification, *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, 1996.

[17] T. Matsui, T. Kanno, S. Furui, Speaker recognition using hmm composition in noisy environments, *Computer Speech* Language, vol. 10, num. 2, 1996, pp. 107–116.

[18] D.-P. Munteanu, S.-A. Toma, Automatic speaker verification experiments using hmm, *2010 8th International Conference on Communications*, 2010.