Journal of Information Technology and Computer Science
Volume 2, Number 1, 2017, pp. 11-18
Journal Homepage: www.jitecs.ub.ac.id

Design of Speaker Verification using Dynamic Time Warping (DTW) on Graphical Programming for Authentication Process

Barlian Henryranu Prasetio[1], Dahnial Syauqy[2]

[1,2] Computer System and Robotics Lab, Faculty of Computer Science, University of Brawijaya, Malang, Indonesia

{barlian@ub.ac.id, dahnial87@ub.ac.id}

Received 13 March 2017; accepted 22 June 2017

Abstract. Authentication mechanism is generally required on systems which need safety and privacy. In common, typed username and password are used and applied in authentication system. However, this type of authentication has been identified to have many weaknesses. In order to overcome the problem, many proposed authentication system based on human voice as unique characteristics of human. We implement Dynamic Time Warping algorithm to compare human voice with reference voice as the authentication process. The testing results show that the speech similarity of the speech recognition average is 86.785%.

1. Introduction

Smart home systems are built based on computer and information technology to control and automate appliances [1]. There are varieties of task that monitors and controls smart home peripheral through the use of computer [2], web browser or smartphone [3]. One of key area in smart home system development is related to its safety. An example is user access limitation to control smart home peripheral. In order to limit the user access, generally authentication process is performed to verify the user with the specific rights. There are varieties of mechanism to perform authentication process e.g. using Personal Identification Number (PIN), password or smart ID cards. However, that kind of authentication mechanism has been studied and explained that they have weaknesses [4].

The use of biometric parameter of human with unique characteristics have been an interesting topics in authentication mechanism [5][6]. One of the popular biometric to be used as authentication is using voice. Human voice is generated mainly by lungs, vocal cords, and articulation. Human voice is known to have unique characteristics from person to person. Therefore, many researches have been proposed to use human voice as authentication input and reference [7][8].

Authentication process which is based on human voice uniqueness is generally related to speaker recognition and speaker verification. Both terms aims to differentiate the speaker from other speaker. However, speaker recognition, or speaker identification aims mainly to recognize who is speaking from set of population. Meanwhile, speaker verification aims mainly to "verify" the input voice whether it is matched with the referenced voice or not. Thus, speaker verification can be used to authenticate and verify the speaker identity as part of the security process [9].

We proposed the design of speaker verification which is based on feature matching using dynamic time warping algorithm which was implemented using graphical programming. In feature matching, we record the main voice as the reference signal and then extract its features. Later in authentication process, another voice as input command will be matched with the reference signal and calculate their degree of similarity.

2. Research Methods

The proposed design system block diagram is shows on figure 1. In the initial step, we store set of voice features which were extracted using MFCC as reference signal. In the feature extraction process itself contain several sub process, such as signal pre-processing and feature extraction itself using Mel Frequency Cepstral Coefficients (MFCC). Later, when the authentication process happens, the stored features will be compared with another input voice in form of sequence of features. The comparison method used Dynamic Time Warping algorithm and the score will be used to verify the correct user.

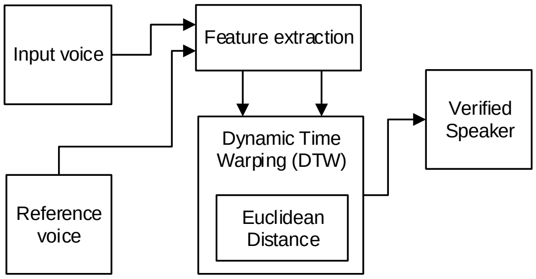


Fig. 1. Block diagram of Feature Matching

3.1 Preprocessing Stages

The preprocessing stages starts with pre-emphasis on voice signal to suppress high frequency parts on the signal by applying FIR filter. Next, the voice signal needs to be separated in small frames since voice signal is known as non-stationary signal. The framing process cuts the voice signal into about 10ms frame segments. Then, for each segment, windowing process will be applied to reduce error from overlapping segment during framing process. Finally, energy detection will be performed on each frame to see whether the frame contains pronunciation or not [10].

3.2 Feature Extraction Stages

The main task in feature extraction stages is to extract feature information from both reference voice and input voice. The feature extraction process was start with calculating Fast Fourier Transform on each frame. The ultimate task on feature extraction is calculating Mel Frequency Cepstral Coefficients (MFCC) on each frame [11]. These coefficients will be stored as voice features.

3.3 Feature Matching and Scoring Stages

In general, every person have different rate of speaking speed. Dynamic Time Warping (DTW) algorithm is known to be able to normalize and find best alignment between two signals. In our case, we will use DTW to find best matching score between reference voice features and input voice features. Figure 2 shows the pseudo code of DTW algorithm. As we can see in the DTW pseudo code, we also need to calculate Euclidean distance between two points, which is the point of the input and reference points.

```
DTW(textsequence,referencesequence)
{
    for(i=0,i<length,j++)
    {
        for(j=0,j<length,j++)
        {
        //calculate Euclidean distance
        len=length(textsequence(i));
        sum=0;
        for (a=0,a<b,a++)
        {  k(a)=(test(a)-reference(a))^2;
           sum=sum+k(a);
        }
        dist(j)=sqrt(sum)
        }
    }
    localcostmatrix=dist;

    n=numberofrow(textsequence);
    r=numberofrow(referencesequence);

22      D=initializearray(0,n,r); //initializearray
23
24      //DTWaligning algorithm:
25      result=DTWalign(localcostmatrix,n,D,r);
26      return result;
27  }
```

Fig. 2. Dynamic Time Warping pseudo code

The DTW function needs two parameters as input; textsequence and referencesequence. For each text sequence, Euclidean distance is calculated (row 5-16) and the process is repeated for all text sequence (loop in row 3). After that, an array is build based on number of row from textsequence and referencesequence (row 22). Finally, it calls another function named DTWalign to calculate the best alignment (row 25).

In the DTW pseudo code, we need to call another function to calculate the best alignment. The function is called “DTWalign”. The pseudo code of the DTW alignment calculation is shown in Figure 3.

```
1   //calculate DTW alignment
2   DTWalign (localcostmatrix,n,D,r)
3   {
4       D(1,1)=localcostmatrix(1,1);
5       for(j=2:r)
6       {D(i,j)=D(1,j-1)+localcostmatrix(1,j);
7       }
8       for(i=2:n)
9       {D(i,j)=D(i-1,1)+localcostmatrix(i,1);
10      }
11      for(i=2:n)
12      {  for(j=2:r)
13          a=D(i-1,j)+localcostmatrix(i,j);
14          b=D(i-1,j-1)+localcostmatrix(i,j);
15          c=D(I,j-1)+localcostmatrix(i,j);
16          D(i,j)=min{[b,a,c]}
17          }
18      }
19      Return D(n,r);
20  }
```

Fig. 3. DTW alignment pseudo code

The DTWalign function needs several parameters as input (row 2). Row 4-18 show the standard procedure to measure minimum distance using Dynamic Time Warping algorithm between both arrays of textsequence and referencesequence. Finally, after finding the minimum best alignment (row 16), it returns back the value (row 19).

3. Result and Discussion

3.1 Code Implementation

We implemented the pseudo code of DTW algorithm in graphical programming using Lab-VIEW. Lab-VIEW is dataflow based graphical programming which was made by National Instruments. Lab-VIEW can automatically compile codes into multiple threads or cores which can makes it runs faster than single thread or sequential operation [12]. In the stored voice features, we recorded and stored one word as an example of command. Later, an input voice as testing command was recorded and extracted as voice features. After getting both voice features, DTW perform its function on both signal. The DTW will normalize the variation of speaking rate by finding the best alignment between two different signals. As part of its operation, DTW will use Euclidean distance to calculate distance between points in both signals. The output of this process is percentage degree of signal alignment. When both signals have best alignment, it means they have higher degree of similarity. Thus, it can be concluded that they were pronounced by the same speaker. Figure 4 shows the implementation of DTW algorithm and Euclidean distance calculation.

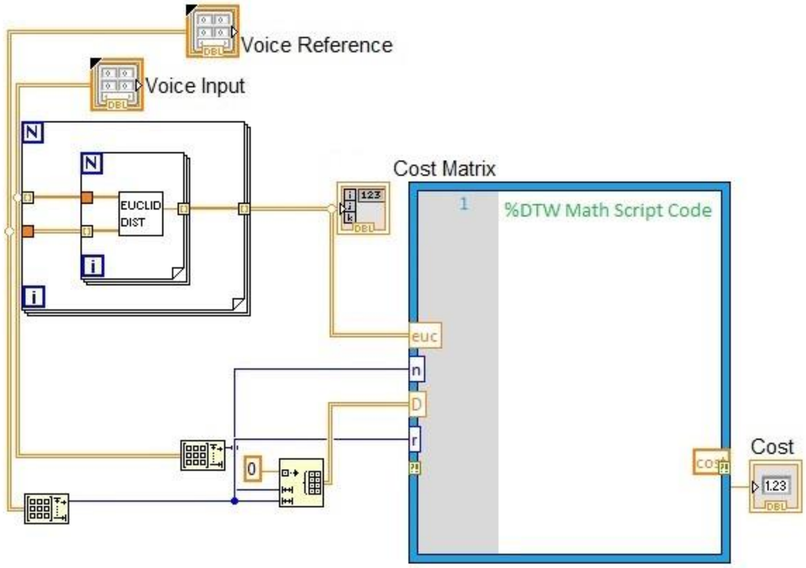
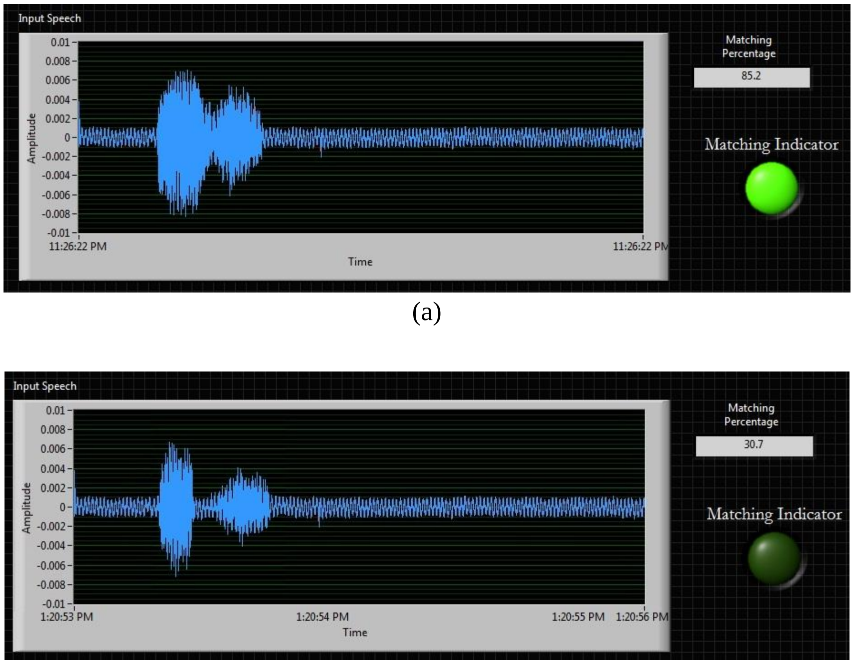


Fig. 4. The implementation of DTW algorithm in graphical programming

Eventually, we also designed the front panel as information for the user how similar their voice with the base reference voice. In our system, we design a minimum threshold for similarity is 80%. Thus, when the similarity percentage is below 80%, the authentication process will be failed. Figure 5 shows result front panel.



(a)

(b)

Fig. 4. Result front panel (a) when it is > 80% similarity (b) when it is < 80% similarity

3.2 Testing and Analysis

The system was testing by connecting a microphone with a computer that has installed lab-VIEW program. The system was tested using two scenarios.

The first scenario, aims to test the system's speech similarity in recognizing speaker voice similarity percentage with the dictionary. This scenario provides the voice input "hello" on the system by a speaker 20 times. The testing results of percentage of speech similarity by the same speaker are shows in table 1.

Table 1. The percentage of Speech Similarity by the same speaker

Speaker1	Similarity (%)
Speech1	86.1
Speech2	88.4
Speech3	85.2
Speech4	89.3
Speech5	81.9
Speech6	86.9

Speech7	91.5
Speech8	90.8
Speech9	84.7
Speech10	85.3
Speech11	86.6
Speech12	89.2
Speech13	82.6
Speech14	88.2
Speech15	87.9
Speech16	87.7
Speech17	85.6
Speech18	85.1
Speech19	84.9
Speech20	87.8
Average	86.785

The second scenario aims to verify the voice. The system was tested by four people who each provide speech input "hello" for 5 times. The system displays the percentage of similarity of sound. The authentication system processes are marked with LED lights. The testing results of percentage of speech similarity by the different speaker are shows in table 2.

Table 2. The percentage of Speech Similarity by the different speaker

Speaker	**Speech**	**Similarity (%)**	**LED ON/OFF**
Speaker1	Speech1	86.1	ON
	Speech2	88.4	ON
	Speech3	85.2	ON
	Speech4	89.3	ON
	Speech5	81.9	ON
Speaker2	Speech1	45.8	OFF
	Speech2	44.9	OFF
	Speech3	48.7	OFF
	Speech4	47.2	OFF
	Speech5	46.2	OFF
Speaker3	Speech1	33.5	OFF
	Speech2	33.1	OFF
	Speech3	35.6	OFF
	Speech4	34.3	OFF
	Speech5	36.1	OFF
Speaker4	Speech1	25.8	OFF
	Speech2	59.3	OFF
	Speech3	24.8	OFF
	Speech4	26.3	OFF
	Speech5	24.7	OFF

4. Conclusions

We have design of speaker verification using dynamic time warping (DTW) on graphical programming for authentication process. The testing results show that the speech similarity of the speech recognition average is 86.785%. This is due to the relatively large noise. During the test, the noise occurs between 15-20% of amplitude sound. Further research is to reduce noise so that the system becomes more accurate.

References

[1] Bhardwaj, P., Manchanda, P., Chahal, P., Chaudhary, P., Singh, R. “A Review Paper On Smart Home Automation”. International Journal of Scientific Research and Management Studies (IJSRMS) Volume 3 Issue 7, pp: 279-283. (2017)

[2] Patchava, V., Kandala, H.B., Babu, P.R. “A Smart Home Automation technique with Raspberry Pi using IoT”. Smart Sensors and Systems (IC-SSS), International Conference on. 21-23 Dec. (2015)

[3] Mowad, M.A.L., Fathy, A., Hafez, A. “Smart Home Automated Control System Using Android Application and Microcontroller”. International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May (2014)

[4] Ravi, S. “Access Control: Principles and Practice”. IEEE communications. pp: 40-46. (1994)

[5] Bhuiyan, A., Hussain, A., Mian, A., Wong, T.Y., Ramamohanarao, K., Kanagasingam, Y. “Biometric authentication system using retinal vessel pattern and geometric hashing”. IET Biometrics. Volume: 6, Issue: 2, 3 (2017)

[6] Sapkale, M., Rajbhoj, S.M. “A biometric authentication system based on finger vein recognition”. Inventive Computation Technologies (ICICT), International Conference on. 26-27 Aug. (2016)

[7] Barbosa, F.G., Silva, W.L.S. “Multiple Support Vector Machines and MFCCs application on voice based biometric authentication systems”. IEEE International Conference on Digital Signal Processing (2015)

[8] Barbosa, F.G., Silva, W.L.S. “Support vector machines, Mel-Frequency Cepstral Coefficients and the Discrete Cosine Transform applied on voice based biometric authentication”. SAI Intelligent Systems Conference (2015)

[9] Singh, N., Khan, R.A, Shree, R. “Applications of Speaker Recognition”. Procedia Engineering 38 pp:3122 – 3126 (2012)

[10] Washani, N., Sharma, S. “Speech Recognition System: A Review”. International Journal of Computer Applications. Volume 115–No 18, April (2015)

[11] Mohan, B.J., Babu, R.N. “Speech recognition using MFCC and DTW”. Advances in Electrical Engineering (ICAEE), International Conference on. 9-11 Jan. (2014)

[12] LabVIEW User Manual, April 2003 Edition, National Instruments (2003)