VOICE RECOGNITION APPLICATION BY USING FISHER'S LINEAR DISCRIMINANT ANALYSIS (FLDA) FEATURE EXTRACTION

[1]Aeri Rachmad*, [2]Devie Rosa Anamisa, [3]Novia Putri Bintari

[1,2,3] Faculty of Engineering, University of Trunojoyo Madura
Jalan Raya Telang PO BOX 2 Kamal, Bangkalan, East Java, Indonesia
*Email: aery_r@yahoo.com

Abstract

In classifying the pattern, the number of learning data used is often very limited, but the number of dimensions are very high. Fisher linear discriminant analysis (FLDA) is a pattern classification method that are widely used in pattern recognition feature extraction and reduction of linear dimensions. FLDA method is able to analyze the data and study the relationship between a set of categorical predictors and response for pattern recognition applications, including speech pattern recognition is used as a command to the system in the presence of employees of the agency. FLDA has the ability to distinguish one pattern with another pattern so that the pattern does not belong to the other so that the pattern of this match only one sound input with voice database which has resulted in data that is best suited for people with a sound level of accuracy that reaches 53.3% for opportunities best. This shows that this method is good enough to be used in the process of the speech recognition.

Keywords: Speech Recognition, Fisher Linear Discriminant Analysis (FLDA), Pattern, Feature Extraction

INTRODUCTION

Speech is major communication between humans naturally and efficiently in order to share information between people in speech (Santosh K, 2010). The process of the speech recognition by human beings began to form as a baby when he was able to hear and he can make a sound. This process unwittingly has done through the learning process, i.e. learning to recognize speech heard. In humans, it is not very difficult to recognize speech which is heard, because humans have the information system which is able to recognize pattern very well. From 2001 to 2010, the accuracy of speech recognition reached 80% and no further progress (Suma Swamy, 2013). Along with the development of technology, humans began to use the technology for speech recognition. But this time, there is no technological device that can recognize spoken language human right. This is due to the difficulty of technological devices to capture the verbal message, then translates and executes the commands contained in this verbal message. DWT (Dynamic Time Warping) method can be used for the data is non-linear in Voice recognition 5 – 10 words with a relatively small error and fast computation (UmaraniJ.Suryawanshi, 2014).

Basically, every human has something unique and it is only owned by himself. This gives rise to the idea of human uniqueness and can serve as identification (AnggoroWicaksono, 2014), (Abdelmajid H. Mansour, 2015). One technique that is based on the identification of human characteristics is the Voice Recognition Attendance System. Selection methods of employee presence in the company uses sound, is due to the direct interaction between users with computers orally. Factors uniqueness of every human type sound, ease of inputting absent, and the accuracy of the voice data reach almost 95% to the runway consideration as well in selecting this system (G. Dahl, 2012). And in the placement of the employee's presence engine installation must be in a separate section and avoid the crowds and noise. One use of speech recognition is beginning to be developed is the use of speech recognition for employee presence engine. Presence system technology has entered the stage where the employee is in the form of voice commands. So it does not need hand key any longer to view the list of attendees sector in institutions, but only by using voice commands to control the machine is able to recognize employees. In the use of speech recognition to record the presence of employees, the required algorithm can be used to perform feature extraction and the speech recognition that allows to obtain a high enough degree of accuracy. Besides, the algorithm must have a speed high enough to recognize voice commands, so the robot can move quickly once ruled by the voice. Computation is needed in the voice recognition (Stephen J. Melnikoff, 2002).

This research made presence system for employees in agencies that are based on natural human characteristic, namely voice, which is used to record the attendance list. The system consists of software with a microphone as the input for generating sound data. The method which is used for the authentication of speech is a FLDA method (Fisher Linear Discriminant Analysis).

METHODOLOGY

FLDA method is an extension of the PCA method (Fitri Damayanti, 2010). By using FLDA method for feature extraction of a sound pattern and the Euclidean Distance as a method of classification and authentication, accuracy rate will

be accurately sounded. In this research, FLDA is proposed to reduce the inter-session variability and increases discrimination speaker. Accurate estimation of this FLDA space of the training dataset is critical to the performance of detection. A typical training dataset, however, does not consist of utterances which are obtained through all sources of interest for each speaker. This has the effect of introducing systematic variations associated with source-speaker speech of the co-variance matrix and produces a complete representation of the in-speaker scatter matrix that is used to FLDA.

In general, Fisher Linear Discriminant Analysis (FLDA) is used to extract a pattern. Fisher Linear Discriminant Analysis (FLDA) is an improved method of Principal Component Analysis (PCA). Disadvantages of the PCA are still the pattern including another pattern that affects the accuracy in extracting a pattern (Desheng Huang, 2009). This method has the advantage in distinguishing one pattern with another pattern so that the most pattern does not include into other patterns (Friday ZinzendoffOkwonu, 2012). Method of Fisher Linear Discriminant Analysis (FLDA) is actually a method of LDA (Linear Discriminant Analysis) because of its ability in the introduction is excellent and there is no doubt, then the method of LDA is often referred to as FLDA (Fisher Linear Discriminant Analysis). Fisher Linear Discriminant Analysis (FLDA) is widely applied for extracting features in a pattern matching process. Basically FLDA method consists of four steps, such as the method of Principal Component Analysis (PCA), Transformation Method of Principal Component Analysis (PCA), the method of Fisher Linear Discriminant Analysis (FLDA) and the transformation of Fisher Linear Discriminant Analysis Method (FLDA).

In the training, there are three types of data required in the operation, such as input data and training data, processing the required data, which provides output data processing software for the execution of the running user. Enter data which is obtained from the preparation of sound data collection. Establishing the sound may starting from the airflow which are produced by the lungs. How it works is much like a piston or a pump which is pressed to generate air pressure. At the time of vocal cord in a state of tension, the air flow will cause the vocal cord vibration and produce speech sounds called voiced speech sound. At the time of vocal cord is in a state of weakness, the air flow goes through a narrow area on the vocal tract and causes turbulence, resulting in a sound that is known as the unvoiced sound. Speech is produced as a series or sequence of the components of its constituent sounds. Each component of the different sounds is created by the difference in the position, shape, and size of the vocal organs of man which can change during the process of speech production.

In addition, a standard microphone is set up to process the sound input into the system is made. The sound input goes through the process of recording the record specification 22050 as sample rate, bit per sample 16 and using mono channel. Subsequent data is saved with the extension (.wav).

The voice data is taken based on the strength of a recorded sound, which is then stored in the form of a matrix. The strength of the sound is represented by the sound indicator. In 1 second, recording indicator sound component produces 280 sound data, because the duration of the sound recording is for 3 seconds then the voice data is for 840 voice data. Visualization results of sound recording and sound reproduction is widened as shown in Figure 1 and Figure 2:

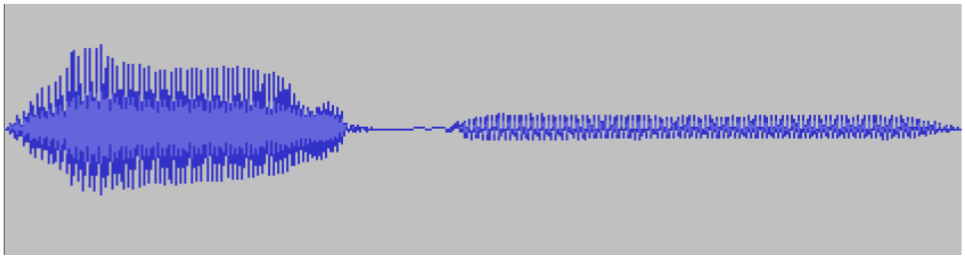

Figure 1. Voice Data nick name

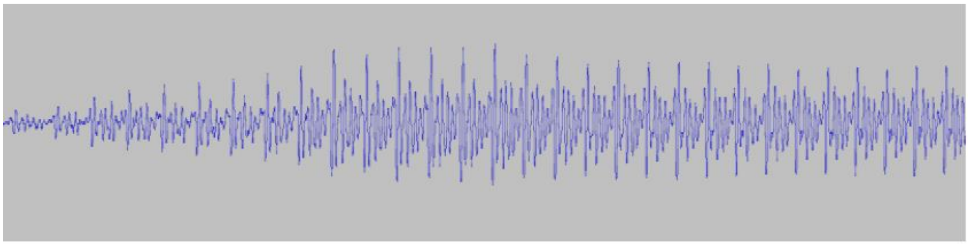

Figure 2. Sound Waves that Expands

Human voice input data consists of 10 classes, each class will be taken as many as 10 data samples, so the data used are 100 voice data. The voice data in the form of time varies signals with relatively slow pace of change. If it is observed at short intervals between 5 to 100 milli-seconds, the practical characteristics are fixed, but if it is observed at a longer visible-interval characteristic, it varies according to the sentence being pronounced. Each line in the figure shows the slice

signal for 100 milliseconds, so the whole of the image shows the speech signal along 500 milliseconds which is shown in Figure 3. The generated output data in this software is a comparison of the results of the recognized voice pattern recognition as the voice of trial data that are in the database by using the Fisher Linear Discriminant Analysis (FLDA).

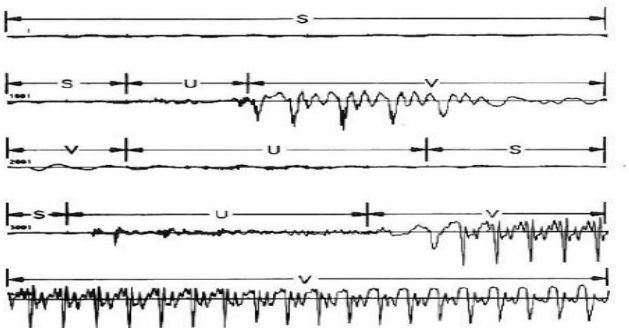

Figure 3. Human Speech Signals (Susetyo Bagas B, 2014)

RESULTS AND DISCUSSION

In the system of FLDA, Voice recognition system for the presence of employees in this study is divided into two sub systems, namely training sub systems and sub system testing, as in Figure 4.

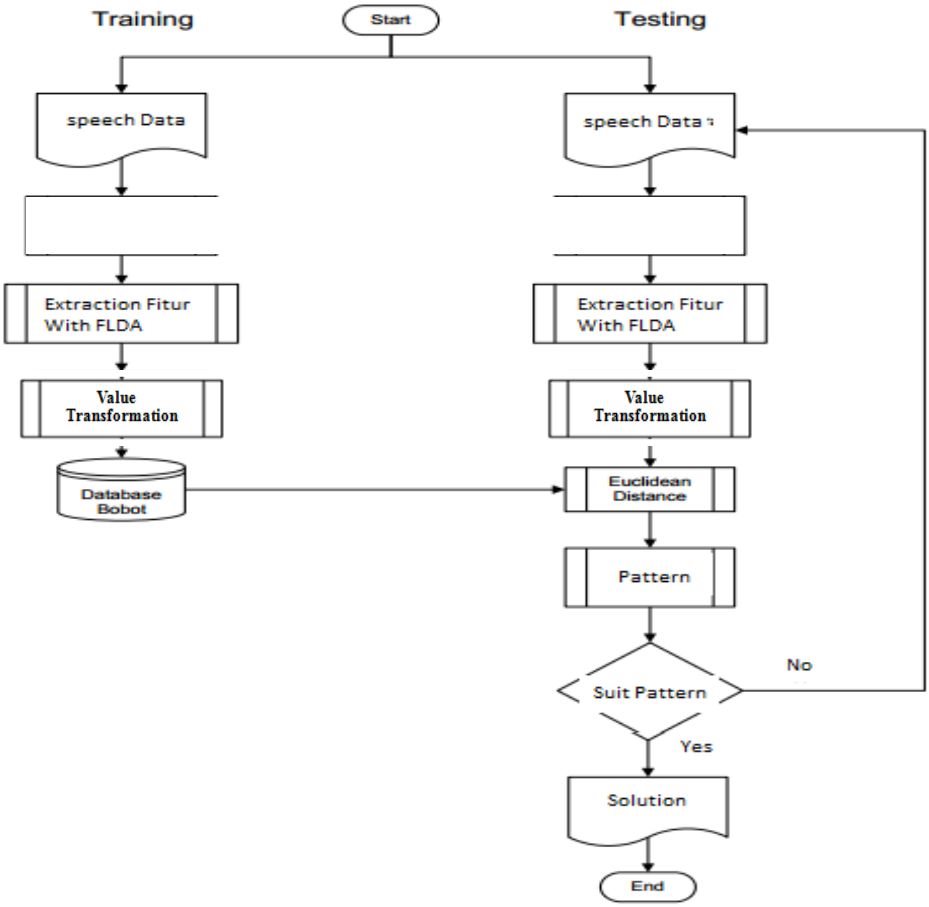

Figure 4. Draft a Model

The process of the normalization is performed on a sound training dataset $x = [x_1, x_2, x_3, ..., x_m]$, it supposes there are m training data size dimensions which are assumed to be n = (widht * high) then the voice data can be formed into a column vector with size (n * 1), shown equation 1. In which each value of the frequency intensity of sound data on an index of all of the training data to n-m. The next column vector normalization = 1 using equation 2.

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (1)$$

$$x = \frac{x_k}{\sqrt{\sum_{k=1}^{N} x_k^2}} \quad (2)$$

After normalization, the results of the matrix is followed by calculating an average matrix to get matrix center, shown in equation 3. Matrix centering can be done by finding the difference matrix dataset with an average matrix before we have to double matrix subtracted average of the total number of voice training (m), as in Equation 4.

$$\mu_{FLDA} = \begin{bmatrix} \frac{x_{11} + x_{12} + \cdots . + x_{1m}}{M} \\ \frac{x_{21} + x_{22} + \cdots . + x_{2M}}{M} \\ \frac{x_{N1} + X_{N2} + \cdots . + X_{NM}}{M} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \quad (3)$$

$$A = X - \mu_{FLDA}$$

$$W_{FLDA} = PCA^T * A$$

$$LD = W_{pca} \quad (4)$$

Voice recognition process is a process of matching the weight of voice which has been trained with the sound stored in the database. They are then searchable within minimum with Euclidean Distance method. This method has been widely used to measure similarity the data with the desired target. Here is the formula to find similarities using Euclidian Distance (Aeri Rachmad, 2015):

Where:

$$\square_{\square} = \sqrt{(\square\square_{1\square} - \square_{1})^2 + (\square\square_{2\square} - \square_{2})^2 + \cdots + (\square\square_{\square\square} - \square_{\square})^2} \quad \textbf{(5)}$$

m : the number of training data

fd_i: Voice Data Training

k : Voice Data Testing

To evaluate the results of FLDA method in this research, it used the accuracy of the percentage of correctness of the data. System testing is done by a collection of recorded sound, where there are 100 speech data which consists of 10 different people with each person per class by performing the recording process 10 times. From 10 speech in each class, there will be 3 types of variations of experimental data.System testing is done by a collection of recorded sound, where there are 100 speech data consists of 10 different people with each person per class by performing the recording process 10 times. From 10 speeches, there will be three types of variations of experimental data. Variations in level 1 used which is seven of data as training data and the data to be three trials in each class, as shown in Table 1 and Table 2 the accuracy of the variation rate of 1.

Tabel 1. Variation Trial Level 1

Variation I																							
Reduce		Training											Testing										
PCA	FLDA	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VII	Class VIII	Class IX	Class X	Total Data	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VII	Class VIII	Class IX	Class X	Total Data
0	0	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	70	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	30
9	1	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	70	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	30
18	2	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	70	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	30
27	3	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	1,2,3,4,5,6,7	70	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	8,9,10	30
		1,2,	1,2,	1,2,	1,2,	1,2,	1,2,	1,2,	1,2,	1,2,	1,2,		8,9,	8,9,	8,9,	8,9,	8,9,	8,9,	8,9,	8,9,	8,9		

Tabel 2. Results of Variation Testing Level 1

Reduce		Σ	True	False	Accuracy (%)
PCA	FLDA				
0	0	30	16	14	53,3
1	1	30	14	16	46,7
2	2	30	13	17	43,3
3	3	30	14	16	46,7
4	4	30	12	18	40
5	5	30	6	24	20
6	6	30	16	14	53,3

However, a variation of level 2 uses five of data as training data and the data 5 as the trial data in each class, it can be seen in table 3 and table 4 is the accuracy of the variations in the level 2.

Tabel 3 Trial of Variation Level 2

Variation II																							
Reduce		Training											Testing										
PCA	FLDA	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VII	Class VIII	Class IX	Class X	Total Data	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VII	Class VIII	Class IX	Class X	Total Data
0	0	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	50	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	50
5	1	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	50	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	50
0	2	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	50	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	50
30	3	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	50	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	50
45	4	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	6,7,8,9,10	50	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	50

Tabel 4 Results of Variation Testing Level 2

Reduce		Class I	Σ	True	False	Accuracy (%)
PCA	FLDA					
0	0	5	50	22	28	44
1	1	5	50	18	32	36
2	2	5	50	22	28	44
3	3	5	50	18	32	36
4	4	5	50	12	38	24

Variation of the 3rd level has used the three data as training data and the data should be seven trials in each class which is shown in table 5 and table 6 shows the accuracy of the variations in the level 3.

Tabel 5. Trial of Variation Level 3

Variation III																							
Reduce		Training											Testing										
PCA	FLDA	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VII	Class VIII	Class IX	Class X	Total Data	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VII	Class VIII	Class IX	Class X	Total Data
0	0	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	30	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	70
27	1	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	30	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	1,2,4,6,8,9,10	70
15	2	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	3,5,7	30	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	1,2,4,6,8,9	70

Tabel 6. Result of Variation Testing Level 3

Reduce		Σ	True	False	Accuracy (%)
PCA	FLDA				
0	0	70	26	44	37,1
1	1	70	24	26	34,3
2	2	70	21	49	30

CONCLUSION

Configuration parameters for FLDA Method in application of speech recognizing in presence of employees in a companywith the seven number on the training data and three for data testing, the highest accuracy at a value of 0 for the reduction of PCA and FLDA approximately 53.3%. They are divided into five training data and five for data testing, the highest accuracy in the value of 0.0 and 2.2 for the reduction of PCA and FLDA raises approximately 44%. Number three is for training data and seven is for the data testing. The highest accuracy comes at a value of 0 for the reduction of PCA and FLDA approximately 37.1%. So, this research can be concluded that if the data were trained more, then the result will be more significant. Recognition at different levels of accuracy was caused by the introduction of sound training data. It is less varied than the test data. In addition, this also affects the recognition, including speech input weak hardware differences and similarities between speech that one does with another speech in the training database.

ACKNOWLEDGMENT

Our thanks go to fellow researchers in the university environment Trunojoyo Madura-Indonesia. The entire team at the lab Multimedia and Networks that have helped the completion of this research and Multimedia Engineering Program and the Network has provided an opportunity to use laboratory equipment.

REFERENCES

G. Dahl, D. Yu, L. Deng, & A. Acero. (2012), “Context-DependentPre-trained Deep Neural Networks for Large Vocabulary SpeechRecognition,” in IEEE Trans. Audio, Speech, and Language Processing. (pp 30-42).

Fitri Damayanti, Agus Zainal Arifin & Rully Soelaiman. (2010) "The introduction of Facial Image Method Using Two-Dimensional Linear Discriminant Analysis and Support Vector Machine". Kursor Journal. 5(3), 147-156.

Anggoro Wicaksono, Sukmawati NE & SatriyoAdhy, Sutikno. (2014) "Speech Recognition Applications Indonesian with Mel-Frequency cepstral method coëficient and Linear Motion Vector Quantization to Control Robot". Proceedings of the National Seminar on Computer Science Undip. (pp 61-66).

Desheng Huang, Yu Quan, Miao He & Baosen Zhou. (2009). "Comparison of linear discriminant analysis methods for the classification of cancer based on gene expression data". Journal of Experimental & Clinical Cancer Research. 28:149.

Friday ZinzendoffOkwonu, Abdul Rahman Othman. (2012). "A Model Classification Technique for Linear Discriminant Analysis for Two Groups". IJCSI International Journal of Computer Science. 9(2).

Susetyo Bagas Bhaskoro, et.al. (2014). Transformasi Pitch Suara Manusia Menggunakan Metode PSOLA. Jurnal TELKOMIKA. 2(2).

UmaraniJ.Suryawanshi, & Prof. Dr. S. R. (2014). Ganorkar, Hardware Implementation of Speech Recognition Using MFCC and Euclidean Distance International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE). 3(8). 11248-11254.

Santosh K. Gaikwad, Bharti W.Gawali,Pravin Yannawar, (2010). A Review on Speech Recognition Technique, International Journal of Computer Applications (0975 – 8887). 10(3).

Suma Swamy & K.V Ramakrishnan. (2013), An Efficient Speech Recognition System, Computer Science & Engineering: An International Journal (CSEIJ). 3(4).

Stephen J. Melnikoff, Steven F. Quigley & Martin J. Russell (Eds). (2002). Speech Recognition on an FPGA Using Discrete and Continuous Hidden Markov Models, 12th International Conference on Field Programmable Logic and Applications FPL. Springer-Verlag Berlin Heidelberg. (pp. 202-211)

Abdelmajid H. Mansour, Gafar Zen Alabdeen Salh, & Khalid A. Mohammed. (2015). Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms. International Journal of Computer Applications (0975 – 8887), 116 (2).

Aeri Rachmad, Muhammad Fuad. (2015). Geometry Algorithm On Skeleton Image Based Semaphore Gesture Recognition, Journal of Theoretical and Applied Information Technology (JATIT). 81(1). 102 - 107.