International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I5P220

Volume 69 Issue 5, 137-146, May 2021

Introduction to Indonesian Syllables Using the LPC Method and the Neural Network of Backpropagation

Mochammad Haldi Widianto #1, Husni Iskandar Pohan[2], Davy Ronald Hermanus[3]

[1] *Informatics Departement, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*
[2] *Informatics Departement, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*
[3] *Informatics Departement, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

[1] mochamad.widianto@binus.ac.id

Abstract *- Sound is the essential component in the development of digital technology today to make human life more manageable. Various voice recognition systems have been developed in multiple countries with multiple languages. This system consists of 4 processes: sound recording, preprocessing, and feature extraction. One method uses Linear Predictive Code (LPC), and the sound classification process uses the Backpropagation Neural Network method. There are 115 syllables and 74 different syllables of the 50 spoken Indonesian words. The total syllables used in Indonesian are 690 syllables from 60 respondents. The accuracy results in the Indonesian syllable recognition system are 100% able to recognize 85 training data from every 20 respondents. The best accuracy is obtained, namely 84% of the 20 respondents who have been tested. Based on the results of the tests that have been carried out, the more training data that is processed in the network, the higher the accuracy of the success obtained.*

Keywords — *Voice recognition, Linear Predictive Code, backpropagation neural network*

I. INTRODUCTION

In this era of rapid technological development, many new technologies make it easier for humans to do their activities, especially in the field of AI (Artificial Intelligence)[1]. in its development, computer systems, especially AI, can perform tasks like humans.

Where previously was done by analog. Computerized voice is a gift from the small, where each human being is given a different voice, having their unique characteristics. With the voice, people can communicate, understand and understand what is being discussed, and one of the most effective ways to convey someone's goals and objectives in sharing information[2],[3], [4] [5], [6].

Sound is the essential component in the development of digital technology today to make human life easier. Therefore, various speech recognition systems or better known as Automatic Speech Recognition (ASR) [7], [8] have been developed in multiple countries with multiple languages. Speech Recognition is a sound identification process based on the words spoken by converting an acoustic signal that is captured by the audio device. Voice recognition [7], [9], [10] is a system used to recognize word commands from a human voice, and then it is translated into data that a computer can understand. Voice recognition can be applied in various fields of life, such as voice-based security systems, voice-based home security systems and voice-based learning systems. Voice-based security systems are more effective and accurate than numbers or letters because they are easy for others to intercept. Likewise, a voice-based learning system makes it easier for a student to learn independently [11].

Indonesia has many studies on voice recognition using other method various methods, among others [12]–[14] and letter detection analysis systems. Voice clustering and using foreign language words as objects of research have been widely studied, among others [15]–[17], .but it is still limited in number and only functions for specific application commands. The level of recognition is influenced by the extraction and classification methods used. In this study, using the Linear Predictive Code (LPC) method for feature extraction and the Backpropagation Neural Network method for sound classification [18]–[22]

The Linear Predictive Code (LPC) method is one of the feature extraction techniques that is often used in extracting voice digital signal features. Sound feature extraction is to convert sound waves into several types of parametric representations that can be processed. The steps taken are Pre-emphasis, Frame Blocking, Windowing, Autocorrelation Analysis and LPC Analysis [23]–[29]. So far, the LPC method is widely used in research related to voice identification. Basically, Indonesian is a language that is often used in communication and is the main language for Indonesian citizens. But everyone has their own voice in accordance with their respective characters and regional accents, which are very influential in Indonesian pronunciation. In the narration of an Indonesian word, each syllable will be accompanied by a breath. This allows for breaks between syllables and results in a different representation of the sound signal of the Indonesian word for each syllable so that the Indonesian speech sound can be recognized at the syllable level.

Knowing the sound pattern makes the Indonesian language learning method easy to recognize by a system. There is not much research on the recognition of Indonesian syllables because Indonesian itself has many syllables compared to other foreign language syllables. Therefore, in this study, the researcher will design a speech recognition system based on Indonesian syllables by implementing the Linear Predictive Code (LPC) method and the Backpropagation Neural Network method to recognize voice signals and the accuracy values obtained in Indonesian syllable speech recognition.

II. RESEARCH METHODOLOGY

In this study, a qualitative approach and the waterfall method were used as a software development approach model with several structured stages: needs analysis, system design, implementation, testing, and maintenance. The algorithm used to build the Indonesian syllable recognition system is the Linear Predictive Code method and the Backpropagation Neural Network method [15], [30].

A. Data Collection

There are 3 stages in data collection in this study:

a) Literature Study

The first stage was carried out at the time of data collection, namely, the literature study. This stage aims to find library sources and get clear information to support the creation of a robust theoretical foundation and the methods to be used in the speech recognition system.

b) Field Study

The second stage, the field study, is collecting data directly into the field using observation to find respondents and word data, documentary studies by voice recording, and focus group discussions.

c) Collecting Data Samples

The third stage is the collection of sound data samples in the Form of consonant verbs, nouns and adjectives in Indonesian. 50 Indonesian words are focused on 1 syllable, 2 syllables and 3 syllables. Total 115 syllables and 74 syllables different from 50 words. The total syllables of 690 syllables from 20 respondents, namely 10 men and 10 women.

III. SYSTEM MODEL AND DESIGN

System design is the initial stage in building an Indonesian syllable recognition system. The design includes a process structure consisting of a voice signal data recording process, a preprocessing, a feature extraction process, and a classification which is divided into 2 process stages, namely the training process and the testing process, which will be represented by simulation software. Then use some software that is used to change the sound file format and cut the sound file, which only happens at the initial stage of the recording process. The structure of the Indonesian syllable recognition system process in general carried out in this study can be seen in Figure 1.

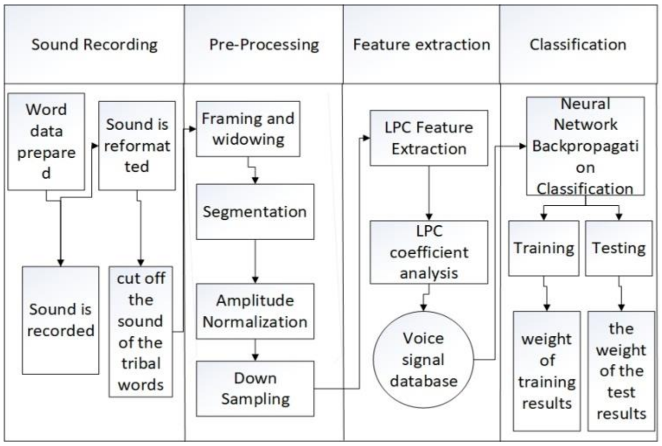

Fig 1. The Process Structure of the Tribal Recognition System Indonesian words

Before recording the sound, data samples were collected as in Table 1. The word data sample consisted of 50 words with 115 syllables to be pronounced during the recording process. Sound recording is done in calm and soundproof conditions. Recorded using the ASUS Zenfone 2 – ZE550ML HandPhone and saved in * .3GPP format. The sound produced from this recording is the word that will become the database in the Indonesian syllable recognition system.

Tabel 1. Data Sample

1 syllable	2 syllable	3 syllable
Cat	I Bu	Be La Jar
Bom	Di A	Be Ker Ja
Lap	Do A	Men Ca Ri
Bor	A Ir	Me Re Ka
Om	A Ku	Me Li Hat
	Ka Mu	Mem Be Ri
	Ka Mi	Ber Ma In
	Per Gi	Ber Ja Lan
	Du Duk	Ber La Ri
	Mi Num	Me Nu Lis
	Ma Kan	Mem Ba Ca
	Ba Ru	Ber Te Mu
	Can Tik	Ber Di Ri
	Sa Kit	Me Mang Gil
	Pa Nas	Ber Ba Gi
	Ba Ik	Me Nyim Pan
	Ma Nis	Me Na Ngis
	Mu Rah	Me Nya Nyi

	Mu Dah	Ter Se Nyum
	Ru Mah	Me Na Rik
	Da Tang	
	To Long	
	Hi Lang	
	Ri Ngan	
	Be Ner	

Before recording the sound, data samples were collected as in Table 1. The word data sample consisted of 50 words with 115 syllables to be pronounced during the recording process. Sound recording is done in calm and soundproof conditions and recorded and saved in * .3GPP format. The sound produced from this recording is the word that will become the database in the Indonesian syllable recognition system. After recording the sound, then change the sound file format to * .Wav format by using the Format Factory application to equalize the sound data file format that will be entered into the system. Data from the sound recording process in * .Wav format will be cut according to the word data that has been provided previously. In this study, the process of cutting sound files was needed to extract the terms contained in the recording file one by one. Missing the sound file is divided into 2 parts, namely cutting the sound manually and cutting the sound automatically. For this stage, you use manual cutting using the Adobe Audition application.

The second process, in Figure 1, is preprocessing, namely, preprocessing of recorded data by reducing the recording effect and amplifying the recorded digital signal. After that, prepare the data in the proper Form for the feature. The extraction process is carried out in two approaches, namely; feature extraction for database creation and feature extraction for entering test data

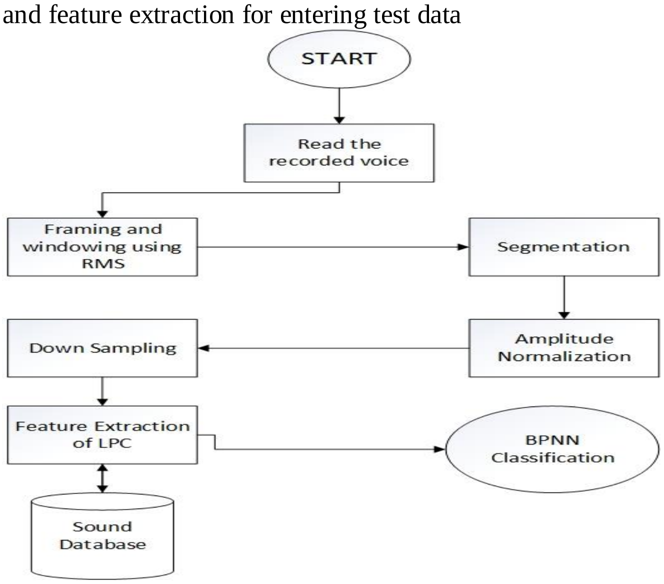

Fig 2. Flowchart of Schematic Pre-processing and Feature Extraction

Flowcharts of preprocessing schemes and feature extraction from the system flow that are made are depicted simply using symbols that are easy to understand.

A. Root Mean Square

The flowchart is shown in Figure 2, starting from reading the * .WAV training voice recording followed by the Framing and windowing process using RMS (Root Mean Square) to see the energy in the sound signal the cutting process can be cut precisely and efficiently. Root Mean Square is the root of the average value of a function squared. To calculate the RMS value or the effective value of a part, the first thing to do is to square the process. As in formula (1)

$$X_{RMS} = \sqrt{\frac{1}{n}.\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 .... + x_n^2}{n}} \qquad (1)$$

Where x RMS is the RMS value, n is the amount of data and xi is the ith data value [31]. In this study, conducting an experiment by entering the window value on the input sound signal between the values 50, 75 and 100 to determine the optimal window value for use in other sound signal samples.

B. Amplitudo Normization

Furthermore, the Amplitude Normalization process is a process used to normalize the degradation of the digital signal sample value due to the difference in the distance between the mouth and the recording microphone. The amplitude normalization process is obtained by dividing all values of the digital signal sample by the maximum absolute value of the digital signal sample. Can be seen in formula 2.

$$x'(n) = \frac{x(n)}{\max(|x|)}, \; 0 \le n \le N-1 \qquad (2)$$

where x '(n) is the value of the normalized amplitude, x (n) is the sample value of the digital signal, max (| x |) ' is the maximum absolute value, and N is the length signal [32].

C. LPC

After the downsampling process, the main process of extracting LPC features is the LPC coefficient analysis process. LPC coefficient analysis is a process to determine the parameter value of Linear Prediction Coding (LPC) by analyzing each input value to obtain the best value. The results of the LPC coefficient analysis are the predicted sound signal results. This study uses the LPC toolbox and analyses the coefficient value of the value 50 and the value of 100 as parameters to determine the best coefficient value. By analyzing the LPC coefficient value, there will be differences in the characteristics of the digital voice signal with a coefficient of 50 and a coefficient of 100. The digital sound signal has a significant difference in the coefficient value that has been determined, and then the coefficient value is optimal for use in other sound signals.

The Backpropagation Neural Network, a sound classification process, is used to classify a certain input signal pattern by improving the weight of the conductor between layers. The classification process is divided into 2 parts: the training process (training) and the testing process (testing). In backpropagation, if the data used is too large, the system will take a long time to find patterns to reach the target. Conversely, if the data is too little, the system cannot recognize the data properly, making it difficult to reach the target. Therefore, data sharing was carried out in this study, namely data for training (74 syllables) and data for testing (115 syllables).

Cat	Per	Ik	La	Lis
Bom	Gi	Nis	Jar	Te
Lap	Du	Rah	Ker	Mang
Bor	Duk	Dah	Ja	Gil
Om	Num	Mah	Men	Nyim
I	Ma	Da	Ca	Pan
Bu	Kan	Tang	Me	Na
Di	Ba	To	Re	Ngis
A	Ru	Long	Li	Nya
Do	Can	Hi	Hat	Nyi
Ir	Tik	Lang	Mem	Ter
Ku	Sa	Ri	Ber	Se
Ka	Kit	Ngan	In	Nyum
Mu	Pa	Be	Lan	Rik
Mi	Nas	Nar	Nu	

Fig 3. Sample Training Data 74 Syllables

The data sample in Table 3 is a sample of data taken from 115 syllables so that 74 different syllables are obtained as in the table above. The sample test data for 115 syllables can be seen in Table 2. Things that need to be considered in the training process are initializing the weights and determining the appropriate parameters to facilitate voice recognition. As shown in Figure 3 below, the Backpropagation Network Architecture used in this study consists of four layers: the input layer, hidden layer 1, hidden layer 2 and the output layer. Input is variable with (In), Hidden layer 1 is variable (Jn) with weight (wij) and bias (oj), Hidden layer 2 is variable (Kn) with weight (wjk) and bias (ok) and in Output layer it is variable (Ln) with weight (wkl) and bias (ol).

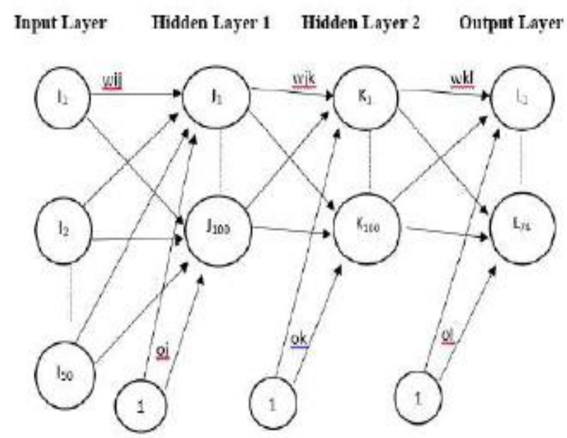

Fig 4. Backpropagation Network Architecture

There is no computation process at the Input layer, only sending the input signal I to the hidden layer. There is a computation process of weight and bias in the hidden layer and the output layer, and the amount of output from the hidden and output layers is calculated based on the binary sigmoid activation function because the expected output is between 0 and 1.

Tabel 2. Backpropagation Network Architecture Parameters

Input Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3
50	100	100	74

Before carrying out the training process, first, determine the parameter value on the network so that it can recognize sounds properly. In Table 2 where the input layer has a value of 50, which is taken from the value of the LPC coefficient, hidden layer 1 and hidden layer 2 has a value of 100 taken from the largest value that exceeds the value of the output layer and the value of 74 output layers is taken based on the trained voice sample data..

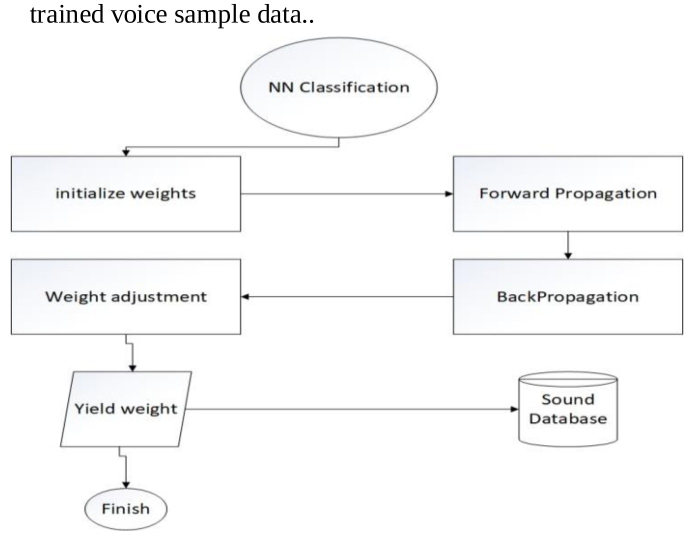

Fig 5. Schema Introduction System Flowchart Training

In the training process, the system will receive input in the form of samples used as training data. The training data will be stored in a database used as a reference in the testing process. The next step is based on a flowchart. Figure 5 above, namely the forward phase (Forward Propagation) process, aims to trace the magnitude of the error. By the backpropagation network architecture, the calculation formula is obtained as in formula 3.

$$yj = \sigma\left(\sum_{i=1}^{n} wij\, xi + bj\right) \tag{3}$$

where yj is the output, σ (sigma) is the symbol of the activation function, n is the number of neurons in the input layer, wij is the weight value of the input layer and hidden layer, xi is the input x, and bj is the bias value. The activation function used is the binary sigmoid activation function. The binary sigmoid activation function has a value in the range of 0 to 1 [15]. It is defined as in formula 4.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4}$$

Where σ () is the symbol of the activation function, x is the value of the output signal from one neuron to be activated, and e is a constant value with the value = 2.718281828. Calculating Error (RMSE) can be seen in formula 5.

$$error = \frac{1}{n}\sum_{i=1}^{n}\left(t_i - p_i\right)^2 \tag{5}$$

where n is the number of data, i is the sequence of data in the database, ti is the target result value, and pi is the predicted value.

Backpropagation in Figure 6 above aims to update the weights by doing back calculations of the output neurons in order to have the appropriate weight value. In this study, updating weights was carried out without using momentum. And adjusting the weight, if the output from the network is different from the expected target, the network will make adjustments to the existing weights. The process will continue until the output on the network and target is the same. Fix a weight (w) based on error (E) with the formula 6.

$$w_{new} = w_{old} - \alpha\frac{\partial E}{\partial w} \tag{6}$$

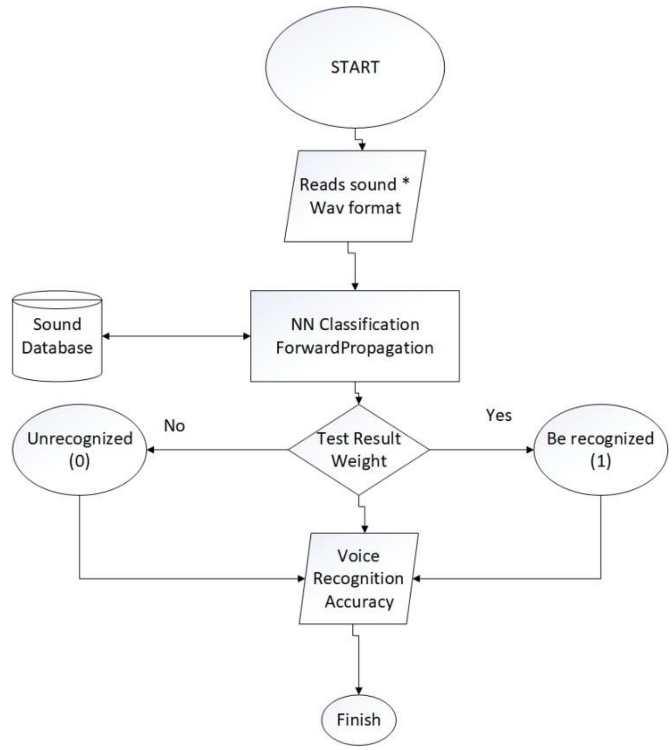

Fig 6. Flowchart of the Testing Scheme Introduction System

The last process is the testing process shown in the flowchart in Figure 6 above. The system will be tested by entering a sample and comparing it with existing training data, and then the results will come out based on the similarity of the training data. The voice data used are samples of testing data for 115 syllables and training data samples of 74 syllables to be recognized according to the actual classification or cannot be recognized at all, and the accuracy of voice recognition is calculated on a percentage scale. Only one test is carried out in the testing phase, namely in the advanced phase, which will produce the test weight. The percentage of system success can be found with the following formula 8:

$$RESULT = \frac{\sum successful\ data}{\sum input\ data} x100\% \tag{7}$$

where Σ successful data is the number of successfully recognized test data, and Σ input data is the total amount of input data that will be tested. With this formula, the researcher can find out the success rate of the system.

IV. EXPERIMENTAL RESULT AND DISCUSSION

The goal to be achieved in this system is a system that can store sound in the form of syllables with various types and variations spoken by each respondent. The system can only cast votes from people who have increased so that if a vote is a cast that is not stored in the database, it cannot be recognized. The voice recognition process steps have been carried out according to the procedure described in Figure 1 or flowchart Figure 2, Figure 4 and Figure 5, implemented in the program, and the results of the process will be seen after the program is run.

The following shows the results of the sound recording process that has been carried out. This sound sample is input which is represented as a speech signal in the form of a matrix

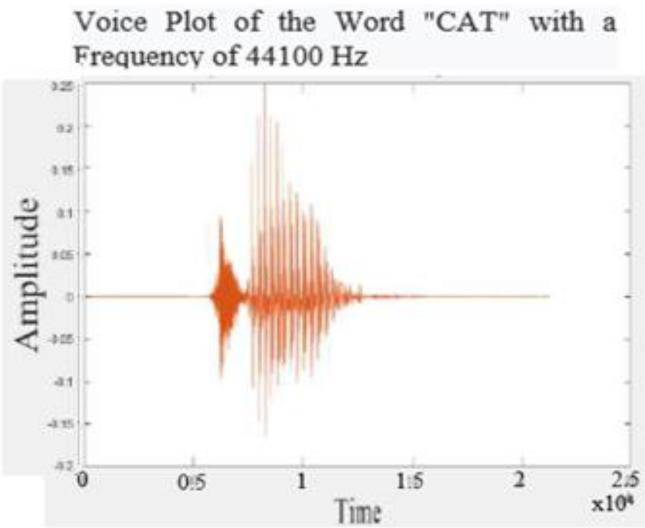

Fig 7. Wav Representation in Graph Form on the word "CAT"

The graph in Figure 7 above, shows the results of the matlab program commands which are represented in graphical Form. Voice signal samples were taken from one of the respondents in the Form of * .WAV files with one of the words "Cat" which had been cut according to the syllable with a frequency of 44100 Hz.

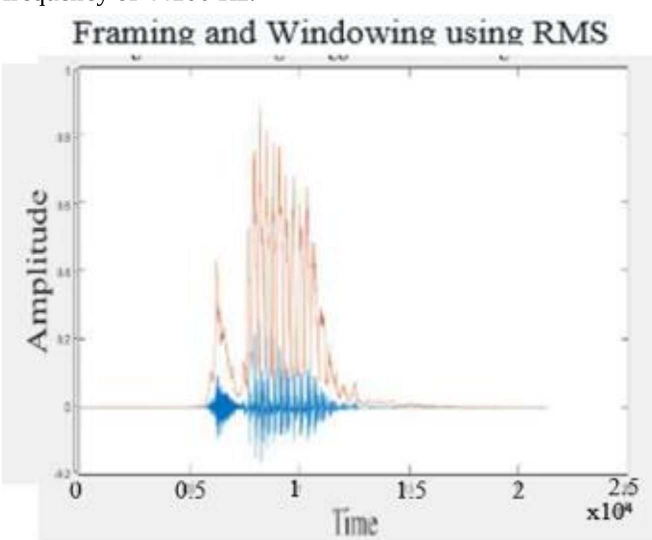

Fig 8. Graph Framing and Windowing Using RMS with a Window Value of 75

The graphic in Figure 8 is one of the framing and windowing process results using RMS with the word "Cat" which is represented in graphic form. Enter the windowing value on the voice input signal, which is 75 that has been determined. Based on this figure, it can be seen that the energy form in the sound signal where the window 75 value looks optimal, not too dense or loose to be used for other sound signal samples. Providing this window value aims to simplify the cutting process at the segmentation stage.

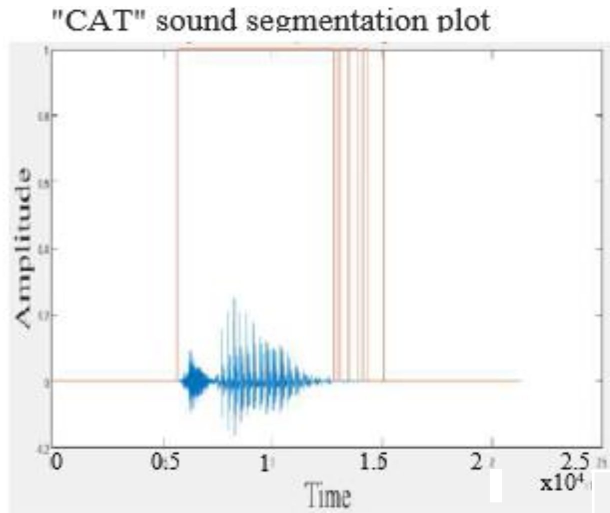

Fig 9. Segmentation Graph of Voice Signal on 'Cat' Word

Figure 9 above is the result of segmentation. It can be seen that determining the threshold value will automatically display the results of separating the sound signal from noise by looking for the energy value & Zero Crossing from the sound sample. The threshold value for each sound sample varies according to the sound condition to be cut.

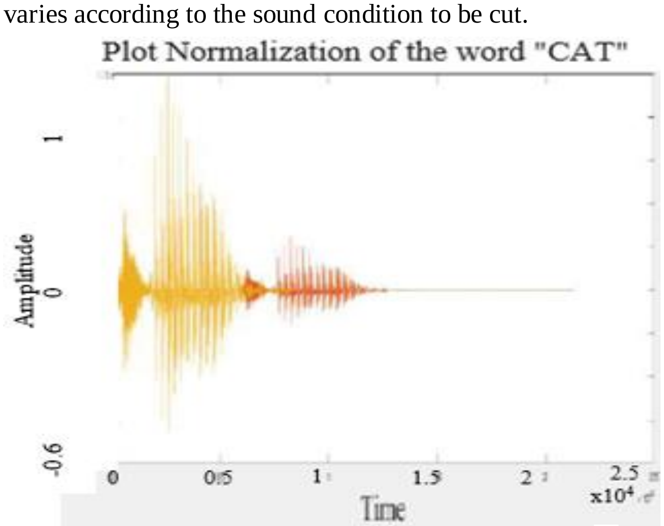

Fig 10. Graph of Normalized Amplitude on the Word 'CAT'

In Figure 10 above, you can see the results of amplitude normalization by normalizing the value degradation sound signal samples so that they have the same amplitude value, namely 1.

Samples of sound signals that have different amplitudes, such as having an amplitude of 0.25 or 0.8, will be normalized to 1 so that the sound signal sample has an equivalent amplitude value of a maximum of 1.

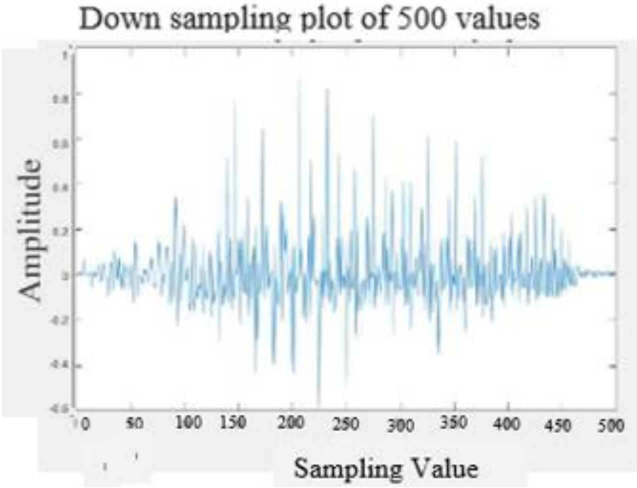

Fig 11 Down Sampling Graph with 500 Sampling Value

The graph in Figure 11 above is the result of lowering the sampling value of one of the sound signals with the word "Cat" taken from female respondents. The sampling value is taken from half the value of the sound signal, namely 500 samplings with an amplitude value of 1. The sound signal does not look solid or very dense so that it can facilitate the extraction of features. The sampling value that has been determined can represent the missing samples.

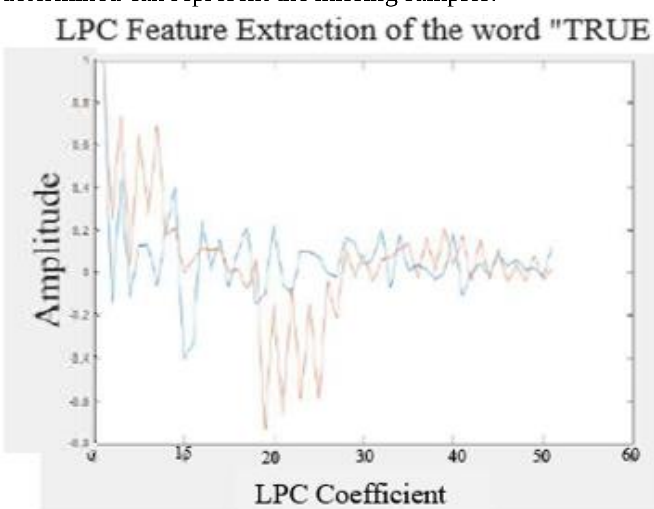

Fig 12. LPC coefficient value 50, downsampling 500 (Respondent 1 male)

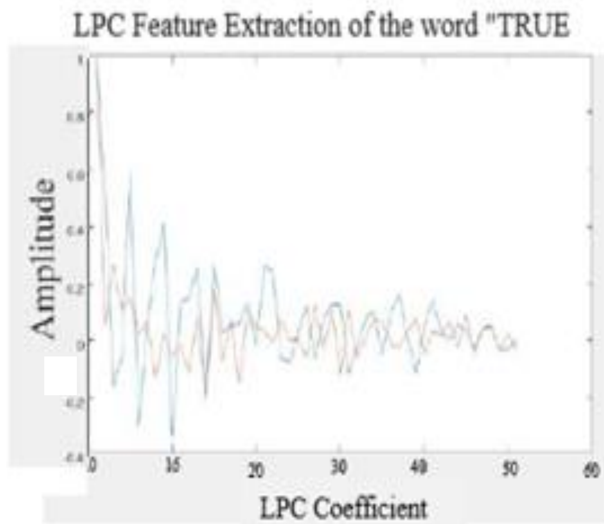

Fig 13. LPC coefficient value 50, downsampling 500 (Respondents 2 women)

Figure 12 and Figure 13 above show the result of feature extraction by providing a coefficient value LPC 50 that has been determined and the Down Sampling value of 500 to get the characteristics of the sound signal. One of the words spoken was the word "True" with 2 different respondents. The LPC feature extraction results represented in the form of a graph show a significant difference between the two sound signals, each of which has different sound signal characteristics. The results of this LPC feature extraction will be a database in training and testing.

A. Training Results

The training results are arranged in a table so that they can be compared to determine the fastest learning rate and epoch conditions. The results of changes in the learning rate and epoch are as shown in Table 5.

Tabel 3. Results of Experiment Examples of Determination of Learning Rate On 1 Training Data

Epoch	Learning Rate	RMSE
>= 3000	0.01	0.0245
	0.02	0.0198
	0.05	0.0149

In Table 3 above, you can see the difference between each learning rate and RMSE value with the same epoch value. It shows that the most optimal learning rate value occurs at a value of 0.05 with an error value (RMSE) of 0.0149, which has the smallest error value compared to the other values. It's just that the value on the epoch is less than optimal because the error results haven't reached the target. The greater the learning rate value, the faster the training process depends on the specified epoch value

Table 4. Results of Training Data from an example of 20 Respondents with the Binary Sigmoid Activation Function

Respondent number	Learning Rate	Epoch	RMSE
1	0.05	5000	0.0123
2	0.05	5000	0.0098
3	0.05	5000	0.0134
4	0.05	5000	0.0097
5	0.05	5000	0.0114
6	0.05	5000	0.0119
7	0.05	5000	0.0099
8	0.05	5000	0.0168
9	0.05	5000	0.0166
10	0.05	5000	0.011
11	0.05	5000	0.0132
12	0.05	5000	0.0092
13	0.05	5000	0.0143
14	0.05	5000	0.0132
15	0.05	5000	0.0096
16	0.05	5000	0.0122
17	0.05	5000	0.0144
18	0.05	5000	0.0097
19	0.05	5000	0.0122
20	0.05	5000	0.0111

After conducting the learning rate experiment on 1 training data, the correct learning rate value will be known. The researcher can see in Table 4 that the training results on 6 training data respondents have the smallest error value that is different. The epoch value takes a value greater than 3000 to 5000 epoch because, at the 5000 epoch, the error limit used is reached, namely 0.01. These parameters provide the most optimal training time compared to others.

B. Test result

In this system, the test is carried out in 2 stages. This training voice test is the first test carried out on the previously trained voice input. This test aims to determine whether the system can recognize or not recognize previously trained sounds. The training data consists of 74 syllables. The test results can be seen in Table 5.

Table 5. Results of Testing Accuracy for 74 Data Samples

Respondent number	Learning Rate	Accuracy
1	0.05	100%
2	0.05	100%
3	0.05	100%
4	0.05	100%
5	0.05	100%
6	0.05	100%
7	0.05	100%
8	0.05	100%
9	0.05	100%
10	0.05	100%
11	0.05	100%
12	0.05	100%
13	0.05	100%
14	0.05	100%
15	0.05	100%
16	0.05	100%
17	0.05	100%
18	0.05	100%
19	0.05	100%
20	0.05	100%

The success rate in the testing process on 74 samples of training data with 6 respondents was 100%. The system can perfectly recognize the data sample. The second stage, testing 115 samples of untrained test data from a sample of 20 respondents, can be seen in Table 6.

Table 6. Results of Testing Accuracy of 115 Data Samples

Respondent number	Learning Rate	Accuracy
1	0.05	65%
2	0.05	67%
3	0.05	69%
4	0.05	68.8%
5	0.05	65.7%
6	0.05	66.3%
7	0.05	66%
8	0.05	69%
9	0.05	67.6%
10	0.05	66.9%

11	**0.05**	**68.2%**
12	**0.05**	**69.3%**
13	**0.05**	**67%**
14	**0.05**	**65.8%**
15	**0.05**	**66.5%**
16	**0.05**	**65.7%**
17	**0.05**	**68.9%**
18	**0.05**	**65.8%**
19	**0.05**	**67.9%**
20	**0.05**	**68.8%**

Based on Table 6 above, the highest accuracy test is at 69%, and the test with the lowest accuracy is at a value of 65%. So based on the results of testing with different levels of accuracy, it shows that the condition of the respondent when they say a word and the data sample that has not been trained is very influencing the percentage value obtained

V. CONCLUSION

Based on the experimental results, it can be concluded that the accuracy of the Indonesian syllable recognition system using the Linear Predictive Code (LPC) method and the Backpropagation Neural Network method has a success rate of voice testing for 74 samples reaching 100% of every 20 respondents. The sound test success rate for the 115 untrained test data reached a high of 68.8%. The condition of the respondent when speaking a word and the untrained data sample greatly affects the percentage value obtained because the more training data that is processed in the network, the higher the accuracy of success obtained (voice signals can be recognized).

REFERENCES

[1] H. Li et al., Improving api caveats accessibility by mining api caveats knowledge graph, Proc. - 2018 IEEE Int. Conf. Softw. Maint. Evol. ICSME 2018, (2018) 183–193.

[2] M. H. Widianto, Analysis of Pharmaceutical Company Websites using Innovation Diffusion Theory and Technology Acceptance Model, Adv. Sci. Technol. Eng. Syst. J., 6(1)(2021) 464–471.

[3] J. Ferreira, H. G. Oliveira, and R. Rodrigues, NLPyport: Named entity recognition with CRF and rule-based relation extraction, CEUR Workshop Proc., 2421(2019) 468–477.

[4] K. Drivers, C. Requirements, and S. Architectures, 6G Technologies: Key Drivers, Core Requirements, System Architectures, and Enabling Technologies, IEEE Veh. Technol. Mag., PP(2019) 1.

[5] İ. Kahramanoğlu, E. D. Ülker, and S. Ülker, Integrating support vector machine (SVM) technique and contact imaging for fast estimation of the leaf chlorophyll contents of strawberry plants, SSRG Int. J. Eng. Trends Technol., 69(3)(2021) 23–28.

[6] H. Jung, Analysis of subthreshold characteristics for top and bottom flat-band voltages of junction less double gate MOSFET, SSRG Int. J. Eng. Trends Technol., 69(3)(2021) 1–6.

[7] A. S. Subramanian et al., Far-Field Location Guided Target Speech Extraction Using End-to-End Speech Recognition Objectives, ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 2020-May, (2020) 7299–7303.

[8] L. Li, D. Wang, Y. Chen, Y. Shi, Z. Tang, and T. F. Zheng, Deep Factorization for Speech Signal, ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., 2018-April, (2018)5094–5098.

[9] C. Fan, J. Yi, J. Tao, Z. Tian, B. Liu, and Z. Wen, Framework for Robust End-to-End, 29(2021) 198–209.

[10] Y. Suh et al., Development of distant multi-channel speech and noise databases for speech recognition by in-door conversational robots, 2017 20th Conf. Orient. Chapter Int. Comm. Coord. Stand. Speech Databases Assess. Tech. O-COCOSDA 2017, no. November, (2018) 1–4.

[11] A. C. Njaka, N. Li, and L. Li, Voice Controlled Smart Mirror with Multifactor Authentication, 2018 IEEE Int. Smart Cities Conf. ISC2 2018, 2019.

[12] G. I. Shamir, Universal source coding for monotonic and fast decaying monotonic distributions, IEEE Trans. Inf. Theory, 59(11) (2013) 7194–7211.

[13] M. Teduh Uliniansyah, H. Riza, A. Santosa, Gunarso, M. Gunawan, and E. Nurfadhilah, Development of text and speech corpus for an Indonesian speech-to-speech translation system, 2017 20th Conf. Orient. Chapter Int. Comm. Coord. Stand. Speech Databases Assess. Tech. O-COCOSDA 2017, 2018-Janua, no. November, (2018)53–57.

[14] H. Prakoso, R. Ferdiana, and R. Hartanto, Indonesian Automatic Speech Recognition system using CMUSphinx toolkit and limited dataset, 2016 Int. Symp. Electron. Smart Devices, ISESD 2016, 283–286, 2017.

[15] J. J. Lasiman and D. P. Lestari, Speech Emotion Recognition for Indonesian Language Using Long Short-Term Memory, 2018 Int. Conf. Comput. Control. Informatics its Appl. Recent Challenges Mach. Learn. Comput. Appl. IC3INA 2018 - Proceeding, (2019) 40–43.

[16] N. B. Wunarso and Y. E. Soelistio, Towards Indonesian speech-emotion automatic recognition (I-SPEAR), arXiv, (2017) 98–101.

[17] H. P. Bluem et al., First Lasing From a High-Power Cylindrical Grating Smith-Purcell Device, IEEE Trans. Plasma Sci., 43(9)(2015) 3176–3184.

[18] L. Goswami, M. K. Kaushik, R. Sikka, V. Anand, K. Prasad Sharma, and M. Singh Solanki, IOT Based Fault Detection of Underground Cables through Node MCU Module, 2020 Int. Conf. Comput. Sci. Eng. Appl. ICCSEA 2020, 2020.

[19] P. N. Saranu, G. Abirami, S. Sivakumar, K. M. Ramesh, U. Arul, and J. Seetha, Theft Detection System using PIR Sensor, Proc. 4th Int. Conf. Electr. Energy Syst. ICEES 2018, (2018) 656–660.

[20] M. Unger, G. Fries, T. Steinecke, C. Waghmare, and R. Ramaswamy, Functional Safety Test Strategy for Automotive Microcontrollers during Electro-Magnetic Compatibility Characterization, EMC COMPO 2019 - 2019 12th Int. Work. Electromagn. Compat. Integr. Circuits, (2019) 49–51.

[21] R. M. Kingsta, A. S. Saumi, and P. Saranya, Design and construction of arduino based pH control system for household waste water reuse, Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019, no. Icoei, (2019) 1037–1041.

[22] V. C. Pinto, P. J. Sousa, V. H. Magalhães, C. F. Araújo, G. Minas, and M. Luís, COST-EFECTIVE LAB-ON-A-CHIP DEVICE FOR SEAWATER PH QUANTIFICATION BY OPTICAL METHODS Microelectromechanical Systems Research Unit (CMEMS - UMinho), University of Minho Campus de Azurém , Guimarães , Portugal Design and fabrication of the microfluidic di," no. June, (2019) 2266–2269.

[23] C. Yan, C. Wei, G. Donghui, and W. Eric, System design of smart card application for tank truck conveying management, 2007 IEEE Int. Work. Anti-counterfeiting, Secur. Identification, ASID, 2(2007) 398–401.

[24] M. Sheth and P. Rupani, Smart Gardening Automation using IoT with BLYNK App," Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019, 2019-April, no. Icoei, (2019) 266–270.

[25] M. S. Hadi, P. Adi Nugraha, I. M. Wirawan, I. Ari Elbaith Zaeni, M. A. Mizar, and M. Irvan, IoT Based Smart Garden Irrigation System, 4th Int. Conf. Vocat. Educ. Training, ICOVET 2020, pp. 361–365, 2020.

[26] D. J. Frp, P. R. Frp, V. J. Frp, E. E. Sdwkln, J. Frp, and P. J. Frp, "6PDUW * DUGHQLQJ $ VVLVWDQFH 6 \ VWHP ZLWK

WKH, (2019) 1–6.
[27] Z. Yuan, X. Zheng, and C. Xue, Modeling and application of digital garden city based on GIS, IET Conf. Publ., vol. 2013, no. 635 CP, (2013) 71–76.
[28] S. Eresheim, R. Luh, and S. Schrittwieser, On the Impact of Kernel Code Vulnerabilities in IoT Devices, Proc. - 2017 Int. Conf. Softw. Secur. Assur. ICSSA 2017, (2018)1–5.
[29] E. Mackensen, J. Klose, A. Rombach, and A. Spitznagel, Energy autonomous automation of Smart Home applications using the example of a wireless Indoor Smart Gardening system, IEEE Int. Conf. Autom. Sci. Eng., vol. 2019-Augus, (2019)1087–1092.
[30] K. S. Hulyalkar S., Deshpande R., Makode K., Implementation of Smartbin Using Convolutional Neural Networks, Int. Res. J. Eng. Technol., 5(4)(2018) 3352–3358.
[31] B. Zhang and R. Sennrich, Root Mean Square Layer Normalization, arXiv, no. NeurIPS, (2019) 1–14.
[32] A. Pisal, R. Sor, and K. S. Kinage, Facial Feature Extraction Using Hierarchical MAX(HMAX) Method, 2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2017, no. figure 2,(2018) 1–5.