Speech Recognition using Linear Predictive Coding (LPC) and Adaptive Neuro-Fuzzy (ANFIS) to Control 5 DoF Arm Robot

W.S. Mada Sanjaya\textsuperscript{1,2}, Dyah Anggraeni\textsuperscript{1,2} and Ikhsan Purnama Santika\textsuperscript{1,2}
\textsuperscript{1}Department of Physics, Faculty of Science and Technology, Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
\textsuperscript{2}Bolabot Techno Robotic Institute, CV. Sanjaya Star Group, Bandung, Indonesia

E-mail: madasws@gmail.com, tsugumikaoru@gmail.com

ABSTRACT. This research shows the implementation of speech recognition to control arm robot. The method to identify the speech recognition using Linear Predictive Coding (LPC) and Adaptive Neuro-Fuzzy Inference System (ANFIS). LPC method used to feature extraction the signal of speech and ANFIS method used to learn the speech recognition. The data learning which used to ANFIS processed are 6 features. The examination system of speech identification using trained and not trained data. The result of the research shows the successful grade for trained speech data is 88.75% and not trained data is 78.78%. Identification of speech recognition system was applied to controlled arm robot based on Arduino microcontroller.

1. Introduction
The automation system, especially in robotics is a system which can help human works [1]. Robotics system can be controlled by human or programmed on robot processor. Speech recognition (or speech control) can be an alternative to help disability. In control system based on speech recognition needed a method to features extraction and learning method as a database (trained data) to identify the features.

They are some method to extract the features of speech, such as Linear Predictive Coding (LPC) method [2] [3] and Mel-Frequency Cepstrum Coefficient (MFCC) [4] [5]. This method can be applied in some field, such as; control smart home [6] [5], control mobile robot [7], control wheelchair [8], control robot [9], [10], biomatrix [11], speaker identification [2], control arm robot [12] [13]. To learning and classify of speech method, have been investigating using Artificial Neural Networks (ANN) [2] [3] [14] [15], Neuro-Fuzzy [6], and other soft computing [16] [17].

In this research will be described an extraction features by using Linear Predictive Coding and learning method to classify the speech by using Adaptive Neuro-Fuzzy Inference System (ANFIS). Finally, this research will be implemented the speech recognition to control arm robot for placing an object.

The paper is organized as follows. In section 2, describe the theoretical background of LPC and ANFIS details. In section 3, describe the method and the system design of the research. In section 4, describe a hardware design of arm robot. In section 5, describe the application of speech recognition in detail. Finally, in Section 6 the concluding remarks are given.
2. Theoretical Background

2.1 Linear Predictive Coding (LPC) Method

A signal processing is an activity to extract a signal information. Linear Predictive Coding (LPC) is a powerful speech analysis technique and facilitating a features extraction which has a good quality and efficient result for computing. In 1978, LPC uses to make a speech synthesis. LPC doing an analysis with predicting a formant decided a formant from signal called inverse filtering, then estimated an intensity and frequency from residue speech signal. Because speech signal has many variations depending on a time, the estimation will do to cut a signal called frame. Procedure to get an LPC coefficient shown in Figure 1.

![LPC Method Diagram](image)

**Figure 1.** LPC Method Diagram

2.1.1 Preemphasis. On processing of speech signal, preemphasis filter needed after sampling process. The filtering purpose is to get a smooth spectral shape of the speech signal. A spectral which have a high value for the low-frequency field and decrease for field frequency higher than 2000 Hz. Preemphasis filter based on the relation of input/output on time domain which is shown by the equation (1),

\[ y(n) = x(n) - ax(n - 1), \]

\[ a \text{ is a constant of preemphasis filter, ordinary have } 0.9 < a < 1.0. \]

2.1.2 Frame Blocking. Frame Blocking. On this process, segmented of speech signal become some frame which overlaps. So that no signal is lost (deletion).

2.1.3 Windowing. Analog signal which converts become digital signal read frame by frame and each frame is windowing with the certain window function. This windowing process purpose to minimize discontinue signal from initial to end of each frame. If window as \( w(n) \), \( 0 \leq n \leq N - 1 \), when \( N \) is total of sample of each frame, thus result of windowing is a signal:

\[ y_1(n) = x_1(n)w(n), 0 \leq n \leq N - 1. \]

Where \( w(n) \) usually use window Hamming with the form:

\[ w(n) = 0.54 - 0.46 \cos \left( \frac{2\pi n}{N-1} \right), 0 \leq n \leq N - 1. \]

2.1.4 Autocorellation Analysis. The next step is autocorrelation analysis toward each frame result by windowing \( y_1(n) \) with equation (4),

\[ r_1(m) = \sum_{n=0}^{N-1-m} y_1(n)y_1(n + m), m = 0,1,2\ldots p, \]

Where \( p \) is ordered from LPC. LPC order which usually used is between 8 until 16.
2.1.5 LPC Analysis. This step will convert each frame from $p+1$ autocorrelation become compilation of “LPC parameter” $a_m = a_m^p$ for $m = 1, 2, ... p$. This compilation becomes LPC coefficient or become other LPC transformation. The formal method to change autocorrelation coefficient become parameter LPC compilation called Durbin method, the form as:

$$E^0 = r(0)$$

$$k_m = \frac{[r(m) - \sum_{j=1}^{m-1} a_j^{(m-1)} r(j-n)]}{E^{(m-1)}} , \quad 1 \leq m \leq p,$$

$$a_m^m = k_m,$$  \hspace{1cm} (6)

$$\alpha_j^m = \alpha_j^{(m-1)} - k_m a_j^{(m-1)}, \quad 1 \leq j \leq m - 1,$$  \hspace{1cm} (7)

$$E^m = (1 - k_m^2)E^{(m-1)}.$$  \hspace{1cm} (8)

With $r(0)$ is a result of autocorrelation, $E(m)$ is an error, $k_m$ is rebound of the coefficient, $\alpha_j^m$ is prediction coefficient for $1 \leq j \leq m$.

2.1.6 LPC Parameter Conversion to LPC Coefficient. LPC coefficient parameter $a_m$ conversion to cepstral coefficient $c_m$ to get the best performance and endure to noise, as the questions:

$$c_m = a_m + \sum_{k=1}^{m-1} (k/m)c_k a_{m-k}, \quad 1 \leq m \leq p$$  \hspace{1cm} (10)

$$c_m = \sum_{k=1}^{m-1} (k/m)c_k a_{m-k}, \quad m > p$$  \hspace{1cm} (11)

This cepstral coefficient is representation coefficient of logarithm spectrum.

2.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

Neuro-fuzzy is relation from fuzzy logic and artificial neural networks. Neuro-fuzzy is the relation between fuzzy logic and artificial neural networks. Neuro-fuzzy system based on inference of fuzzy system which tests using learning algorithm which derivate from artificial neural networks system. Thus, neuro-fuzzy has an advantage which has by the interference fuzzy system and artificial neural networks system. From the learning capability, the neuro-fuzzy system can be called as ANFIS (Adaptive Neuro-Fuzzy Inference System).

![Figure 2. ANFIS Roger Jang [18]](image)

As shown as Figure 2, the fuzzy structure which applied is the fuzzy model by Takagi-Sugeno-Kang. In ANFIS system consist 5 layers, described below;
2.2.1 Layer 1: Fuzzification. The output from node $i$ on layer 1 can be noticed as $O_{1,i}$. Each node on layer 1 is an adaptive node, with the output:

$$O_{1,i} = \mu_{A_i}(x), \quad i = 1,2$$

$$O_{1,i} = \mu_{B_i}(y), \quad i = 1,2$$

Where $x$ and $y$ is an input value for the node, $A_i$ and $B_i$ are fuzzy membership. So, the function of each node on layer 1 is to generate the degree of a membership.

2.2.2 Layer 2: Product. Noticed as $\pi$. It is also known as rule layer. The function of each node on this layer is to calculate activation strange (firing strength) on each rule as the product from all input or as a t-norm (triangular norm) operator:

$$O_{2,i} = w_i = \mu_{A_i}(x) \Delta \mu_{B_i}(y), \quad i = 1,2 ,$$

So,

$$w_1 = \mu_{A_1}(x) \text{ AND } \mu_{B_1}(y) ,$$

$$w_2 = \mu_{A_2}(x) \text{ AND } \mu_{B_2}(y) .$$

The output of this layer works as weight function.

2.2.3 Layer 3: Normalization. Noticed as $N$. This layer has a non-adaptive characteristic which has a function to calculate a ratio between firing strength on rule to it toward firing strength to sum from all the rule:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} , \quad i = 1,2 .$$

2.2.4 Layer 4: Defuzzification. Each node in this layer have adaptive characteristic, with the function:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$$

Where $\bar{w}_i$ is a normalized firing strength from layer 3 and $\{x + q_i y + r_i\}$ are the parameter of membership on fuzzy Takagi-Sugeno first order model.

2.2.5 Layer 5: Total Output. Represented by a single summation node, Symbolized as $\Sigma$. This layer has a function to aggressing all output on layer 4 (summation from all input signal):

$$O_{5,i} = \Sigma \bar{w}_i f_i = \sum_i \frac{w_i f_i}{\Sigma w_i}$$

Totally, the five-layer will build as an adaptive-networks which functionally equivalent to fuzzy Takagi-Sugeno first order model.

3. Method and System Design

The tools which used in this research are microphone 54db+2 db (which have specification: impedance 32 ohm at 1 kHz sensitivity102db/mW), 5 DoF arm robot based Arduino microcontroller. The personal computer as speech processing using MATLAB® software. The steps in this research generally show in Figure 3:
Figure 3. General scheme system

The first process is making training data, consist of features extraction process using LPC and learning process using ANFIS that will build a database as fuzzy logic formulation. The second process is the testing system, such as LPC features extraction, classifying ANFIS data process, and implementation to control arm robot for placing an object using speech command.

4. Hardware Design

Figure 4 is the schematic and realization of arm robot which used in this research. This arm robot consists of five motor servo component which connects with Arduino microcontroller.
Arm robot circuit is shown in Figure 5, each servo must have a supply using the 5-volt battery which has 100 mA of current in order get a better result. Each servo ground connected to the ground on Arduino microcontroller. Arm robot consists of 5 motor servo called 5 DoF (Degree of Freedom); Servo1 connect to pin 8 (base, rotate horizontally), Servo2 connect to pin 9 (work as shoulder, rotate vertically), Servo3 connect to pin 10 (work as elbow, rotate vertically), Servo4 connect to pin 11 (work as wrist, rotate horizontally), and Servo5 connect to pin 12 (work as gripper to place an object).
5. Result and Discussion

5.1 IDE Arduino Program Test
To test an IDE Arduino program, have been investigating by simulation using Proteus. The result shows a suitable between program and simulation output shown in Figure 6.

Figure 6. Serial test by IDE Arduino program to control 5 DoF arm robot.

5.2 Features Extraction Database using LPC

We use 6 features of speech signal using LPC method toward respondents with speaking a word (in Indonesian) “ambil” (pick) and “simpan” (place) with each of 10 sample, can be seen in the following table:

| No | Feature 1  | Feature 2  | Feature 3  | Feature 4  | Feature 5  | Feature 6  | Target |
|----|------------|------------|------------|------------|------------|------------|--------|
| 1  | -1.76619   | -0.38658   | 1.854785   | 1.101728   | -2.29431   | -2.42487   | 1      |
| 2  | -1.75599   | -0.5623    | 2.043627   | 1.520301   | -2.65515   | -3.35814   | 1      |
| 3  | -2.13159   | 0.095945   | 2.878341   | -0.25327   | -4.7726    | 0.496327   | 1      |
| 4  | -2.60417   | 1.248259   | 2.836023   | -2.65658   | -3.70216   | 5.616098   | 1      |
| 5  | -2.26193   | 0.756091   | 2.494955   | -1.85409   | -3.3322    | 4.04501    | 1      |
| 6  | -2.52144   | 1.082503   | 2.82768    | -2.35296   | -4.16399   | 5.577152   | 1      |
| 7  | -2.08075   | 0.570233   | 1.791693   | -0.97266   | -1.79269   | 1.671262   | 1      |
| 8  | -2.24344   | 0.76278    | 1.953595   | -0.82145   | -2.60754   | 1.466322   | 1      |
| 9  | -2.39999   | 0.617584   | 3.218707   | -1.45934   | -5.45294   | 3.365111   | 1      |
| 10 | -1.93844   | -0.02257   | 2.184292   | 0.417742   | -3.35865   | -0.9442    | 1      |
| 11 | -1.51965   | -0.31074   | 1.035648   | 0.670428   | -0.92408   | -0.82099   | 0      |
On Table 1 show the differences between word “ambil” and “simpan”. Speech “ambil” as target “1”, then speech “simpan” as target “0”. This features will become an input for ANFIS method as database for Fuzzy logic.

5.3 Speech Recognition System Test
The test system results shown on Table 2 that the average speech recognition accuracy rate respondents trained data (in database) are 88.75% and the respondents not trained (outside database) data produces an accuracy rate of 78.78%. Errors on speech recognition can be influenced by the accent / intonation speech input and noise on the environment.

| Sample | Word | Target | Recognition Trained Data | Recognition Not-Trained Data |
|--------|------|--------|--------------------------|-----------------------------|
|        |      |        | Output  | Error   | Output  | Error   |
| 1      | “Ambil” | 1 | 0.97226  | 0.02774 | 1.40514 | 0.40514 |
|        | “Simpan” | 0 | -0.027366 | 0.027366 | 0.0781117 | 0.0781117 |
| 2      | “Ambil” | 1 | 1.31593  | 0.31593 | 1.00856 | 0.00856 |
|        | “Simpan” | 0 | 0.109141 | 0.109141 | -0.21035 | 0.21035 |
| 3      | “Ambil” | 1 | 0.87573  | 0.12427 | 0.573719 | 0.426281 |
|        | “Simpan” | 0 | 0.0438407 | 0.0438407 | -0.112783 | 0.112783 |
| 4      | “Ambil” | 1 | 1.06282  | 0.06282 | 1.08353 | 0.0853 |
|        | “Simpan” | 0 | 0.0520409 | 0.0520409 | 0.158865 | 0.158865 |
| 5      | “Ambil” | 1 | 0.766645 | 0.233355 | 0.80136 | 0.19864 |
|        | “Simpan” | 0 | 0.128364 | 0.128364 | 0.437931 | 0.437931 |

Figure 7 shown a GUI (Graphic User Interface) when tested the word "ambil" and "simpan". When the word "ambil" recorded, the system will recognize with output value 0.953285 approach to "1", so that the computer will transmit serial data as characters "1" to the microcontroller to drive the robot arm in a position to pick up objects. Meanwhile, when the word "simpan" recorded, then the system will recognize the output amounted to 0.232289 which is close to the value "0", so that the computer will transmit serial data as character '0' to the microcontroller to drive the robot arm in a position to place objects. Generally, the tool works fine 100% because of rounding value.
Figure 7. GUI MATLAB interface to control arm robot using speech recognition.

6. Conclusion
In this research can be concluded that the system works successfully suitable the speech command. The results obtained speech recognition system implementation of LPC and ANFIS method have a high average accuracy rate of speech recognition, which is 88.75% of the respondents trained data and 78.78% of the respondents not trained data. In the experiment result to control the 5 DoF of arm robot based on Arduino microcontroller to pick and place an object works well.

Acknowledgement
The authors would like gratefully acknowledge the financial support from DIPA UIN Sunan Gunung Djati Bandung.

Reference
[1] al Jazari I a R 1974 The Book of Knowledge of Ingenious Mechanical Devices Trans and annotated by Donald R. Hill Dordrecht: Reidel
[2] Daqrouq K, Morfaq A, Ajour M and Alkhateeb A 2013 Wavelet LPC With Neural Network for Speaker Identification System," WSEAS Transaction on Signal Processing 9 4 216-226
[3] Thiang and Wijoyo S 2011 Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot International Conference on Information and Electronics Engineering 6 179-183
[4] Abinayaa B, Arun D, Darshini B and Nataraj C 2015 Voice Command Based Computer Application International Journal of Innovative Research in Science, Engineering and Technology 4 4 57-63
[5] A. M. Soe, M. M. Latt, H. M. Tun and Z. Mi, "Electronic Control System Of Home Appliances Using Speech Command Words," International Journal of Scientific & Technology Research, vol. 4, no. 6, pp. 323-329, 2015.
[6] Sanjaya W S M and Salleh Z 2014 Implementasi Pengenalan Pola Suara Menggunakan Mel_Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inferenze System (ANFIS) sebagai Kontrol Lampu Otomatis Al-Hazen Journal of Physics 1 1 44-54
[7] Kulji B, János S and Tibor S 2007 Mobile robot controlled by voice International Symposium on Intelligent Systems and Informatics 5 189-192
[8] Srishti, Jain P, Shalu and Singh S 2015 Design and Development of Smart Wheelchair using Voice Recognition and Head Gesture Control System International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 4 5 4790-4798

[9] Koo Y m, Kim G B, Jang S C, Lee W S, Kim H-G and Han S H 2015 A Study on Travelling Control of Mobile Robot by Voice Command International Conference on Control, Automation and System 15 13-16

[10] Punchihewa A and Arshad Z M 2011 Voice Command Interpretation for Robot Control International Conference on Automation, Robotics and Applications 5 90-95

[11] Rudrapal D, Smita Das S D and Kar N D N 2012 Voice Recognition and Authentication as a Proficient Biometric Tool and its Application in Online Exam for P.H People International Journal of Computer Applications 39 12 6-12

[12] Varalakshmi M and Raju N N 2014 Design of Speech Controlled Pick and Place Robot with Wireless Zigbee Technology International Journal of Scientific Engineering and Technology Research 3 20 4062-4066

[13] JadHAV K S and Gaikwad S M 2015 Writing Robotic Arm by Speech Recognition International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 4 6 4983-4990

[14] Rai N and Rai B 2014 An ANN Based Mobile Robot Control Through Voice Command Recognition Using Nepali Language International Journal of Applied Control, Electrical and Electronics Engineering (IJACEEE) 2 4 13-22

[15] Srivastava N 2014 Speech Recognition using Artificial Neural Network International Journal of Engineering Science and Innovative Technology 3 3 406-412

[16] Thiang and Wijaya D 2009 Limited Speech Recognition for Controlling Movement of Mobile Robot Implemented on ATmega162 Microcontroller in International Conference on Computer and Automation Engineering 347-350

[17] El- emary I M M, Fezari M and Hamz 2011 Hidden Markov model/Gaussian mixture models(HMM/GMM) based voice command system: A way to improve the control of remotely operated robot arm TR45 Scientific Research and Essays 6 2 341-350

[18] Jang R, Sun C T and Mizutani E 1997 Neuro-Fuzzy and Soft Computing, Prentice-Hall, Inc.