Classification of EMG Signals from forearm Muscles As Automatic Control Using Naive Bayes

Adi Dwi Irwan Falih[1], Adhi Dharma W[1,2], Surya Sumpeno[1,2]
[1]Department of Electrical Engineering,
[2]Department of Computer Engineering,
Institut Teknologi Sepuluh Nopember (ITS).
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
Email: falih15@mhs.ee.its.ac.id, adhiosa@ee.its.ac.id, surya@ee.its.ac.id

***Abstract*— The wheelchair is still a mobility aids commonly used by patients with muscle weakness or stroke patients. Some stroke patients, having constraints in moving a joystick or controlling an electric wheelchair due to muscle limitations of their hands Myo-armband, as wearable device that have an Electromyogram sensor can be used as an alternative in controlling the electric device like wheelchair more easily. The Electromyography Research (EMG) on feature of particular muscle activation pattern which has correlation with a motion contributes inspiration to be applied as motion control media on electric wheelchair. Classification process of EMG will be a new alternative to control wheelchair movement for user or patient who hasn't latitude to move their limb and just able to do easy motion using their forearm. The stages of this project is detecting signal in the muscle using EMG, extracting feature of muscle response in time domain base, and be classified by Naïve Bayes, the dataset classification is pinned in raspberry and output to arduino controller to be used as output motion in motor of electric wheelchair. The result of this research is classification of MAV feature, Peak number, RMS and Gradient Magnitude in 275 stream of muscle data show that detected and correctly can be discriminate 90.18%, thus, a sum of 248 instances and wrongly 9.8182% a sum of 27 instances.**

Keywords— Electromyography, forearm muscle, Naive Bayes, Raspberry, Arduino

I. Introduction

The use of a wheelchair is mandatory for the disabled and stroke sufferers who are unable to move the leg muscles. Electric wheelchairs are the best option to help their mobility. Stroke patients who have limitations in controlling wheelchairs with joystick steering, require specific steering alternatives to their abilities.

This motorized wheelchair is a solution as automatically, so patients can do their activities independently. But this motorized wheelchair remains a problem for user who has deformed hand and disabilities in moving their hands to operate and to reach joystick in electric wheelchair.

Based on the results of a survey conducted by the Product Development Team of Mechanical Engineering Department of ITS to several places such as hospitals and institutions of disability [1], it was found that the appreciation of people with disabilities will need a wheelchair that is more attractive and automatic is very big. This is because they are less satisfied with wheelchairs now incomplete (58.3%), less comfortable (25%) and for difficult operation (4.1%). They also stated that manual wheelchairs are now not able to make them perform activities like normal people (12.6%)

Requirement of alternative control of the electric wheelchair should be developed in order to help users with limitations which have been mentioned above. Electromyography (EMG), which is one of monitoring method and recording the muscle electrical signals have been widely used in the medical field as well as applications related to human activity [2]. In the medical field, EMG is used to diagnose disorders performance of nerve and muscle. Also, EMG has function as a replacement of motion organ robotically. In addition, EMG can detect diseases related with muscle weakness [3], etc. In non-medical fields, EMG has been applied / implemented for human and computer interaction. For example, a game application, industry and robot-based Security, and applications that replace the function of human organ.

Thalmic Labs is one of the companies occupied in development of human interaction/communication with the computer. One of Thalmic Labs' products is Myo Armband. Myo Armband is a new technology shaped a bracelet which is containing of EMG sensor, sensitive IMU sensor containing three-axis gyroscope, three-axis accelerometer sensor, and three-axis magnetometer for requirements of human interaction with the computer. In this research, the specification of EMG possessing 8 EMG channel sensor with data frame rate EMG 200 Hz, can result stream data of certain muscles to be processed and classified whether for medical needs or communication with (control) the computer [4].

Based on capability of EMG sensor to detect electrical signals that is transmitted by the nervous system when the muscle contracts, output data of EMG is expected to be treated with Naive Bayes algorithm and used for the development of applications to control the movement of an electric wheelchair so as to help users who have limited mobility to control an electric wheelchair.

The study is also aimed to obtain a pattern of muscle activation response [2], so that the output of the pattern classification can be used as a specific output (this study uses the robot in the implementation of the classification output). That pattern is expected to apply generally on a variety of different patient conditions.

II. METHOD

EMG research is experiment-based stages in order to classify or recognize a pattern of electrical signals from human muscles. EMG signal variations form subjectively on muscle fibers [5]. In general the stages performed in this study as illustrated in the diagram below.

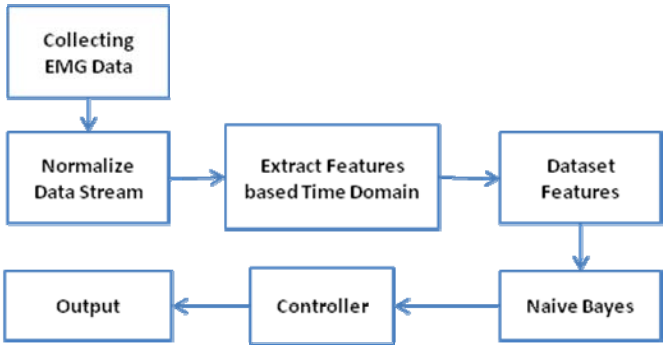

Fig. 1. General Method

A. Collecting EMG Data

In this research the process of taking EMG data through Myo Armband is done by 11 normal people voluntarily. EMG data retrieval is done on their right arm, where each person performs 5 motion poses. Hand posed poses include wrist poses, stretching pose, waving in pose, waving out pose and relaxed pose (not doing muscle contraction).

TABLE I. FIVE POSE TO BE CLASSIFY

Pose	Count of Sample	Subject	Sum of data
wrist	5	11	55
stretching	5	11	55
wave In	5	11	55
wave Out	5	11	55
relax	5	11	55
Total Gerakan			275

The motion sampling process is done for five times by each person, so the total data of motion data retrieval is 275 sample data (11 people * 5 poses * 5 times). Data retrieval is done by placing the position of myo armband logo on the outer / posterior side of the arm as shown in Fig. 2.

The data stream taking at each pose is done for approximately 30 seconds, with idle breaks in every 3 seconds then performing poses and maintaining muscle contraction position for approximately 3 seconds for 5 movements. The movements performed are illustrated at Fig. 3.

Fig. 2. Using position with Myo Armband *(taken and modified https://www.myo.com/)*

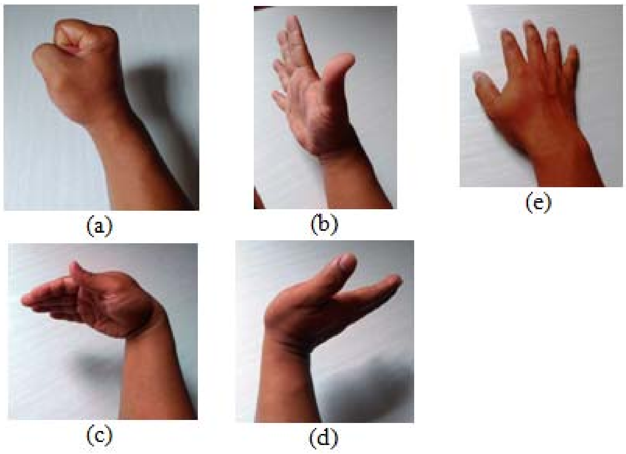

Fig. 3. This photo is taken on the right arm of person with various poses according to the research done. (a) wrist pose (b) stretch pose (c) wavein pose (d) waveout pose (e) relax pose

When placement of myo armband, also recorded muscle dominance correlated / in direct contact with channel sensor EMG.

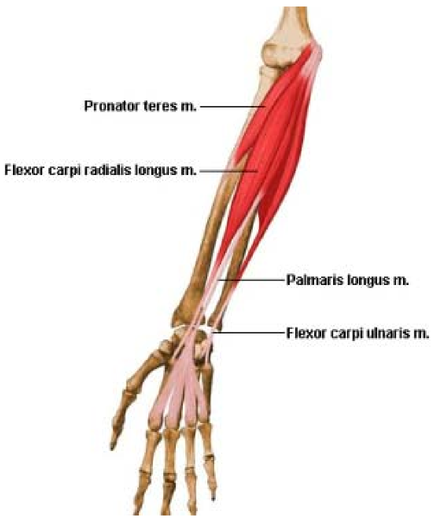

Fig. 4. Flexor Superficial Layer

(http://depts.washington.edu/msatlas/117.html)

Based on muscle image at the Fig. 4 and Fig. 5 above, the placement of Myo Armband in this study can be mapped and concluded that the muscle dominance at the time of capture EMG data for measurement in each channel from Myo Armband is as Table II.

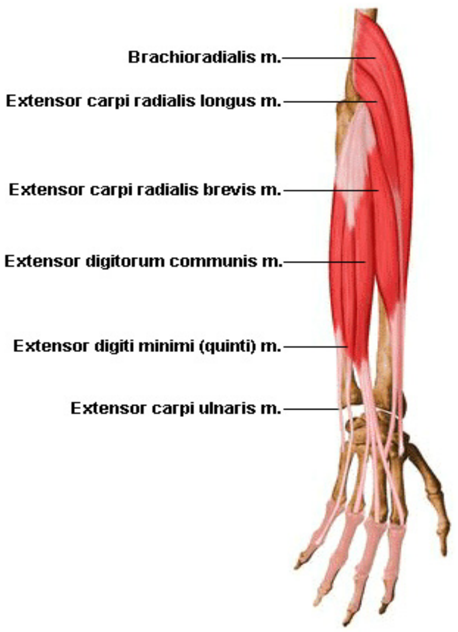

Fig. 5. Long Extensor (http://depts.washington.edu/msatlas/117.html)

TABLE II. MAPPING OF EACH CHANNEL PLACEMENT ON MUSCLE

Channel	Dominant muscle being recorded
CH4	Pronator teres m.
CH3	Brachioradialis m. dan Extensor Carpi radialis longus m
CH2	Extensor Carpi Radialis Brevis m. dan Extensor digitorum communis m (Extensor)
CH1, CH8, and CH7	Flexor Carpi Ulnaris m.
CH6 and CH5	Flexor Carpi Radialis Longus m

Step of retrieving an EMG data stream using the Java API library for Thalmic's Myo Device

B. *Normalize Data Stream*

After the data retrieval process is done then, the data on each channel carried out the process of normalization (this process is by absolute each data point into a positive integer).

During normalization process, smoothing average is done to eliminate noise in data stream in each channel. Then the data stream is crop manually based on observations on the amplitude that is considered when the muscle contraction (using threshold) to be instance dataset.

Data Stream instances in channel k was given from data retrieval process, were processed using a threshold of 20 mV [1], then continued by smoothing process each point in the signal to make data clear, and the most algorithms are based on the "shift and multiply" technique

$$B_k = f(x_t) = \frac{x_{(t-1)} + x_t + x_{(t+1)}}{3} \qquad (1)$$

where :

x_t = is the point of data stream at time sequence domain
B_k = is instance / vector data result from smoothing process

By replaces point data with the average of adjacent points, where m is a positive integer called the smooth width or commonly called convolution

C. *Extract EMG Features Based Time Domain*

Some feature domains in EMG have been defined through various studies, including frequency domain feature (FD), time domain feature (TD), and time frequency feature (TFD). [5, 6, 7, 8]

In this study used feature extraction on the time domain cause [9], that The commonly used in [3]

1) Calculate the sum of peak (ρ) from data stream EMG,

$$P = \sum_{t}^{t+T} [f(x_t)] \qquad (2)$$

where

$$f(x_t) = \begin{cases} 1, & (x_{(t-1)} < x_t) \, and \, (x_{(t+1)} < x_t) \\ 0, & otherwise \end{cases} \qquad (3)$$

2) Calculate Mean Absolute Value (MAV) of each sample at each Channel [1]

$$MAV = \frac{1}{N} \sum_{n=1}^{N} |x_n| \qquad (4)$$

3) Calculate *Willison Amplitude* (WAMP) of each sample at each Channel

$$WAMP = \sum_{t}^{T+t} f\,|(x_{t+1} - x_t)| \qquad (5)$$

where

$$f(x_t) = \begin{cases} 1, & (x) > threshold \\ 0, & otherwise \end{cases} \qquad (6)$$

4) Calculated Root Mean Square (RMS) of each sample at each Channel

RMS is statistical features used that approach in traditional time domain analysis and it calculated after smoothing process before normalizing the data stream of EMG.

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^{N} {x_n}^2} \qquad (7)$$

where

N : Number of data sequences
X_n : value of data at index n

D. Features Dataset

After getting value from those feature, arranged value to the matrix like Table III.

TABLE III. DATASET FEATURES FORMAT

Sample Number	P_i	MAV_i	$WAMP_i$	RMS_i	*...i*	*Class*
1	.					
2	.					
...						
	further samples are arranged into rows					

Each feature on each channel is arranged into columns or attributes with an instance feature value as a row, it is used as an input to the Naive Bayes classifier.

TABLE IV. ACTIVATION CHANNEL ON WRIST POSE

Subject	wrist pose							
	CH1	CH 2	CH 3	CH 4	CH 5	CH 6	CH 7	CH 8
a	v	v	v	v	v	v	v	v
b	v	v	v	v	v		v	v
c	v	v	v	v	v		v	v
d	v		v	v	v		v	v
e	v	v	v	v	v	v	v	v
f	v	v	v	v	v	v	v	v
g	v	v	v	v	v	v	v	v
h	v	v	v	v	v		v	v
i	v	v	v	v	v	v	v	v
j	v	v	v	v	v	v	v	v
k	v	v	v	v	v		v	v

Before performing the classification process with naive bayes, mapping of muscle activation on each channel on all subjects has been done based on observations of EMG raw data. Observation is done based on the dominance of muscle activeness during contraction. Mapping was conducted to determine channel variations in the combination of active muscle in pose motion to be classified. The mapping results are also used to support the selection of subjects while training on the dataset. mapping of muscle activation in each channel for each pose motion can be seen in the Table IV until Table VII

On the dataset Table IV. seen the number of dominant active channels on wrist pose are **all** channels. Known based on the percentage of the maximum number of active subject muscles on each channel.

TABLE V. ACTIVATION CHANNEL ON STRETCH POSE

Subject	stretch pose							
	CH1	CH 2	CH 3	CH 4	CH 5	CH 6	CH 7	CH 8
a	v	v	v	v				v
b	v	v	v	v	v			v
c	v	v	v	v			v	v
d		v	v	v			v	v
e	v	v	v	v				v
f	v	v	v	v	v	v	v	v
g	v	v	v	v			v	v
h	v	v	v	v	v			v
i	v	v	v	v				v
j	v	v	v	v	v		v	v
k	v	v	v	v			v	v

Then at stretch pose , some active channels are **Ch1, Ch2, Ch3, Ch4, Ch7,** and **Ch8**. Special on the relaxed pose, the number of dominant active channels on this pose are **none**, cause there is no muscle contraction in the related channel and the stream data of recording muscle by EMG mostly zero after going through the threshold

TABLE VI. ACTIVATION CHANNEL ON WAVE IN POSE

Subject	wave in pose							
	CH1	CH 2	CH 3	CH 4	CH 5	CH 6	CH 7	CH 8
a	v	v	v			v	v	v
b	v	v	v	v		v	v	v
c	v	v	v			v	v	v
d		v	v			v	v	v
e	v	v	v			v	v	v
f	v		v			v	v	v
g	v		v			v	v	v
h	v	v	v	v		v	v	v
i	v	v	v	v		v	v	v
j	v	v	v	v			v	v
k	v	v				v	v	v

On the Table VI. seen the number of dominant channels on wave in pose are **Ch1,Ch2,Ch3,Ch6,Ch7,**and **Ch8**.

TABLE VII. ACTIVATION CHANNEL ON WAVE OUT POSE

Subject	wave out pose							
	CH1	CH 2	CH 3	CH 4	CH 5	CH 6	CH 7	CH 8
A	v	v	v	v	v			
B	v	v	v	v	v	v		
C	v	v	v	v	v			
d		v	v	v				
e	v	v	v	v		v		
f	v	v	v	v	v		v	
g		v	v	v	v	v		
h	v	v	v	v	v			v
i	v	v	v	v	v			
j	v	v	v	v	v	v		
k		v	v	v				

On the dataset Table VII. seen the number of maximum number or dominant active channels on wave out pose are **Ch1, Ch2, Ch3, Ch4,** and **Ch5**.

E. Naive Bayes

Calculate the weight classification with Naive Bayes, based on the normal data distribution, on still dataset

$$y' = \arg\max P(y) \prod_{k=1}^{n} P(x_k \mid y) \tag{8}$$

we can use Maximum A Posteriori (MAP) estimation to estimate P(y) and $P(x_k \mid y)$; the former is then the relative frequency of class in the training set [10].

$$P(x_k \mid C_i) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2} \right) \qquad (9)$$

where μ_y is mean feature attribute on each class, and σ^2 is variance feature

F. Controller

The controller used to connect the myo arm band and electric wheelchair is Raspberry Pi and Arduino, where the flow of EMG data **stream** from Myo Arm Band is processed and classified by raspberry pi, which then outputs the result of classification into output for the Arduino in regulating the motion of the motor in the wheelchair

G. Experimental Setup

Experiments in this study used the parable of the electric wheelchair prototype by using a robot, which in general the mechanism of the data communication flow to become ouput in the form of motor movement is as in Fig. 6.

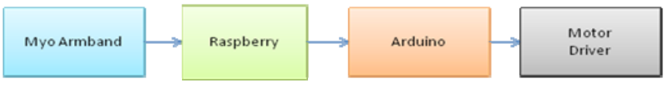

Fig. 6. Device Setup of Data Communication

Raspberry [11] use in this study is as a controller to receive raw EMG data from Myo Armband and calculate bayes method in the operating system. Then forward the results of classification to the new data, on arduino for subsequent use as a motor trigger in the movement of the robot.

Using the mechanism as shown in Fig. 6, it is expected that 60% of the work can be completed, which is capable of moving an electric wheelchair system with EMG data. The next 40% is to design the mechanical and electrical specifications of the wheelchair so as to be able to move the user's load.

III. RESULT

From this research, the test results using weka tool with naive bayes algorithm to feature dataset, done with some condition, Evaluation of the various conditions is tabulated in a confusion matrix in order to conclude the best stages with optimal classification results.

The first result of all instance dataset at the classification process is done on all features in all channels. the data test used cross-validation at 10 folds, give result more than 87% Correctly Classified Instances. it were able to predict 210 instances from 275 instances in the dataset

TABLE VIII. CLASSIFICATION SUMMARY OF ALL CHANNEL FEATURE, USED CROSS VALIDATION 10 FOLDS

Summary	Percentage %	Instances
Correctly Classified Instances	87.2727%	240
Incorrectly Classified Instances	12.7273 %	35
Total Instances from 11 subject		**275**

from the confuse Table IX known that (d) class can be higher classified than the other class with value of 94.5%, Because the relaxed pose (d) does not have an active channel while recording the data stream, and it becomes a unique feature of the other poses, so that the classifier can classify pose from datasets.

TABLE IX. CONFUSION MATRIX IN %

Classified As	a	b	C	d	e
a = wrist	81.8	18.2	0.0	0.0	0.0
b = stretch	10.9	89.1	0.0	0.0	0.0
c = wave in	5.5	3.6	90.9	0.0	0.0
d = relax	0.0	1.8	3.6	94.5	0.0
e = wave out	9.1	10.9	0.0	0.0	80.0

The second test is done by evaluate all dataset to naive bayes model. the results obtained, as in the [table 12]

TABLE X. SUMMARY RESULT OF DATA TEST USED ALL DATASET FOR EACH CLASS

Summary	Percentage %	Instances
Correctly Classified Instances	90.1818%	248
Incorrectly Classified Instances	9.8182 %	27
Total Instances		**275**

The classification on wrist pose and stretch pose has an error with the result in both, wrist poses are still recognized as stretch pose and vice versa. It can happen because almost all the channels are active on both poses as shown in the Table V and Table VI.

TABLE XI. CONFUSION MATRIX IN % OF EVALUATION FROM ALL DATASET

Classified As	a	b	c	d	e
a = wrist	81.8	18.2	0.0	0.0	0.0
b = stretch	9.1	90.9	0.0	0.0	0.0
c = wave in	5.5	0.0	94.5	0.0	0.0
d = relax	0.0	0.0	0.0	100.0	0.0
e = wave out	5.5	10.9	0.0	0.0	83.6

The third test is by performing the subject's elimination on the tabulated feature dataset, using the function (and) according to the active channel in each pose. Then the number of datasets is reduced to the Table XII.

After evaluation on naive bayes classifier, on all features in selected subject by the data test used cross-validation at 10 folds, give result increased than ever before became 90,4348% correctly classified that shown in Table XIII.

TABLE XII. DATASETS AFTER THE ELIMINATION OF SUBJECTS NOT MATCHING THE ACTIVE CHANNELS IN EACH POSE

Pose	Count of Sample	The Subject used	Sum of data
Wrist	5	[a, e, f, g, i, j]	30
Stretching	5	[c, f, g, j, k]	25
wave In	5	[a, b, f, h, i]	25
wave Out	5	[a, b, c, f, h, i, j]	35
Relax	5	11 (use all in class before)	55
Total Instance			170

Then the dataset is generated into a classifier model to control the wheel chair

TABLE XIII. SUMMARY RESULT ON SELECTED SUBJECT DATASET

Summary	Percentage %	Instances
Correctly Classified Instances	93.5294 %	159
Incorrectly Classified Instances	9.5652 %	11
Total Instances from 11 subject		**170**

IV. DISCUSSION

The steps of this method in this study need to be improved to obtain better classification results. The percentage of correct results is not very high at 93.5%, because it can be caused by a sample data variant of the subject, whereas for EMG data itself, it is specific data on a particular subject, which depends on aspects of muscle mass, motion ability, subject habits and so on. In doing the subject Habit movement means how the dominant muscle combination when doing the pose motion. In this case the process of recording the muscle activation of channels that can represent the motion of the pose

The accuracy of the classification results is also influenced by the laying of EMG channels from Myo armband device. The channel position must be precise so that the recorded data stream can be the correct data input on the classifier.

The feature selection method like as [12] needs to be applied so that the classifier is able to find the characteristic feature differences between different pose, the selection of the feature selection method must be precise in order to produce the expected classification.

The selection of datasets correlated with the corresponding class can determine the result of better classification in order for the output of a controller to be precision as expected.

REFERENCES

[1] M. Reaz, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological Procedures Online,* pp. 11-35, 2006.

[2] Q. Wu, X. Chen, L. Ding, C. Wei, H. Ren, R. Law and H. Dong, "Classification of EMG Signals by BFA-Optimized GSVCM for Diagnosis of Fatigue Status," *IEEE Transactions on Automation Science and Engineering,* vol. 14, pp. 915-930, 2017.

[3] S. V, "Biosignals offer potential for direct interfaces and health monitoring," *Pervasive Computing, IEEE,* vol. 3(1), p. 99–103, 2004.

[4] Morwati, Pengenalan Citra Huruf Alphabet Tulisan Tangan Menggunakan Metode Naive Bayes Classifier, Malang: UIN Maulana Malik Ibrahim, 2014.

[5] Thalmic Labs, "Thalmic Developer," 2015. [Online]. Available: https://developer.thalmic.com/docs/api_reference/platform/getting-started.html. [Accessed 1 April 2016].

[6] N. Nurhazimah, M. Azizi Abdul Rahman and S.-I. , "A Review of Classification Techniques of EMG," *Sensor,* vol. 16, p. 1304, 2016.

[7] "Tsai, A.C.; Hsieh, T.H.; Luh, J.J.; Lin, T.T. A comparison of upper-limb motion pattern recognition using EMG".

[8] P. Angkoon, P. Pornchai and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications,* p. 7420–7431, 2012.

[9] D. S. Putra, A. D. Wibawa and M. H. Purnomo, "Classification of EMG during walking using principal component analysis and learning vector quantization for biometrics study," in *International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Lombok, 2016.

[10] A. Wibawa, N. V. J. H. J. B. R. D. and G. V. , "Musculoskeletal modeling of human lower limb during normal walking, one-legged forward hopping and side jumping: Comparison of measured EMG and predicted muscle activity patterns," *Journal of Biomechanics,* vol. 49, no. 15, pp. 3660-3666, 2016.

[11] The Raspberry Pi Foundation, "Raspberry pi," The Raspberry Pi Foundation, [Online]. Available: https://www.raspberrypi.org/. [Accessed 20 April 2017].

[12] Liem, Y. K. Hatta, P. S. M. and I. T. M. , "Rancang Bangun Kursi Roda Elektrik Menggunakan Perintah Suara Berbasis Aplikasi Android," *JURNAL TEKNIK POMITS,* vol. 1, pp. 1-6, 2012.