

# PERANCANGAN DAN IMPLEMENTASI PENGENALAN UCAPAN INSTRUKSI PADA SISTEM KENDALI DENGAN METODA KLASIFIKASI JARINGAN SYARAF TIRUAN PROPAGASI BALIK

## DESIGN AND IMPLEMENTATION OF SPEECH INSTRUCTION RECOGNITION FOR CONTROL SYSTEM WITH NEURAL NETWORK BACKPROPAGATION CLASSIFIER

Agung Chrisyancandra Mobonguni<sup>1</sup>, Erwin Susanto<sup>2</sup>, Fiky Yosef Suratman<sup>3</sup>

<sup>1,2,3</sup>Prodi S1 Teknik Elektro, Fakultas Teknik Elektro, Universitas Telkom

<sup>1</sup>agungmobo@students.telkomuniversity.ac.id, <sup>2</sup>erwinelektro@telkomuniversity.ac.id, <sup>3</sup>fysuratman@telkomuniversity.ac.id

### ABSTRAK

Pada tugas akhir ini, penulis telah mengimplementasikan pengenalan ucapan untuk sistem kendali pada helm asisten dengan metoda klasifikasi Jaringan Syaraf Tiruan propagasi balik terhadap perintah yang diberikan berupa sinyal audio manusia. Pemrosesan sinyal audio dilakukan dengan mengekstrak ciri sinyal menggunakan algoritma Mel Frequency Ceptral Coefficient (MFCC). Implementasi dilakukan pada Single Board Computer sebagai media komputasi pada kesatuan sistem otomasi tertanam pada helm asisten. Hasil yang diperoleh adalah performansi pengenalan ucapan dengan Jaringan Syaraf Tiruan dengan akurasi pengujian real-time sebesar 85% dengan durasi rata-rata eksekusi sekitar 1.34775 detik dalam menghasilkan pemicu aktivasi aktuator pada helm asisten berupa motor servo sebagai penggerak kaca helm dan LED sebagai penerangan.

Kata Kunci: Jaringan Syaraf Tiruan, MFCC, Single Board Computer, helm asisten

### ABSTRACT

In this final assignment, author has done to implement speech recognition to the control system on the helmet assistant with the Neural Network Back Propagation classifier of human's audio signals. Audio signal processing is done by using algorithms to extract signal characteristics by Mel Frequency Ceptral Coefficient (MFCC). Implementation is done on a Single Board Computer as computational media on the unity of the automation system is embedded in the helmet assistant. The results is the performance of instruction speech recognition in the classification for simple control system to activating the actuator on the helmet assistant such as servo motor to driving visor and LED as lighting. The Best result of speech recognition is achieved with average real time response for 1.34775 seconds with accuracy 85% for produce a trigger for activation servo motor as visor driver and LED as lightning.

Keywords: Artificial Neural Network, MFCC, Single Board Computer, helmet assitant

### 1. Pendahuluan

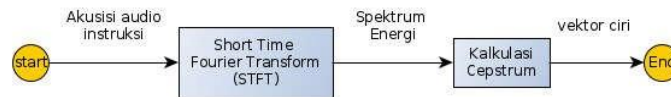
Soft Computing merupakan salah satu bidang keilmuan menggabungkan Fuzzy Logic, Neuro Computing, Evolutionary Computation, dan Probabilistic Computing [1]. Jaringan Syaraf Tiruan merupakan salah satu disiplin ilmu dalam soft computing yang digunakan untuk melakukan pencocokan terhadap masukan yang baru berdasarkan hasil belajar akan pengenalan data. Dalam pengenalan ucapan, JST berperan untuk menentukan kelas keluaran dari data masukan.

Tugas Akhir ini bertujuan untuk mengimplementasikan sistem kendali cerdas berbasis pengenalan pola dengan algoritma Jaringan Syaraf Tiruan Propagasi Balik, terutama pola audio manusia berupa instruksi perintah untuk menggerakkan aktuator sederhana pada sistem tertanam (embedded system). Dengan algoritma Mel Frequency Ceptral Coefficients (MFCC) sebagai langkah ekstraksi ciri audio.

### 2. Dasar Teori

#### 2.1 Mel Frequency Ceptral Coefficients (MFCC)

Mel Frequency Ceptral Coefficients merupakan suatu algoritma yang memungkinkan sinyal audio untuk direpresentasikan dalam bentuk yang ekuivalen dan sederhana. Hasil manipulasi sinyal ini yang disebut dengan ciri dari sinyal audio masukan.



Gambar 1. Diagram blok proses ekstraksi ciri dengan MFCC

Ada dua bagian besar dalam proses MFCC, yaitu Short Time Fourier Transform (STFT) dan kalkulasi nilai cepstrum.. Proses Short Time Fourier Transform ditunjukkan pada persamaan (1)

$$X_l[k] = \sum_{n=-N/2}^{N/2-1} w[n]x[n+lH]e^{-j\frac{2\pi}{N}kn} \quad (1)$$

$w[n]$  adalah fungsi window,  $l$  adalah indeks frame,  $H$  adalah ukuran panjang sampel pada frame,  $k$  adalah indeks frekuensi,  $n$  adalah indeks waktu diskrit,  $N$  adalah ukuran DFT dan  $e^{-j\frac{2\pi}{N}kn}$  adalah fungsi sinusoidal

kompleks. Short Time Fourier Transform menghasilkan himpunan spektrum dari satu sinyal keseluruhan. Analisis STFT dilakukan untuk sinyal yang bersifat time variant seperti sinyal audio yang dihasilkan oleh manusia. Spektrum hasil STFT (respon magnitude) tiap frame difilter menggunakan bank filter band pass filter segitiga. Karena sifat audio manusia berupa logaritmik, maka untuk mendapatkan nilai bank filter mel dilakukan transformasi Mel dengan persamaan (2) untuk mengkonversi himpunan nilai frekuensi pembentuk bank filter sehingga diperoleh himpunan nilai frekuensi dengan rentang logaritmik. Hasil Transformasi ini digunakan untuk membentuk bank filter segitiga.

$$f_{mel}(f) = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \quad (2)$$

Hasil perkalian respon magnitude dan bank filter mel, diambil nilai logaritmanya. Kemudian, nilai logaritma ini dilakukan Discrete Cosine Transform (DCT) untuk memperoleh cepstrum dengan jumlah tiap vektor 12 atau 13 koefisien teratas dari hasil DCT .

## 2.2 Jaringan Syaraf Tiruan Propagasi Balik

Jaringan Syaraf Tiruan merupakan suatu metoda pembelajaran mesin untuk dapat mengenali pola dan menghasilkan nilai keluaran berdasarkan bobot jaringan pada proses belajar. Salah satu metoda Jaringan Syaraf Tiruan adalah propagasi balik (back propagation). Back propagation akan mengestimasi error dengan gradient descent terhadap tiap-tiap jaringan yang berada pada arsitektur Jaringan Syaraf Tiruan tersebut. Hasil estimasi ini mempengaruhi nilai bobot jaringan yang siap untuk memprediksi pola masukan yang baru.

Untuk setiap neuron pada lapisan tersembunyi akan menerima masukan dari neuron lapisan masukan dikalikan dengan bobot jaringan yang memasuki neuron tersebut dan dijumlahkan. Begitu pula dengan neuron pada lapisan keluaran akan menerima masukan dari neuron lapisan tersembunyi dikalikan bobot jaringan yang memasuki neuron tersebut dan menjumlahkannya. Hasil keluaran dilakukan fungsi aktivasi dan menghitung error. Apabila error belum mencapai target, maka dilakukan estimasi error fungsi Sum Square Error (SSE) neuron lapisan keluaran. SSE merupakan cost function yang berbentuk kuadratik parabola.

$$\text{Sum Square Error (SSE)} = \frac{1}{2} \sum_{n=0}^{N-1} (\text{target Neuron Output}_k - \text{Output Neuron}_k)^2 \quad (3)$$

dimana  $n$  adalah pola yang dilatih,  $N$  adalah banyaknya neuron output dan  $k$  adalah indeks neuron output

## 2.3 Single Board Computer

Single Board Computer (SBC) merupakan perangkat komputer yang dikemas dalam satu papan PCB yang berisi mikroprosesor, memori dan antarmuka I/O[4]. Pada Tugas Akhir ini, penulis menggunakan Raspberry B+ sebagai media komputasi algoritma yang digunakan dan sistem kendali pada tugas akhir ini. SBC ini memiliki Prosesor ARM 11 dengan kapasitas RAM sebesar 512 MB.

2.5 Sistem Kendali Lingkak Terbuka

Sistem lingkak terbuka merupakan sistem yang tindakan pengendaliannya tergantung pada keluarannya[5]. Kelas keluaran ditentukan oleh proses belajar Jaringan Syaraf Tiruan dengan memperoleh bobot jaringan pada arsitekturnya dan mengambil bobot hasil belajar untuk digunakan proses pengujian.



Gambar 2. Diagram blok sistem kendali lingkak terbuka sistem yang diimplementasikan

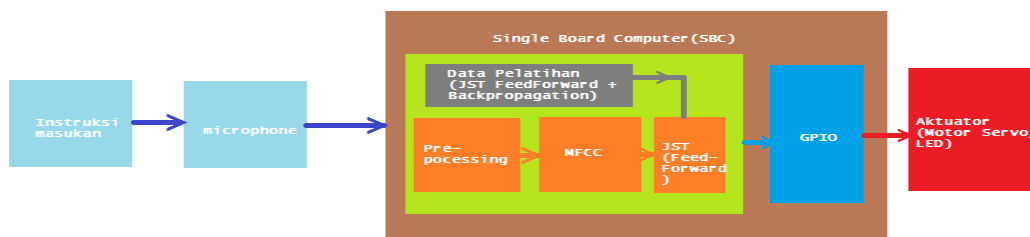
3. Perancangan Sistem

Diagram blok sistem yang dirancang ditunjukkan pada Gambar 3. Proses pertama yang dikerjakan adalah akuisisi suara masukan. Suara masukan direkam melalui microphone dalam rentang waktu yang ditentukan selama 2 detik dengan frekuensi sampling 16 kHz, 1 kanal (mono) dan tipe data sample integer 16. Hasil rekaman disimpan dalam format .wav untuk pemrosesan selanjutnya. Proses kedua adalah pre-processing sinyal akuisisi dengan menggunakan high pass filter digital sesuai persamaan (3)[6]

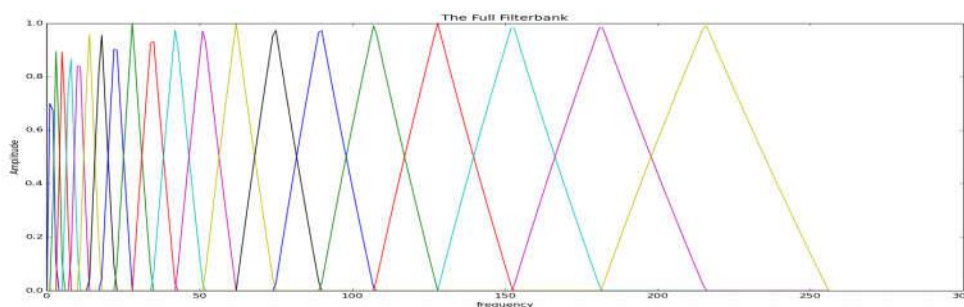
$$H(z) = 1 - az^{-1} \tag{3}$$

Dengan parameter  $a$  dipilih 0.97

Proses Ketiga adalah mencari nilai koefisien Mel dari proses MFCC. Keluaran dari proses ini adalah kumpulan array dengan panjang tiap array 12. Proses Keempat adalah proses pelatihan koefisien Mel yang telah diperoleh melalui algoritma Jaringan Syaraf Tiruan. Koefisien Mel yang telah ditetapkan untuk tiap perintah dilatih dan hasil latih pada tingkat tertentu menghasilkan nilai bobot jaringan yang akan digunakan untuk komputasi masukan baru dan mengklasifikasi informasi masukan tersebut. Total elemen dalam array hasil ekstraksi ciri, diambil sebanyak 2000 elemen dimulai dari index 0-1999.



Gambar 3. Diagram blok sistem keseluruhan

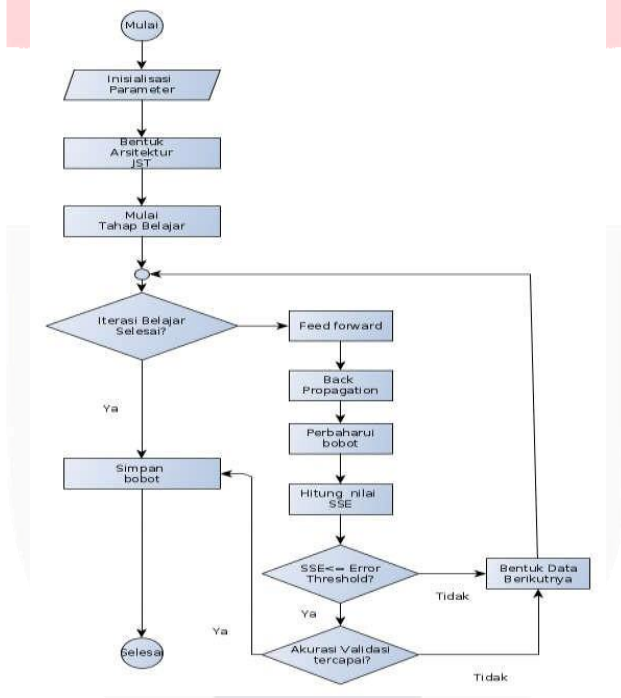


Gambar 4. 20 bank filter yang digunakan untuk memfilter respon magnitude pada tiap frame. Skala bank filter dimulai dari 0 Hz hingga 7000 Hz skala mel. Rentang ini dinormalisasi menjadi 0 Hz - 256 Hz agar mencakup rentang sampel pada respon magnitude tiap frame

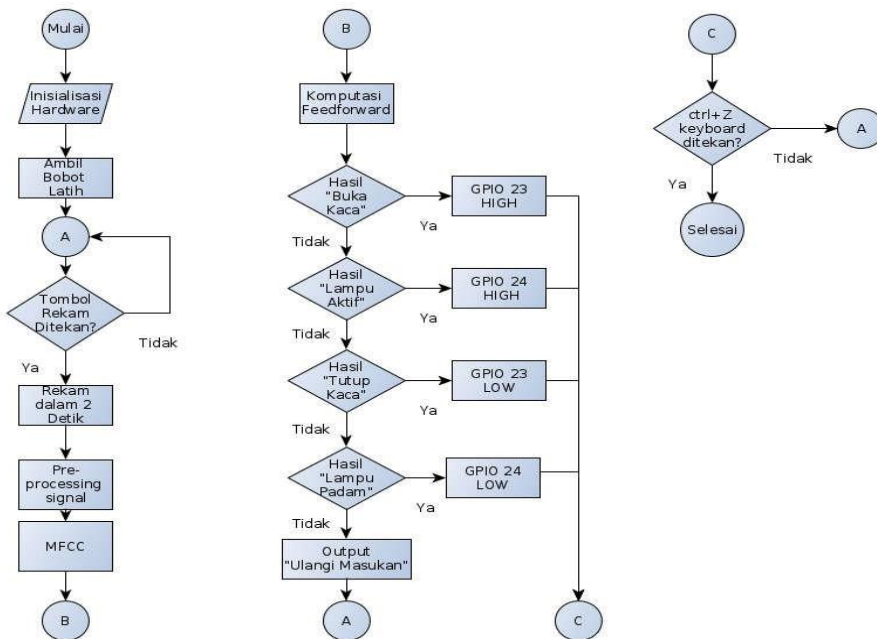
Tabel 1 Bentuk perintah masukan beserta bentuk keluaran

Perintah	Keluaran neuron output[0]	Keluaran neuron output[1]	Keluaran neuron output[2]	Keluaran neuron output[3]
“BUKA KACA”	1	0	0	0
“LAMPU AKTIF”	0	1	0	0
“TUTUPKACA”	0	0	1	0
“LAMPU PADAM”	0	0	0	1

Sebanyak 4 perintah untuk tugas output yang berbeda beda pada aktuator. Setelah melalui proses latih, bobot yang telah dihasilkan digunakan pada proses perambatan maju pada aplikasi utama dan menentukan kelas keluaran yang menentukan pin GPIO single board computer mana yang aktif atau non-aktif. Proses pembelajaran pola audio instruksi ditunjukkan pada Gambar 4 dan proses kerja real-time program Jaringan Syaraf Tiruan ditunjukkan pada Gambar 5.



Gambar 4. Diagram alir kerja proses belajar pengenalan pola suara di JST Back propagation

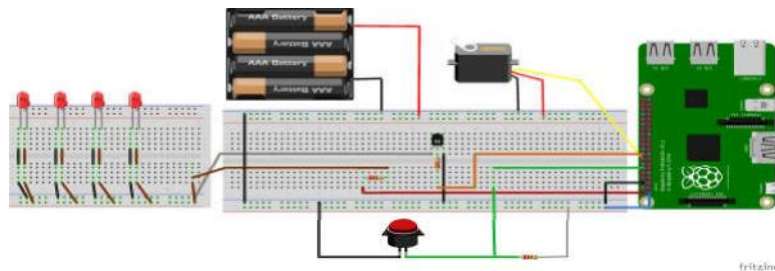


Gambar 5. Diagram alir kerja sistem pengujian perintah suara

### 3.1 Perancangan Perangkat Keras

#### 3.1.1 Perancangan Rangkaian Elektronika

Tiap-tiap komponen elektronika dihubungkan dengan pin GPIO Raspberry pi B+ yang ditunjukkan pada Gambar 6. Pemetaan tiap komponen pada GPIO Raspberry Pi ditunjukkan pada Tabel 2:



Gambar 6. Rangkaian elektronika sistem

Tabel 2 Pemetaan kabel dan aktuator pada GPIO

Komponen	GPIO
Kontrol Servo	23
Tombol Rekam	18
LAMPU LED*	24
GND	6

Push button digunakan untuk memulai merekam suara. Catu daya 6V DC digunakan pada motor servo dan himpunan LED menggunakan catu daya 5V dengan resistansi 100 Ohm antara catu daya LED dan himpunan LED. Transistor NPN 2N222A digunakan sebagai driver untuk aktivasi lampu led yang dihubungkan ke GPIO 24.

#### 3.3.2 Perancangan Mekanik

Perancangan mekanik dilakukan pada helm bertipe half face dengan menempatkan aktuator motor servo dan lampu led berada diatas helm. Untuk single board computer dan catu daya motor servo, diletakan di bagian belakang helm. Rangkaian aktuator dengan pin GPIO single board computer dihubungkan dengan kabel jumper dan dengan media bread board.



Gambar 7. Bentuk helm asisten yang diimplementasikan. (a) tampak samping, (b) Tampak depan, (c) tampak belakang

## 4. Pengujian

### 4.1 Hasil Pengujian Pengenalan Ucapan

Dilakukan perbandingan antara arsitektur Jaringan Syaraf Tiruan dan pemilihan learning rate sebagai pembandingan performansi terbaik. Dilakukan pelatihan dengan iterasi sebanyak 500 kali, parameter hidden layer sebanyak 1 layer, fungsi aktivasi sigmoid dengan faktor  $\sigma = 1$ , learning rate = 0.1 dan 0.001, momentum = 0.1

dan treshold pada fungsi aktivasi sigmoid sebesar 0.70. Implementasi program ekstraksi ciri dan Jaringan Syaraf Tiruan dilakukan dengan pemrograman Python dan kakas Numpy, Scipy, Cython, dan PyAudio.

#### 4.2.1 Arsitektur 2000-13-4

Proses pembelajaran pola dengan distribusi 180 data menjadi 140 data latih, 20 data validasi dan 20 data uji.



Hasil yang diperoleh pada pengujian di Raspberry pi B+ ditunjukkan pada Tabel 3 dan Tabel 4:

Tabel 3. Confusion matrix pengujian real-time dengan learning rate 0.1

80 kali percobaan (20 kali tiap instruksi)		Percobaan			
		Dikenali BUKA KACA	Dikenali LAMPU AKTIF	Dikenali TUTUP KACA	Dikenali LAMPU PADAM
Kelas Keluaran	BUKA KACA	16 (80%)	0	0	0
	LAMPU AKTIF	0	19 (95%)	0	0
	TUTUP KACA	0	0	18 (90%)	0
	LAMPU PADAM	1 (5%)	0	0	18 (90%)

Pembelajaran pola selesai pada iterasi ke - 11 dan Sum Square Error data latih 0.0008752805 dan Sum Square Error data validasi 0.0008631879. Perintah “BUKA KACA” tidak dikenali sebanyak 3 kali, perintah “LAMPU AKTIF” tidak dikenali sebanyak 1 kali, perintah “TUTUP KACA” tidak dikenali sebanyak 2 kali dan “LAMPU PADAM” tidak dikenali sebanyak 1 kali dan dikenali “BUKA KACA” sebanyak 1 kali. Kecepatan rata-rata eksekusi dalam menghasilkan kelas keluaran pada arsitektur ini selama 1.348 detik

Tabel 4. Confusion matrix real-time dengan learning rate 0.001

80 kali percobaan (20 kali tiap instruksi)		Percobaan			
		Dikenali BUKA KACA	Dikenali LAMPU AKTIF	Dikenali TUTUP KACA	Dikenali LAMPU PADAM
Kelas Keluaran	BUKA KACA	13 (65%)	3 (15%)	1 (5%)	3 (15%)
	LAMPU AKTIF	0	19 (95%)	0	0
	TUTUP KACA	5 (25%)	0	13 (65%)	1 (5%)
	LAMPU PADAM	0	0	0	20 (100%)

Pembelajaran pola selesai pada iterasi ke -17 dan Sum Square Error data latih 0.00105 dan Sum Square Error data validasi 0.00019. Perintah “BUKA KACA” tidak dikenali sebanyak 10 kali, perintah “LAMPU AKTIF” tidak dikenali sebanyak 3 kali, perintah “TUTUP KACA” tidak dikenali sebanyak 3 kali dan perintah “LAMPU PADAM” tidak dikenali sebanyak 3 kali. Kecepatan rata-rata eksekusi dalam menghasilkan kelas keluaran pada arsitektur ini selama 1.292475 detik.

#### 4.2.2 Arsitektur 2000-89-4

Dengan distribusi data yang sama seperti arsitektur 2000-13-4, hasil yang diperoleh pada pengujian di Raspberry pi B+ ditunjukkan pada Tabel 5 dan Tabel 6 :

Tabel 5. Confusion matrix pengujian real-time dengan learning rate 0.1

80 kali percobaan (20 kali tiap instruksi)		Percobaan			
		Dikenali BUKA KACA	Dikenali LAMPU AKTIF	Dikenali TUTUP KACA	Dikenali LAMPU PADAM
Kelas Keluaran	BUKA KACA	17 (85%)	0	0	0
	LAMPU AKTIF	0	18 (90%)	0	0
	TUTUP KACA	0	0	16 (80%)	0

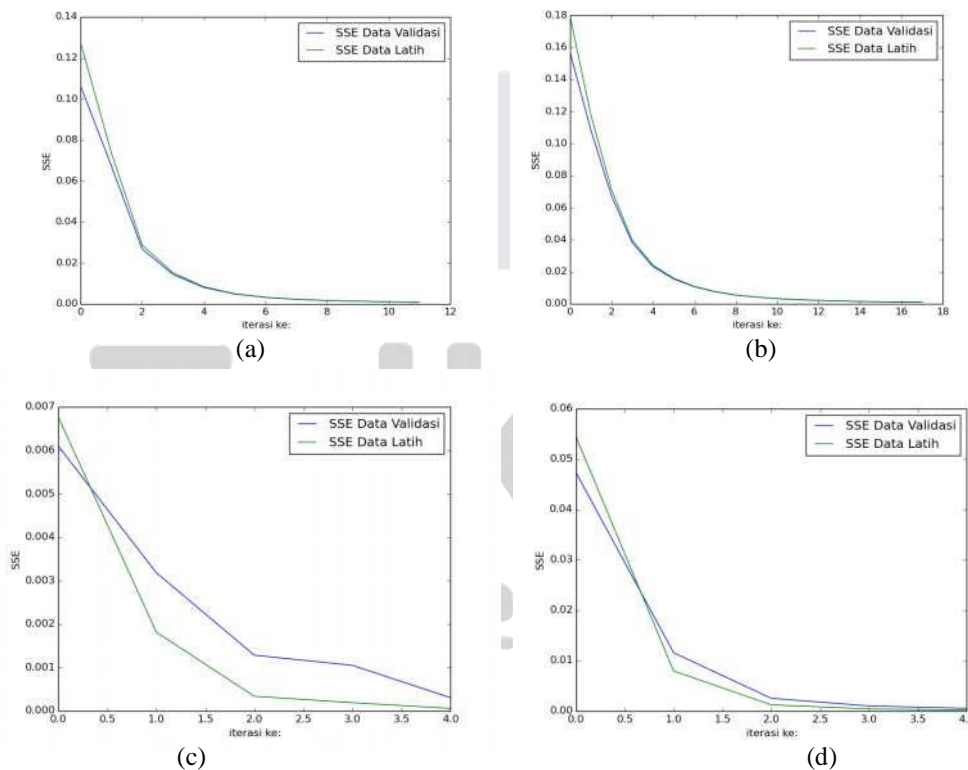
	LAMPU PADAM	0	0	0	20 (100%)
--	-------------	---	---	---	--------------

Pembelajaran pola selesai pada iterasi ke -4 dan Sum Square Error data latih 0.00019 dan Sum Square Error data validasi 0.00105. Perintah “BUKA KACA” tidak dikenali sebanyak 3 kali, perintah “LAMPU AKTIF” tidak dikenali sebanyak 2 kali, dan perintah “TUTUP KACA” tidak dikenali sebanyak 4 kali. Kecepatan rata-rata eksekusi dalam menghasilkan kelas keluaran pada arsitektur ini selama 4.2961 detik.

Tabel 6 Confusion matrix pengujian real-time dengan learning rate 0.001

80 kali percobaan (20 kali tiap instruksi)		Percobaan			
		Dikenali BUKA KACA	Dikenali LAMPU AKTIF	Dikenali TUTUP KACA	Dikenali LAMPU PADAM
Kelas Keluaran	BUKA KACA	7 (35%)	0	0	0
	LAMPU AKTIF	0	18 (90%)	0	0
	TUTUP KACA	0	0	18 (90%)	0
	LAMPU PADAM	0	0	0	17 (85%)

Pembelajaran pola selesai pada iterasi ke - 4 dan Sum Square Error data latih 0.0002099 dan Sum Square Error data validasi 0.0005639. Perintah “BUKA KACA” tidak dikenali sebanyak 13 kali, perintah “LAMPU AKTIF” tidak dikenali sebanyak 2 kali, perintah “TUTUP KACA” tidak dikenali sebanyak 2 kali dan perintah “LAMPU PADAM” tidak dikenali sebanyak 3 kali. Kecepatan rata-rata eksekusi dalam menghasilkan kelas keluaran pada arsitektur ini selama 4.2965 detik. Grafik SSE data latih dan data validasi pembelajaran untuk semua jenis arsitektur ditunjukkan pada gambar 8.



Gambar 8. Sum Square Error Data Latih dan Data Uji. Garis biru adalah SSE data latih dan garis hijau adalah SSE data validasi. (a) Arsitektur 2000-13-4 dengan learning rate 0.1, (b) Arsitektur 2000-13-4 dengan learning rate 0.001, (c) Arsitektur 2000-89-4 dengan learning rate 0.1, (d) Arsitektur 2000-89-4 dengan learning rate 0.1

## 5. Kesimpulan

Proses pembelajaran Jaringan Syaraf Tiruan berhasil diimplementasikan pada Single Board Computer Raspberry B+ dengan kaskas pemrograman yang berbeda dari tugas akhir dengan topik speech recognition menggunakan Jaringan Syaraf Tiruan sebelumnya. Tugas akhir ini mampu melakukan proses secara real time dengan hasil terbaik dalam waktu 1.34775 detik pada arsitektur 2000-13-4 learning rate 0.1 dengan akurasi 85% pada pengujian real-time .

Arsitektur dengan jumlah neuron pada lapisan tersembunyi dengan jumlah 13 neuron memiliki kecepatan eksekusi rata-rata 1.261 detik pengenalan real-time lebih cepat 3 detik dibandingkan dengan jumlah neuron sebanyak 89 dengan kecepatan rata-rata 4.2963 detik.

## Daftar Pustaka

- [1] Zadeh, Lotfi. 1997. "What is Soft Computing". Soft Computing. Springer-Verlag Germany/USA 1997
- [2] Suyanto. 2008. "Soft Computing, Membangun Mesin Ber-IQ Tinggi". Bandung. INFORMATIKA
- [3] Sutojo, T., dkk. 2011. "Kecerdasan Buatan". Yogyakarta. Andi Publisher
- [4] Rosch, Winn. 1999. "Hardware Bible Fifth Edition". ISBN0-7897-1743-3. pp.50-51
- [5] Huang, Xuedong., dkk. 2001. "Spoken Language Processing. A Guide to Theory, Algorithm and System Development". USA. Prentice Hall
- [6] Jyh-Shing Roger Jang, "Audio Signal Processing and Recognition," available at the links for real-time courses at the author's homepage at <http://www.cs.nthu.edu.tw/~jang>
- [7] Sirko, Mihael Pitz, Ralf Schlüter, Hermann Ney. "Computing Mel-Frequency Cepstral Coefficients on the Power Spectrum", Aachen – University of Technology, 52056 Aachen, Germany.
- [8] Hunter, J. D. 'Matplotlib: A 2D graphics environment.' Computing In Science Engineering, 9:90-95, 2007
- [9] Distefano, Joseph, dkk. "Sistem Pengendalian dan Umpan Balik". 1996. Jakarta. Penerbit Erlangga.
- [10] Lyons, James. 2012. <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/> [Diakses pada tanggal 24 Juni 2015]
- [11] Serra, Xavier. 1989. "A System For Sound Analysis/Transformation/Synthesis Based on A Deterministic Plus Stochastic Decomposition". USA. Stanford University.
- [12] Daqrouq, Khaled, dkk. 2010. "Speaker identification System-based Mel Frequency and Wavelet Transform using Neural Network Classifier". European Journal of Scientific Research, 4:515-525.