IMPLEMENTASI KUNCI BERBASIS SUARA MENGGUNAKAN METODE MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

Implementation of Voice Recognition Based Key Using Mel Frequency Cepstral Coefficient (MFCC)

Muhammad Nashih Rabbani[1], Achmad Rizal, S.T., M.T.[2], Dr. Ing. Fiky Yosef Suratman, S.T., M.T.[3]

[1,2,3]Prodi S1 Teknik Elektro, Fakultas Teknik Elektro, Universitas Telkom

[1] rabbani.email1@gmail.com, [2]achmadrizal@telkomuniversity.ac.id, [3]fysuratman@telkomuniversity.ac.id

Abstrak

Pada dasarnya setiap individu menghasilkan suara yang berbeda-beda, walaupun seseorang dapat menirukan suara tersebut namun suara yang dihasilkan tidak identik dengan suara yang ditiru. Sistem biometrik adalah sistem untuk melakukan identifikasi dengan menganalisa karakteristik fisik dan perilaku.

Tugas Akhir ini membuat suatu sistem keamanan suara berbasis mikro komputer yang diimplementasikan menjadi kunci. Tugas Akhir ini menggunakan metode MFCC sebagai ekstraksi ciri dan K-NN sebagai klasifikasi cirinya.

Pada penelitian Tugas Akhir ini telah berhasil membuat sistem pengenalan pembicara dengan tingkat akurasi terbaik sebesar 87.5% dan 1.80277 detik dengan menggunakan K = 5 dalam implementasi pembuka kunci menggunakan suara.

Kata kunci : *Mel-Frequency Cepstral Coefficient* (MFCC), *K-Nearest Neighbor* (K-NN), biometrik suara, kunci suara, Euclidean Distance.

Abstract

Basically, each individual produces a different sound, although one can imitate the sound, but the sound produced is not identical with that inimitable voice. Biometric system is a system for the identification by analyzing the physical characteristics and behavior.

This Final Project is to create a system of micro computer based voice security that is implemented into the key. This Final Project using MFCC as feature extraction and K-NN as classification characteristics.

In this Final Project has managed to make the speaker recognition system with the best accuracy rate of 87.5% and 1.80277 seconds using K = 5 on implementation of unlock using voice.

Keywords : MFCC (Mel Frequency Cepstral Coefficient), KNN (K-Nearest Neighbor), Speaker Recognition, Euclidean Distance, Biometric.

1. PENDAHULUAN

.

Di era globalisasi saat ini teknologi sudah berkembang pesat oleh sebab itu tuntutan di aspek keamanan dan privasi semakin meningkat. Walaupun teknologi keamanan sudah berkembang pesat, namun ada beberapa aspek yang masih perlu ditingkatkan. Itu semua bertujuan agar segala sesuatunya lebih mudah, aman, dan handal. Kunci sekarang yang masih banyak digunakan yaitu menggunakan kunci konvensional, *smart card*, dan kode pin. Beberapa kunci tersebut dirasa kurang efisien dikarenakan banyak hal yang mungkin bisa terjadi seperti, kunci hilang, lupa menyimpan, kunci tertinggal, dan terkadang lupa berapa kode pin yang harus dimasukkan.

Biometrik merupakan sebuah teknologi yang mengenali sebuah individu berdasarkan ciri fisiologis atau karakteristik perilaku[1]. Teknologi ini memiliki dua fase yaitu, identifikasi dan verifikasi. Identifikasi berfungsi untuk menentukan identitas seseorang. Verifikasi berfungsi untuk menerima atau menolak identitas yang didapatkan oleh seseorang[2]. Teknologi biometrik dirasa cukup praktis dan efisien untuk diterapkan sebagai kunci keamanan. Salah satu ciri yang dapat dikenali yaitu dengan suara yang diucapkan oleh seseorang.

Pada dasarnya setiap individu menghasilkan suara yang berbeda-beda, walaupun seseorang dapat menirukan suara tersebut namun suara yang dihasilkan tidak identik dengan suara yang ditiru. *Speaker recognition* (pengenalan pembicara) merupakan salah satu teknologi biometrik yang dapat mengenali identitas seseorang dari suaranya. Dengan teknologi biometrik menggunakan suara sistem membuka kunci akan lebih efisien, tidak lupa, dan tidak mudah untuk dipalsukan karena kunci tersebut terdapat pada diri seseorang.

Maka dari itu dibuatlah sebuah sistem biometrik yang memanfaatkan suara manusia sebagai masukan yang kemudian akan diidentifikasi dan diverifikasi. Dengan menggunakan Raspberry pi 2 model B memiliki *processor quad core*, ram 1 GB, dan memiliki ruang penyimpanan yang besar karena sistem membutuhkan ruang data yang besar untuk menampung *database* dan *processor* yang mampu melakukan komputasi yang berat. Proses ekstraksi

ciri yang digunakan menggunakan metode *Mel Frequency Cepstral Coefficient* (MFCC). Proses klasifikasinya menggunakan metode K *Nearest Neighbor* (K-NN).

2. TINJAUN PUSTAKA DAN PERANCANGAN

2.1 Biometrik

Identifikasi biometrik mengacu mengidentifikasi sebuah individu berdasarkan fisiologis atau karakteristik perilaku (biometrik pengidentifikasi). Karena banyak karakteristik fisiologis atau perilaku yang khas untuk setiap orang, pengidentifikasi biometrik lebih handal dan lebih mampu mengidentifikasi seseorang daripada hanya ingatan manusia[1].

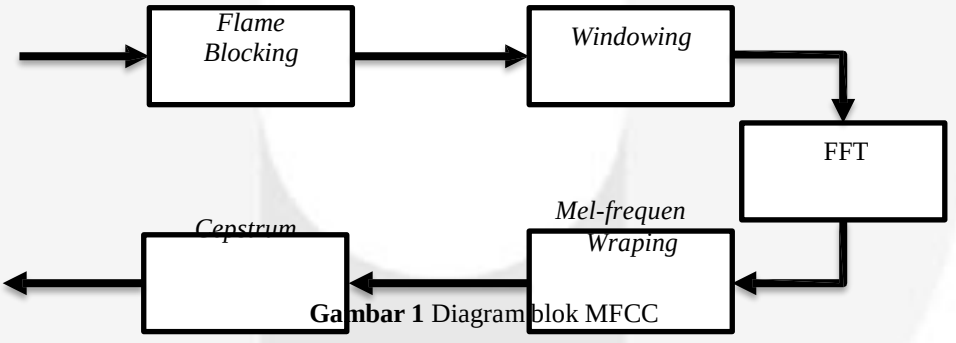
2.2 *Speaker Recognition*

Speaker recognition merupakan pengenalan pembicara, proses yang bertujuan mengetahui siapa yang berbicara. Pengenalan pembicara dapat diklasifikasikan ke dalam tiga tahap yaitu identifikasi, deteksi dan verifikasi. Identifikasi pembicara merupakan proses untuk menentukan identitas pembicara melalui suara yang telah diucapkan, sedangkan deteksi pembicara merupakan proses penemuan suara pembicara dari sekumpulan suara, dan verifikasi pembicara merupakan proses untuk memverifikasi kesesuaian suara pembicara dengan identitas yang diklaim oleh pembicara. Pengenalan pembicara lebih menitikberatkan pada pengenalan suara pembicara dan tidak pada pengenalan ucapan pembicara[5].

Semua sistem identifikasi melalui dua proses penting yaitu *feature extration* dan *feature matching*. *Feature extraction* merupakan proses mengekstraksi data hasil akuisisi sehingga dihasilkan data yang berdimensi lebih kecil, yang nantinya digunakan untuk merepresentasikan tiap-tiap pembicara. *Feature matching* menyangkut prosedur aktual yang mengidentifikasi pembicara yang tidak dikenal dan membandingkan fitur ekstraksi suara yang dimasukan dengan salah satu dari himpunan pembicara yang telah dikenal[7].

2.3 *Mel Frequency Cepstral Coefficient* (MFCC)

Mel Frequency Cepstral Coefficient (MFCC) merupakan salah satu metode ekstraksi ciri yang digunakan dalam bidang pengolahan suara. Metode ini digunakan untuk sebuah proses yang mengkonversikan sinyal suara menjadi beberapa parameter. Ekstraksi representasi parametrik terbaik sinyal akustik merupakan tugas penting untuk menghasilkan kinerja pengenalan yang lebih baik. Efisiensi dari tahap ini adalah penting untuk tahap berikutnya karena hal itu mempengaruhi perilakunya. MFCC didasarkan pada persepsi pendengarn manusia yang tidak dapat mendengar suara frekuensi lebih dari 1 kHz dengan kata lain, di MFCC didasarkan pada variasi dikenal dari telinga manusia bandwidth yang kritis dengan frekuensi. MFCC memiliki dua jenis filter yang spasi linear pada frekuensi rendah di bawah 1000 Hz dan logaritmik di atas 1000 Hz[13]. Hasil akhir proses MFCC yaitu mendapatkan nilai cepstrum. Cepstrum merupakan invers transformasi *fourier* dari spektrum energi[14]. Proses MFCC secara umum ditunjukkan pada Gambar 1.



Gambar 1 Diagram blok MFCC

MFCC terdiri dari lima langkah komputasi. Setiap langkah memiliki fungsi dan matematika pendekatan seperti yang dibahas secara singkat sebagai berikut:

1. *Flame Blocking*

   Proses segmentasi sampel bicara yang diperoleh dari analog konversi digital ke dalam bingkai kecil dengan panjang dalam kisaran 20 sampai 40 ms. Sinyal suara dibagi menjadi *frame* sampel N. *Frame* yang berdekatan dipisahkan oleh M (M<N). Panjang *frame* yang digunakan mempengaruhi hasil dalam analisis spektral. Proses *frame bloking* dilakukan sampai mencangkupi seluruh sinyal. Untuk menghindari hilangnya ciri dan karakteristik suara *overlapping* dilakukan sebagai perpotongan antar setiap *frame*. Panjang *overlapping* yang biasa digunakan yatu 30% sampai 50% dari panjang *frame*.
2. *Windowing*

Proses windowing dilakukan pada setiap *frame* bertujuan agar meminimumkan terjadi hilangnya informasi pada sinyal suara. Hamming window digunakan karena mempunyai hasil yang baik dalam menyaring sinyal yang akan dianalisis. Karena pada proses *frame blocking* dapat menyebabkan sinyal menjadi diskontinuitas. Oleh karena itu proses window pada setiap *frame* dilakukan.

3. *Fast Fourier Transform*

Untuk mengkonversi setiap *frame* sampel N dari domain waktu ke domain frekuensi. *Discrete Fourier Transform* (DFT) adalah metode untuk mengkonversi setiap *frame* sampel N dari domain waktu ke domain frekuensi. FFT merupakan metode transformasi *fourier* dengan proses lebih cepat. Rumus transformasi *fourier* terdapat pada persamaan (2.4) :

$$X(k) = FFT[h(n) * x(n)] = X(k) * X(k) \quad (2.4)$$

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{2\pi}{N}} \quad ; k = 0,1,2,\dots,N-1 \quad (2.5)$$

Keterangan:

$X(k)$ = Output DFT
N = Jumlah sampel yang akan diproses
$x(n)$ = Nilai sampel sinyal
k = variabel frekuensi diskrit

4. *Mel Frequency Wraping*

Frekuensi yang berkisar di spektrum FFT adalah sinyal yang sangat luas dan suara tidak mengikuti skala linier. *Filterbank* adalah salah satu bentuk dari filter yang dilakukan dengan tujuan untuk mengetahui ukuran energi dari *Frequency band* tertentu dalam sinyal suara. Frekuensi besarnya masing-masing penyaring tanggapan ini berbentuk segitiga dan sama dengan kesatuan di pusat frekuensi dan berkurang secara linier menjadi nol pada pusat frekuensi dua filter yang berdekatan. Kemudian, masing-masing filter *output* adalah jumlah dari yang disaring komponen spektral. Setelah itu persamaan (2.3) digunakan untuk menghitung mel untuk diberikan frekuensi f di Hz. Frekuensi skala mel terbagi menjadi dua, frekuensi linier yang berada di bawah 1000 Hz dan bentuk logaritmik berada di atas 1000 Hz.

$$X(Mel) = [2595 * \log 10 (1 + \frac{f}{700})] \quad (2.6)$$

Keterangan:
F(Mel) = Fungsi *Mel Scale*
f = Frekuensi

5. *Discrete Cosine Transform (Cepstum)*

Ini adalah proses untuk mengkonversi log mel spektrum menjadi domain waktu menggunakan *Discrete Cosine Transform* (DCT). Hasil konversi disebut mel frekuensi cepstrum koefisien. Set koefisien disebut vektor akustik. Oleh karena itu, setiap masukan ucapan diubah menjadi urutan vektor akustik.

$$C_n = \sum_{k=1}^{K} (\log S_k) \cos [n (k - \frac{1}{2}) \frac{\pi}{K}] ; n = 1,2,\dots,K \quad (2.7)$$

Keterangan:

S_k = keluaran dari proses filterbank pada indeks k
K = jumlah koefisien yang diharapkan

2.4 *K-Nearest Neighbor* (K-NN)

Ini merupakan metode yang nonparametik, menandai titk data baru, dengan menemukan titik terdekat dari data pelatihan. Untuk menemukan titik terdekat, digunakan pengukuran jarak berdasarkan kesamaan. Pengklasifikasian oleh K-NN yang utama dijelaskan oleh jumlah dari tetangga. Parameter ini mendefinisikan beberapa jenis efisiensi identifikasi, atau akurasi. Hal ini tidak mudah untuk mendefinisikan dan untuk penerapan yang berbeda baik untuk menggunakan jumlah tetangga yang berbeda.

1. Euclidean Distance

jarak minimum dari tes sinyal suara untuk masing-masing pelatihan sinyal suara di *training set* dihitung untuk menemukan kategori K-NN dari kumpulan data pelatihan. Pengukuran *Euclidean Distance* $d_E(x,y)$ digunakan untuk menghitung jarak antara pelatihan dan pengujian sinyal suara terdiri dari fitur N.

$$d(x,y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \quad (2.8)$$

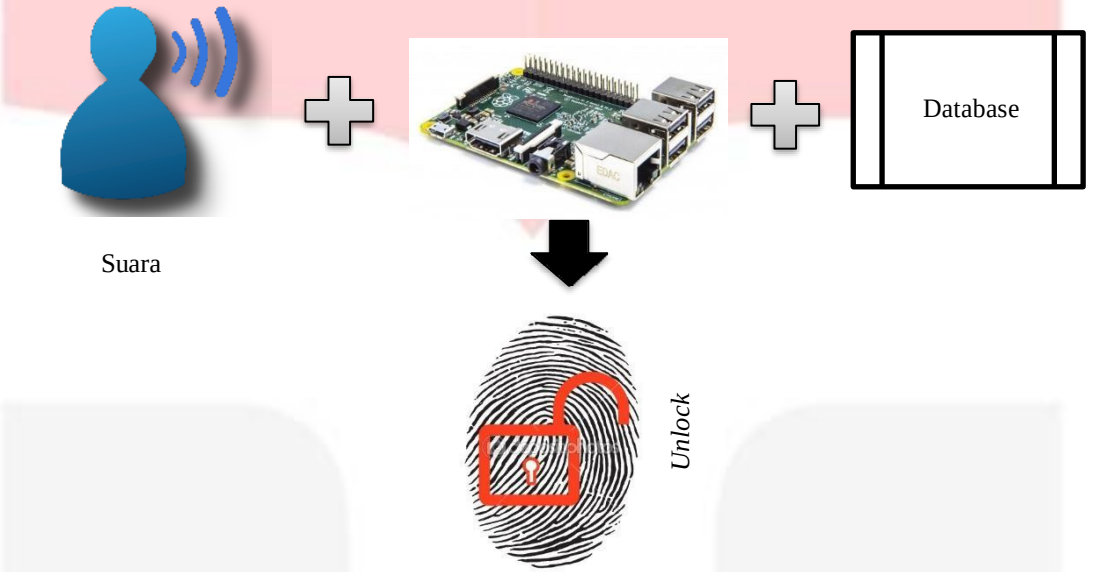
Keterangan:

$d_E(x,y)$: jarak skalar antara dua buah vektor x dan y dari matriks D dimensi

i : jumlah data ke n
N : jumlah data latih
x : data *training*
y : data *testing*

2.6 Model Sistem

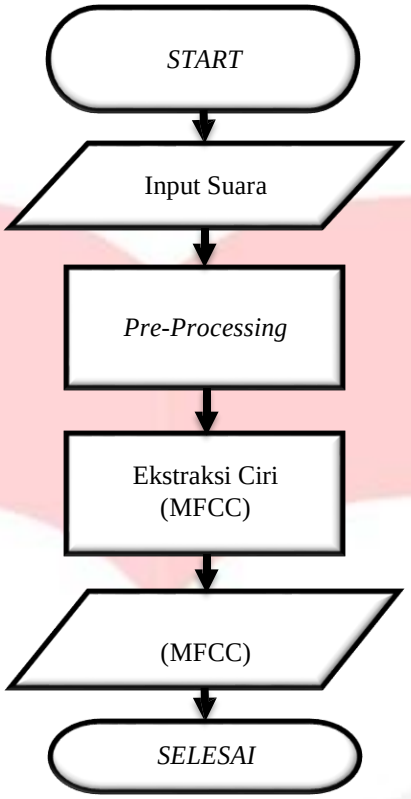
Sistem dirancang agar dapat mengenali masukan suara dan mengenali siapa yang berbicara. Sistem terbagi ke dalam dua fase yaitu fase *training* dan fase *testing*. Fase *training* merupakan tahap pendaftaran dan memodelkan suara pembicara. Memodelkan suara pembicara dilakukan untuk mendapatkan ciri dan karakteristik suara kemudian menyimpannya ke dalam *database*. Fase *testing* merupakan tahap yang mengenali suara yang berbicara apakah cocok dengan *database* atau tidak. Proses identifikasi dilakukan dengan cara menghitung jarak terdekat (*nearest neighbor*) dengan *database*. Sistem yang dirancang pada Tugas akhir ini dimodelkan pada Gambar 2.



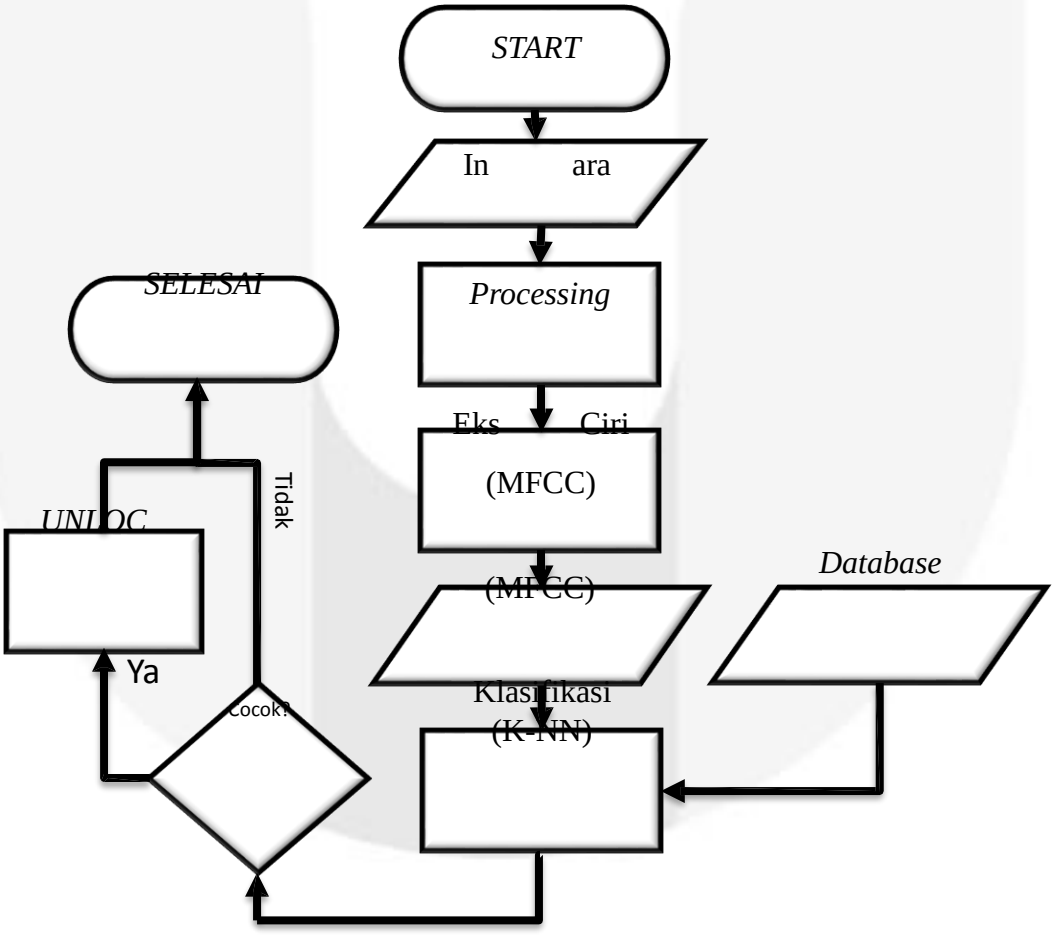
Gambar 2 Model sistem

2.7 Diagram Blok Sistem

Pada tahap tersebut hasil data akan disimpan menjadi *database* yang kemudian akan dicocokkan di tahap *testing*.



Gambar 3 Diagram blok latih



Gambar 4 Diagram blok uji

Proses *training* yang dijelaskan pada Gambar 3 bertujuan untuk mendaftarkan ciri individu dan kemudian disimpan ke dalam *database*. Pada tahap *testing* yang dijelaskan pada Gambar 4 data suara perekaman akan dicocokkan dengan suara di *database* yang bertujuan untuk memverifikasi pembicara.

3. PENGUJIAN DAN ANALISIS

3.1 Pengujian dan Analisis *Non Real Time*

1. Pengaruh jumlah mel bank filter yang digunakan

Tabel 1 Akurasi berdasarkan perbedaan jumlah filter

Filter	Jumlah Percobaan	Dikenali	Tidak Dikenali	Akurasi (%)
12	20	12	8	60%
32	20	14	6	70%
40	20	16	4	80%

Hasil pengujian akurasi menggunakan perbedaan jumlah filter ditunjukkan dalam Tabel 1. Analisis dari hasil pengujian sistem dengan mengubah parameter jumlah filter bank, bahwa nilai filter bank dapat mempengaruhi tingkat akurasi sistem. Semakin banyak filter maka, semakin tinggi tingkat akurasi sistem tersebut.

2. Pengaruh jumlah koefisien yang digunakan

Tabel 2 Akurasi berdasarkan perbedaan jumlah koefisien

Koefisien	Jumlah Percobaan	Dikenali	Tidak Dikenali	Akurasi (%)
10	20	13	7	60%
13	20	16	4	80%
16	20	15	5	75%

Hasil pengujian akurasi menggunakan perbedaan jumlah koefisien ditunjukkan dalam Tabel 2. Analisis dari hasil pengujian sistem dengan mengubah parameter jumlah koefisien, bahwa nilai koefisien dapat mempengaruhi tingkat akurasi sistem. Koefisien merupakan keluaran yang nantinya akan menjadi masukan data yang akan diuji. Banyaknya koefisien akan berpengaruh pada data secara keseluruhan. Terlalu sedikit koefisien tidak cukup untuk mewakili data yang akan dikenali, karena proses pengenalan akan semakin sulit untuk berhasil. Terlalu banyak akan membuat ciri semakin tidak jelas, karena jika terlalu banyak maka tidak terlalu berpengaruh. Jumlah koefisien terbaik adalah 13.

3.2 Pengujian dan Analisis *Real Time*

1. Pengujian kecepatan respon alat

Kecepatan waktu yang diamati adalah kecepatan komputasi pengendali utama mengenali data.

Tabel 3 pengujian kecepatan respon alat

Percobaan	Kecepatan
1	2,23124
2	2,11352
3	2,42211
4	2,45808
5	2,43683
6	2,13684
7	1,92358
8	2,69031
9	2,79034
10	2,67793
Rata-rata	2,388078

Kecepatan waktu yang diamati adalah kecepatan komputasi pengendali utama mengenali data. Kecepatan sistem untuk memverifikasi suara ditunjukkan pada Tabel 3. Dari percobaan yang dilakukan didapat bahwa rata-rata dalam 10 kali percobaan alat membutuhkan waktu sekitar 2,388078 detik untuk melakukan komputasi. Perbedaan waktu komputasi disebabkan oleh perbedaan data masukan dari tiap-tiap percobaan.

2. Pengujian data

Pengujian dilakukan oleh tiga orang yang mempunyai *database* dan satu orang yang tidak mempunyai *database*. Pada Tabel 4 diperlihatkan nilai banyaknya data yang terverifikasi.

Tabel 4 Hasil pengujian

Pembicara	Jumlah percobaan	Dikenali sebagai			Tidak dikenali
		S1	S2	S3	
S1	40	34	0	2	4
S2	40	0	35	1	4
S3	40	3	1	34	2
AS	40	2	3	3	32

- Pembicara 1

Percobaan yang dilakukan oleh pembicara 1 sebanyak 40 kali percobaan dan dikenali dengan benar sebanyak 34 kali.

$$\text{Akurasi} = \frac{34}{40} \text{ x } 100\% = 85\%$$

$$\text{Tingkat Error} = 100\% - 85\% = 15\%$$

- Pembicara 2

Percobaan yang dilakukan oleh pembicara 2 sebanyak 40 kali percobaan dan dikenali dengan benar sebanyak 35 kali.

$$\text{Akurasi} = \frac{35}{40} \text{ x } 100\% = 87{,}5\%$$

$$\text{Tingkat Error} = 100\% - 87{,}5\% = 12{,}5\%$$

- Pembicara 3

Percobaan yang dilakukan oleh pembicara 3 sebanyak 40 kali percobaan dan dikenali dengan benar sebanyak 34 kali.

$$\text{Akurasi} = \frac{34}{40} \text{ x } 100\% = 85\%$$

$$\text{Tingkat Error} = 100\% - 85\% = 15\%$$

- Pembicara asing

Percobaan yang dilakukan oleh pembicara asing sebanyak 40 kali percobaan dan dikenali dengan benar sebanyak 32 kali.

$$\text{Akurasi} = \frac{32}{40} \text{ x } 100\% = 80\%$$

$$\text{Tingkat Error} = 100\% - 85\% = 15\%$$

4. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan pada sistem pengenalan pebicara menggunakan MFCC sebagai ekstraksi ciri dan KNN sebagai klasifikasi, maka dapat diambil kesimpulan sebagai berikut :

1. *Mel-Frequency Cepstral Coefficient* adalah metode yang baik untuk ekstraksi fitur pada pengenalan suara.
2. Banyaknya *filter* dan koefisien pada MFCC dapat mempengaruhi akurasi dan waktu komputasi.
3. Setiap orang memliliki ciri suara yang berbeda-beda.
4. Percobaan pengujian akurasi memiliki akurasi terbaik 87,5% yang dilakukan oleh pembicara 2.
5. Semakin banyak data yang dibandingkan maka akurasi sistem akan semakin meningkat.
6. Semakin banyak data yang akan dibandingkan, maka proses komputasi akan semakin lama.

DAFTAR PUSTAKA

[1] I. S. Magazine and P. Ibm, "Biometric Recognition :," no. November, 2015.
[2] Anil Jain, Lin Hong, Sharath Pankanti, "BIOMETRIC," vol. 43, no. 2, 2000.
[3] L. Feng, "Speaker Recognition," 2004.
[4] D. Rhomanzah, "Sistem Kecerdasan Buatan Untuk Robot Asisten Berbasis Algoritma Case Base

Reasoning," 2015.
[5] C. E. Ho, "Speaker recognition system," *Proj. Report. Calif. Calif. Inst. Technol.*, 1998.
[6] A. Mudry, "Speaker Identification using Wavelet Transform," *Tesis Master Eng. Ontario Ottawa-carlet. Inst. Electr. Eng.*, 1997.

[7] K. Agustini, "Biometrik Suara Dengan Transformasi," no. 1, pp. 49–57, 1994.
[8] "Gudang Linux Indonesia," *16 September*, 2013. [Online]. Available: http://gudanglinux.com/glossary/sbc-single-board-computer/. [Accessed: 26-Nov-2015].
[9] "Raspberry pi 2 Model B," *February*, 2015. [Online]. Available: https://www.raspberrypi.org/products/raspberry-pi-2-model-b/. [Accessed: 26-Nov-2015].
[10] "Raspberry pi 2 on sale now at $35," *February*, 2015. [Online]. Available: https://www.raspberrypi.org/blog/raspberry-pi-2-on-sale/. [Accessed: 26-Nov-2015].
[11] "Fungsi Mikrofon Komputer." [Online]. Available: http://prokomputer.com/fungsi-mikrofon-komputer/. [Accessed: 26-Nov-2015].
[12] H. S. Manunggal, "Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara Dengan Menggunakan Analisa MFCC Feature Extraction.," *Tugas Akhir Sarj. pada Jur. Tek. Inform. Fak. Teknol. Ind. Univ. Kristen Petra Surabaya*, 2005.
[13] L. Muda, M. Begam, and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," *J. Comput.*, vol. 2, no. 3, pp. 138–143, 2010.
[14] A. V Oppenheim and R. W. Schefer, *Discrete Time Signal Processing*, vol. 1999. 1999.
[15] J.-S. R. Jang, *"Audio Signal Processing and Recognition."* available at the links for on-line courses at the autho's homepage at http://www.cs.nthu.edu.tw/~jang.
[16] J. K. Matúš PETEJA, "Speaker Identification," 2012.
[17] M. Hariharan, L. S. Chee, O. C. Ai, and S. Yaacob, "Classification of speech dysfluencies using LPC based parameterization techniques," *J. Med. Syst.*, vol. 36, no. 3, pp. 1821–1830, 2012.
[18] K. Rahayu, B. Hidayat, and S. Wibowo, "Analisis Dan Simulasi Sistem Penerjemah Kata Berbahasa Bali Ke Bahasa Inggris Berbasis Speech To Text Secara Real Time Menggunakan Metode Klasifikasi HMM."
[19] "Some tips on assignment 5 (MFCC)." [Online]. Available: http://www.comp.nus.edu.sg/~duanzy/mfcc.html. [Accessed: 16-Sep-2016].