

**SISTEM VERIFIKASI PENUTUR MENGGUNAKAN METODA MEL
FREQUENCY CEPSTRAL COEFFICIENTS-VECTOR QUANTISATION
(MFCC-VQ) SERTA SUM SQUARE ERROR (SSE) DAN PENGENALAN
KATA MENGGUNAKAN METODA LOGIKA FUZZY**

Oleh :

Atik Charisma ⁽¹⁾

⁽¹⁾Dosen Jurusan Teknik Elektro

Fakultas Teknologi Industri Institut Teknologi Padang

ABSTRAK

Metoda Mel Frequency Cepstral Coefficients-Vector Quantization (MFCC-VQ) dapat digunakan dalam sistem verifikasi penutur dan Logika Fuzzy dalam sistem pengenalan kata. Proses ekstraksi ciri sinyal wicara menggunakan metoda Mel Frequency Cepstral Coefficients (MFCC) nantinya akan menghasilkan vektor akustik sinyal wicara. Vector Quantization (VQ) digunakan untuk membentuk vektor akustik yang spesifik untuk tiap penutur. Pada pengenalan atau verifikasi, Sum Square Error digunakan untuk mencocokkan penutur tak dikenal dengan penutur dalam filebase berdasarkan error terkecil. Dalam penelitian ini, sistem digunakan untuk memverifikasi penutur dan mengenali kata, yaitu merah, biru, dan hijau dalam Bahasa Indonesia. Sistem ini telah diuji dengan membandingkan tingkat keberhasilan verifikasi penutur antara sumber suara yang digunakan sebagai filebase dan pemodelan kata dengan sumber suara yang tidak digunakan sebagai filebase. Untuk 10 kali pengucapan pada masing-masing pengujian diperoleh persentase keberhasilan verifikasi penutur yang baik. Pada pengujian sumber suara yang digunakan sebagai filebase, rata-rata persentase keberhasilan verifikasi adalah 70% dengan rata-rata persentase kebenaran pengenalan kata sebesar 87.5%, sedangkan pengujian sumber suara yang tidak digunakan sebagai filebase memperoleh persentase rata-rata keberhasilan verifikasi sebesar 78.3%.

Kata kunci : *Mel-Frequency Cepstral Coefficients (MFCC), Vector Quantization, Sum Square Error,*

ABSTRAC

The Method of Mel-Frequency Cepstral Coefficients Vector Quantization (MFCC-VQ) can be used in the speaker verification system and Fuzzy Logic in word recognition system. The process of feature extraction of speech signal using Mel Frequency Cepstral Coefficients (MFCC) vectors will produce acoustic speech signal. Vector quantization (VQ) is used to form the specific acoustic vector for each speaker. The introduction or verification, Sum Square Error is used to match unidentified speakers with speakers in filebase by the smallest error. In this research, the system is used to verify the speaker and recognize the word, namely red, blue, and green in Indonesian. This system has been tested by comparing the success rates between sound source speaker verification are used as filebase and modeling to the sound source said that is not used as filebase. To 10 times the pronunciation of each test the percentage of success obtained a good speaker verification. On testing the sound source used as filebase, the average percentage of verification success was 70% with an average percentage of correctness of 87.5% word recognition, while testing the sound sources that are not used as filebase obtain an average percentage of 78.3% successful verification.

Key Words : *Mel-Frequency Cepstral Coefficients (MFCC), Vector Quantization, Sum Square Error, .*

1. Pendahuluan

Kemajuan teknologi memberikan banyak andil dalam kontribusinya akan
Jurnal Teknik Eletro ITP, Volume 2 No. 2; Juli 2013

kemudahan pengenalan informasi pada suatu objek oleh sebuah mesin. Dalam melakukan kegiatan sehari-hari baik itu kegiatan pribadi

maupun kegiatan yang berhubungan dengan pekerjaan, saat ini telah banyak dibantu oleh hasil kemajuan teknologi dan informasi itu sendiri. Perkembangan teknologi telah terjadi di berbagai bidang. Salah satu bidang yang sedang dalam pengembangan adalah *Artificial Intelligent* atau Kecerdasan Buatan. Aplikasi dari teknologi ini adalah *Biometric Recognition*. Salah satu jenis teknologi biometrik adalah pengenalan suara (*voice recognition*). Suara yang dimaksudkan adalah sinyal suara yang dihasilkan langsung dari indera bicara pengakses.

Pengenalan suara (*voice recognition*) memiliki dua bagian, yaitu :

Pengenalan suara (*voice recognition*) memiliki dua bagian, yaitu pengenalan pembicara (*speaker recognition*) dan pengenalan ucapan (*speech recognition*). *Pengenalan pembicara (speaker recognition)* merupakan proses pengenalan suara berdasarkan orang yang berbicara. *Adapun pengealan ucapan (speech recognition)* adalah proses identifikasi suara berdasarkan kata yang diucapkan.

Teknik pengolahan suara (*speech processing*) dengan sistem pengenalan sumber suara (*speaker recognition*) dikembangkan berdasarkan prinsip sistem pendengaran manusia. Sistem ini digunakan untuk mengenali suara yang masuk dengan membandingkannya dengan suara yang tersimpan sebelumnya pada memori sistem tersebut. Sistem ini memungkinkan untuk menggunakan suara sebagai identifikasi atau pembuktian diri seseorang.

Hampir semua sistem pengenalan suara melakukan ekstraksi parameter untuk menampilkan bentuk sinyal dari kata yang diucapkan. Salah satu dari ekstraksi parameter dapat dilakukan dengan menggunakan *Mel Frequency Cepstrum Coefficients* (MFCC). Pemodelan ucapan manusia dalam suatu sistem pengenalan ucapan dapat dilakukan dengan menggunakan logika *fuzzy*.

1.1 Tujuan Penelitian

Penelitian ini bertujuan untuk merancang sistem verifikasi atau pengenalan penutur menggunakan *Mel Frequency Cepstral Coefficient-Vector Quantization* (MFCC-

VQ) serta Sum Square Error (SSE) dan pengenalan kata menggunakan Logika *Fuzzy*, yang dapat memverifikasi penutur sekaligus mengenali kata secara langsung dengan waktu proses pengenalan yang minimal.

2. Teori Dasar

2.1. Konsep Dasar Pengenalan Suara

Pengolahan suara (*speech processing*) dapat dikategorikan menjadi 3 bagian, yakni *analysis*, pengenalan (*recognition*), dan *coding*. Adapun pengenalan (*recognition*) terbagi lagi menjadi 3 bagian, yakni *speech recognition*, *speaker recognition* dan *language recognition*.

a. Pengenalan Pembicara (*Speaker Recognition*)

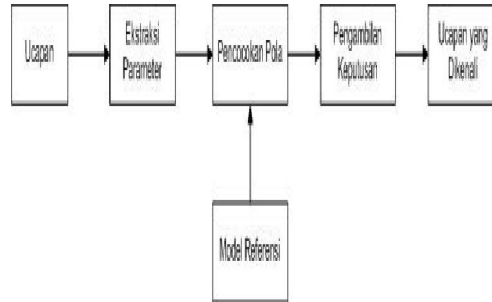
Pengenalan pembicara (*speaker recognition*) adalah suatu proses yang bertujuan untuk mengenali siapa yang sedang berbicara berdasarkan informasi yang terkandung dalam gelombang suara yang diinputkan. Teknik ini memungkinkan menggunakan suara penutur untuk memverifikasi identitas wicara dan mengontrol layanan seperti menekan nomor telepon dengan suara (*voice dialing*), perbankan dengan telepon, belanja melalui telepon, layanan akses melalui basis data (*database*), layanan informasi, surat dengan suara (*voice mail*), kontrol keamanan area rahasia, dan akses jarak jauh dengan komputer.

b. Pengenalan Ucapan (*Speech Recognition*)

Pengenalan ucapan didefinisikan sebagai proses pengubahan sinyal suara ke bahasa (*linguistic*) mesin dalam bentuk data digital (biasanya berupa teks sederhana). Dengan kata lain, pengenalan suara menyatakan kemampuan untuk mencocokkan pola dari yang didapatkan atau diperoleh perbendaharaan kata terhadap sinyal suara ke dalam bentuk yang tepat. Pengertian lainnya, pengenalan ucapan adalah suatu proses di mana komputer (jenis mesin lainnya) dapat mengenal kata-kata yang diucapkan oleh manusia. Proses ini disebut juga mengartikan ucapan manusia dalam komputer.

Sistem pengenalan suara (*voice recognition*) merupakan gabungan dari sistem pengenalan pembicara (*speaker recognition*)

dan pengenalan ucapan (*speech recognition*). Secara umum suatu sistem pengenalan ucapan, yang merupakan bagian dari pengenalan suara, terdiri atas dua proses utama. Proses pertama adalah ekstraksi parameter dan proses kedua adalah pencocokan pola. Proses ini dapat dilihat pada Gambar 2.1 :

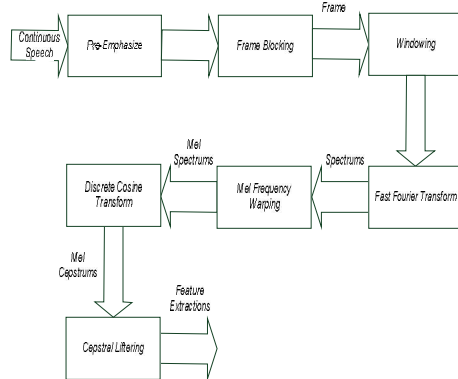


Gambar 2.1 Blok Diagram Sistem Pengenalan Ucapan

Proses sistem pengenalan suara pada sistem ini dimulai dengan mengambil sinyal suara manusia yang telah direkam dengan sebuah *microphone*. Sinyal ini di-*input*-kan ke komputer melalui *sound card* untuk mengubah sinyal analog menjadi sinyal digital agar sistem mudah untuk memproses lebih lanjut.

2.2. Mel Frequency Cepstrum Coefficients (MFCC)

Mel Frequency Cepstrum Coefficients (MFCC) merupakan salah satu metode yang banyak digunakan dalam bidang *speech processing*, baik itu *speech recognition* maupun *speaker recognition*. Metode ini digunakan untuk melakukan ekstraksi parameter, sebuah proses yang mengonversikan sinyal suara menjadi beberapa parameter.



Gambar 2.2 Blok Diagram MFCC

Tahap-tahap ekstraksi parameter menggunakan metode MFCC adalah sebagai berikut :

1. Pre-emphasis

Pre-emphasis merupakan salah satu jenis filter yang mempertahankan frekuensi-frekuensi tinggi pada sebuah spektrum, yang umumnya tereliminasi pada saat proses produksi suara.

2. Frame Blocking

Karena sinyal suara terus mengalami perubahan akibat adanya pergeseran artikulasi dari organ produksi vokal, sinyal suara harus diproses secara *short segment* (*short frame*).

3. Windowing

Tahap selanjutnya adalah me-*window* setiap *frame* yang bertujuan untuk meminimasi ketidakkontinuan sinyal pada permulaan dan akhir dari tiap-tiap *frame*.

4. Fast Fourier Transform (FFT)

Inti dari transformasi fourier adalah menguraikan sinyal ke dalam komponen-komponen bentuk sinus yang berbeda-beda frekuensinya.

5. Mel Frequency Wrapping

Tahap ini merupakan proses pemfilteran dari spektrum setiap *frame* yang diperoleh dari tahapan sebelumnya dengan menggunakan sejumlah *M filter* segitiga.

6. Discrete Cosien Transform (DCT)

Untuk mendapatkan nilai *cepstrum* MFCC, maka *mel* frekuensi tersebut harus ditransformasikan kembali menjadi *domain* waktu menggunakan metode Discrete Cosien Transform (DCT).

7. Vector Kuantisasi

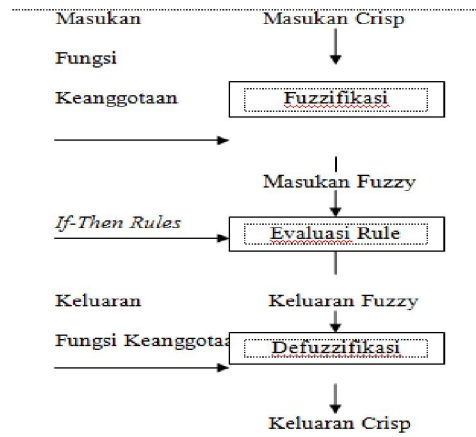
Vektor kuantisasi adalah proses memetakan vektor-vektor dari ruang vektor besar menjadi jumlah terbatas daerah ruang *vector*. Masing-masing daerah disebut kluster dan dapat direpresentasikan oleh pusatnya yang disebut *codeword*. Kumpulan dari semua *codeword-codeword* disebut *codebook*.

8. Sum Square Error (SSE)

Proses pematchingan dalam SSE adalah dengan mencari eror terkecil dari suara yang masuk dengan suara-suara yang telah mempunyai ciri (nilai) yang terdapat dalam *code book*.

2.3. Logika Fuzzy

Metode fuzzy menggunakan pernyataan - pernyataan jika – maka (*if-then*) untuk menyatakan hubungan antara input dengan output sistem.



Gambar 2.3 Konsep dari logika Fuzzy

Tahapan-tahapan logika fuzzy adalah :

1. Fuzzyfikasi

Proses fuzzifikasi adalah proses perubahan masukan *crisp* menjadi masukan *fuzzy*. Untuk mentransformasikan masukan *crisp* menjadi masukan *fuzzy*, diperlukan Fungsi Keanggotaan untuk setiap input. Proses fuzzifikasi mengambil nilai masukan *crisp* dan membandingkannya dengan Fungsi Keanggotaan yang telah ada untuk menghasilkan harga masukan *fuzzy*.

2. Evaluasi Rule

Pada tahap ini dilakukan evaluasi tiap *rule* dengan input yang dihasilkan dari proses fuzzifikasi.

3. Defuzzyfikasi

Pada proses defuzzifikasi, semua nilai keluaran *fuzzy* yang dihasilkan proses evaluasi *rule* dikombinasikan dengan kealuaran fungsi Keanggotaan untuk mendapatkan keluaran sesuai sistem yang diinginkan.

3. Langkah-langkah Penelitian

Program yang dirancang dalam tugas akhir ini terbagi atas 5 (lima) algoritma pokok, yaitu:

1. Algoritma ekstraksi parameter MFCC. Algoritma ini digunakan untuk memperoleh nilai koefisien MFCC dari sinyal suara yang di-*input*-kan.
2. Algoritma pengenalan pembicara (*speaker recognition*). Pengenalan pembicara bertujuan untuk melakukan proses verifikasi pembicara berdasarkan Sum Square Error.
3. Algoritma pelatihan atau pemodelan menggunakan *Logika Fuzzy* Tujuannya ialah untuk mendapatkan model dari ucapan yang di-*input*-kan.
4. Algoritma pengujian atau pengenalan *Logika Fuzzy* dan pengambilan keputusan.

4. Hasil Penelitian

Pada penelitian ini, ada dua jenis pengujian yang dilakukan, yaitu :

1. Pengujian kemampuan sistem memverifikasi penutur dan kemampuan sistem mengenali kata (warna) yang diucapkan penutur tersebut. Pengujian ini dilakukan oleh penutur yang sama dengan *filebase* sebanyak 10 kali untuk masing-masing kata “merah”, “biru, dan “hijau”.
2. Pengujian kemampuan sistem memverifikasi penutur terhadap sumber suara penutur lain sebanyak 2 orang.

Tabel 6.1 Hasil Pengujian Verifikasi Penutur dengan Penutur yang Sama dengan *Filebase*

Kata	Jumlah Pengucapan	Jumlah Pengenalan Pembicara Dikenali	Jumlah Pengenalan Kata yang Benar	Jumlah Kesalahan Pengenalan Pembicara	Jumlah Kesalahan Pengenalan Kata
Merah	10	4	3	6	1
Biru	10	9	9	1	-
Hijau	10	8	7	2	1

Tabel 6.2 Hasil Pengujian verifikasi Penutur dengan Penutur yang Berbeda dengan *Filebase*.

Kata	Jumlah Pengucapan	Jumlah Pengenalan Pembicara Dikenali	Jumlah Pembicara yang Tidak Dikenali
Merah	20	2	18
Biru	20	7	13
Hijau	20	4	16

5. Penutup

5.1. Kesimpulan

Dari penelitian yang telah dilakukan, dapat disimpulkan bahwa :

1. Penentuan *Sum Square Error* sebagai nilai *threshold* sangat berpengaruh pada keberhasilan sistem dalam memverifikasi penutur. Semakin tinggi nilai *threshold*, maka semakin tinggi keberhasilan sistem dalam mengenali penutur yang sama dengan *filebase* dan semakin rendah keberhasilan sistem dalam memverifikasi penutur lainnya.
2. Pengujian sistem dengan sumber suara yang sama dengan *filebase* memiliki rata-rata keberhasilan atau kebenaran untuk verifikasi penutur sebesar 70% dan 87.5% untuk pengenalan kata.
3. Pengujian sistem dengan sumber suara yang berbeda dengan *filebase* memiliki rata-rata keberhasilan dalam memverifikasi penutur adalah sebesar 78.3%

5.2. Saran

Penelitian ini hanya menggunakan satu suara *filebase* sehingga belum mewakili karakter suara melalui verifikasi penutur dan pengenalan kata, penulis menyarankan beberapa hal sebagai berikut :

1. Melalui sistem ini diharapkan dapat memverifikasi penutur dengan *filebase* lebih dari satu suara, sehingga sistem

dapat mengenali lebih banyak penutur untuk penelitian selanjutnya.

2. Pada penelitian selanjutnya diharapkan dapat mengenali lebih dari tiga kata.
3. Pada masa yang akan datang, sistem ini diharapkan dapat ditingkatkan persentase keberhasilannya dalam memverifikasi penutur. Tidak hanya menggunakan *Sum Square Error (SSE)* sebagai acuan dalam memverifikasi suara yang diinputkan, tapi juga menambahkan metoda lain, misalnya algoritma yang memiliki proses pembelajaran, seperti *Neural Network* ataupun *Genetic Algorithm*.

DAFTAR PUSTAKA

- [1] Fitrilina.2005. *Pengenalan Ucapan Jenis Isolated Word Recognition Berdasarkan Koefisien Prediksi Linier dengan Menggunakan Continuous Hidden Markov Model Tipe Bakis*. Tugas Akhir. Padang: Teknik Elektro Universitas Andalas.
- [2] Gu,Liang. *Perceptual Harmonic Cepstral Coefficients as the Fron-End for SpeechRecognition*.Signal Compression LaboratoryResearch.Project.http://scl.ece.ucsb.edu/html/prpat_3.htm.
- [3] Mustofa, Ali.2007. *Sistem Pengenalan Penutur dengan Metode Mel- frequency Wrapping*. Jurnal Teknik Elektro Vol. 7, No. 2, September 2007: 88 – 96.
- [4]Noel,Mike.*PrinciplesofSpeakerRecognition* .<http://cslu.cse.ogi.edu/HLTsurvey/ch1nod e47.html>
- [5] Rabiner, Lawrence, dan Biing-Hwang Juang. 1993. *Fundamentals of Speech Recognition*. New Jersey: Prentice Hall International, Inc.
- [6] Silvana, Meza.2006. *Optimalisasi Bobot Jaringan Syaraf Tiruan Menggunakan Algoritma Genetik dalam Identifikasi Suara*. Tugas Akhir. Padang: Teknik Elektro Universitas Andalas