

# APLIKASI PENDETEKSI EMOSI MANUSIA MENGGUNAKAN METODE MFCC DAN DTW

**Youllia Indrawaty Nurhasanah<sup>[1]</sup>, Mira Musrini Barmawi<sup>[1]</sup>, Adrian Hafiz David<sup>[1]</sup>**

<sup>[1]</sup> Jurusan Teknik Informatika, Fakultas Teknologi Industri, Itenas Bandung  
 youllia@itenas.ac.id, sangkuriang69@gmail.com, adrianyuki16@gmail.com

## ABSTRACT

*Classification of sound is one of the fields of Artificial Intelligent namely on the field of science Pattern Recognition, classification analysis based on the emotion was created in order for the computer to recognize the sound of angry, sad, happy, disgust and fear, by using voice as a voice recognition system or voice matching. This study that was made about the recognition of human emotions from the speaker by saying that depend on the text and the speaker, in the process of the recognition, Mel frequency cepstral Coefficients (MFCC) algorithm is being used. This technique is used to perform feature extraction process of the speech signal. While the process of matching the voice signal to be tested is using Dynamic Time wraping algorithm (DTW). Dynamic Mechanical Time wraping (DTW) is intended to make difference accommodations in time between the recording process when testing with the database that available on the reference signal. MFCC process will convert voice signals into several vectors which are useful for voice recognition process. Feature vector of the MFCC results will then be compared with the feature vector which are stored in the database using DTW.*

**Key Word :** Definition Of Voice, Emotion, MFCC, DT

## 1. Pendahuluan

Emosi merupakan suatu kondisi mental seseorang yang dapat mendorongnya untuk melakukan suatu tindakan atau berekspresi yang dapat dipicu dari dalam atau luar dirinya. Dalam kehidupan sehari-hari sangat penting untuk memahami kondisi emosional seseorang dengan emosi tertentu. Emosi juga merupakan salah satu aspek penting bagi kehidupan. Emosi seseorang dapat diketahui salah satunya dari ekspresi wajah, namun terkadang ekspresi wajah seseorang tidak sesuai dengan apa yang sedang dialaminya. Maka dari itu, untuk mempermudah dalam mengetahui emosi yang sedang dialami oleh seseorang, dibuatlah aplikasi untuk mendeteksi emosi seseorang berdasarkan suara menggunakan metode *Mel Frequency Cepstral Coefficients* (MFCC) dan *Dynamic Time Wrapping* (DTW).

## 1.2. Rumusan Masalah

Adapun masalah yang dirumuskan dalam penelitian ini sebagai berikut:

1. Bagaimana proses *Speech Recognition* dapat menentukan tingkatan emosi dengan menggunakan algoritma MFCC dan DTW.
2. Bagaimana cara algoritma MFCC untuk mengambil ekstraksi suara menjadi tingkatan emosi.
3. Bagaimana cara melakukan pencocokan data suara dengan algoritma DTW.

## 1.3. Tujuan

Berdasarkan perumusan masalah tujuan dari penelitian ini yaitu untuk mengembangkan penggunaan *Voice Recognition* dalam mendeteksi emosi melalui ekstraksi suara dengan menggunakan algoritma MFCC dan pencocokan menggunakan DTW.

## 1.4. Batasan Masalah

Dalam penelitian ini batasan masalah adalah sebagai berikut:

1. Input yang dikenali adalah suara user yaitu emosional marah, senang, sedih, bingung atau ketakutan.
2. Suara analog diubah menjadi digital dengan format \*.wav
3. Menerapkan *word dependent* dengan mengucapkan kalimat tertentu.
4. Penggunaan aplikasi hanya bisa pada umur tergolong dewasa dengan rentang usia 18-24 tahun
5. *Sampling rate* yang digunakan adalah 16000Hz
6. Bahasa yang digunakan dalam proses latihan dan proses uji adalah Bahasa Indonesia dan diucapkan oleh orang Indonesia.

## 2. Landasan Teori

### 2.1 Voice Recognition

*Voice Recognition* adalah suatu sistem yang dapat mengidentifikasi seseorang melalui suaranya, pada saat saya mencari pengertian dari *voice recognition* terdapat juga pengertian *Speech Recognition* yang

hampir sama secara fungsinya. Tapi terdapat perbedaan antara keduanya, *Voice Recognition* mengidentifikasi siapa yang berbicara, tetapi *Speech Recognition* mengidentifikasi apa yang diucapkan.

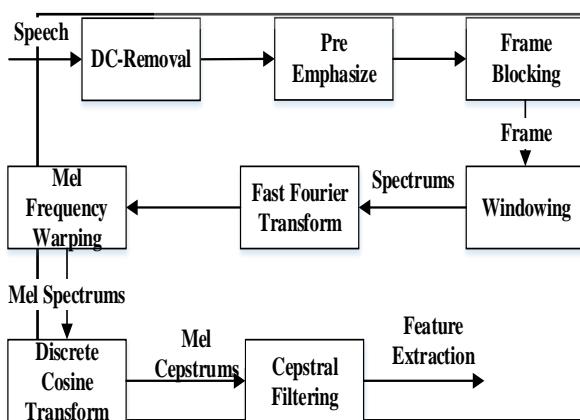
### 2.1.1 Speech Recognition

*Speech Recognition* adalah proses identifikasi suara berdasarkan kata yang diucapkan dengan melakukan konversi sebuah sinyal akustik, yang ditangkap oleh *audio device* (perangkat input suara). *Speech Recognition* juga merupakan sistem yang digunakan untuk mengenali perintah kata dari suara manusia dan kemudian diterjemahkan menjadi suatu data yang dimengerti oleh komputer.

### 2.1.2 Mel Frequency Cepstrum Coefficient (MFCC)

MFCC merupakan salah satu metode yang digunakan untuk melakukan *feature extraction*, sebuah proses yang mengkonversikan signal suara menjadi beberapa parameter. Beberapa keunggulan dari MFCC, antara lain:

1. Mampu untuk menangkap karakteristik suara yang sangat penting bagi pengenalan suara, atau dengan kata lain dapat menangkap informasi-informasi penting yang terkandung dalam signal suara.
  2. Menghasilkan data seminimal mungkin, tanpa menghilangkan informasi-informasi penting yang dikandungnya.
  3. Mereplikasi organ pendengaran manusia dalam melakukan persepsi terhadap signal suara.
- Tahapan Proses *Mel Frequency Cepstrum Coefficient* ditunjukkan pada Gambar 1.



Gambar 1 Tahap proses MFCC

#### 1. DC-Removal

Remove DC Components bertujuan untuk menghitung rata-rata dari data sampel suara, dan mengurangkan nilai setiap sampel suara dengan nilai rata-rata tersebut. Tujuannya adalah mendapat

normalisasi dari data suara input. Untuk penghitungan menggunakan *Dc-Removal* dapat diperlihatkan pada Persamaan (1) sebagai berikut:

$$y[n] = x[n] - \bar{x}, \quad 0 \leq n \leq N-1 \quad \dots \quad (1)$$

Keterangan :

$y[n]$  = sampel signal hasil DC removal

$x[n]$  = sampel signal asli

$\bar{x}$  = nilai rata-rata sampel signal asli.

$N$  = panjang signal

#### 2. Pre-Emphasize

*Pre-emphasize Filtering* merupakan salah satu jenis *filter* yang sering digunakan sebelum sebuah *signal* diproses lebih lanjut. *Filter* ini mempertahankan frekuensi-frekuensi tinggi pada sebuah spektrum, yang umumnya tereliminasi pada saat proses produksi suara.

Rumus *pre-emphasize* dapat dilihat pada persamaan (2) berikut:

$$y[n] = s[n] - \alpha s[n-1] \quad \dots \quad (2)$$

Keterangan :

$y[n]$  = signal hasil *pre-emphasize filter*

$s[n]$  = signal sebelum *pre-emphasize filter*

#### 3. Frame Blocking

*Frame* diartikan sebagai bingkai, dan *Blocking* diartikan sebagai pembagian menjadi beberapa blok. Jadi *frame blocking* adalah proses pembagian sinyal menjadi beberapa *frame*. Tujuan *frame blocking* adalah untuk memudahkan untuk melihat titik-titik sample pada sinyal ketika sinyal tersebut di perbesar. Panjang *frame* yang digunakan untuk pemrosesan sinyal adalah 10-30 ms. Untuk nilai default nya adalah 20 ms.

Rumus untuk menghitung jumlah *frame* dapat dilihat pada persamaan (3) berikut:

$$( \frac{(1-N)}{M} + 1 ) \quad \dots \quad (3)$$

Keterangan:

$I$  = *Sample rate*

$N$  = *Sample Point* (*Sample rate* \* waktu framing (s))

$M$  =  $N/2$

#### 4. Windowing

Proses ini dibutuhkan untuk mengurangi terjadinya kebocoran *spectral* atau *aliasing* pada sinyal. Proses *windowing* yang dipakai adalah *hamming window*, proses tersebut dapat dituliskan dalam persamaan (4) berikut:



$$D = \sqrt{(x_1 + y_1)^2 + (x_2 + y_2)^2} \quad \dots \dots \quad (9)$$

Keterangan :

$X$  = Sinyal uji

$Y$  = Sinyal Latih

## 2.2 Teori Emosi Manusia

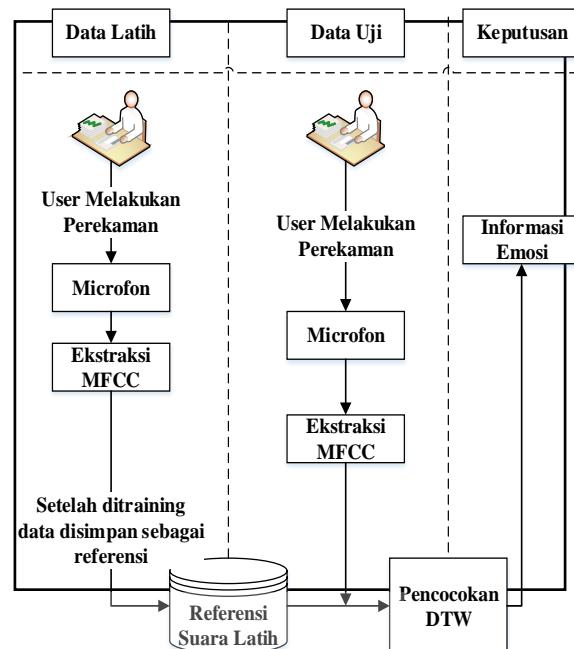
Kata Emosi berasal dari dua bahasa yaitu bahasa perancis *emotion* yang berarti kegembiraan dan dari bahasa latin *emovere* yang berarti bergerak menjauh. Emosi adalah perasaan intens yang ditujukan kepada seseorang atau sesuatu dalam arti lain emosi adalah reaksi terhadap seseorang atau kejadian.

Arti kata emosi diatas menyiratkan bahwa kecenderungan bertindak merupakan hal yang mutlak dalam emosi. Kebanyakan ahli yakin bahwa emosi lebih cepat berlalu daripada suasana hati. Emosi merupakan reaksi terhadap rangsangan dari luar dan dalam diri individu. Emosi dapat ditunjukkan ketika merasa senang mengenai sesuatu, marah kepada seseorang, ataupun takut terhadap sesuatu.

## 3. Analisis Dan Perancangan

Gambar umum sistem dapat dilihat pada Gambar 2. Pada proses awal yang dilakukan untuk mendeteksi emosi seseorang adalah melakukan proses data latih. Proses data latih digunakan untuk menyimpan referensi jenis emosi. Pada proses data latih pertama user melakukan proses perekaman suara menggunakan mikrofon kemudian suara yang masuk dilakukan ekstraksi ciri menggunakan metode MFCC, setelah itu didapat ciri suara dari data latih yang akan disimpan di database data latih.

Setelah dibuat data referensi data latih maka selanjutnya dilakukan data uji (pengujian). Pada data uji memiliki tahap yang hampir sama seperti proses pada data latih. Pada proses data uji ini user kembali melakukan perekaman suara berbentuk .wav. Hasil dari data rekaman suara tersebut akan menjadi data uji yang diekstraksi menggunakan MFCC. Setelah didapatkan ciri suara dari data uji, dilakukan pencocokan menggunakan metode DTW. Dengan menghitung jarak dari kedua vektor antara data yang diuji dengan data yang berada di dalam data referensi



Gambar 3 Workflow Cara kerja Sistem

## 4. Pengujian

Pada sub bab ini akan dijelaskan tentang pengujian sistem pada aplikasi yang dibuat.

### 4.1 Pengujian Sistem

Pada pembahasan pengujian sistem ini, pengujian dilakukan dengan dua cara yaitu pengujian secara Tidak Langsung dan pengujian Langsung.

#### 4.1.1. Pengujian Tidak Langsung

Pengujian ini dilakukan dengan menggunakan 10 data suara menggunakan bahasa Indonesia. Hasil pengujian tidak langsung (*Tidak Langsung*) untuk deteksi emosi emosi marah pada 1 orang dapat dilihat pada Tabel 1.

Tabel 1 Hasil Pengujian Tidak Langsung Menggunakan Sample Suara Alfira untuk Deteksi Emosi Marah

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Marah1	Marah
	Marah2	Marah
	Marah3	Sedih
	Marah4	Marah
	Marah5	Muak
	Marah6	Muak
	Marah7	Marah
	Marah8	Muak
	Marah9	Takut
	Marah10	Marah

Tingkat keberhasilan dalam (%) untuk deteksi emosi marah dapat dihitung menggunakan Persamaan (10) :

$$\text{Validitas} = \frac{\text{Hasil yang sesuai}}{\text{Banyaknya hasil uji coba}} \times 100\% \quad (10)$$

Sehingga dapat diketahui nilai keberhasilan dari deteksi emosi marah adalah:

$$\frac{5}{10} \times 100\% = 50\%$$

Hasil pengujian Tidak Langsung untuk deteksi emosi sedih dapat dilihat pada tabel 2.

Tabel 2 Hasil Pengujian Tidak Langsung Menggunakan Sample Suara Alfira untuk Deteksi Emosi Sedih

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Sedih1	Sedih
	Sedih2	Muak
	Sedih3	Muak
	Sedih4	Muak
	Sedih5	Sedih
	Sedih6	Sedih
	Sedih7	Sedih
	Sedih8	Sedih
	Sedih9	Sedih
	Sedih10	Sedih

Presentase keberhasilan deteksi emosi sedih adalah:

$$\frac{7}{10} \times 100\% = 70\%$$

Hasil pengujian Tidak Langsung untuk deteksi emosi sedih dapat dilihat pada tabel 3.

Tabel 3 Hasil Pengujian Tidak Langsung Menggunakan Sample Suara Alfira untuk Deteksi Emosi Takut

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Takut1	Takut
	Takut2	Sedih
	Takut3	Muak
	Takut4	Sedih
	Takut5	Takut
	Takut6	Takut
	Takut7	Takut
	Takut8	Senang
	Takut9	Takut
	Takut10	Takut

Presentase keberhasilan deteksi emosi takut adalah:

$$\frac{6}{10} \times 100\% = 60\%$$

Hasil pengujian Tidak Langsung untuk deteksi emosi sedih dapat dilihat pada tabel 4.

Tabel 4 Hasil Pengujian Tidak Langsung Menggunakan Sample Suara Alfira untuk Deteksi Emosi muak

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Muak1	Muak
	Muak2	Sedih
	Muak3	Muak
	Muak4	Sedih
	Muak5	Sedih
	Muak6	Senang
	Muak7	Muak
	Muak8	Muak
	Muak9	Takut
	Muak10	Muak

Presentase keberhasilan deteksi emosi muak adalah:

$$\frac{5}{10} \times 100\% = 50\%$$

Hasil pengujian Tidak Langsung untuk deteksi emosi sedih dapat dilihat pada tabel 5.

Tabel 5 Hasil Pengujian Tidak Langsung Menggunakan Sample Suara Alfira untuk Deteksi Emosi Senang

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Senang1	Senang
	Senang2	Takut
	Senang3	Takut
	Senang4	Marah
	Senang5	Senang
	Senang6	Senang
	Senang7	Muak
	Senang8	Senang
	Senang9	Senang
	Senang10	Senang

Presentase keberhasilan deteksi emosi senang adalah:

$$\frac{6}{10} \times 100\% = 60\%$$

#### 4.2.1. Pengujian Langsung

Pengujian ini dilakukan dengan menggunakan 10 data suara menggunakan bahasa Indonesia. Hasil pengujian langsung (*Langsung*) untuk deteksi emosi marah pada 1 orang dapat dilihat pada Tabel 6.

Tabel 6 Hasil Pengujian *Langsung* Menggunakan Sample Suara Vine untuk Deteksi Emosi Marah

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Marah1	Marah
	Marah2	Senang
	Marah3	Muak
	Marah4	Marah
	Marah5	Senang
	Marah6	Marah
	Marah7	Senang
	Marah8	Senang
	Marah9	Senang
	Marah10	Muak

Presentase keberhasilan deteksi emosi marah adalah:

$$\frac{3}{10} \times 100\% = 30\%$$

Hasil pengujian *Langsung* untuk deteksi sedih pada 1 orang dapat dilihat pada Tabel 7.

Tabel 7 Hasil Pengujian *Langsung* Menggunakan Sample Suara Vine untuk Deteksi Emosi Sedih

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Sedih1	Sedih
	Sedih2	Takut
	Sedih3	Takut
	Sedih4	Takut
	Sedih5	Takut
	Sedih6	Takut
	Sedih7	Marah
	Sedih8	Takut
	Sedih9	Takut
	Sedih10	Sedih

Presentase keberhasilan deteksi emosi sedih adalah:

$$\frac{2}{10} \times 100\% = 20\%$$

Hasil pengujian *Langsung* untuk deteksi takut pada 1 orang dapat dilihat pada Tabel 8.

Tabel 8 Hasil Pengujian *Langsung* Menggunakan Sample Suara Vine untuk Deteksi Emosi takut

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Takut1	Takut
	Takut2	Takut
	Takut3	Takut
	Takut4	Takut
	Takut5	Senang
	Takut6	Takut
	Takut7	Takut
	Takut8	Takut
	Takut9	Marah
	Takut10	Marah

Presentase keberhasilan dari deteksi emosi takut adalah:

$$\frac{7}{10} \times 100\% = 70\%$$

Hasil pengujian *Langsung* untuk deteksi muak pada 1 orang dapat dilihat pada Tabel 9.

Tabel 9 Hasil Pengujian *Langsung* Menggunakan Sample Suara Vine untuk Deteksi Emosi Muak

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Muak1	Muak
	Muak2	Muak
	Muak3	Muak
	Muak4	Muak
	Muak5	Marah
	Muak6	Marah
	Muak7	Marah
	Muak8	Muak
	Muak9	Muak
	Muak10	Marah

Presentase keberhasilan deteksi emosi muak adalah

$$\frac{6}{10} \times 100\% = 60\%$$

Hasil pengujian *Langsung* untuk deteksi senang pada 1 orang dapat dilihat pada Tabel 10.

Tabel 10 Hasil Pengujian *Langsung* Menggunakan Sample Suara Vine untuk Deteksi Emosi Senang

Kalimat Yang Diuji	Data Suara Yang Diuji (*.wav)	Hasil Uji
Saya Ikut Serta Dalam Kegiatan Itu	Senang1	Senang
	Senang2	Senang
	Senang3	Senang
	Senang4	Senang
	Senang5	Senang
	Senang6	Sedih
	Senang7	Senang
	Senang8	Senang
	Senang9	Takut
	Senang10	Senang

Presentase keberhasilan deteksi emosi senang adalah:

$$\frac{8}{10} \times 100\% = 80\%$$

## 5. PENUTUP

Berdasarkan hasil pengujian terhadap program dan pembahasan sebelumnya, dapat diambil kesimpulan sebagai berikut.

1. Dengan mengacu kepada Tabel 1, Tabel 2, Tabel 3, Tabel 4, Tabel 5 Maka dapat ditarik kesimpulan bahwa tingkat akurasi aplikasi pendekripsi emosi bervariasi yaitu tingkat akurasi untuk emosi Marah sebesar 50%, untuk emosi Sedih 70%, untuk emosi Takut sebesar 60%, untuk emosi Muak sebesar 50%, dan untuk

- emosi senang sebesar 60% dengan keputusan akhir berdasarkan pencocokan pada database bahwa emosi yang didapat adalah emosi Sedih.
2. Dengan Mengacu kepada Tabel 1, Tabel 2, Tabel 3, Tabel 4, Tabel 5 maka didapatkan nilai rentang frekuensi berdasarkan ciri – ciri emosi:
- Marah = 1910 s/d 3.715,7
  - Sedih = 4173,4 s/d 8718,4
  - Takut = 1444,8 s/d 2.066,1
  - Muak = 1964 s/d 3212
  - Senang = 3857,2 s/d 4881
3. Dengan mengacu kepada Tabel 6, Tabel 7, Tabel 8, Tabel 9, Tabel 10 Maka dapat ditarik kesimpulan bahwa tingkat akurasi aplikasi pendekripsi emosi bervariasi yaitu tingkat akurasi untuk emosi Marah sebesar 30%, untuk emosi Sedih 20%, untuk emosi Takut sebesar 70%, untuk emosi Muak sebesar 60%, dan untuk emosi senang sebesar 80% dengan keputusan akhir berdasarkan pencocokan pada database bahwa emosi yang didapat adalah emosi Senang.
4. Dengan mengacu kepada Tabel 6, Tabel 7, Tabel 8, Tabel 9, Tabel 10 maka didapatkan nilai rentang frekuensi berdasarkan ciri – ciri emosi:
- Marah = 446 s/d 5.598
  - Sedih = 1.357 s/d 2.080,3
  - Takut = 2.995,6 s/d 11.990,5
  - Muak = 10533 s/d 68274,4
  - Senang = 5872,4 s/d 11328,5
5. Pada aplikasi pendekripsi emosi ini perbedaan kalimat tidak dapat mempengaruhi hasil akurasi melainkan beberapa faktor eksternal seperti *noise* dan emosi yang dibuat-buat oleh user
- [6] Fadhilah Syafira, Agus Buono, dan Bib Paruhum Silalahi, 2014. *Pengenalan Suara Paru-Paru dengan MFCC sebagai Ekstraksi Ciri dan Backpropagation sebagai Classifier*. Bogor. Institut Pertanian Bogor.
- [7] Haryanto, S.Pd. 2009. *Pengertian Emosi*. <http://belajarpsikologi.com/pengertian-emosi/>, (Diakses pada 26 September 2015).
- [8] Jans Hendry. 2012. *Dynamic Time Warping Algorithm*. (Diakses Pada 12 Oktober).
- [9] Sri Waluyanti. 2008. Pengertian Sinyal Dasar Audio. (Diakses Pada 12 Oktober).
- [10] Rena Latifa , 2012. *Psikologi Emosi:Ekspresi Emosi*. <https://sites.google.com/a/uinjkt.ac.id/renalatifa/article/s/bab-4-psikologi-emosi-ekspresi-emosi> (Diakses Pada 17 November).
- [11] Reza Dio Nugraha, Youllia Indrawaty, Rio Korio Utomo. 2014. *Pengenalan Pola Ucapan Kata Menggunakan Metode Dynamic Time Warping(DTW) Berbasis Multimedia Interaktif*. Bandung. Institut Teknologi Nasional.
- [12] Reza Rianto, Youllia Indrawaty, Andriana Zulkarnain. 2014. *Pengembangan Pembelajaran Kata Dalam Bahasa Indonesia Menggunakan Multimedia Interaktif Dan Speech Recognition*. Bandung. Institut Teknologi Nasional.

## REFERENSI

- [1] Ali Mustofa. 2007. *Sistem Pengenalan Penutur dengan Metode Mel-Frequency Wraping*. Malang. Universitas Brawijaya.
- [2] Angga Setiawan, Achmad Hidayatno, R. Rizal Isnanto . 2011. *Applikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) Untuk mengoperasikan kursor computer*. Semarang. Universitas Dipenogoro.
- [3] Atik Charisma. 2013. *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*. Padang. Institut Teknologi Padang.
- [4] Darma Putra, Adi Resmawan. 2011. *Verifikasi Biometrika Suara Menggunakan Metode MFCC Dan DTW*. Bali. Universitas Udayana.
- [5] Ersa Triansyah dan Youllia Indrawaty. 2015. *Implementasi Metode Pattern Recognition Untuk Pengenalan Ucapan Huruf Hijayah*. Bandung. Institut Teknologi Nasional.