

IMPLEMENTASI METODE *PATTERN RECOGNITION* UNTUK PENGENALAN UCAPAN HURUF HIJAIYYAH

¹⁾ Ersa Triansyah, ²⁾ Youllia Indrawaty N
^{1,2)} Jurusan Teknik Informatika. Fakultas Teknologi Industri
Institut Teknologi Nasional
¹⁾ ersa.triansyah@yahoo.com, ²⁾youllia@itenas.ac.id

Abstrak

Pattern recognition memiliki kemampuan untuk mengenali suara dengan melakukan pengenalan pola suara melalui fitur-fitur sinyal suara yang kemudian dilakukan pengenalan pola melalui perbandingan pola suara uji dengan suara referensi. Untuk mendapatkan fitur-fitur sinyal suara, diperlukan metode untuk mengekstraksi sinyal suara sehingga fitur-fitur sinyal suara yang dibutuhkan terpenuhi. MFCC (*Mel Frequency Cepstral Coefficients*) merupakan alternatif metode untuk melakukan ekstraksi sinyal yang menghasilkan koefisien cepstral dari sinyal suara. Koefisien cepstral sinyal suara dari hasil ekstraksi tersebut, kemudian dilakukan perbandingan kesesuaian antara suara uji dan suara referensi. DTW (*Dynamic Time Warping*) salah satu algoritma untuk dapat melakukan perbandingan koefisien tersebut. Dalam kasus pengenalan ucapan huruf *hijaiyyah* umumnya dilakukan secara *talaqqi* (belajar intensif) antar seorang guru dengan murid, penilaian yang dilakukan bersifat subjektif berdasarkan kemampuan indera dari seorang guru, untuk itu aplikasi pengucapan huruf *hijaiyyah* merupakan salah satu alternatif untuk mengenali dan menguji kesesuaian ucapan secara objektif melalui penghitungan matematis dengan melakukan pengenalan pola suara. Dari pengujian yang telah dilakukan, dari 6 orang yang diuji melakukan pengucapan 29 huruf 3 tanda baca dan pengulang sebanyak 5 kali menghasilkan persentase kecocokan suara mencapai di atas 90 %, nilai *threshold* 1,3

Kata kunci: *Speech Recognition, Pattern Recognition, MFCC, DTW, Hijaiyyah*

Abstract

Pattern recognition has ability to recognize voice by voice pattern recognition through voice signal features which then carried out voice pattern recognition through comparison of tester's voice pattern with a reference voice. To get the sound signal features, it needs a method for extracting sound signal so that required sound signals features are fulfilled. MFCC is an alternative method to perform signal extraction which is produce cepstral coefficients of the sound signal. Cepstral coefficients of sound signal from the extraction then will be compared by the match between tester voice and reference voice. DTW is one of algorithm to do a comparison of the coefficients. In the case of introducing hijaiyyah generally talaqqi (intensively) conducted between a teacher and students, the appraisal is subjective based on the sensory capabilities of the teacher, therefore hijaiyyah pronunciation application is an alternative to identify and test the suitability of speech objectively through mathematical calculations by performing voice pattern recognition. From the testing that has been done, from 6 people tested do pronunciations 29 letters and punctuation repeater 3 to 5 times the yield percentage matches the sound reaches above 90%, a threshold value of 1.3.

Keywords: *Speech Recognition, Pattern Recognition, MFCC, DTW, Hijaiyyah*

I. PENDAHULUAN

Pattern Recognition adalah sebuah metode pengenalan pola yang merupakan bagian dari kecerdasan buatan. Metode ini memiliki kemampuan untuk mengambil data-data mentah seperti sinyal suara, lalu dari sinyal tersebut dilakukan sebuah pengambilan keputusan. Data yang diperoleh dari

pengambilan keputusan diolah dengan cara membandingkannya dengan data yang sudah dikenali oleh sistem.

Dalam tahap membandingkan sinyal suara yang telah dikenali sistem (sinyal referensi) dengan sinyal suara asli, proses ini dilakukan dengan menggunakan bantuan algoritma *Dynamic Time Warping* (DTW). Algoritma DTW merupakan algoritma pemrograman dinamis yang efisien dalam memecahkan masalah mengenai proses pengenalan pola yang tugasnya mencocokkan sinyal berdasarkan normalisasi waktu maupun pola.

Untuk mendapatkan normalisasi sinyal suara berdasarkan waktu maupun pola, *Mel Frequency Cepstral Coefficient* (MFCC) merupakan alternatif metode untuk mengekstraksi sinyal suara menjadi nilai koefisien yang dapat dieksekusi ke dalam normalisasi waktu maupun pola. Salah satu sinyal suara yang dapat diekstraksi oleh MFCC yaitu sinyal suara dari pengucapan huruf hijaiyyah

Salah satu pembelajaran pengucapan huruf hijaiyyah ini bisa dilakukan dengan menguji pengucapan seseorang. Dengan mendengarkan suara seseorang yang sedang mengucapkan huruf hijaiyyah lalu dicek huruf yang di ucapkan seseorang tersebut sesuai atau tidak sesuai. Namun penilaian ini bersifat subjektif dimana seseorang menilai ucapan dari hasil pendengarannya. Adapun penilaian agar bersifat objektif bisa dilakukan dengan menghitung rasio kecocokan sinyal suara yang sudah sesuai dengan kaidah bacaannya dengan suara yang diucapkan seseorang sehingga nilai rasio menentukan sinyal suara yang diucapkan sesuai atau tidak sesuai.

Dari beberapa hal yang telah dipaparkan maka perlu dibuat sebuah aplikasi yang dapat membantu seseorang yang ingin belajar membaca Al-Quran secara benar dengan mengenalkan dan menguji pengucapan huruf-huruf hijaiyyah yang sesuai dengan teknik bacaannya secara objektif. Sehingga seseorang yang ingin belajar pengucapan huruf hijaiyyah dapat mengetahui, memahami dan menguji pengucapannya sesuai atau tidak sesuai.

II. KAJIAN LITERATUR

II.1 Speech Processing

Speech Processing (pemrosesan ucapan) adalah metode mengekstrak informasi yang diinginkan dari sebuah sinyal suara. Untuk memproses sebuah sinyal dengan sebuah computer digital, sinyal harus

dihadirkan dalam bentuk digital sehingga sinyal tersebut dapat digunakan oleh sebuah computer digital (Rabiner, 1993). Awalnya, gelombang suara akustik diubah ke suara sebuah sinyal digital untuk *voice processing*. Sinyal digital disini adalah sinyal analog yang telah melalui proses *sampling*, *quantization*, dan *encoding*.

II.2 Speech Recognition

Speech Recognition (Pengenalan Ucapan), atau yang sering disebut dengan *Automatic Speech Recognition* (ASR) adalah suatu pengembangan teknik dan sistem yang memungkinkan komputer untuk menerima masukan berupa kata yang diucapkan (Retnawati, Wirawan, Wijati, 2009). Teknologi ini merupakan suatu perangkat untuk mengenali dan memahami kata-kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu untuk mengidentifikasi kata-kata tersebut. Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan atau dapat dibaca oleh perangkat teknologi sebagai sebuah komando untuk melakukan suatu pekerjaan.

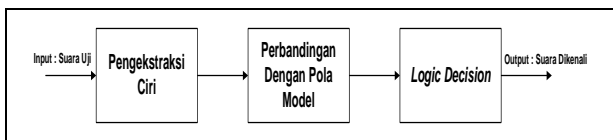
Speech Recognition merupakan salah satu upaya agar suara dapat dikenali atau diidentifikasi sehingga dapat dimanfaatkan untuk berbagai aplikasi. Pengenalan ucapan dapat dibedakan ke dalam tiga bentuk pendekatan, yaitu pendekatan akustik-fonetik (*the acoustic-phonetic approach*), pendekatan kecerdasan buatan (*the artificial intelligence approach*), dan pendekatan pengenalan-pola (*the pattern recognition approach*) (Putra, dkk, 2011)

II.3 Pattern Recognition

Sebuah kamus inggris mendefinisikan sebuah pola (*pattern*) sebagai satu contoh atau model sesuatu yang dapat disalin. Sebuah pola adalah juga tiruan sebuah model. Tetapi disaat menjelaskan berbagai tipe objek dalam dunia fisik dan abstrak, definisi yang muncul dalam ingatan - sebuah pola adalah setiap antar hubungan data (analog atau digital), kejadian dan/atau konsep yang dapat dibedakan. Bentuk wajah, sebuah meja, urutan nada sebaik musik, tema sebuah sajak atau simponi, jejak yang dibuat partikel pada pelat fotografik, kesemuanya merupakan tipe yang berlainan dari pola-pola. Jadi, pengenalan sebuah wajah, sebaik musik, lukisan, perkataan dicetak, kesemuanya adalah masalah pengenalan pola.

Pengenalan pola langsung mencakup pengenalan visual dan aural spasial (karakter, gambar, sidik jari) dan temporal (gelombang, ucapan), dimana seseorang membutuhkan bantuan alat pengindra (sensor).

Pattern Recognition adalah sub topik dari *machine learning* pada *computer science*. *Pattern recognition* dapat didefinisikan sebagai cara untuk mengambil data mentah (raw data) dan mengambil sebuah aksi dari kategori data-data tersebut (Ronando dan Irawan, 2012). Pada blok diagram pengenalan pola (Gambar 1), bahwa dalam pengenalan suara diperlukan suara uji, pengekstraksi ciri, perbandingan dengan pola model, *logical decision*, dan keluaran suara yang dikenali.



Gambar 1 Blok Diagram Pattern Recognition (Sumber: Syaiful R (L2F001644))

Penjelasan setiap blok diagramnya sebagai berikut :

1. Pengekstraksi Ciri merupakan proses mendapatkan sederetan besaran pada bagian sinyal masukan untuk menetapkan pola pembelajaran atau pola uji.
2. Perbandingan Dengan Pola Model merupakan pola uji yang akan dikenali dan dibandingkan dengan setiap kelas pola referensi. Kesamaan besaran antara pola uji dengan setiap pola referensi akan dihitung.
3. *Logic Decision* merupakan langkah untuk menentukan kelas pola referensi mana yang paling cocok untuk pola uji berdasarkan klasifikasi pola.

II.4 Mel Frequency Cepstrum Coefficients (MFCC)

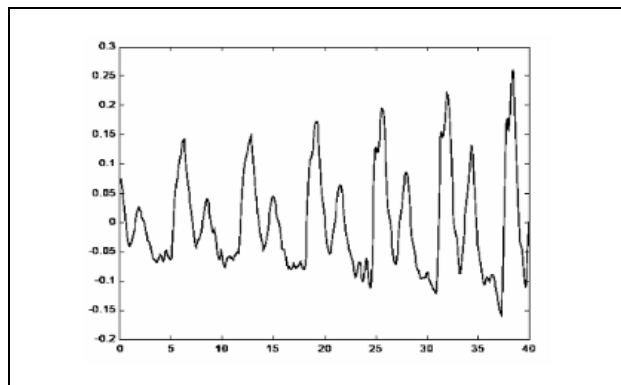
MFCC (*Mel Frequency Cepstrum Coefficients*) merupakan salah satu metode yang banyak digunakan dalam bidang *speech technology*, baik *speaker recognition* maupun *speech recognition*. Metode ini digunakan untuk melakukan *feature extraction*, sebuah proses yang mengkonversikan *signal* suara menjadi beberapa parameter (Resmawan, 2011). Beberapa keunggulan dari metode ini adalah:

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.

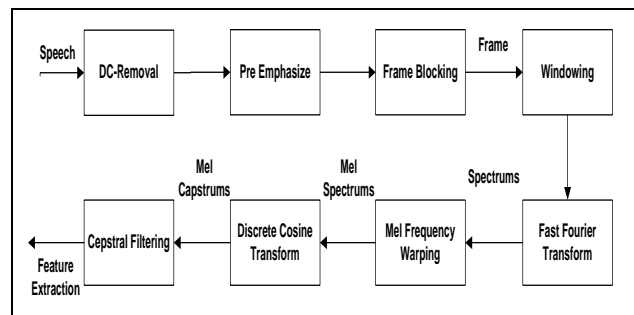
2. Menghasilkan data seminimal mungkin, tanpa menghilangkan informasi-informasi penting yang dikandungnya.
3. Mereplikasi organ pendengaran manusia dalam melakukan persepsi terhadap *signal* suara.

Contoh dari *signal* suara dapat dilihat pada Gambar 2. Pengujian yang dilakukan untuk periode waktu yang cukup pendek akan menunjukkan karakteristik *signal* suara yang *stationary*. Tetapi bila dilakukan dalam periode waktu yang lebih panjang karakteristik *signal* suara akan terus berubah sesuai dengan kata yang diucapkan



Gambar 2 Contoh signal suara (Sumber: Manuggal, 2005)

MFCC *feature extraction* sebenarnya merupakan adaptasi dari sistem pendengaran manusia, dimana *signal* suara akan difilter secara linier untuk frekuensi rendah (dibawah 1000 Hz) dan secara logaritmik untuk frekuensi tinggi (diatas 1000 Hz). Gambar 3 merupakan *block diagram* untuk MFCC.



Gambar 3 Blok Diagram MFCC (Sumber: Adi Resmawan, 2011)

II.5 Dynamic Time Warping (DTW)

Satu masalah yang cukup rumit dalam *speech recognition* (pengenalan wicara) adalah proses perekaman yang terjadi seringkali berbeda durasinya, biarpun kata atau kalimat yang diucapkan sama. Bahkan untuk satu suku kata yang sama atau vocal yang sama seringkali proses perekaman terjadi dalam durasi yang berbeda. Sebagai akibatnya proses matching antara sinyal uji dengan sinyal referensi (*template*) seringkali tidak menghasilkan nilai yang optimal.

Sebuah teknik yang cukup populer di awal perkembangan teknologi pengolahan sinyal wicara adalah dengan memanfaatkan sebuah teknik *dynamic-programming* yang juga lebih dikenal sebagai *Dynamic Time Warping (DTW)* (Resmawan, 2011).

Teknik DTW ini ditujukan untuk mengakomodasi perbedaan waktu antara proses perekaman saat pengujian dengan yang tersedia pada *template* sinyal referensi. Prinsip dasarnya adalah dengan memberikan sebuah rentang '*steps*' dalam ruang (dalam hal ini sebuah frame-frame waktu dalam sample, frame-frame waktu dalam template) dan digunakan untuk mempertemukan lintasan yang menunjukkan *local match* terbesar (kemiripan) antara time frame yang lurus. Total '*similarity cost*' yang diperoleh dengan algorithm ini merupakan sebuah indikasi seberapa bagus *sample* dan *template* ini memiliki kesamaan, yang selanjutnya akan dipilih *best-matching template*.

DTW (*DynamicTime Warping*) adalah metode untuk menghitung jarak antara dua data *time series*. Keunggulan DTW dari metode jarak yang lainnya adalah mampu menghitung jarak dari dua vektor data dengan panjang berbeda.

Jarak DTW diantara dua vektor dihitung dari jalur pembengkokkan optimal (*optimal warping path*) dari kedua vektor tersebut. Ilustrasi pencocokan dengan metode DTW ditunjukkan pada Gambar 4.

Dari beberapa teknik yang digunakan untuk menghitung DTW, salah satu yang paling handal adalah dengan metode pemrograman dinamis. Jarak DTW dapat dihitung dengan persamaan sebagai berikut :

Persamaan:

$$D(U, V) = \gamma(m, n)$$

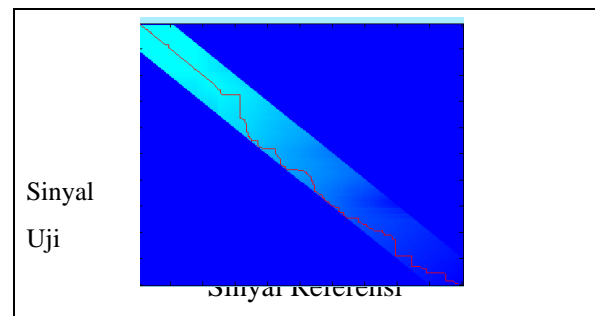
$$\gamma(m, n) = d_{base}(u_i, v_j) + \min \begin{cases} \gamma(i-1, j) \\ \gamma(i-1, j-1) \\ \gamma(i, j-1) \end{cases}$$

Keterangan :

$$\gamma(0,0) = 0, \gamma(0, \infty) = \infty, \gamma(\infty, 0) = 0, \gamma(\infty, \infty) = \infty$$

$$(i = 1,2,3...m; j = 1,2,3...n)$$

Kolom dengan nilai $\gamma(i, j) (1 \leq i \leq m, 1 \leq j \leq n)$ dinamakan matriks jarak terjumlahkan.



Gambar 4 Ilustrasi DTW

Gambar 5 adalah contoh matriks jarak terjumlahkan.

Dalam *speech recognition*, DTW merupakan metode yang paling sering digunakan khususnya dalam proses pencocokan suara.

		Sinyal Uji (V)					
		0	3	6	0	6	1
Sinyal	2	4	5	21	25	41	42
	5	29	8	6	31	26	42
Refrensi (U)	2	33	9	22	10	26	27
	5	58	13	10	35	11	27
	3	67	13	19	19	20	15

Gambar 5 Ilustrasi matrik jarak terjumlahkan (cumulative distance matrix) antara 2 vektor. U={2,5,2,5,3}, V={0,3,6,0,6,1}

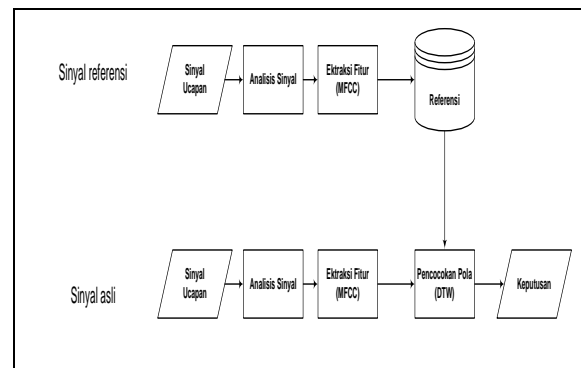
Tabel 1 Penghitungan DTW Menggunakan Matriks
(Sumber : Adi resmawan, 2010)

	=Kolom 1	Kumulatif	=Kumulatif	
$(0-2)^2$	4	0+4	4	→Minimum
$(0-5)^2$	25	4+25	29	
$(0-2)^2$	4	29+4	33	
$(0-5)^2$	25	33+25	58	
$(0-3)^2$	9	58+9	67	
	=Kolom 2	Kumulatif	=Kumulatif	
$(3-2)^2$	1	1+4	5	→Minimum
$(3-5)^2$	4	4+4	8	
$(3-2)^2$	1	1+8	9	
$(3-5)^2$	4	4+9	13	
$(3-3)^2$	0	0+13	13	
	=Kolom 3	Kumulatif	=Kumulatif	
$(6-2)^2$	16	16+5	21	
$(6-5)^2$	1	1+5	6	→Minimum
$(6-2)^2$	16	16+6	22	
$(6-5)^2$	1	1+9	10	
$(6-3)^2$	9	9+10	19	
	=Kolom 4	Kumulatif	=Kumulatif	
$(0-2)^2$	4	21+4	25	
$(0-5)^2$	25	25+6	31	
$(0-2)^2$	4	4+6	10	Minimum
$(0-5)^2$	25	25+10	35	
$(0-3)^2$	9	9+10	19	
	=Kolom 5	Kumulatif	=Kumulatif	
$(6-2)^2$	16	16+25	41	
$(6-5)^2$	1	1+25	26	
$(6-2)^2$	16	16+10	26	
$(6-5)^2$	1	1+10	11	→Minimum
$(6-3)^2$	9	9+11	20	

	=Kolom 6	Kumulatif	=Kumulatif	
$(1-2)^2$	1	1+41	42	
$(1-5)^2$	16	16+26	42	
$(1-2)^2$	1	1+26	27	
$(1-5)^2$	16	16+11	27	
$(1-3)^2$	4	4+11	15	→Minimum

III. CARA KERJA SISTEM

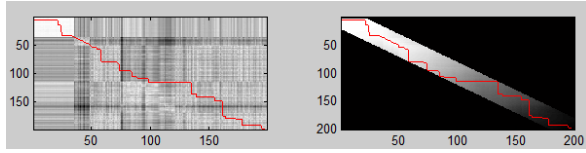
Untuk menggambarkan langkah-langkah dan urutan prosedur dari aplikasi pengenalan ucapan ini secara umum maka dibuat sebuah model kerja sistem aplikasi secara umum seperti pada Gambar 6. Langkah pertama dalam aplikasi pengenalan ucapan yaitu menerima masukan sinyal ucapan dari *microphone* yang kemudian di analisis sinyal tersebut dengan melakukan konversi dari sinyal ucapan analog menjadi digital. Setelah di konversi langkah selanjutnya menganalisa sinyal digital dengan dilakukan normalisasi sinyal tersebut sesuai kebutuhan kerja yang selanjutnya akan digunakan untuk melakukan ekstraksi sinyal menggunakan teknik *Mel-Frequency Cepstrum Coefficients* (MFCC) sehingga menghasilkan fitur-fitur ekstraksi sinyal yang diperlukan dalam proses pencocokan input sinyal ucapan asli dengan sinyal ucapan referensi di *database* yang sudah dilakukan analisis sinyal dan fitur ekstraksi.



Gambar 6 Sistem Kerja Aplikasi Pengenalan Ucapan

Hasil proses pencocokan (Gambar 7) yang dilakukan dengan cara mencocokkannya menggunakan bantuan algoritma *Dynamic Time Warping* (DTW). Nilai perbandingan sinyal ucapan yang dihasilkan dari

hasil perbandingan sinyal ucapan kemudian dijadikan sebagai acuan dalam menghasilkan sebuah keputusan.



Gambar 7. Hasil Proses DTW

IV. MFCC

Tahapan-tahapan yang digunakan dalam proses *feature extraction* MFCC sebagai berikut:

1. Sinyal ucapan di rekam melalui *microphone* yang setelah di rekam di konversi dari analog menjadi digital. Data sinyal suara diperoleh dengan cara merekam suara melalui mikrofon yang dihubungkan dengan komputer. Perakaman suara di dalam aplikasi menggunakan frekuensi sampling standar 44100Hz. Suara dengan format *.wav* ini bisa menggunakan 16 bits/sample dan 1 untuk *channel mono*. Durasi suara yang direkam apabila lebih pendek lebih mudah untuk diambil perbedaan fiturnya. Dalam analisis ini digunakan contoh durasi rekaman yang diambil adalah 2 detik.

$$X = FS \times dt(\text{detik}) \times \left(\frac{\text{bit}}{8}\right) \times j$$

$$X = 44100 \times 2 \times \left(\frac{16}{8}\right) \times 1 = 176400 \text{ byte}$$

2. Melakukan proses *DC-Removal* untuk mendapatkan nilai normalisasi dari sampel data suara masukan. Masukan dari proses ini adalah sampel suara dalam bentuk array dan outputnya adalah array sample suara yang sudah dinormalisasi dengan sampel nilai sampel data 10.

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

$$\bar{x} = \frac{10+12+14+15}{4} = 12.75$$

$$n = 10 - 12.75 = -2.75$$

3. Melakukan proses *Pre-emphasize* untuk memperbaiki sinyal dari gangguan *noise*, sehingga dapat meningkatkan tingkat akurasi dari proses *feature extraction*

$$y[n] = s[n] - \alpha s[n-1]$$

$$Y = (-0.75 - (-2.75 * 0, 97)) = 1.9175$$

4. Melakukan proses *Frame Blocking* bertujuan untuk membagi sampel sinyal menjadi beberapa *frame slot* dengan panjang tertentu

$$\text{Jumlah frame} = ((I-N)/M)+1$$

$$N = 88200 * 0,02 = 1764 \text{ Sample Point}$$

$$M = 882$$

Sehingga Jumlah Frame :

$$((I-N)/M) + 1 = ((88200-1764)/882)+1 = 99$$

5. Melakukan proses *Windowing* untuk mengurangi efek *aliasing* pada ujung-ujung *frame* yang dihasilkan oleh proses *frame blocking*.

$$W = 0.54 - 0.46 \cos \frac{2\pi n}{N-1}$$

$$W = 0.54 - 0.46 \cos \frac{2 \times 3.14 \times 0}{1764 - 1} = 0.08$$

6. Melakukan proses FFT (*Fast Fourier Transform*) merubah *signal digital* dari *time domain* ke *frequency domain*. sehingga diperoleh sampel signal dalam frekuensi domain.

$$F = 1/4[-0.44(\cos((2\phi * 0^0)/4)) - j\sin(\cos(2\phi * 0^0)/4) + [0.09(\cos((2\phi * 0^1)/4)) - j\sin(\cos(2\phi * 0^1)/4) + [0.26(\cos((2\phi * 0^2)/4)) - j\sin(\cos(2\phi * 0^2)/4) + [0.46(\cos((2\phi * 0^3)/4)) - j\sin(\cos(2\phi * 0^2)/4) = -0.0925 + 0j = 0.0925$$

7. Melakukan proses *Mel Frequency Warping* atau proses *filterbank* untuk untuk mendapatkan nilai diskrit dari sinyal untuk proses DCT (*Discrete Cosine Transform*

$$H = 2595 * \log(1 + 1000/700) / 0.0925 / 2 = 21621.44$$

$$S = 0.0925 * 21621.44 = 1999.15$$

$$Y[i] = \sum_{j=1}^N S[j] H_i[j]$$

8. Melakukan proses DCT untuk mendapatkan *mel cepstrums*. Proses DCT adalah langkah terakhir dari proses *feature extraction*. Melalui proses ini akan diperoleh beberapa nilai yang merupakan dimensi dari vector yang bersangkutan

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

$$C_n = \sum_{k=1}^K \left((\log S) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\phi}{K} \right] \right)$$

$$C_n = \log(19991.15) \cos \left[\frac{0 \left(0 - \frac{1}{2} \right) 3.14}{1} \right]$$

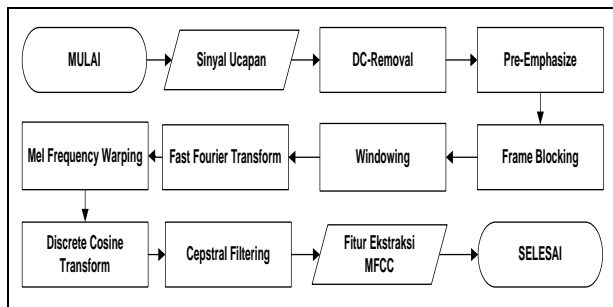
$$C_n = 0.0026$$

9. Melakukan proses *Cepstral Liftering*, bertujuan menghaluskan *spectrum* sinyal dari *mel cepstrums*.

$$W = \{1 + L/2 \sin(n*\phi/L)\}$$

$$W = 0.0026 * 1/2 * \sin(3.14/1) = 0.20$$

10. *Spectrum* sinyal yang telah dihasilkan bernilai *cepstral coefficient* digunakan sebagai fitur ekstraksi.



Gambar 8 Flowchart MFCC
(Sumber: Adi Resmawan, 2011)

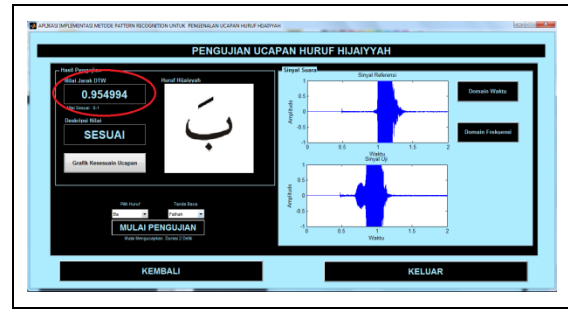
11. Persentasi Match dihitung berdasarkan jumlah yang sesuai dibagi dengan jumlah keseluruhan pengujian lalu di 100%.

$$\text{Matching} = \frac{\text{Hasil Yang Sesuai / Benar}}{\text{Total Seluruh Uji Coba}} \times 100 \%$$

V. IMPLEMENTASI

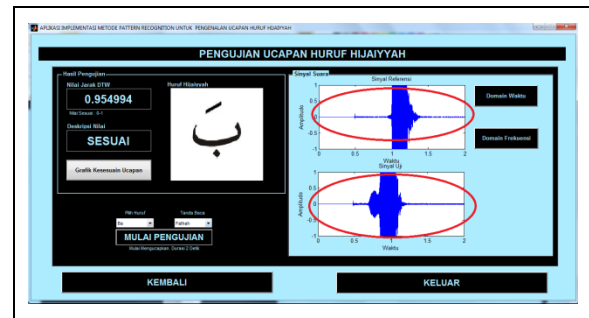
Pada implementasi fungsionalitas jarak kesesuaian ini dapat dilakukan pengguna mendapatkan informasi seberapa besar nilai jarak kesamaan antara suara yang di referensi dengan suara yang diucapkan pengguna setelah melakukan pengujian ucapan. Hasil dari nilai jarak kesesuaian terlihat pada Gambar 9.

Pada implementasi fungsionalitas sinyal domain waktu ini dapat di lakukan pengguna mendapatkan informasi sinyal suara yang diucapkan berdasarkan domain waktu setelah melakukan pengujian ucapan. Terlihat pada Gambar 10.



Gambar 9 Nilai Jarak Kesesuaian (DTW)

Pada implementasi fungsionalitas sinyal domain waktu ini dapat di lakukan pengguna mendapatkan informasi sinyal suara yang diucapkan berdasarkan domain waktu setelah melakukan pengujian ucapan. Terlihat pada Gambar 11

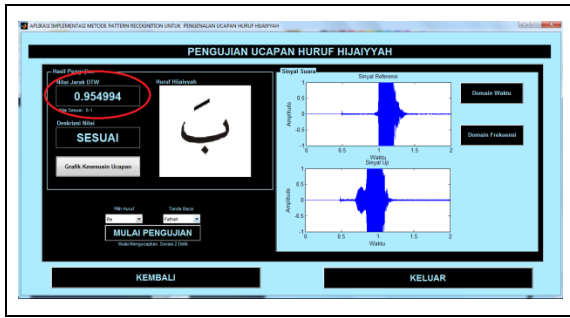


Gambar 10 Sinyal Suara Domain Waktu

Pada implementasi fungsionalitas jarak kesesuaian ini dapat dilakukan pengguna mendapatkan informasi seberapa besar nilai jarak kesamaan antara suara yang di referensi dengan suara yang diucapkan pengguna setelah melakukan pengujian ucapan. Hasil dari nilai jarak kesamaan terlihat pada Gambar 12



Gambar 11 Sinyal Suara Domain Frekuensi



Gambar 12 Nilai Jarak DTW

VI. PENGUJIAN

Dari pengujian yang dilakukan kepada 6 orang yang terdiri dari 3 pria dan 3 wanita. Hasil Pengujian dihasilkan nilai kecocokan pada Tabel 2 terhadap P1, Tabel 3 terhadap P2, Tabel 4 terhadap P3 untuk hasil uji kepada pria dewasa dan Tabel 5 terhadap P4, Tabel 6 terhadap P5, dan Tabel 7 terhadap P6 untuk hasil uji kepada wanita dewasa. Jumlah total pengujian setiap orang dilakukan sebanyak 435 kali untuk 29 huruf 3 tanda baca dan 5 kali pengulangan masing-masing di setiap huruf dan tanda bacanya dengan nilai ambang batas kesesuaian jarak maksimal sebesar 1.

Tabel 2 Hasil Uji Pria Dewasa untuk threshold 1 terhadap P1

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
L1	Dendri	435	69	15.8
L2	Hambali	435	89	20.4
L3	Aditya	435	90	20.6

Tabel 3 Hasil Uji Pria Dewasa untuk threshold 1 terhadap P2

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
L1	Dendri	435	81	18.6
L2	Hambali	435	94	21.6
L3	Aditya	435	63	14.4

Tabel 4 Hasil Uji Pria Dewasa untuk threshold 1 terhadap P3

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
L1	Dendri	435	70	16.0

L2	Hambali	435	87	20
L3	Aditya	435	54	12.4

Tabel 5 Hasil Uji Wanita Dewasa untuk threshold 1 P4

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
W1	Tita	435	63	14.4
W2	Nada	435	56	12.8
W3	Dinda	435	71	16.3

Tabel 6 Hasil Uji Wanita Dewasa untuk threshold 1 P5

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
W1	Tita	435	74	17.0
W2	Nada	435	72	16.5
W3	Dinda	435	67	15.4

Tabel 7 Hasil Uji Wanita Dewasa untuk threshold 1 P6

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
W1	Tita	435	77	17.7
W2	Nada	435	81	18.6
W3	Dinda	435	64	14.7

Dari pengujian yang dilakukan kepada 6 orang yang terdiri dari 3 pria dan 3 wanita. Hasil Pengujian dihasilkan nilai kecocokan pada Tabel 8 terhadap P1, Tabel 9 terhadap P2, Tabel 10 terhadap P3 untuk hasil uji kepada pria dewasa dan Tabel 11 terhadap P4, Tabel 12 terhadap P5, dan Tabel 13 terhadap P6 untuk hasil uji kepada wanita dewasa. Jumlah total pengujian setiap orang dilakukan sebanyak 435 kali untuk 29 huruf 3 tanda baca dan 5 kali pengulangan masing-masing di setiap huruf dan tanda bacanya dengan nilai ambang batas kesesuaian jarak maksimal sebesar 1.1.

Tabel 8 Hasil Uji Pria Dewasa untuk threshold 1.1 terhadap P1

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
L1	Dendri	435	215	49.4
L2	Hambali	435	220	50.5
L3	Aditya	435	203	46.6

Tabel 9 Hasil Uji Pria Dewasa untuk threshold 1.1 terhadap P2

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentase Match (%)
----	-----------	------------------	--------------	----------------------

L1	Dendri	435	207	47.5
L2	Hambali	435	163	37.4
L3	Aditya	435	197	45.2

Tabel 10 Hasil Uji Pria Dewasa untuk threshold 1.1 terhadap P3

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
L1	Dendri	435	211	48.5
L2	Hambali	435	237	54.4
L3	Aditya	435	210	48.2

Tabel 11 Hasil Uji Wanita Dewasa untuk threshold 1.1 P4

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
W1	Tita	435	201	46.2
W2	Nada	435	233	53.5
W3	Dinda	435	214	49.1

Tabel 12 Hasil Uji Wanita Dewasa untuk threshold 1.1 P5

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
W1	Tita	435	193	44.3
W2	Nada	435	216	49.6
W3	Dinda	435	183	42.0

Tabel 13 Hasil Uji Wanita Dewasa untuk threshold 1.1 P6

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
W1	Tita	435	221	50.8
W2	Nada	435	187	42.9
W3	Dinda	435	199	45.7

Dari pengujian yang dilakukan kepada 6 orang yang terdiri dari 3 pria dan 3 wanita. Hasil Pengujian dihasilkan nilai kecocokan pada Tabel 14 terhadap P1, Tabel 15 terhadap P2, Tabel 16 terhadap P3 untuk hasil uji kepada pria dewasa dan Tabel 17 terhadap P4, Tabel 18 terhadap P5, dan Tabel 19 terhadap P6 untuk hasil uji kepada wanita dewasa. Jumlah total pengujian setiap orang dilakukan sebanyak 435 kali untuk 29 huruf 3 tanda baca dan 5 kali pengulangan masing-masing di setiap huruf dan tanda bacanya dengan nilai ambang batas kesesuaian jarak maksimal sebesar 1.2.

Tabel 14 Hasil Uji Pria Dewasa untuk threshold 1.2 terhadap P1

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
L1	Dendri	435	245	56.3
L2	Hambali	435	231	53.1
L3	Aditya	435	243	55.86

Tabel 15 Hasil Uji Pria Dewasa untuk threshold 1.2 terhadap P2

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
L1	Dendri	435	238	54.7
L2	Hambali	435	215	49.4
L3	Aditya	435	223	51.2

Tabel 16 Hasil Uji Pria Dewasa untuk threshold 1.2 terhadap P3

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
L1	Dendri	435	241	55.4
L2	Hambali	435	263	60.4
L3	Aditya	435	247	56.7

Tabel 17 Hasil Uji Wanita Dewasa untuk threshold 1.2 P4

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
W1	Tita	435	233	53.5
W2	Nada	435	201	46.2
W3	Dinda	435	251	57.7

Tabel 18 Hasil Uji Wanita Dewasa untuk threshold 1.2 P5

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
W1	Tita	435	228	52.4
W2	Nada	435	243	55.8
W3	Dinda	435	215	49.4

Tabel 19 Hasil Uji Wanita Dewasa untuk threshold 1.2 P6

ID	Pengujian	Jumlah Pengujian	Jumlah Match	Persentas i Match (%)
W1	Tita	435	262	60.2
W2	Nada	435	251	57.7
W3	Dinda	435	213	48.9

Dari pengujian yang dilakukan kepada 6 orang yang terdiri dari 3 pria dan 3 wanita. Hasil Pengujian dihasilkan nilai kecocokan pada Tabel 19 terhadap P1, Tabel 20 terhadap P2, Tabel 21 terhadap P3 untuk hasil uji kepada pria dewasa dan Tabel 22 terhadap P4, Tabel 23 terhadap P5, dan Tabel 24 terhadap P6 untuk hasil uji kepada wanita dewasa. Jumlah total pengujian setiap orang dilakukan sebanyak 435 kali untuk 29 huruf 3 tanda baca dan 5 kali pengulangan masing-masing di setiap huruf dan tanda bacanya dengan nilai ambang batas kesesuaian jarak maksimal sebesar 1.3.

Tabel 19 Hasil Uji Pria Dewasa untuk threshold 1.3 terhadap P1

ID	Penguji an	Jumlah Pengujian	Jumlah Match	Persentasi Match (%)
L1	Dendri	435	314	72.1
L2	Hambali	435	351	80.6
L3	Aditya	435	363	83.4

Tabel 20 Hasil Uji Pria Dewasa untuk threshold 1.3 terhadap P2

ID	Penguji an	Jumlah Pengujian	Jumlah Match	Persentasi Match (%)
L1	Dendri	435	361	82.9
L2	Hambali	435	354	81.3
L3	Aditya	435	401	92.1

Tabel 21 Hasil Uji Pria Dewasa untuk threshold 1.3 terhadap P3

ID	Pengujian n	Jumlah Pengujian n	Jumlah Match	Persentasi Match (%)
L1	Dendri	435	407	93.5
L2	Hambali	435	411	94.4
L3	Aditya	435	370	85.0

Tabel 22 Hasil Uji Wanita Dewasa untuk threshold 1.3 P4

ID	Pengujian n	Jumlah Pengujian n	Jumlah Match	Persentasi Match (%)
W1	Tita	435	331	76.0
W2	Nada	435	326	74.9
W3	Dinda	435	357	82.0

Tabel 23 Hasil Uji Wanita Dewasa untuk threshold 1.3 P5

ID	Pengujian n	Jumlah Pengujian n	Jumlah Match	Persentasi Match (%)
W1	Tita	435	385	88.5
W2	Nada	435	271	62.2
W3	Dinda	435	357	82.0

Tabel 24 Hasil Uji Wanita Dewasa untuk threshold 1.3 P6

ID	Pengujian n	Jumlah Pengujian	Jumlah Match	Persentasi Match (%)
W1	Tita	435	378	86.8
W2	Nada	435	345	79.3
W3	Dinda	435	391	89.88

VII. KESIMPULAN

Dari pengembangan aplikasi pengenalan ucapan huruf hijaiyyah ini ada beberapa yang dapat disimpulkan sebagai berikut :

1. Aplikasi mampu melakukan pengenalan ucapan huruf hijaiyyah menggunakan metode sebagai fitur ekstraksi MFCC dan algoritma DTW untuk pencocokan dua sinyal ucapan asli maupun referensi.
2. Hasil pengujian dilakukan kepada 6 orang pengguna yang terdiri dari 3 pria dan 3 presentase nilai similitarty cost dengan threshold 1.3 paling optimal nilai kecocokanya yang mencapai diatas 94%.

REFERENSI

- Darma, Putra ,dkk. 2011. “Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW”. Universitas Udayana
- Rabiner, Lm et.al. 1993. “Fundamental of Speech Recognition”. Prentice Hall PTR
- Retnawati, Wirawan, Endang Widjiati. 2009. “Kompresi Audio Secara Terdistribusi Pada Microphone Array”. ITS. Universitas Kristen Petra
- Resmawan, A, 2010. Verifikasi Suara Menggunakan Metode MFCC dan DTW.
- Ronando, E., dan Irawan, M.I, 2012. Pengenalan Ucapan Kata Sebagai Pengendali Gerakan Robot Lengan Secara Real-Time dengan Metode Linear Predictive Coding – Neuro Fuzzy