

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 Huruf Hijaiyah

Huruf hijaiyah atau huruf arab terdiri dari 28 huruf dimulai dari huruf **ا** (hamzah) hingga huruf **ق** (ya) seperti dapat dilihat pada Tabel 2.1. Huruf ini juga dipakai dalam kitab suci Al-Qur'an, sehingga jika seseorang ingin mempelajari membaca Al-Qur'an maka harus dimulai dengan mempelajari huruf hijaiyah. Huruf hijaiyah memiliki kekhususan tersendiri dalam pelafalannya, karena huruf ini memiliki makhraj dan sifat yang membedakan antara huruf-huruf satu dengan lainnya. Tidak ada dua huruf hijaiyah yang memiliki makhraj dan sifat yang sama, jika ada dua huruf yang makhrajnya sama maka dapat dipastikan sifatnya berbeda begitu pula sebaliknya. Jika ada dua huruf yang memiliki makhraj dan sifat yang sama maka dipastikan bahwa dua huruf tersebut merupakan satu huruf yang sama (Bisri, 2016).

Tabel 2.1 Huruf Hijaiyah

No	Huruf	Nama
1	ا	Hamzah
2	ب	Ba
3	ت	Ta
4	ث	Tsa
5	ج	Jim
6	ح	Ha
7	خ	Kha
8	د	Dal
9	ذ	Dzal
10	ر	Ra
11	ز	Zai
12	س	Sin
13	ش	Syin
14	ص	Shad
15	ض	Dhad
16	ط	Tha
17	ظ	Zha
18	ع	'Ain
19	غ	Ghain
20	ف	Fa
21	ق	Qaf

No	Huruf	Nama
22	ك	Kaf
23	ل	Lam
24	م	Mim
25	ن	Nun
26	و	Wau
27	ه	Hha
28	ي	Ya

2.1.1 Makhraj Huruf

Secara morfologi makhraj berasal dari kata bahasa arab yaitu *kharaja* (خرج) yang bermakna keluar. Makhraj huruf dapat didefinisikan sebagai tempat keluarnya huruf (Bisri, 2016). Makhraj huruf terbagi atas lima macam yaitu:

3. *Al-Jauf* (الجوف), artinya rongga atau lubang mulut,
4. *Al-Halq* (الحلق), artinya tenggorokan,
5. *Al-Lisan* (اللسان), artinya lidah,
6. *Asy-Syafatani* (الشفاتان), artinya dua bibir,
7. *Al-Khaisyum* (الخيضوم), artinya lubang pangkal hidung.

2.2 Audio

Audio dalam Kamus Besar Bahasa Indonesia adalah sesuatu yang dapat didengar atau alat peraga yang bersifat dapat didengar. Selain itu audio didefinisikan sebagai media yang menyajikan informasi dalam bentuk suara yang ditangkap oleh indra pendengaran (Riana, 2012). Tingkat kekuatan audio dipengaruhi oleh amplitudo dan frekuensi, amplitudo adalah tingkatan tekanan suara, sedangkan frekuensi adalah jumlah gelombang dalam satuan detik. Amplitudo diukur dengan satuan desibel (dB), sementara frekuensi diukur dengan satuan hertz (Hz). Waktu yang diperlukan dalam terbentuknya satu gelombang audio disebut sebagai perioda (T) dan diukur dalam satuan detik (s). Pendengaran manusia hanya mampu mendengarkan gelombang suara dengan rentang frekuensi tertentu yaitu antara 20 Hz sampai 20.000 Hz (Binanto, 2010).

Gelombang audio awalnya berbentuk analog, agar dapat diproses oleh komputer maka sinyal analog diubah menjadi sinyal digital. Audio yang telah mengalami perubahan bentuk dari analog ke digital disebut juga dengan audio

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

digital (Dofat, 2016). Proses mengubah audio analog ke format digital dilakukan menggunakan *Audio to Digital Converter* (ADC). Audio digital agar kembali dapat didengar oleh manusia harus diubah kembali ke bentuk sinyal analog dengan menggunakan *Digital to Analog Converter* (DAC). File audio digital dapat diubah, direplikasikan, dan diproses oleh komputer. Kualitas audio rekaman ditentukan oleh *sample rate* yaitu seberapa sering sample diambil, *sample rate* dihitung dalam satuan hertz (Hz).

2.3 Pengenalan Ucapan (*Automatic Speech Recognition*)

Pengenalan ucapan atau *Automatic Speech Recognition* (ASR) adalah proses mengubah ucapan menjadi teks. Proses ini melalui dua tahapan, yaitu mengubah sinyal suara ucapan menjadi sebuah rentetan simbol menggunakan model akustik, kemudian rentetan simbol diubah menjadi teks menggunakan model bahasa (Rao dan Vuppala, 2014). Pengenalan ucapan memiliki peran penting dalam perkembangan teknologi komunikasi manusia ke manusia dan komunikasi manusia ke komputer (Yu dan Deng, 2015).

2.4 Format *File Audio*

File audio digital memiliki beberapa format yang biasa digunakan untuk menyimpan sinyal suara hasil rekaman diantaranya adalah WAV, AIFF, MP3, dan AAC (Truesdell, 2007). Masing-masing format memiliki perbedaan seperti yang dijelaskan di bawah ini:

1. WAV

Format *file* WAV atau disebut WAVE singkatan dari *waveform*, adalah format *file* yang dikembangkan oleh Microsoft dan IBM untuk menyimpan audio berdasarkan *Pulse Code Modulation* (PCM). *File* yang disimpan dengan format ini tidak terkompresi.

2. AIFF

Format *file* AIFF (*Audio Interchange File Format*) dikembangkan oleh Apple untuk menyimpan *file* audio di perangkat komputer Apple. *File* dengan format ini mirip dengan format WAV, menggunakan PCM yang tidak terkompresi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. MP3

Format MP3 adalah yang paling populer digunakan untuk *file* audio. Dengan format ini *file* audio dikompresi untuk menghilangkan frekuensi yang tidak dapat didengar oleh manusia. Hasil kompresi ini menyebabkan *file* audio dengan format MP3 jauh lebih kecil dibandingkan WAV dan AIFF.

4. AAC

AAC adalah singkatan dari *Advanced Audio Coding*, merupakan format *file* audio yang digunakan oleh Apple pada penjualan musik yang dilakukannya di iTunes. Sama seperti format MP3, format AAC juga melakukan kompresi terhadap *file* audio namun memiliki kualitas yang lebih baik dan ukuran *file* yang lebih kecil dibandingkan format MP3.

2.5 Ekstraksi Ciri

Ekstraksi ciri atau ekstraksi fitur adalah bagian dari tahapan awal pemrosesan suara agar dapat dikenali. Pada tahapan ini sinyal suara akan mengalami beberapa tahapan sehingga menghasilkan keluaran akhir berupa parameter-parameter ciri suara yang dapat digunakan pada tahapan klasifikasi. Ekstraksi parameter terbaik yang merepresentasikan sinyal suara adalah tugas penting untuk menghasilkan pengenalan pengucapan yang baik (Dave, 2013).

Parameter yang digunakan sebagai nilai masukan yang diproses pada tahapan ekstraksi ciri yaitu nilai amplitudo dan jumlah getaran dari sinyal suara yang akan diproses. Jumlah getaran merupakan hasil kali dari frekuensi *sample rate* dalam satuan Hertz (Hz) dengan lama durasi sinyal dalam satuan detik (s) pada Persamaan (2.1). *Sample rate* yang digunakan dalam pengenalan suara manusia cukup dengan 16.000 Hz.

$$N = f \times t \tag{2.1}$$

Keterangan:

N = jumlah getaran

f = frekuensi

t = waktu

Selanjutnya nilai amplitudo tadi diproses menggunakan metode ekstraksi ciri sinyal. Beberapa metode yang biasa digunakan untuk ekstraksi ciri sinyal suara diantaranya *Mel-Frequency Cepstrum Coefficients* (MFCC) dan *Linear Predictive Coding* (LPC). Dikarenakan suara manusia secara alami tidak bersifat linear LPC tidak menjadi pilihan yang baik untuk pengenalan suara. MFCC merupakan metode yang memiliki konsep perhitungan logaritmik menyerupai konsep indra pendengaran manusia sehingga memiliki kemampuan yang lebih baik dalam pengenalan suara (Dave, 2013).

2.6 Mel Frequency Cepstrum Coefficients (MFCC)

Mel Frequency Cepstrum Coefficients (MFCC) adalah salah satu teknik ekstraksi ciri suara dalam pengenalan suara. Ekstraksi ciri dengan MFCC yaitu dengan menghitung koefisien *cepstrum* yang berdasarkan kepada indra pendengaran manusia. MFCC terdiri dari beberapa tahapan yaitu *Pre-Emphasis*, *Hamming Window*, *Fast Fourier Transform* (FFT), *Mel-scale filterbank*, *Log*, dan *Discrete Cosine Transform* seperti yang dapat dilihat pada Gambar 2.1 (Tan dan Lindberg, 2008). Keluaran akhir dari tahapan ini adalah berupa vektor MFCC. Selain tahapan tadi ada beberapa tahapan yang dapat meningkatkan kualitas keluaran ekstraksi ciri yaitu reduksi *noise*. Adapun rincian masing-masing tahapan adalah sebagai berikut:



Gambar 2.1 Tahapan MFCC (Tan dan Lindberg, 2008)

1. *Pre-Emphasis*

Sebelum melewati filter, sinyal terlebih dahulu melalui tahapan *frame blocking* dimana sinyal dibagi menjadi *frame-frame* yang berdurasi pendek yaitu dengan rentang 10 milidetik hingga 50 milidetik. Tahapan *pre-emphasis* terjadi proses melewatkan sinyal melalui sebuah filter yang menekankan pada frekuensi sinyal yang lebih tinggi. Proses ini akan meningkatkan energi dari sinyal pada frekuensi tinggi. Sinyal suara

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

masukan, $s(n)$, akan menghasilkan keluaran berupa sinyal suara yang telah di-*pre-emphasis*, $x(n)$, dengan dihitung menggunakan Persamaan (2.2):

$$x(n) = s(n) - \alpha s(n - 1) \quad (2.2)$$

2. *Hamming Window*

Hamming window dilakukan untuk mengubah sinyal-sinyal yang telah dibagi dalam bentuk *frame-frame* dan telah melewati filter *pre-emphasis*, $x_i(n)$, menjadi vektor ciri dengan Persamaan (2.3):

$$h(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N-1}, \quad 0 \leq n \leq N - 1 \quad (2.3)$$

Keterangan:

- $h(n)$ = *Hamming Window*
- N = jumlah sample pada *frame*
- n = urutan sampel

Tiap sinyal hasil *windowing* dihitung nilainya dengan Persamaan (2.4):

$$y(n) = x(n) \times h(n), \quad 0 \leq n \leq N - 1 \quad (2.4)$$

Keterangan:

- $y(n)$ = sinyal hasil *windowing*
- $x(n)$ = nilai sinyal sebelum *windowing*
- $h(n)$ = nilai *window*

3. *Fast Fourier Transform (FFT)*

Tahapan ini bertujuan untuk mengubah tiap *frame* sebanyak N sampel dari domain waktu menjadi domain frekuensi. Tiap *frame* yang telah melewati tahapan *windowing* akan ditransformasikan dengan menggunakan Persamaan (2.5):

$$X_n = \sum_{k=0}^{N-1} X_k e^{-2\pi jkn/N} \quad (2.5)$$

Keterangan:

- N = jumlah segmen *frame*
- X_k = nilai sinyal ke k
- n = 0,1,2,3,...,N-1
- k = 0,1,2,3,...,N-1
- j = bilangan imajiner

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. *Mel-scale filterbank*

Nilai frekuensi masing-masing *frame* yang diperoleh melalui tahapan FFT tidak bersifat linear. Untuk mengubah nilai frekuensi ke skala *mel* dapat dilakukan dengan menggunakan Persamaan (2.6):

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2.6)$$

Keterangan:

f = frekuensi keluaran FFT

dan dapat kita konversikan kembali menjadi frekuensi dengan Persamaan (2.7):

$$f = 700(10^{m/2595} - 1) \quad (2.7)$$

Untuk membentuk *mel filterbank* dilakukan mencari nilai kenaikan linear (*delta*) antara nilai minimum dan maksimum dan jumlah filter atau merupakan jumlah koefisien MFCC yang dihasilkan dengan Persamaan (2.8) berikut:

$$\text{delta} = (\text{mel}_{\max} - \text{mel}_{\min}) / \text{filter} + 1 \quad (2.8)$$

Keterangan:

delta = jarak linear *mel filterbank*

mel_{\max} = frekuensi *mel* maksimum

mel_{\min} = frekuensi *mel* minimum

filter = jumlah filter

5. *Discrete Cosine Transform (DCT)*

Tahapan terakhir yaitu untuk mendapatkan koefisien *Mel-Frequency Cepstrum*. Koefisien didapatkan dengan menggunakan Discrete Cosine Transform yang dihitung dengan Persamaan (2.9):

$$C_n = \sum_{k=1}^K m(k) \cos \left(\frac{n\pi(k-0.5)}{K} \right) \quad (2.9)$$

Keterangan:

C_n = MFCC

K = jumlah koefisien

k = 1,2,3,...,K

m = nilai skala mel

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Berdasarkan penelitian yang dilakukan sebelumnya, (Permana dan Negara, 2011), (Rahman, 2017), dan (Faqih, 2017) bahwa jumlah koefisien MFCC yang digunakan akan mempengaruhi performa dalam pengenalan suara. Ketiga penelitian tersebut menemukan performa terbaik dengan menggunakan jumlah koefisien MFCC sebanyak 30, 30, dan 25.

2.7 Manhattan Distance

Manhattan Distance atau dikenal juga dengan *Taxicab Geometry* atau *City Block Distance* digagas pertama kali oleh Hermann Minkowski. *Manhattan Distance* digunakan untuk penyelesaian menghitung jarak yang ditempuh untuk melewati blok-blok kota Manhattan. Perhitungan yang dilakukan adalah dengan menghitung selisih absolut antara dua titik atau vektor dengan menggunakan Persamaan (2.10) (Ponmoli dan Selvamuthukumar, 2014).

$$d = \sum_{i=1}^n |a_i - b_i| \tag{2.10}$$

Keterangan:

- d = Nilai jarak
- a = vektor a
- b = vektor b
- n = panjang vektor

Manhattan Distance sering dikaitkan dengan *Euclidean Distance*, namun pada dasarnya kedua persamaan ini memiliki perbedaan. *Euclidean Distance* menghitung jarak terpendek antara dua titik secara langsung, sementara *Manhattan Distance* memperhitungkan jumlah jarak yang harus ditempuh dari satu titik untuk mencapai titik lain dengan mempertimbangkan dimensi antara dua titik tersebut (Akila dan Chandra, 2013)

2.8 Pengukuran Tingkat Akurasi

Setelah pengujian selesai dilakukan selanjutnya dilakukan pengukuran tingkat akurasi pengujian, dalam penelitian ini pengukuran tingkat akurasi adalah dengan *confusion matrix*. *Confusion matrix* adalah matriks yang menampilkan informasi mengenai hasil aktual dan hasil yang diklasifikasikan pada tahapan

klasifikasi (Liu, 2008). Lebih jelas dapat dilihat pada Tabel 2.2 dan Persamaan (2.11).

Tabel 2.2 Confusion Matrix

		Kelas Klasifikasi	
		Positif	Negatif
Kelas Sebenarnya	Positif	TP	TN
	Negatif	FP	FN

Keterangan:

TP (*True Positive*) : Jumlah hasil klasifikasi yang benar dari sampel positif

FN (*False Negative*) : Jumlah hasil klasifikasi yang salah dari sampel positif

FP (*False Positive*) : Jumlah hasil klasifikasi yang salah dari sampel negatif

TN (*True Negative*) : Jumlah hasil klasifikasi yang benar dari sampel negatif

Pengukuran akurasi dapat dilakukan dengan Persamaan (2.11):

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (2.11)$$

2.9 Penelitian Terkait

Penelitian yang telah dilakukan berkaitan tentang pengenalan huruf hijaiyah, penggunaan metode ekstraksi ciri MFCC, dan penggunaan pengukuran jarak dengan *Manhattan Distance* menjadi referensi dalam penelitian ini. Penelitian terkait yang pernah dilakukan sebelumnya dapat dilihat pada Tabel 2.3 berikut ini.

Tabel 2.3 Daftar Penelitian Terkait

No	Peneliti	Tahun	Judul	Metode	Kesimpulan
1	Wahidah, dkk.	2012	<i>Makhraj Recognition Using Speech Processing</i>	MFCC dan Mean Square Error	Akurasi pengenalan hingga 100%
2	Zainon, dkk.	2013	<i>Speech quality based On Arabic Pronunciation Using MFCC And LDA: Investigating The Emphatic Consonants</i>	MFCC dan LDA	Akurasi tertinggi 92,5%

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Peneliti	Tahun	Judul	Metode	Kesimpulan
3	Akila dan Chandra	2013	<i>Slope Finder – A Distance Measure For DTW based Isolated Word Speech Recognition</i>	DTW dan Manhattan Distance	Manhattan Distance dapat digunakan dalam klasifikasi pengenalan suara
4	Ponmoli dan Selvamuthukumar	2014	<i>Analysis of Face Recognition using Manhattan Distance Algorithm with Image Segmentation</i>	Manhattan Distance	Akurasi yang diperoleh dengan menggunakan Manhattan Distance mencapai 97%, lebih baik jika dibandingkan dengan Euclidean Distance
5	Gautama, Purwanto, dan Purboyo	2015	Analisis Pengaruh Manhattan Distance Pada Algoritme Clustering Isodata (Self-Organizing Data Analysis Technique) Untuk Sistem Deteksi Anomali Trafik	Manhattan Distance dan Algoritme Clustering Isodata	Manhattan Distance memiliki performa yang lebih baik dari sisi waktu dibandingkan Euclidean Distance
6	Wisesty, dkk.	2017	<i>A Classification of Marked Hijaiyah Letters Pronunciation Using Hidden Markov Model</i>	MFCC dan LPC	Akurasi tertinggi 90%
7	Rahman	2017	Implementasi Mel-Frequency Cepstral Coefficients Dan Chebyshev Distance Untuk Pengenalan Suara Ucapan Makhraj Huruf Hijaiyah	MFCC dan Chebyshev Distance	Akurasi tertinggi 71,42% pada penujian dengan 7 kelas dan 30 koefisien MFCC