

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

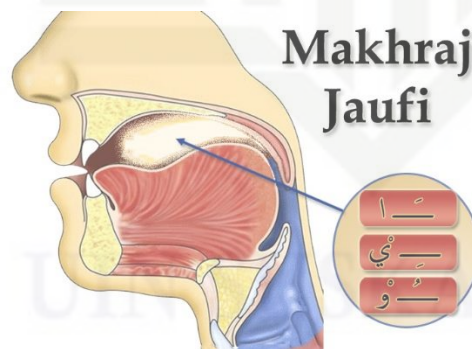
2.1 Huruf Hijaiyah berdasarkan Makharijul Huruf

Huruf Hijaiyah atau huruf abjad merupakan huruf penyusun pada Al-Qur'an. Harokat merupakan tanda baca yang digunakan untuk memperjelas pelafalan/pengucapan huruf. *Makharijul Huruf* merupakan tempat keluarnya huruf pada waktu huruf tersebut dibunyikan (Annuri, 2010).

Makharijul Huruf menurut Imam Ibnu Jazari dibagi menjadi 17 *makharijul huruf*. Pada penelitian Tugas Akhir ini, penulis akan mengenali 5 kelompok *makharijul huruf* serta harokat. Berikut Pembagian Kelompok *Makharijul Huruf* tersebut (Annuri, 2010):

1. Al-Jaufi (مَوْضِعُ الْجَوْفِ)

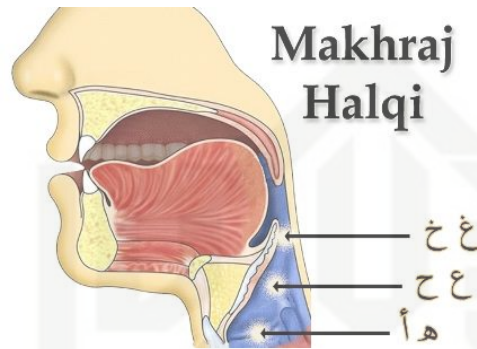
Makhraj huruf yang keluar dari rongga mulut adalah huruf-huruf mad sebagai berikut: membuka mulut dengan sempurna (اَ), menurunkan bibir bagian bawah (يَ), dan memonyongkan dua bibir (وُ).



Gambar 2.1 Kelompok Rongga Mulut
(Sumber: <http://www.celikalquran.com>)

2. Al-Halqi (مَوْضِعُ الْحَلْقِ)

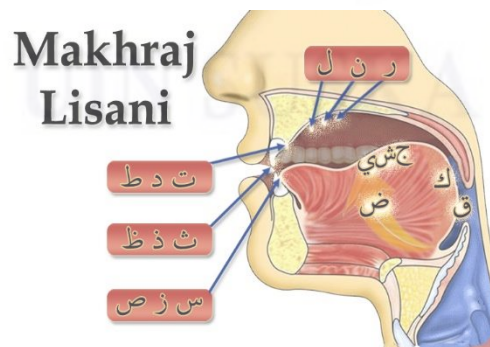
Makhraj huruf yang keluar dari tenggorokan adalah huruf-huruf sebagai berikut: keluar dari tenggorokan bawah (هـ ء), keluar dari tenggorokan tengah (ع ح), dan keluar dari tenggorokan atas (غ خ).



Gambar 2.2 Kelompok Tenggorokan
 (Sumber: <http://www.celikalquran.com>)

3. Al-Lisani (مَوْضِعُ اللِّسَانِ)

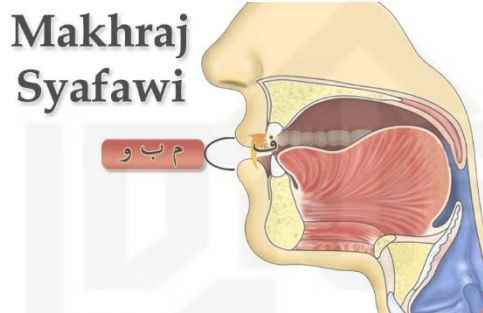
Makhraj huruf yang keluar terletak di lidah sebagai berikut: pangkal lidah (ق ك), tengah lidah dengan langit-langit (ي ش ج), sisi lidah bertemu gigi geraham (ض), ujung lidah dengan langit-langit depan (ل), ujung lidah dibawah makhraj lam (ن), ujung lidah (ر), ujung lidah dengan pangkal gigi seri (ط د ت), ujung lidah bertemu ujung gigi seri atas (ظ ذ ث), dan ujung lidah bertemu ujung gigi seri bawah (ص س ز).



Gambar 2.3 Kelompok Lidah
 (Sumber: <http://www.celikalquran.com>)

4. Al-Syafataini/Syafawi (مَوْضِعُ الشَّفَتَيْنِ)

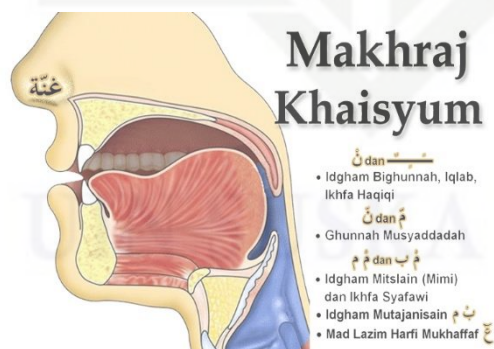
Makhraj huruf yang keluar dari kedua bibir sebagai berikut: bibir bawah bagian dalam bertemu ujung gigi atas (ف), dua bibir secara tertutup (ب م), dan dua bibir membentuk bulatan (و)



Gambar 2.4 Kelompok Dua Bibir
 (Sumber: <http://www.celikalquran.com>)

5. Al-Khaisyum (مَوْضِعُ الْخَيْشُومِ)

Makhraj huruf yang keluar dari rongga hidung disebut *ghunnah* (dengung) sebagai berikut: nun/mim yang ditasydid (نّ / مّ), nun/mim mati (diikuti oleh huruf iqlab, idgham bighunnah, dan ikhfa'), dan mim mati yang diikuti huruf ba'/mim.



Gambar 2.5 Kelompok Rongga Hidung
 (Sumber: <http://www.celikalquran.com>)

2.2 Speech Recognition

Speech recognition adalah suatu kegiatan yang melibatkan proses identifikasi suara, pengelompokan suara berdasarkan kesamaan ciri yang dimiliki (research.ibm.com).

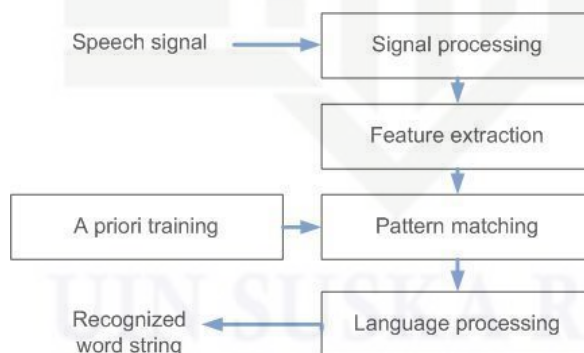
Speech recognition dibagi menjadi dua kategori, yaitu:

1. Berdasarkan sumber inputan

Speech recognition dibagi menjadi dua, yaitu *speech recognition* sumber tak bebas dan *speech recognition* sumber bebas. Yang dimaksud dari sumber tak bebas adalah suara tersebut yang dijadikan masukan merupakan suara yang dijadikan data *sample* latih yang tersimpan di *database*. Berbeda dengan sumber bebas yang mana suara masukan berasal dari suara yang berbeda dari data *sample* yang dilatih ke sistem.

2. Berdasarkan metode pendeteksian

Speech recognition dibagi menjadi dua, yaitu *isolated-word speech recognition* dan *continuous speech recognition*. Jenis pertama hanya mendeteksi satu kata pada setiap *running*, sedangkan jenis kedua mendeteksi *continuous speech* atau beberapa kata/kalimat yang diucapkan sumber suara.



Gambar 2.6 Blok Diagram *Speech Recognition* Secara Umum

Secara umum, sistem *speech recognition* memproses sinyal suara yang masuk dan menyimpannya dalam bentuk digital. Hasil proses digitalisasi tersebut kemudian dikonversi dalam bentuk spectrum suara yang akan dianalisis dengan cara membandingkannya dengan pola suara pada *database* sistem. Adapun dua acuan utama yang dibutuhkan, yaitu:

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.

1. Ekstraksi Ciri (*feature extraction*)

Tujuan proses ini adalah untuk mengkonversi sinyal suara menjadi beberapa jenis representasi parametrik (pada tingkat informasi yang jauh lebih rendah). Data suara masukan dipilah-pilah dan diproses satu per satu berdasarkan urutannya. Proses pemilahan ini bertujuan untuk proses analisis dapat dilakukan secara paralel.

2. Pencocokan Ciri

Tujuan proses ini adalah untuk membandingkan atau mencocokkan data suara dari sinyal masukan yang telah melalui proses ekstraksi ciri dengan data *sample* latih yang ada dalam *database*. Hasil dari pencocokan ciri ini yang akan menjadi keluaran sistem.

Perancangan sistem *speech recognition* melalui dua fase secara garis besar, yaitu fase pelatihan dan fase pengujian. Pada fase pelatihan, sistem akan menerima inputan sebagai suatu data *sample* yang akan dijadikan data latih. Data latih yang disimpan sistem akan dijadikan acuan dalam fase pengujian. Dan pada fase pengujian, masukan sistem akan dibandingkan dengan data latih yang ada dan kemudian didapatkan hasil keluaran berdasarkan kemiripan dari klasifikasi data latih.

2.3 Ekstraksi Ciri

Tujuan dari proses ini bertujuan untuk mengonversi gelombang suara, menggunakan alat pemrosesan sinyal digital (PSD), menjadi set tertentu (pada tingkat informasi yang jauh lebih rendah) untuk analisis lebih lanjut. Proses ekstraksi ciri membutuhkan inputan berupa nilai amplitudo dan nilai amplitudo tersebut memiliki jumlah getaran. Nilai frekuensi yang digunakan adalah frekuensi *sample rates* dalam Hz dan waktu yang digunakan adalah lamanya gelombang suara dalam detik. Berikut persamaan menghitung jumlah getaran dari nilai amplitudo:

$$\text{Jumlah getaran} = f \times t \tag{2.1}$$

f = frekuensi

t = waktu

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.

Karena sinyal ucapan memiliki karakteristik yang sama dalam interval waktu yang singkat, maka *short-time analysis* merupakan cara paling umum untuk mengarakterisasi sinyal suara. Saat ini ada tiga metode yang cocok digunakan untuk mengekstraksi ciri sinyal suara dalam *speech recognition* yaitu *Linear Prediction Coding* (LPC), *Zero Crossing with Peak Amplitudes* (ZCPA) dan *Mel Frequency Cepstral Coefficients* (MFCC). MFCC merupakan metode ekstraksi ciri yang memiliki tingkat akurasi paling tinggi dan membutuhkan waktu paling singkat dibandingkan ketiga metode ekstraksi tersebut. (Riyanto dan Sutejo, 2014).

2.4 *Mel Frequency Cepstral Coefficients* (MFCC)

Mel Frequency Cepstral Coefficients (MFCC) merupakan suatu metode yang digunakan dalam pemrosesan sinyal suara sehingga sinyal suara yang diolah memiliki ciri-ciri tertentu yang dapat dibedakan oleh sistem. Dalam hal ini, ciri-ciri yang dihasilkan oleh MFCC berupa koefisien, koefisien *Mel* frekuensi.

Beberapa kelebihan dari metode ini antara lain:

1. Dapat menangkap informasi-informasi penting yang terkandung di dalam sinyal suara
2. Menghasilkan data seminimal mungkin tanpa menghilangkan informasi-informasi penting yang dikandungnya
3. Mereplikasi telinga manusia dalam memproses sinyal suara

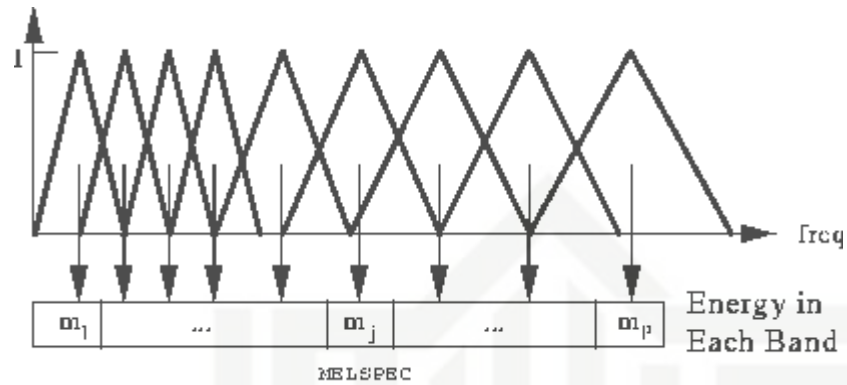
Analisis suara pada *mel-frequency* didasarkan pada persepsi pendengaran manusia, karena telinga manusia telah diamati dapat berfungsi sebagai *filter* pada frekuensi tertentu (Buono, 2009). *Filter-filter* tersebut memiliki jarak yang tidak seragam pada sumbu frekuensi, yaitu banyak *filter* pada daerah frekuensi rendah dan sedikit pada daerah frekuensi tinggi. Hal tersebut dikarenakan oleh sensitifitas pendengaran manusia yang tidak merata. Sehingga MFCC memproses sinyal secara logaritmik.

Filter-filter tersebut memiliki respon frekuensi yang membentuk *Band Pass Filter* (BPF) segitiga, dan spasi antar bandwidthnya ditentukan oleh interval *mel-frequency* yang konstan. MFCC dapat digunakan untuk mengekstrak data

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.

sinyal suara sehingga bisa didapatkan ciri yang terdapat pada data sinyal suara tersebut.



Gambar 2.7 MFCC Filter Bank

Metode MFCC memiliki beberapa tahapan, yaitu:

1. *Frame Blocking*

Sinyal suara manusia termasuk sinyal yang tidak stabil. Tetapi sinyal suara dapat diasumsikan sebagai sinyal yang stabil pada skala waktu 10-30 ms. *Frame blocking* merupakan proses pembagian sinyal menjadi beberapa frame yang lebih kecil agar sinyal lebih mudah untuk diproses selanjutnya (Syafria, 2014). Proses ini dilakukan dengan cara manual yakni dengan memotong masing-masing suara menjadi satu huruf menggunakan *software* Audacity.

2. *Windowing*

Proses *windowing* bertujuan untuk mengurangi terjadinya kebocoran spektral atau aliasing yang merupakan suatu efek dari timbulnya sinyal baru yang memiliki frekuensi yang berbeda dengan sinyal aslinya. Efek tersebut dapat terjadi karena rendahnya sampling rate atau karena proses *frame blocking* yang menyebabkan sinyal menjadi diskontinu. Berikut persamaan menghitung $w(n)$ sebagai *window* dan berikutnya menghitung $y(n)$ sebagai *windowing*:

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right], \quad 0 \leq n \leq N - 1 \quad (2.2)$$

$w(n)$ = Window

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.

N = Jumlah *sample* pada masing-masing *frame*

n = 0, 1, 2, 3, ..., $N - 1$

$$y(n) = x(n) \times w(n), 0 \leq n \leq N - 1 \quad (2.3)$$

$y(n)$ = Sinyal hasil *widowing sample* ke- n

$x(n)$ = Nilai *sample* ke- n

$w(n)$ = Nilai *window* ke- n

n = 0, 1, 2, 3, ..., $N - 1$

N = Jumlah *sample* dalam *frame*

3. FFT (*Fast Fourier Transform*)

Untuk mendapatkan sinyal dalam domain frekuensi dari sebuah sinyal diskrit (domain waktu), salah satu metode transformasi Fourier yang digunakan adalah FFT. FFT dilakukan terhadap masing-masing *frame* dari sinyal yang telah di-*windowing* dengan persamaan:

$$f(n) = \sum_{k=0}^{N-1} Y_k e^{-2\pi jkn/N} \quad (2.4)$$

$f(n)$ = Frekuensi

Y_k = Sinyal masukan dari proses *windowing*

N = Jumlah *sample* pada masing-masing *frame*

n = 0, 1, 2, 3, ..., $N - 1$

k = 0, 1, 2, 3, ..., $N - 1$

j = Bilangan Imajiner

4. *Mel-Frequency Wrapping*

Skala *Mel-Frequency* adalah frekuensi yang linier di bawah 1 kHz dan logaritmik di atas 1 kHz. Skala Mel dapat diperoleh dengan pendekatan persamaan:

$$mel(n) = 2595 * \log_{10}(1 + f(n)/700) \quad (2.5)$$

$mel(n)$ = Mel-filter

n = 0, 1, 2, 3, ..., $N - 1$

f = Frekuensi hasil FFT

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.

$$Freq(m) = 700(10^{\frac{mel(m)}{2595}} - 1) \quad (2.6)$$

$Freq(m)$ = Frekuensi mel
 $m = 0, 1, 2, 3, \dots, N - 1$

$$S(l) = f(l) \times Freq(l) \quad (2.7)$$

$S(l)$ = Sinyal hasil *mel-frequency wrapping*
 $l = 0, 1, 2, 3, \dots, N - 1$
 $f(l)$ = Frekuensi hasil FFT
 $Freq(l)$ = Frekuensi mel

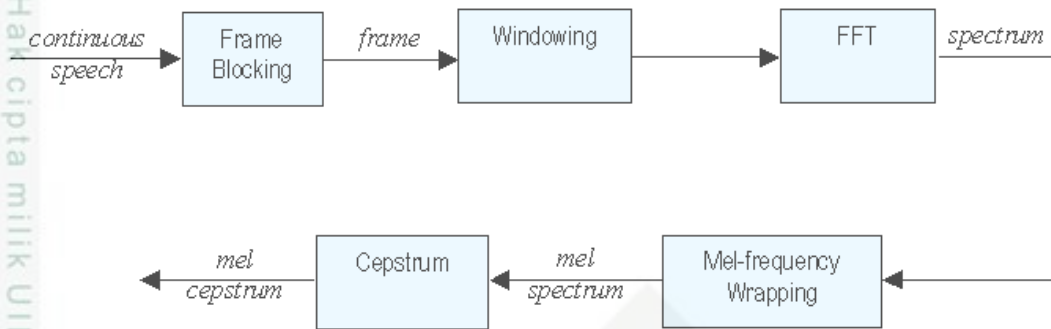
5. *Cepstrum*

Mel-Frequency Cepstrum kemudian didapat dari *invers DCT (Discrete Cosine Transform)* untuk mendapatkan kembali sinyal dalam domain waktu. Hasilnya disebut sebagai *Mel-Frequency Cepstral Coefficient (MFCC)*. Maka koefisien *mel-frequency* sebagai berikut:

$$C_n = \sum_{p=1}^K (\log S_p) \cos \left[n \left(p - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad (2.8)$$

$n = 1, 2, 3, \dots, K$
 C_n = MFCC
 S_p = Sinyal hasil *mel-frequency wrapping*
 $p = 0, 1, 2, 3, \dots, N-1$
 K = Jumlah koefisien

Dimana C_n adalah hasil akumulasi dari kuadratik magnitude FFT yang dikalikan dengan *Mel-Filter Bank*. Setelah itu didapatlah MFCC. Pada sistem pengenalan suara, biasanya hanya 13 *cepstrum* koefisien pertama yang digunakan.



Gambar 2.8 Blok Diagram MFCC

(Sumber: http://minhdo.ece.illinois.edu/teaching/speaker_recognition)

2.5 Chebyshev Distance

Chebyshev Distance merupakan suatu perhitungan jarak dari dua buah titik dalam sebuah *vector space*. *Chebyshev Distance* sering juga disebut *Chessboard Distance* (Jarak Papan Catur) karena dalam permainan catur dibutuhkan jumlah minimum untuk raja bergerak. Menurut Kardi (2015) *Chebyshev Distance* adalah mengukur besaran nilai mutlak dari koordinat pada sepasang objek.

$$d_{xy} = \max_k |x_{ik} - y_{jk}| \quad (2.9)$$

Dimana d_{xy} adalah jarak *Chebyshev* dari titik x ke titik y . Nilai \max_k adalah selisih maksimum dari objek x_{ik} dan y_{jk} . Semakin besar nilai d_{xy} , maka semakin besar kesamaan antara kedua objek atau kasus tersebut.

2.6 Perhitungan Akurasi dan Tingkat Error

Pada tahapan pengujian selesai, maka akan dilakukan perhitungan nilai tingkat akurasi dan nilai tingkat error. Dalam penelitian ini menggunakan perhitungan akurasi *confusion matrix* (Ratih, 2009 dikutip oleh Andrian, 2012).

Confusion matrix merupakan alat yang menyatakan jumlah data dalam pengujian benar diklasifikasikan pada kelas yang sesuai dan jumlah data pengujian salah diklasifikasikan. Berikut contoh *confusion matrix* untuk klasifikasi biner ditunjukkan pada Tabel 2.1 berikut:

Tabel 2.1 Confusion Matrix

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan :

TP (*True Positive*) = Jumlah dokumen dari kelas 1 yang benar dan diklasifikasi senagai kelas 1.

FP (*False Positive*) = Jumlah dokumen dari kelas 0 yang salah dan diklasifikasi senagai kelas 1.

FN (*False Negative*) = Jumlah dokumen dari kelas 1 yang salah dan diklasifikasi senagai kelas 0.

TN (*True Negative*) = Jumlah dokumen dari kelas 0 yang benar dan diklasifikasi senagai kelas 0.

Berikut persamaan untuk menghitung akurasi :

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (2.10)$$

$$Error = 100\% - Akurasi \quad (2.11)$$

2.7 Penelitian Terkait

Berikut penelitian-penelitian terkait *speech recognition* yang telah membahas metode MFCC dan *Chebyshev Distance*.

2.7.1 Identifikasi Pembicara Independent Text Pada Data Close-Set dengan Menggunakan Mel Frequency Cepstral Coefficients

Topik pengenalan pembicara tersebut diangkat mengingat keterbatasan manusia dalam mengenali suara manusia yang berbagai macam serta banyak suara yang hampir sama antar sesama manusia. Seperti halnya dalam identifikasi siapa yang berbicara pada rekaman percakapan dijadikan bukti dalam sebuah persidangan.

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.

Penelitian tersebut menguji tingkat akurasi dari identifikasi suara MFCC dan SOM memiliki tingkat akurasi tertinggi pembicara wanita 54.4%, pembicara pria 75.6%, dan rata-rata keduanya adalah sekitar 62.5% (Permana, 2011).

2.7.2 Perbandingan Metode Ekstraksi Ciri Suara MFCC, PS-ZCPA, dan LPC

Pada proses ekstraksi ciri pada pengenalan suara muncul masalah-masalah seperti keakuratan dalam ekstraksi/pengenalan, sehingga perlu dilakukan pengkajian lagi metode apa yang mempunyai keakuratan yang tinggi dan waktu yang singkat. Terdapat tiga metode yang dapat digunakan untuk melakukan ekstraksi ciri suara, yaitu MFCC, PS-ZCPA, dan LPC.

Penelitian tersebut menghasilkan tingkat akurasi pada pengenalan suara MFCC 85.3%, PS-ZCPA 38.5%, dan LPC 82.3% serta menghasilkan waktu ekstraksi ciri MFCC 0.092, LPC 0.152, dan PS-ZCPA 27.38. Kesimpulan bahwa MFCC memiliki tingkat akurasi tinggi dan membutuhkan proses waktu yang singkat (Riyanto dan Sutejo, 2014).

2.7.3 A Comparative Study of K-Means Algorithm by Different Distance Measures

Pada bidang *data mining*, *clustering* merupakan hal yang sangat penting. Data diklasifikasikan menjadi *cluster* dengan cara mengelompokkan berdasarkan kesamaan dan ketidaksamaan. Pada penelitian ini Algoritma K-Means adalah sebuah metode penting yang digunakan untuk menemukan *cluster*. Implementasinya sangat sederhana dan cepat. Data yang digunakan pada penelitian ini adalah data bunga. Berbagai metode pengukuran jarak digunakan seperti, *Euclidean Distance*, *Manhattan Distance*, dan *Chebyshev Distance* untuk menganalisa hasil dari banyak iterasi, akurasi dan nilai error. Penelitian tersebut menghasilkan kesimpulan bahwa metode *Chebyshev Distance* memiliki akurasi lebih baik dengan rata-rata 86%, dan nilai *error* 0 (Kouser dan Sunita, 2013).

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.

2.7.4 Pengenalan Ucapan Huruf Hijaiyah Menggunakan Mel Frequency Cepstral Coefficients (MFCC) dan Hidden Markov Model (HMM)

Mempelajari huruf hijaiyah merupakan tahap paling awal dalam proses pembelajaran Al-Qur'an. Dalam proses tersebut diperlukan bimbingan pengejar secara khusus yang mampu mengenalkan dan mengajarkan huruf hijaiyah. Dengan menggunakan *speech recognition system*, diharapkan peran seorang pengajar Al-Qur'an dapat tergantikan sehingga proses belajar dapat dilakukan secara mandiri dan semakin efektif.

Penelitian tersebut menggunakan metode *Mel Frequency Cepstral Coefficients* (MFCC) untuk ekstraksi ciri suara dan menggunakan metode *Hidden Markov Model* (HMM) untuk pengenalan suara. Adapun hasil yang didapat dari penelitian tersebut rata-rata persentase akurasi terbaik sekitar 67.75% dan untuk pengembangan penelitian tersebut lebih lanjut disarankan untuk meneliti hingga tingkat kebenaran *makharijul huruf* (Fauzi, 2013).