

Hak Cipta Dilindungi Undang-Undang

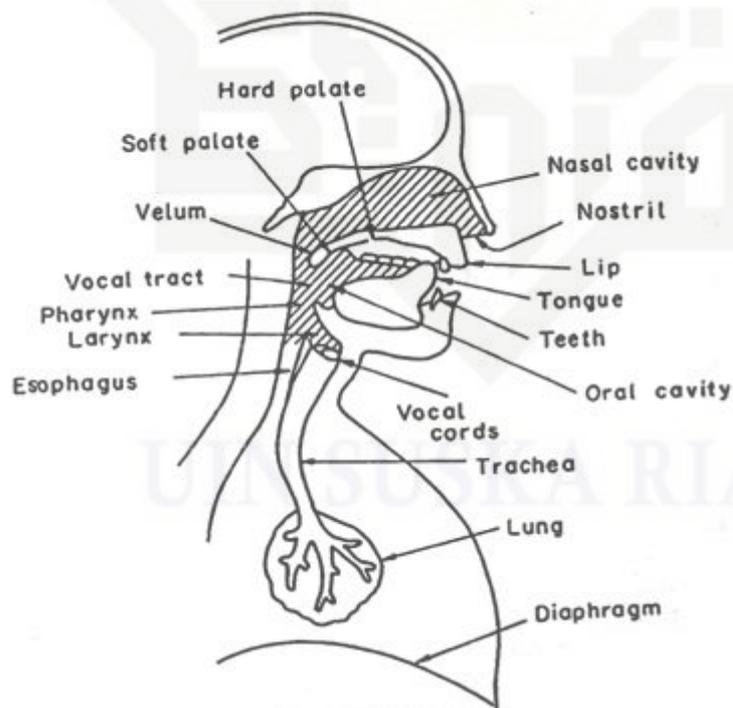
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan 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

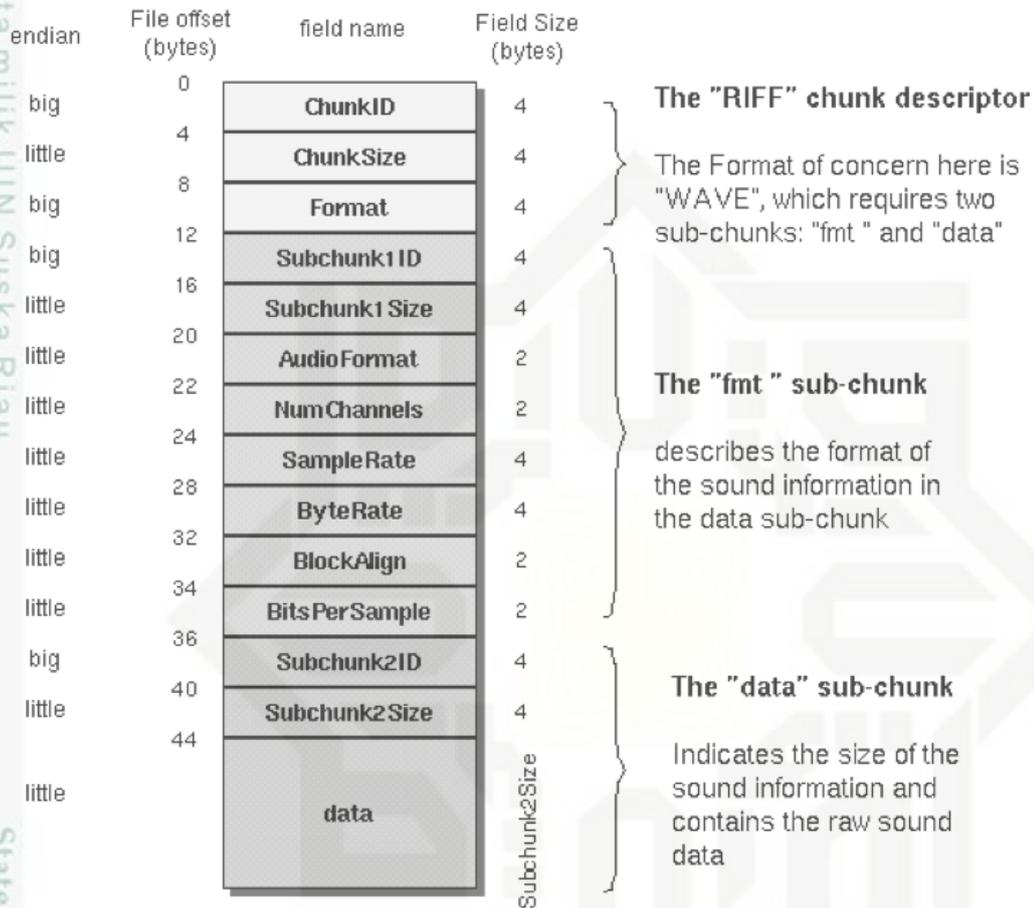
2.1 Biometrik Suara

Berbicara merupakan cara alami manusia untuk berkomunikasi antar individu. Tiap individu memiliki karakteristik suara yang berbeda – beda, sehingga kita dapat mengenali perbedaan suara tiap individu. Beragamnya suara yang dihasilkan tiap individu merupakan kombinasi dari gerakan fisik rahang, lidah, laring, ketebalan pita suara dan resonansi di bagian hidung. Dikarenakan sifat fisik saluran suara, laring dan organ suara lainnya, tidak ada suara manusia yang terdengar identik (Karpov, 2011).



Gambar 2. 1 Organ Pembentuk Suara Manusia
(Saquib, Salam, Nair, & Pandey, 2011)

The Canonical WAVE file format



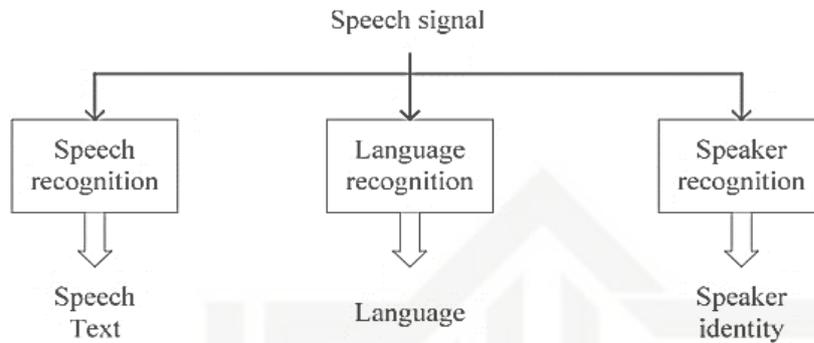
Gambar 2. 3 Struktur File WAV (Yoo et al., 2012)

Kualitas audio digital tergantung pada *sampling rate*, *bit depth* serta pilihan kompresi (*lossy* atau *lossless*). *Sampling rate* adalah banyaknya gelombang suara dalam satu detik (satunya Hz). *Bit depth* adalah ukuran *sample* atau kisaran angka yang yang digunakan untuk menyatakan *sample*. *Sampling rate* dan *bit depth* menentukan resolusi dari sebuah file audio.

2.3 Pengenalan Pembicara

Terdapat perbedaan antara *speech recognition* dan *speaker recognition*. *Speech recognition* adalah mengenali kata atau kalimat yang sedang diucapkan. Sedangkan *speaker recognition* adalah mengenali siapa yang sedang berbicara.

Pengenalan pembicara merupakan proses identifikasi seseorang dari karakteristik suaranya.



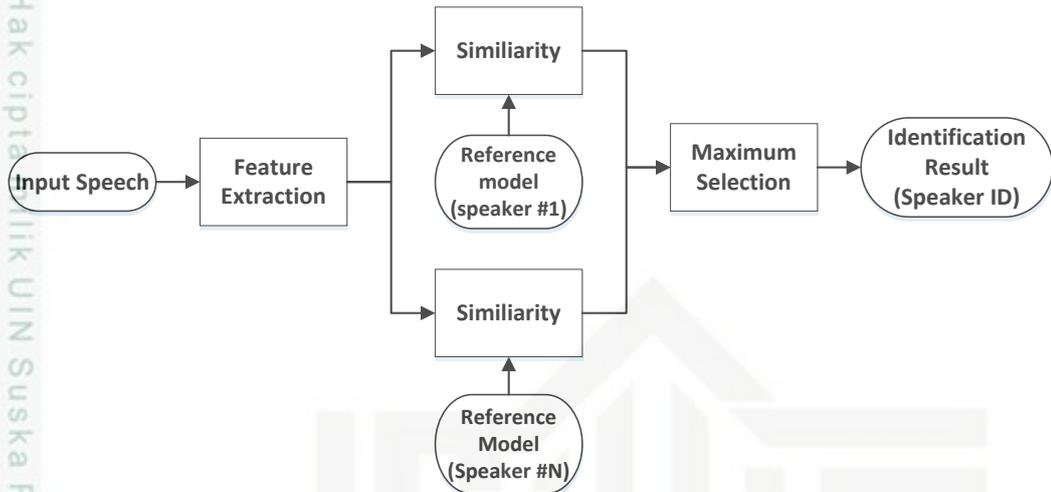
Gambar 2. 4 Pembagian Sinyal Suara (Feng, 2004)

Pengenalan pembicara terbagi menjadi dua tujuan, yaitu identifikasi dan verifikasi. Pada identifikasi, suara dari pembicara yang tidak diketahui akan dianalisis kemudian dibandingkan ke contoh suara dari beberapa pembicara. Sedangkan pada proses verifikasi contoh suara dari pembicara yang tidak diketahui yang diklaim oleh seseorang akan dibandingkan. Jika perbandingannya cocok, maka klaim akan diterima.

Ada dua jenis identifikasi pembicara, yaitu *closed-set* dan *open-set*. Identifikasi *closed-set*, suara dari pembicara uji akan dibandingkan terhadap semua model pembicara yang tersedia kemudian akan ditentukan kecocokan yang terdekat (tidak ada skema penolakan). Sedangkan pada identifikasi *open-set* merupakan kombinasi dari identifikasi *closed-set* dan verifikasi pembicara. Pertama kali akan dilakukan identifikasi *closed-set* dan akan didapatkan ID dari pembicara yang ditentukan dari kecocokan terdekat, kemudian akan dilakukan verifikasi pembicara. Jika ID pembicara lolos pada tahap verifikasi, maka suara yang diuji diterima sebagai pemilik asli. Jika verifikasi gagal, maka suara yang diuji tidak memiliki kecocokan terhadap model pembicara (Beigi, 2011).

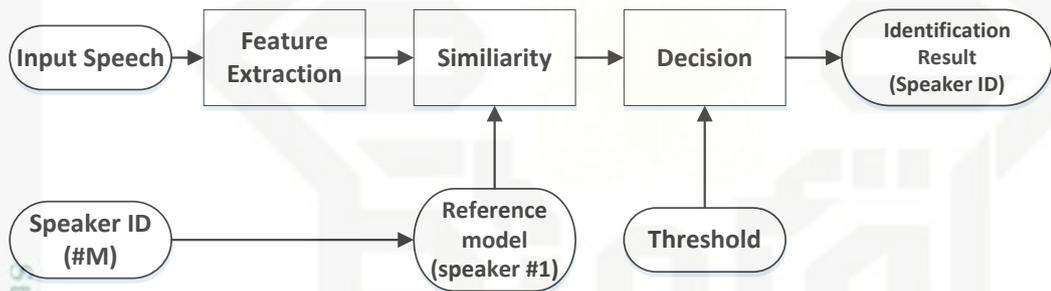
Hak Cipta Dilindungi Undang-Undang

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.



Gambar 2. 5 Struktur Identifikasi Pembicara

(www.ifp.illinois.edu/~minhdo/teaching/speaker_recognition/speaker_recognition.html)



Gambar 2. 6 Struktur Verifikasi Pembicara

(www.ifp.illinois.edu/~minhdo/teaching/speaker_recognition/speaker_recognition.html)

Pengenalan pembicara memiliki dua fase, yaitu :

1. Tahap training, suara pembicara akan direkam lalu diekstraksi ciri dan dijadikan voice print.
2. Tahap verifikasi/testing, contoh suara akan dibandingkan dengan voice print pada tahap sebelumnya.

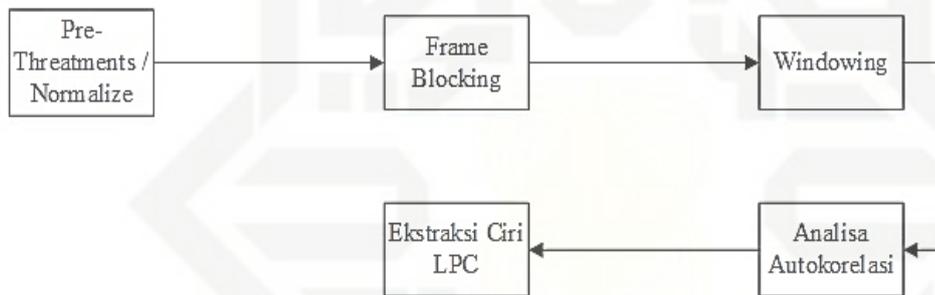
Berdasarkan kalimat yang diucapkan, pengenalan pembicara terbagi menjadi dua, yaitu :

1. *Text-dependent*, yaitu kalimat/kata yang diucapkan harus sama pada saat training dan testing. Biasanya digunakan untuk kata sandi berbasis suara
2. *Text-independent*, yaitu kalimat/kata yang diucapkan tidak harus sama. *Text-independent* bisa digunakan tanpa diketahui oleh pembicara, sehingga

cocok diaplikasikan untuk mengenali identitas pelaku kriminal atau yang lainnya.

2.4 Linear Predictive Coding (LPC)

Ekstraksi ciri merupakan proses untuk mendapatkan parameter sinyal suara. Salah satu metode untuk ekstraksi ciri adalah *Linear Predictive Coding* (LPC). LPC merupakan metode ekstraksi ciri yang membutuhkan waktu lebih cepat dibandingkan metode ekstraksi ciri lainnya seperti *Mel Frequency Cepstral Coefficients* (MFCC) dan PS-ZCPA (Riyanto & Sutejo, 2014).

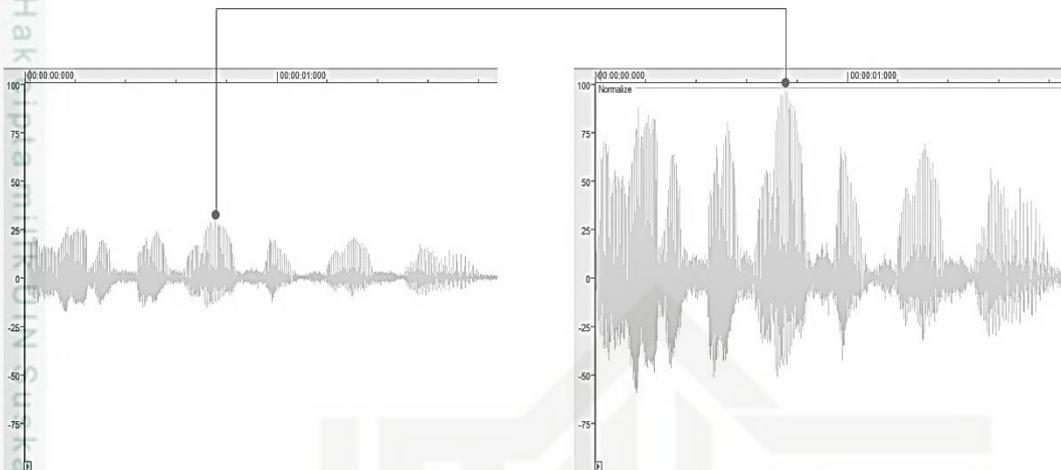


Gambar 2.7 Blok Diagram LPC

Langkah – langkah proses ekstraksi ciri menggunakan LPC untuk mendapatkan koefisien LPC adalah sebagai berikut :

2.4.1 Normalisasi / Preemphasis

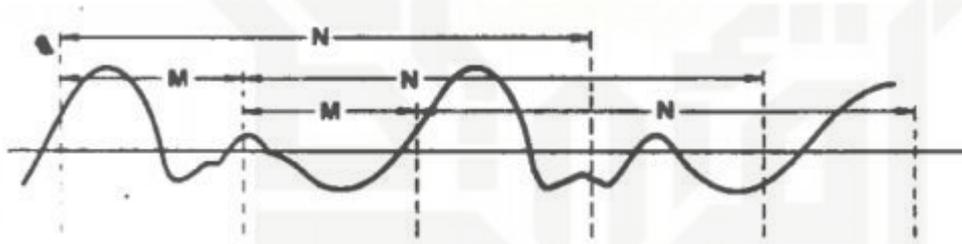
Sinyal suara pertama kali akan melewati tahap normalisasi sebelum diproses ketahap selanjutnya. Pada tahapan ini, untuk mendeteksi ada atau tidaknya suara. Sinyal yang tidak terdeteksi adanya suara akan dihapus dan juga akan menghapus *white noise*. *White noise* adalah sinyal *noise* yang teratur/konstan dalam frekuensi tertentu. Tujuan utama pada tahap ini adalah untuk meningkatkan frekuensi yang lebih tinggi untuk meratakan spektrum dengan cara mengecek apakah sinyal tersebut merupakan *gain* tertinggi atau terendah. Gambar 2.8 berikut menampilkan perbedaan sinyal suara sebelum dan setelah dinormalisasi.



Gambar 2. 8 Normalisasi Pada Sinyal Suara

2.4.2 Frame Blocking

Pada tahap ini akan membagi sinyal suara kedalam beberapa *frame* yang terdiri dari N *sample* suara dan dipisahkan secara *overlap* (tumpang tindih) oleh M *sample*. *Overlap* dimaksudkan untuk mencegah tidak adanya informasi pada *frame* tersebut (Feng, 2004).



Gambar 2. 9 Frame Blocking (Rabiner & Juang, 1993)

2.4.3 Windowing

Windowing bertujuan untuk meminimalkan diskontinuitas pada ujung awal dan ujung akhir pada setiap *frame* dengan men-*taper* sinyal menuju nol pada ujung – ujungnya (Irmawan, Hikmarika, Sari, & Tammimi, 2014). Terdapat beberapa teknik *windowing*, seperti : *Hanning*, *Hamming*, *Blackman* dan *Rectangular window*. Penelitian ini menggunakan *Blackman window*. *Blackman window* didefinisikan persamaan berikut (Podder et al., 2014):

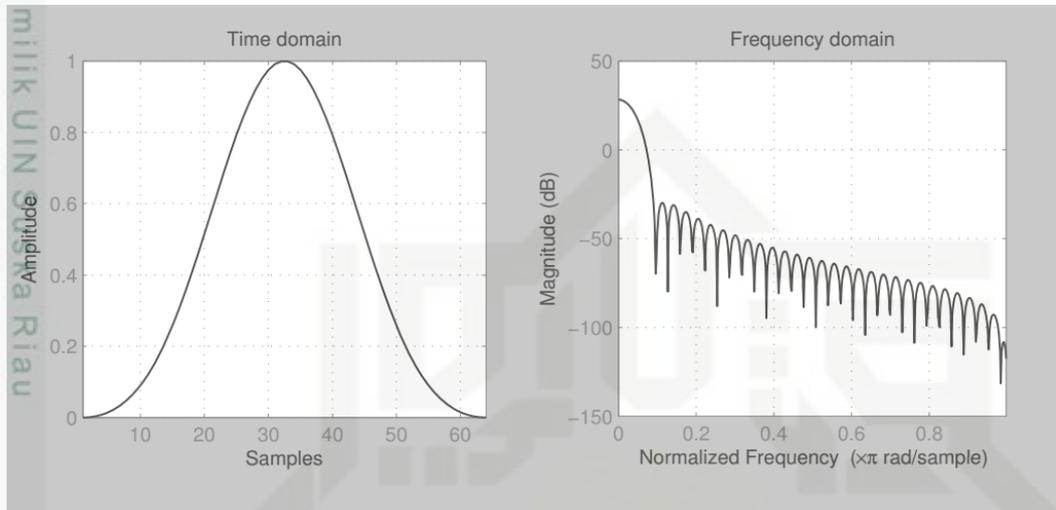
$$w(n) = a_0 - a_1 \cos\left(\frac{2\pi n}{N-1}\right) + a_2 \cos\left(\frac{4\pi n}{N-1}\right) \quad (2.1)$$

Umumnya nilai $\alpha = 0,16$ dimana a_0, a_1, a_3 merupakan konstanta yaitu

$$a_0 = 7938/18608 \approx 0,42659,$$

$$a_1 = 9240/18608 \approx 0,49656,$$

$$a_2 = 1430/18608 \approx 0,076849$$



Gambar 2. 10 Contoh Grafik *Blackman Windowing* (Beigi, 2011)

2.4.4 Analisis Autokorelasi

Selanjutnya setiap *frame* sinyal akan dilakukan analisis autokorelasi untuk mendapatkan nilai analisis autokorelasi. Analisis autokorelasi bertujuan untuk menprediksi *error* pada sinyal. Persamaannya sebagai berikut :

$$R_{yy}(l) = \sum_{n \in Z} y(n)\bar{y}(n-l) \quad (2.2)$$

dimana nilai $y(n)$ merupakan nilai dari sebuah *window* ke- n dan l merupakan *lag* atau sinyal yang akan dicari nilai autokorelasinya.

2.4.5 Ekstraksi Ciri LPC

Proses selanjutnya adalah analisis LPC, yaitu mengubah setiap *frame* autokorelasi menjadi parameter LPC atau yang biasa disebut sebagai koefisien LPC. Tahapan untuk mencari koefisien LPC dikenal dengan metode Levinson-Durbin (Beigi, 2011), yang dijabarkan dengan :

$$W(0) = S(n) \quad (2.3)$$

$$W(n) = W(n) * w(n) \quad (2.4)$$

$$E^{(0)} = r(0) \quad (2.5)$$

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.

pengujian terhadap pengaruh parameter seperti jumlah iterasi, jumlah koefisien MFCC dan jumlah data latih terhadap ketepatan identifikasi pembicara. Hasil pengujian didapatkan bahwa jumlah koefisien MFCC dan jumlah iterasi dapat meningkatkan ketepatan identifikasi pembicara dengan hasil tertinggi sebanyak 80% dan akurasi rata – rata sebesar 44%.

2. **Perbandingan Metode Ekstraksi Ciri Suara MFCC, ZCPA dan LPC (Riyanto & Sutejo, 2014)**

Penelitian ini membandingkan tiga metode ekstraksi , yaitu *Mel Frequency Cepstral Coefficients* (MFCC), *Zero Crossings with Peak Amplitudes* (ZCPA) dan *Linear Predictive Coding* (LPC). Hasil yang didapatkan dari penelitian ini adalah MFCC memiliki tingkat keakuratan yang lebih baik.

3. ***Comparative Performance Analysis of Hamming , Hanning and Blackman Window* (Podder et al., 2014)**

Penelitian ini bertujuan untuk membandingkan performa dari *Hamming*, *Hanning* dan *Blackman window*. Penelitian ini dilakukan menggunakan simulasi MATLAB. Hasil dari penelitian ini adalah *Blackman window* memberikan performa yang lebih unggul dibandingkan dari *Hamming* dan *Hanning window*.

4. **Comparison of the Impact of Some Minkowski Metrics on VQ/GMM Based Speaker Recognition (Hanilçi & Ertaş, 2011)**

Penelitian ini mengevaluasi dampak dari Minkowski metric (Euclidean, City Block, dan Chebychev distances) terhadap performa dari vector quantization (VQ) dan Gaussian mixture model (GMM) berdasarkan closed-set text-independent pada pengenalan pembicara untuk mengukur tingkat pengenalan dan akurasi.

Hasil dari penelitian ini adalah yang berbasis VQ dan GMM, *city-block distance* memberikan tingkat identifikasi yang paling tinggi, dan *chebyshev distance* memberikan tingkat verifikasi yang terbaik.