

Hak Cipta Diindungi Undang-Undang

1. Diarangi 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.

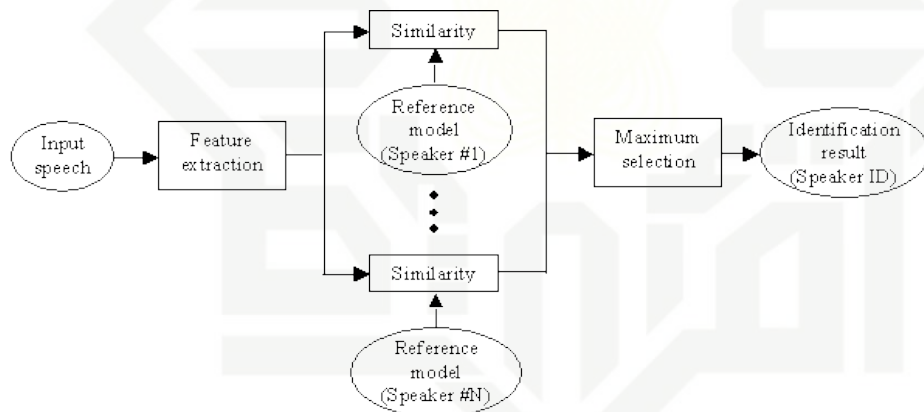
2. Diarangi mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

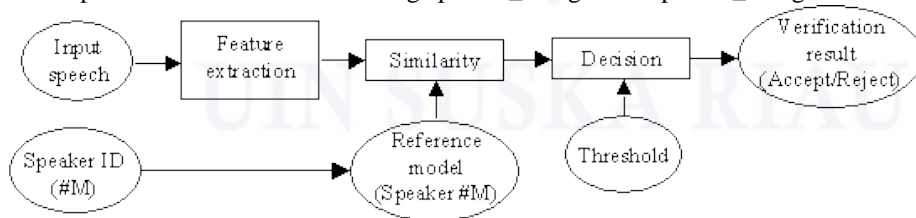
2.1 Pengenalan Pembicara

Pengenalan pembicara (*speaker recognition*) adalah semua kegiatan yang melibatkan identifikasi pembicara, berdasarkan suaranya dan pengelompokan pembicara berdasarkan kesamaan dari suara mereka (research.ibm.com). Menurut Alex S. Park pengenalan pembicara memiliki dua tujuan, yaitu untuk identifikasi dan verifikasi. Identifikasi pembicara adalah proses menentukan pembicara dari data-data suara yang telah terdaftar sebelumnya sedangkan verifikasi pembicara adalah proses penentuan untuk menerima atau menolak identitas yang diklaim oleh pembicara.



Gambar 2.1 Struktur Identifikasi Pembicara

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



Gambar 2.2 Struktur Verifikasi Pembicara

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

Menurut Homayoon Beigi, ada dua tipe identifikasi pembicara yaitu, *closed-set* dan *open-set*. Identifikasi pembicara yang bersifat *closed-set* akan membandingkan suara dari pembicara dengan data-data suara yang telah dipelajari

Hak Cipta Diindungi Undang-Undang

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

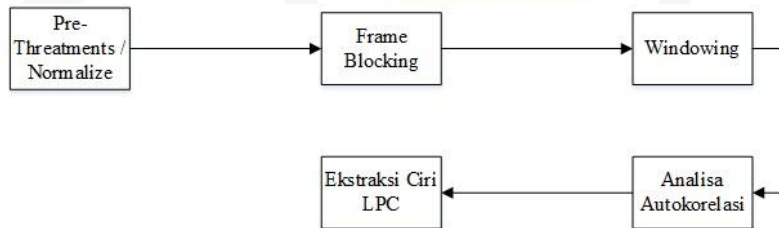
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

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

sedangkan identifikasi pembicara yang bersifat *open-set* bisa membandingkan suara yang belum dipelajari sebelumnya.

2.2 Linear Predictive Coding

Bagian terpenting dari pengenalan suara adalah ekstraksi ciri (*feature extraction*). Kegunaan dari ekstraksi ciri adalah mengurangi informasi-informasi dari sensor dan mengkonversi informasi-informasi yang penting dari signal untuk pengenalan pola agar menghasilkan format yang lebih sederhana dengan kelas-kelas yang jelas. *Linear Predictive Coding* (LPC) melakukan analisis dengan cara memperkirakan *formant*, memisahkan *formant* dari sinyal, yang dinamakan proses *inverse filtering*, lalu mengestimasi intensitas dan frekuensi dari sinyal percakapan yang tersisa, yang disebut *residue*. Karena sinyal percakapan bervariasi seiring waktu, estimasi tersebut dilakukan untuk setiap potongan kecil dari sinyal, yang dinamakan *frame*.



Gambar 2.3 Blok Diagram LPC (Priyanga, 2011)

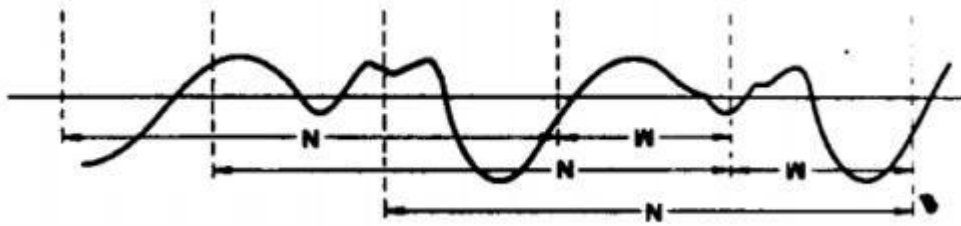
2.2.1 Pre-Threatments / Normalize

Sinyal suara digital akan melewati tahap *pre-threatments* atau tahap *normalize* sebelum diproses lebih lanjut. *Pre-Threatments* merupakan sebuah tahap untuk mendeteksi adanya aktivitas suara atau tidak. Dengan mendeteksi aktivitas suara pada sinyal maka akan dilakukan penghapusan terhadap waktu yang tidak memiliki aktivitas suara (*remove silence*) dan menghapus *white noise*.

Setelah sinyal suara digital melalui tahap *Pre-Threatments*, sinyal suara akan memasuki tahap *normalize* atau tahap normalisasi. Pada tahap ini sinyal akan dinormalisasi dengan mencari nilai tertinggi dari *gain* pada *sample* suara pada angka 1.0 ataupun -1.0.

2.2.2 Frame Blocking

Sinyal yang telah dilakukan *pre-treatments*, kemudian dibagi menjadi sejumlah *frame*. Tiap *frame* berisi N *sample*, dengan jarak setiap *frame* dipisahkan sebanyak M *sample*. Proses *frame blocking* bisa dilihat pada gambar 2.4.



Gambar 2.4 Frame Blocking (Rabiner & Juang, 1993)

Dapat diperhatikan pada gambar, *frame* pertama mempunyai panjang N *sample*. *Frame* kedua dimulai M *sample* setelah *frame* pertama dan terjadi *overlap* sebanyak $N - M$ *sample*, begitu juga dengan *frame* ketiga yang dimulai dari $2M$ *sample* dan terjadi *overlap* sebanyak $N - 2M$ *sample*. Proses ini akan berlanjut sampai semua *frame* dikunjungi. Parameter yang harus diperhatikan untuk mendapatkan *frame* yang *overlap* adalah ukuran $M \leq N$, dan hasil estimasi koefisien LPC akan berkorelasi antar *frame*.

2.2.3 Windowing

Tahap selanjutnya dalam proses LPC adalah meminimalisir diskontinuitas sinyal suara pada awal dan akhir setiap *frame*. *Windowing* dapat dilakukan untuk memudahkan *smoothing* pada setiap *frame*. *Window* yang biasa digunakan dalam LPC adalah *Hamming Window*, yang mempunyai persamaan sebagai berikut:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N - 1 \quad (2.1)$$

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

N = Ukuran *window*

2.2.4 Autocorrelation Analysis (Analisa Autokorelasi)

Setiap *frame* yang telah dilakukan proses *windowing* kemudian dilakukan *autocorrelation analysis* untuk mendapatkan nilai *autocorrelation analysis*. Adapun persamaannya bisa dilihat pada rumus dibawah ini:

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.

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

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

2.2.5 Ekstraksi Ciri LPC

Proses selanjutnya adalah ekstraksi ciri dengan menggunakan LPC *Analysis*, yaitu merubah setiap *frame* dari *autocorrelation* menjadi himpunan LPC parameter, yang berupa koefisien LPC dan koefisien *error* yang kemudian akan diiterasikan sebanyak jumlah *sample* pada data suara. Metode formal untuk melakukan konversi LPC parameter menjadi koefisien LPC dikenal dengan metode *Durbin*, dan dapat dijabarkan dengan:

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

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

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

$$k_i = \frac{\{r(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} r(i-j)\}}{E^{(i-1)}} \quad (2.6)$$

$$\alpha_i^{(i)} = k_i \quad (2.7)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad (2.8)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad (2.9)$$

$$G(i) = (\alpha(i) + \alpha(i-1))/Q \quad (2.10)$$

- $S(n)$ = nilai *sample* data suara ke- n
- $W(n)$ = nilai *window* ke- n
- $w(n)$ = nilai *hamming window* ke- n
- $E(i)$ = nilai *Error* pada LPC ke- i
- $\alpha(i)$ = nilai koefisien LPC ke- i
- $G(i)$ = nilai vector ciri suara ke- i

2.3 Chebyshev Distance

Chebyshev Distance merupakan perhitungan jarak dari dua buah titik dalam sebuah *Vector Space*. *Chebyshev Distance* biasa juga disebut *Chessboard Distance*



Hak Cipta Diindungi 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.

(Jarak Papan Catur). Menurut Kardi (2015) *Chebyshev Distance* adalah mengukur besarnya nilai mutlak dari koordinat pada sepasang objek.

$$d_{xy} = \max_k |x_{ik} - y_{jk}| \quad (2.11)$$

Dimana d_{xy} adalah jarak *Chebyshev* dari titik x ke titik y . Semakin kecil nilai d_{xy} , maka semakin besar kesamaan antara kedua obyek atau kasus tersebut.

2.4 Unified Modeling Language (UML)

Unified Modeling Language (UML) adalah sebuah “bahasa“ yang telah menjadi standar dalam industri untuk visualisasi, merancang dan mendokumentasikan sistem piranti lunak, UML menawarkan sebuah standar untuk merancang model sebuah sistem (Dharwiyanti & Wahono, 2003).

UML adalah bahasa pemrograman yang membangun sebuah program berdasarkan grafik atau gambar untuk memvisualisasikan sampai pada membangun sebuah sistem dengan basis objek. UML tidak hanya bahasa pemrograman visual, namun juga dapat secara langsung dihubungkan ke berbagai bahasa pemrograman seperti *Java*, *C++*, *Visual Basic*, dan *database* yang berbasis objek, begitu juga dengan pendokumentasian seperti *requirements*, arsitektur, *design*, *source code*, *project plan*, *test* dan *prototype*.

UML membutuhkan konsep sebuah bahasa model dan mempelajari 3 elemen utama dari UML seperti *building block*, aturan *building block* diletakkan secara bersamaan, dan mekanisme umum.

2.5 Pembahasan Penelitian Terdahulu

Penelitian ini bukanlah penelitian pertama yang membahas tentang identifikasi pembicara dengan metode LPC dan metode *Chebyshev Distance*. Terdapat penelitian-penelitian terdahulu yang juga sudah membahas metode LPC dan metode *Chebyshev Distance* ini namun dengan ruang lingkup dan studi kasus yang berbeda. Penelitian terdahulu yang pernah dilakukan dapat dilihat sebagai berikut.

2.5.1 Deteksi Emosi Melalui Pengenalan Suara Menggunakan *Linear Predictive Coding* (LPC) dan *Hidden Markov Model* (HMM)

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

Salah satu bentuk interaksi antara manusia-komputer atau manusia-mesin adalah ucapan. Ucapan terdiri dari kata yang diucapkan dengan cara tertentu. Berdasarkan informasi yang terkandung, ucapan dapat dibagi menjadi *linguistic information* dan *paralinguistic information*. *Paralinguistic information* mengacu kepada pesan tersirat yang terkandung di dalam ucapan, seperti emosi dari pengucap. Emosi yang akan diidentifikasi pada penelitian tersebut adalah senang, sedih, marah, jijik, dan netral. Emosi tersebut biasa disebut *archetypal emotion*.

Sinyal suara direpresentasikan dengan beberapa ciri, yaitu: fitur *Linear Predictive Coding* (LPC), turunan orde pertama koefisien *cepstral*, *fundamental frequency*, dan energi. *Hidden Markov Model* (HMM) digunakan untuk klasifikasi emosi. Performa dari sistem akan diukur berdasarkan akurasi dalam ketepatan deteksi emosi (Yahya, 2012).

Hasil penelitian tersebut menunjukkan bahwa rata-rata akurasi yang dihasilkan sistem adalah 60.65% dan yang terbaik adalah 70.70% dalam mengenali 5 jenis emosi. Hasil penelitian juga menunjukkan bahwa fitur LPC yang dikombinasikan dengan *fundamental frequency* dan energi akan menghasilkan akurasi yang lebih tinggi dalam pendeteksian emosi melalui suara.

Pada penelitian tersebut objek yang digunakan berupa emosi seseorang yang bersifat *text-dependent*. Pada penelitian ini penulis akan menggunakan objek berupa suara seseorang dan bersifat *text-independent* serta penulis tidak menggunakan HMM sebagai *classifier*, tetapi penulis menggunakan *Chebyshev Distance* sebagai *classifier* dikarenakan classifier sebelumnya merupakan sebuah jaringan syaraf tiruan yang mana membutuhkan data training yang banyak sehingga akan memakan banyak waktu dalam pengujian.

2.5.2 Perancangan Sistem Pengenalan Suara Untuk Identifikasi Personal Menggunakan Metoda Ekstraksi LPC dan Algoritma Propagasi Balik Berbasis *Neural Network*

Tujuan dari penelitian tersebut adalah pembangunan sistem identifikasi dengan konsep ekstraksi ciri suara (*speech*) dan *cross-correlation* ciri suara. Metoda ekstraksi ciri menggunakan *Linear Predictive Coding* (LPC) dan teknik



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.

crosscorrelation antar ciri menggunakan *Neural Network* dengan algoritma propagasi balik. LPC digunakan karena menyediakan kompresi dan ekstraksi parameter ciri suara dengan sangat baik dan telah digunakan secara luas. Penggunaan *Neural Network* ditujukan untuk mendapatkan akurasi yang tinggi dari hasil *cross-correlation*.

Sistem identifikasi tersebut dibangun dengan menggunakan MATLAB untuk antarmuka dan seluruh badan sistem. Penggunaan konfigurasi sistem ini didukung dengan GUI menggunakan MATLAB sehingga dapat digunakan oleh user secara lebih praktis. *User* dapat memilih untuk melakukan input baru secara langsung, identifikasi, play suara, dan tampilan sinyal asli dengan estimasi dalam satu axe untuk 300 sample pertama. Dengan media suara sebagai identifikasi, dari hasil uji akurasi telah dapatkan tingkat akurasi 20% (Priyangga, 2011).

Penulis tidak akan menggunakan seluruh langkah yang ada dikarenakan ada beberapa langkah yang tidak relevan dengan metode LPC. Penelitian tersebut juga memiliki tingkat akurasi yang masih tergolong rendah yaitu 20%. Pada penelitian kali ini sistem yang dibangun diharapkan mempunyai tingkat akurais yang lebih tinggi dari penelitian sebelumnya.

2.5.3 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 begitu banyak ragamnya serta banyak suara yang hampir sama antara manusia satu dengan manusia lainnya. Dalam kehidupan sehari-hari pengenalan pembicara juga sangat penting, contohnya dalam mengidentifikasi siapa yang berbicara pada rekaman percakapan yang dijadikan bukti dalam sebuah persidangan.

Penelitian tersebut menguji tingkat akurasi dari identifikasi suara menggunakan MFCC (*Mel Frequency Cepstral Coefficient*) dan SOM (*Self Organizing Map*). MFCC digunakan sebagai ekstrasi ciri dari sinyal suara. SOM digunakan sebagai *cluster* pada vektor ciri yang didapatkan dari MFCC. Hasil penelitian menunjukkan bahwa SOM bisa digunakan sebagai identifikasi


Hak Cipta Diindungi 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.

pembicara dengan tingkat akurasi tertinggi untuk pembicara wanita adalah 54,4%, pembicara pria 75,6% dan rata-rata keduanya adalah 62,5% (Permana, 2011). Pada penelitian kali ini metode ekstraksi ciri yang digunakan adalah metode LPC yang bertujuan juga untuk membandingkan tingkat akurasi metode LPC terhadap metode MFCC.

2.5.4 *A Comparative Study of K Means Algorithm by Different Distance Measures*

Clustering merupakan bagian yang sangat penting pada penelitian dibidang *data mining*. Data dikelompokkan menjadi *cluster* dengan sebuah cara yang mengelompokkan berdasarkan kesamaan dan ketidaksamaan. Algoritma K-Means adalah sebuah metode penting untuk menemukan *cluster*. Implementasinya sangat sederhana dan cepat. Pada penelitian tersebut algoritma K-Means digunakan pada data bunga. Berbagai metode pengukuran jarak digunakan seperti, *Euclidean Distance*, *Manhattan Distance*, dan *Chebyshev Distance* untuk menganalisa hasil dari banyaknya iterasi, akurasi, dan nilai *error*. Penelitian tersebut menunjukkan bahwa metode *Chebyshev Distance* lebih baik daripada metode lainnya dengan rata-rata akurasi 86%, banyaknya iterasi adalah sebanyak 30, dan nilai *error* 0.