

SISTEM AKSES NILAI HASIL STUDI MENGGUNAKAN PASSWORD SUARA**Wahju Adi Prijono**

Abstrak: Suara dapat digunakan manusia untuk saling berkomunikasi, memberikan informasi maupun perintah. Namun banyak persoalan yang muncul ketika suara dimanfaatkan dalam sebuah sistem karena setiap manusia mempunyai ciri suara yang berbeda. Pada Penelitian ini, sinyal suara dimanfaatkan untuk mengakses nilai hasil studi. Sinyal suara diolah terlebih dahulu dalam beberapa tahap, yaitu *pre-processing*, *pre-Emphasis*, *frame blocking*, *windowing*, *FFT* (*Fast Fourier Transform*) dan *LPC* (*Linier Predictive Coding*) yang terdiri dari analisis autokorelasi, analisis *LPC* dan parameter *LPC* ke koefisien *cepstrum*. Sedangkan metode yang digunakan pada proses pengenalan suara adalah *DTW* (*Dynamic Time Warping*). Pada proses pengenalan suara, koefisien *cepstrum* sinyal masukan dibandingkan dengan koefisien *cepstrum* sinyal standar. Perangkat keras yang digunakan adalah *PC* (*Personal Computer*), *sound card*, mikropon dan *speaker*, sedangkan perangkat lunak yang digunakan adalah *Tcl/Tk*, *snack* dan bahasa *C++*. Kata yang diuji adalah kata “Komputer”, “Cantik”, “Basket”, “Malang” dan “Juni”. Berdasarkan hasil pengujian dan analisis, keberhasilan sistem pengenalan suara dengan sinyal masukan yang diucapkan oleh pengucap standar adalah 58%, sedangkan keberhasilan sistem pengenalan suara dengan sinyal masukan yang diucapkan oleh pengucap lain yang bukan standar adalah 99%. Tingkat keberhasilan sistem pengenalan suara dipengaruhi oleh beberapa faktor, yaitu kondisi kesehatan penutur, pola suara yang diucapkan oleh penutur, tipe *hardware* yang digunakan dan keadaan lingkungan.

Kata kunci: *Front-end detection*, *pre-Emphasis*, *frame blocking*, *windowing*, *Fast Fourier Transform*, *Linier Predictive Coding*, analisis autokorelasi, analisis *LPC*.

Pada *Speech Recognition* sinyal suara dianalisis dan diolah sehingga didapatkan besaran tertentu. Untuk suara yang berbeda akan didapatkan besaran yang berbeda, perbedaan ini dapat disebabkan karena pengucapan fonem, kondisi emosi, noise dan faktor-faktor lainnya. Sistem *Speech Recognition* dapat dioperasikan pada dua mode yang berbeda. Pertama adalah mode belajar. Pada mode ini, sistem akan dilatih menggunakan sejumlah kata atau kalimat yang memenuhi suatu kriteria tertentu dan disimpan sebagai standar. Kedua adalah pengenalan suara. Pada mode ini, setiap kata atau kalimat yang ingin dikenali akan dianalisis polanya. Berdasarkan hasil perbandingan dengan standar, pengambil keputusan akan mengidentifikasi kata atau kalimat yang diucapkan tersebut.

Pada Penelitian ini akan dibuat suatu sistem pengenalan wicara yang digunakan untuk membuka aplikasi daftar nilai dengan menggunakan password suara. Pembahasan Penelitian ditekankan pada:

1. Bagaimana blok diagram sistem pengolahan sinyal suara.
2. Bagaimana *Linier Predictive Coding* (*LPC*) diterapkan pada sistem pengolahan sinyal suara.
3. Bagaimana mendapatkan code book dengan menggunakan *Linier Predictive Coding* (*LPC*).
4. Bagaimana mengenali suara manusia dengan menggunakan *DTW* (*Dynamic Time Warping*) untuk membuka aplikasi daftar nilai hasil studi dengan password suara.

Batasan-batasan yang diberikan dalam penulisan Penelitian ini adalah sebagai berikut:

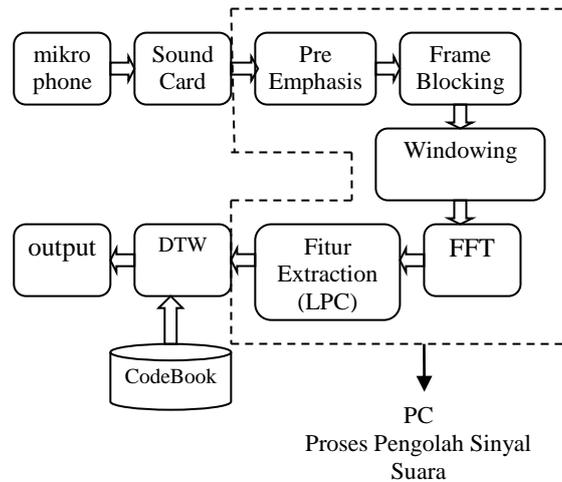
1. Tidak memperhitungkan noise.
2. Hanya bisa mengakses satu aplikasi dalam waktu yang sama

3. Objek perekam bersifat dependent speaker dan berjumlah 5 orang.
4. Terdapat 5 kata yang dapat dikenali, yaitu “Komputer”, “Cantik”, “Malang”, “Basket”, “Juni”.
5. Tidak membahas detail hardware yang digunakan.

Sinyal wicara dimanfaatkan sebagai perintah untuk membuka aplikasi daftar nilai hasil studi menggunakan password suara dan menganalisis keluaran dari masing-masing proses pengolahan suara.

Pengolahan Sinyal Suara

Sinyal suara diolah dengan LPC (*Linier Predictive Coding*) sehingga didapatkan ekstraksi sinyal suara yang digunakan untuk membedakan suara manusia yang satu dengan manusia yang lain.



Sistem yang akan direncanakan adalah sebagai berikut :

Gambar 1. Blok Diagram Proses Pengolah Sinyal Suara

Sumber: Perencanaan

Proses Sampling

Pada proses sampling dilakukan konversi sinyal waktu kontinu menjadi sinyal waktu diskrit. Berdasarkan pada teori sampling Nyquist:

$$F_s \geq 2 \times F_{maks}$$

Keterangan:

F_s = Frekuensi sampling

F_{maks} = Frekuensi sinyal analog tertinggi.

Pre-Emphasis

Pre-Emphasis merupakan proses pemfilteran dengan filter FIR (*Finite Impulse Response*) orde satu. Hal ini dilakukan untuk meratakan *cepstrum* sinyal suara.

$$s'(n) = s(n) - bs(n-1)$$

Keterangan:

$s'(n)$ = keluaran *Pre-Emphasis*

$s(n)$ = masukan *Pre-Emphasis*

$s(n-1)$ = masukan *Pre-Emphasis* sebelumnya

b = koefisien *Pre-Emphasis* ($0.9 \leq b \leq 1.0$)

Frame Blocking

Pada *frame blocking* sinyal suara dibagi menjadi beberapa frame sinyal kecil. Sinyal suara dibagi menjadi frame-frame dengan masing-masing frame memuat N

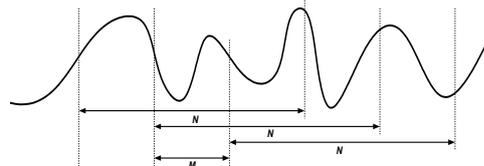
cuplikan kata dan frame-frame yang berdekatan dipisahkan sejauh M cuplikan, dimana $N \approx 3 \times M$, seperti ditunjukkan pada Gambar 2. Jika didefinisikan $x_l(n)$ adalah frame ke- l , dan terdapat L frame maka didapatkan nilai dari sinyal yang baru adalah sebagai berikut:

$$x_l(n) = s'(M \bullet l + n)$$

Dimana

$$n = 0, 1, 2, \dots, N-1$$

$$l = 0, 1, 2, \dots, L-1$$



Gambar 2. Bentuk Sinyal yang di *Frame Blocking*

Sumber: Rabiner, L. R., Juang, B. H. *Fundamental of Speech Recognition*

Windowing

Windowing digunakan untuk meminimalkan discontinuitas pada ujung awal dan ujung akhir setiap frame.

$$x_l'(n) = x_l(n) \times w(n) \quad 0 \leq n \leq N-1$$

Keterangan:

$x_l'(n)$ = Sinyal suara hasil windowing ke- n untuk frame ke- l

$x_l(n)$ = Sinyal suara ke- n untuk frame ke- l

$w(n)$ = Fungsi windowing

N = Jumlah data dalam satu frame

n = Urutan data ke- n

Fast Fourier Transform

Transformasi *Fourier* dilakukan untuk mentransformasikan sinyal dari domain waktu ke domain frekuensi.

Jika $x(t)$ adalah sinyal periodik, maka dengan transformasi *fourier* akan didapatkan:

$$X(F) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$$

Sinyal periodik $x(t)$ dapat diperoleh kembali dengan invers transformasi *fourier*:

$$x(t) = \int_{-\infty}^{\infty} X(F) e^{j\omega t} dF$$

Dengan mengubah variable waktu (t) dan frekuensi (F) kedalam bentuk diskrit akan didapatkan:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\omega_0 kn}$$

Keterangan:

$X(k)$ = Sinyal dalam domain frekuensi

$x(n)$ = Sinyal masukan dalam domain waktu

k = $0, 1, 2, \dots, N-1$

N = Jumlah sampel

n = Sampel ke- dari jumlah sample

ω_0 = frekuensi fundamental

$$= 2\pi/\text{sampling rate} = 2\pi/N$$

Fast fourier Transform digunakan agar didapatkan perhitungan yang lebih cepat. Pada algoritma FFT radik 2, data sebanyak N buah titik dibagi menjadi 2 bagian, yaitu

data dengan indeks 0 sampai dengan $N/2-1$ dan data dengan indeks $N/2$ sampai dengan $N-1$.

LPC (Linier Predictive Coding)

Analisis Autokorelasi

Masing-masing frame sinyal yang telah diwindow di autokorelasikan dengan menggunakan persamaan :

$$r(m) = \sum_{n=0}^{N-1-m} x'(n)x'(n+m), \quad m=0,1,\dots,p$$

Keterangan:

$r(m)$ = nilai autokorelasi orde ke m

$x'(n)$ = nilai data window urutan data ke-n

m = orde *Linier Predictive Coding*

Nilai *autokorelasi* tertinggi (p) adalah orde dari analisis LPC. Pada proses analisis *autokorelasi*, *autokorelasi* ke-0 $R_1(0)$ adalah energi dari frame ke-1. Energi dari frame tersebut digunakan untuk sistem deteksi suara.

Analisis LPC

Pada analisis LPC digunakan metode Durbin dengan algoritma sebagai berikut:

$$E^{(0)} = r(0)$$

$$k_i = \{r(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} r(i-j)\} / E^{(i-1)}, \quad 1 \leq i \leq p$$

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

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

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

Setelah menyelesaikan persamaan 2.9 sampai 2.13, maka persamaan diselesaikan secara rekursif untuk $i=1,2,\dots,p$

Penyelesaian akhir digunakan :

$$\alpha_m = \alpha_j = \alpha_j^{(p)}, \quad 1 \leq j \leq p$$

Keterangan:

$E^{(0)}$ = energi pada frame

$r(i)$ = nilai autokorelasi orde ke i

α_m = koefisien *Linier Predictive Coding*

k_m = koefisien PARCOR

Parameter LPC ke Koefisien Cepstrum

Rekursif yang digunakan adalah :

$$c_0 = \ln \sigma^2$$

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad 1 \leq m \leq p$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad m > p$$

Keterangan:

c_m = koefisien *cepstrum* orde ke-m

σ = penguatan dalam model *Linier Predictive Coding*

a_m = koefisien *Linier Predictive Coding*

Koefisien *cepstrum* adalah koefisien transformasi *fourier* yang merepresentasikan spektrum *log magnitude*. Koefisien *cepstrum* lebih tahan terhadap noise jika digunakan pada pengenalan suara daripada penggunaan koefisien LPC (*Linier Predictive Coding*).

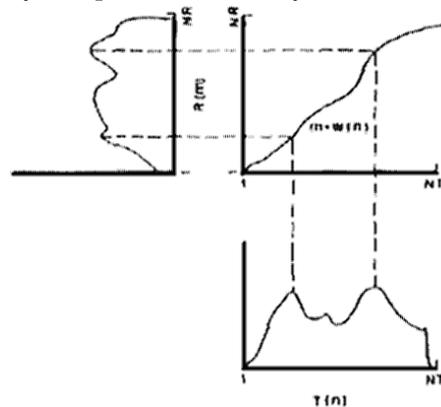
Tabel 2. Nilai Parameter Analisis yang Biasa Digunakan Pada Sistem Pengenalan Suara.

Parameter	Fs = 6.67 kHz	Fs = 8 kHz	Fs = 10 kHz
N	300 (45msec)	240 (30msec)	300 (30msec)
M	100 (15 msec)	80 (10msec)	100 (10msec)
p	8	10	10

Sumber: Rabiner, L. R., Juang, B. H. *Fundamental of Speech Recognition*

Dynamic Time Warping

Dynamic Time Warping adalah pengukuran jarak antara sinyal standar dan sinyal input. Hal yang diukur adalah berupa deretan nilai hasil dari LPC *Cepstrum* dalam bentuk kolom dan baris (i,j) yang disimpan dalam satu file. Output yang dihasilkan adalah jarak antara hasil sinyal *cepstrum* dari sinyal standar dan sinyal masukan.



Gambar 3. Pola Sinyal Dalam Proses DTW

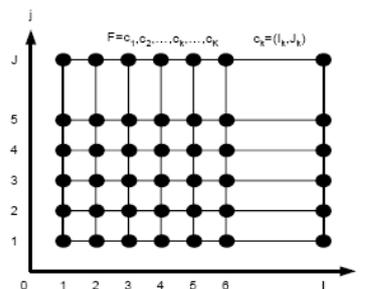
Sumber: David Mel Airlie. *A Voice Controlled Window Manager for the X-Windowing System*

Diasumsikan dua rentetan waktu untuk vektor utama yang dibandingkan sebagai berikut:

$$A = a_1, a_2, \dots, a_I$$

$$B = b_1, b_2, \dots, b_J$$

Dengan menganggap rentang bidang antara A dan B seperti ditunjukkan pada gambar 2.4, fungsi pelengkungan waktu yang sesuai menunjukkan waktu antara rentetan A dan B dapat dipresentasikan sebagai rentetan titik-titik pada bidang $c=(i,j)$, sebagai berikut:



Gambar 4. Plane Jaringan untuk *Dynamic Programming*

Sumber: Furui, Sadaoki. *Digital Speech Processing, Syntesis and Recognition*
 Penjumlahan jarak dari awal sampai akhir dari rentetan sepanjang F adalah:

$$D(F) = \frac{\sum_{k=1}^K d(ck)wk}{\sum_{k=1}^K wk}$$

$$\sum_{k=1}^K wk = I + J$$

Maka akan didapatkan persamaan:

$$D(F) = \frac{1}{I + J} \sum_{k=1}^K d(ck)wk$$

Keterangan:

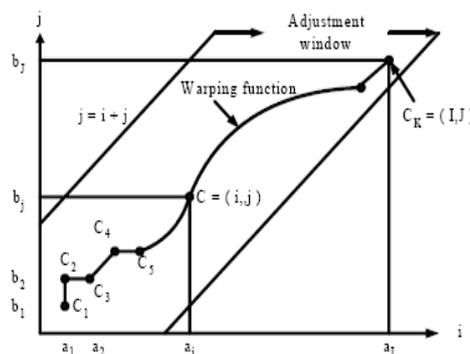
D(F) : Penjumlahan jarak dari awal sampai akhir dari rentetan sepanjang F

d(ck) : Jarak spektral antara dua vektor utama a1 dan b1

wk : Fungsi weighting

I : Panjang vektor frame I

J : Panjang vektor frame J



Gambar 5. DTW 2 Rentetan Waktu AdanB

Sumber: Furui, Sadaoki. *Digital Speech Processing, Syntesis and Recognition*

Persamaan diatas dapat diminimalisasi. Penjumlahan rentetan \$c_1, c_2, \dots, c_k (c_k=(i,j))\$ adalah:

$$g(ck) = g(i, j) = \min_{c_1, \dots, c_{k-1}} \left[\sum_{k=1}^k d(ck)wk \right]$$

Tabel 3. Local Constraints, Slope Weights dan DP Recursion Formula

Local Constraints and Slope Weights	DP Recursion Formula
<p>The diagram shows a small grid with three points. A diagonal arrow labeled '1' points from the bottom-left to the top-right. A horizontal arrow labeled '1' points from the bottom-left to the top-right. A vertical arrow labeled '1' points from the bottom-left to the top-right. A diagonal arrow labeled '2' points from the bottom-left to the top-right.</p>	$g(i, j) = \min \left\{ \begin{array}{l} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{array} \right\}$

Sumber: Rabiner, L. R., Juang, B. H. *Fundamental Of Speech Recognition*

Berdasarkan tabel 3. maka jarak antara dua rentetan waktu A dan B adalah:

$$D(F) = \frac{g(i, j)}{I + J} \quad (2.24)$$

Awal dan Akhir Sinyal suara

Front End Detection

Digunakan untuk menentukan batasan suatu sinyal, yaitu letak sinyal awal dan akhir dari suatu frame berdasarkan power.

$$P = \sum_{i=0}^{i=N} \sqrt{X_i^2}$$

Keterangan:

P = Power sinyal suara (dB)

X_i = Frame ke-i (dB)

N = Jumlah sampel per frame

i = Sampel ke-i

Nilai power digunakan untuk membedakan *voice* atau bukan. Dari nilai power diatas didapatkan nilai rata-rata, dengan menambahkan standar deviasi, didapatkan nilai awal dan akhir dari suatu frame.

voice > mean + standar deviasi

METODE

Studi Literatur

Sampling, pre-Emphasis, frame blocking, windowing, FFT (Fast Fourier Transform), LPC (*Linier Predictive Coding*), DTW (*Dynamic Time Warping*).

Pengambilan Sampel Ucapan.

Proses perekaman suara dilakukan dengan menggunakan software yang telah dibuat dengan Tcl/Tk 8.3 dan Snack 2.2. Frekuensi sampling diset sebesar 10 KHz dan disimpan dalam bentuk .wav. Pengambilan sampel ucapan dilakukan sebanyak 5 orang dengan kata yang berbeda-beda. Masing-masing orang mengucapkan kata yang sama sebanyak 10 kali untuk dijadikan standar.

Pengolahan Sinyal Suara

Data sampel berbentuk .wav diolah pada PC dalam beberapa tahap, yaitu sampling, *pre-Emphasis*, *frame blocking*, *windowing*, FFT dan LPC (*Linier Predictive Coding*).

Perbandingan Sinyal Suara

Koefisien *cepstrum* sinyal standar dan sinyal masukan dibandingkan dengan *Dynamic Time Warping*. Keluarannya berupa jarak antara koefisien *cepstrum* sinyal masukan dan sinyal standar.

Perancangan Sistem Pengenalan Suara

Perancangan sistem menjelaskan tentang perancangan sistem pada tiap blok.

- **Mikropon**

Mikropon yang digunakan adalah jenis global A 80, dengan frekuensi kerja sebesar 20 Hz – 20 KHz.

- **Sound card**

Pada penulisan Penelitian ini digunakan *sound card on board*, yaitu *sound card* yang menempel langsung pada *mother board* komputer.

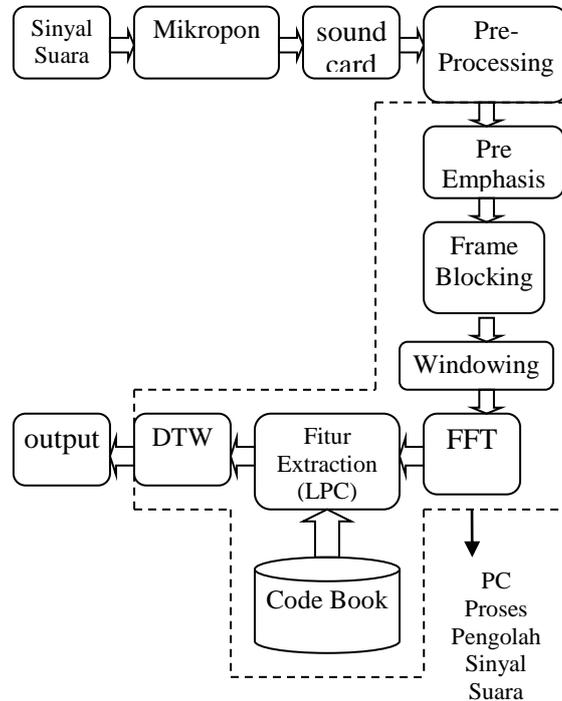
- **Pre-Processing**

Recording (Perekaman)

Sinyal suara disimpan dalam bentuk .wav dengan frekuensi sampling 10 KHz. Proses perekaman dilakukan dengan menggunakan mikropon dan *soundcard* yang ada pada PC. Perangkat lunak Tcl/Tk 8.3 dan Snack 2.2 digunakan untuk proses perekaman sinyal suara.

Front end Detection

Front end Detection adalah suatu proses yang digunakan untuk menentukan letak awal dan akhir suatu sinyal dengan menggunakan nilai power.



Gambar 6. Blok Diagram Pengenalan Suara
Sumber: Perancangan

- **Pengolahan Sinyal Suara.**

Pre-Emphasis

Pre-Emphasis adalah suatu proses pemfilteran dengan filter FIR orde 1. *Pre-Emphasis* digunakan agar noise yang disebabkan oleh perangkat PC dapat diredam.

Frame Blocking

Frame blocking adalah suatu proses yang digunakan untuk membagi sinyal suara kedalam beberapa frame. Seperti pada gambar 2.8 masing-masing frame memuat N cuplikan kata dan frame-frame yang berdekatan dipisahkan sejauh M cuplikan, dimana $N \approx 3 \times M$. Berdasarkan tabel 2.1 ukuran dari *frame blocking* ditetapkan sebesar 30 ms, hal ini sesuai dengan batasan waktu stasioner yaitu 5-100 ms untuk setiap frame, waktu stasioner ini diperlukan agar sinyal lebih mudah untuk dianalisis. Berdasarkan tabel 2.1 ditetapkan ukuran frame sebesar 30 ms atau 300 sampel dengan jarak antar frame sebesar 10 ms atau 100 sampel.

Windowing

Windowing digunakan untuk meminimalkan diskontinuitas pada ujung-ujung frame. Window yang digunakan adalah window hamming.

Fast Fourier Transform

Fast Fourier Transform digunakan untuk menyederhanakan komputasi data. Pada perancangan ini digunakan FFT 512 titik karena data yang akan di FFT adalah 300 data. Hasil sinyal yang di FFT merupakan suatu sinyal yang simetris, sehingga diambil 256 titik sedangkan 256 titik yang lain tidak digunakan karena berupa pencerminan saja.

Metode LPC (Linier Predictive Coding)

LPC adalah suatu metode yang digunakan dalam pengenalan suara untuk mendapatkan ciri dari suatu sinyal suara.

Analisis Autokorelasi

Analisis *autokorelasi* digunakan untuk menyederhanakan data berdasarkan orde LPC. Sinyal yang berada dalam domain waktu dan telah diwindowing dicari korelasinya dengan orde LPC sebesar 10 yaitu orde yang sesuai dengan tabel 2.1, sehingga dari 300 data dapat disederhanakan menjadi 10 data. Hal ini berarti dalam 1 frame akan didapatkan 10 data yang dapat mewakili keseluruhan data dalam 1 frame.

Analisis LPC

Analisis LPC digunakan untuk mendapatkan parameter sinyal suara, yaitu koefisien LPC. Hal ini dilakukan dengan cara mengkonversi nilai *autokorelasi* ke parameter LPC sehingga didapatkan koefisien LPC.

LPC Parameter ke Koefisien Cepstrum

Koefisien *cepstrum* adalah parameter sinyal yang baik, karena lebih tahan terhadap noise jika digunakan pada pengenalan suara daripada parameter LPC yang lain.

- **Dynamic Time warping.**

Dynamic Time warping adalah suatu cara untuk membandingkan pola sinyal suara. Sinyal yang dibandingkan adalah sinyal hasil *cepstrum* dari sinyal masukan dengan sinyal standar.

HASIL DAN PEMBAHASAN

Pengujian dan Analisis Proses Pengolahan Sinyal Suara.

Pre-Processing.

Recording (Perekaman).

Terdapat 5 jenis kata yang digunakan, yaitu “Komputer”, “Cantik”, “Malang”, “Basket”, “Juni”. Kata tersebut dipilih berdasarkan asumsi bahwa *password* yang digunakan masing-masing orang berbeda-beda, jenis kata yang digunakan sebagai *password* dibatasi 2-3 suku. Hal ini dilakukan agar proses pengenalan suara dapat bekerja lebih baik. Pengambilan sampel ucapan tersebut diambil dari 5 orang. Masing-masing orang mengucapkan kata yang sama sebanyak 10 kali untuk dijadikan standar. Dalam pengambilan sinyal masukan, pelafalan setiap kata diusahakan sama karena sangat berpengaruh pada proses pengenalan suara.

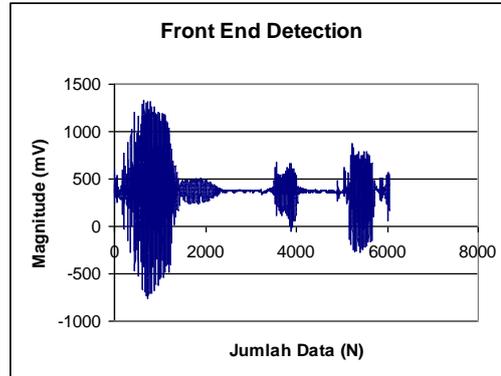


Gambar 7. Tampilan Hasil Perekaman Kata “Komputer”

Sumber: Hasil Perancangan dengan Menggunakan Tcl/Tk 8.3 dan Snack

Front end Detection.

Pada Gambar 8 Sinyal awal dan akhir dipotong berdasarkan nilai power, sehingga *noise* pada awal dan akhir sinyal suara dapat dihilangkan.

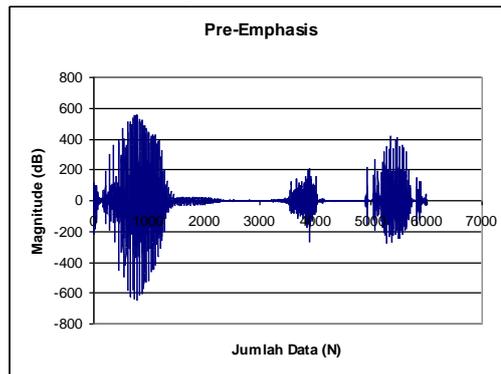


Gambar 8. Data Hasil *Front end Detection* Kata "Komputer"

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack

Pre-Emphasis.

Gambar 9 adalah gambar sinyal suara setelah difilter. *Cepstrum* sinyal suara setelah difilter lebih rata dibandingkan dengan sebelum difilter. Noise yang disebabkan oleh perangkat *Personal Computer (PC)* dapat diredam.

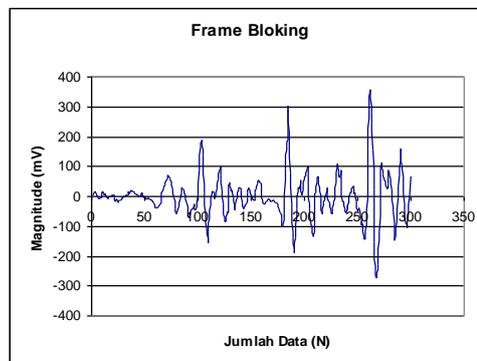


Gambar 9. Data Hasil *Pre-Emphasis* Kata "Komputer"

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack

Frame Blocking.

Berdasarkan Gambar 10 dapat dilihat bahwa dalam 1 frame terdiri dari 300 data. Bentuk sinyal yang telah di *frame blocking* lebih jelas daripada sinyal sebelum di *frame blocking*, sehingga lebih mudah untuk dianalisis.

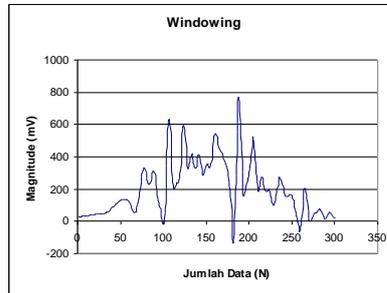


Gambar 10. Data Hasil *Frame Blocking* Kata "Komputer" ke 0.01 second

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack

Windowing

Seperti pada Gambar 11 dapat diketahui bahwa window hamming menyebabkan sinyal yang disampel lebih halus pada ujung-ujung frame. Hal ini membuktikan bahwa efek diskontinuitas pada ujung-ujung frame dapat dikurangi melalui proses *windowing*.

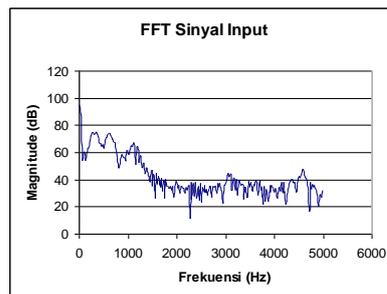


Gambar 11. Data Hasil *Windowing* Kata “Komputer” ke 0.01 second

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack

FFT (*Fast Fourier Transform*)

Pada perancangan ini digunakan FFT 512 titik karena data yang akan di FFT adalah 300 data. Hasil sinyal yang di FFT merupakan suatu sinyal yang simetris, sehingga diambil 256 titik sedangkan 256 titik yang lain tidak digunakan karena berupa pencerminan saja. Dari Gambar 12 dapat dilihat bahwa range frekuensi yang digunakan adalah 0-5000 Hz. Frekuensi tersebut telah sesuai dengan frekuensi sampling pada perancangan.



Gambar 12. Data Hasil *Fast Fourier Transform* Sinyal Masukan Kata “Komputer”

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack

LPC (*Linier Predictive Coding*)

Analisis Autokorelasi

Pada Penelitian ini digunakan orde 10. Hal ini berarti dalam 1 frame terdapat 10 data yang dapat mewakili 300 data.

Tabel 4. Nilai Autokorelasi Frame ke-1

Nilai Autokorelasi Frame ke-1
16624372
16522402
16255809
15875009
15457314
15067266
14749808
14513898
14348306
14240878
.....

Sumber: Perancangan

Energi pada tiap frame ditunjukkan oleh data pertama dari masing-masing frame pada analisis *autokorelasi*, sehingga nilai dari data pertama lebih besar dari nilai data lainnya.

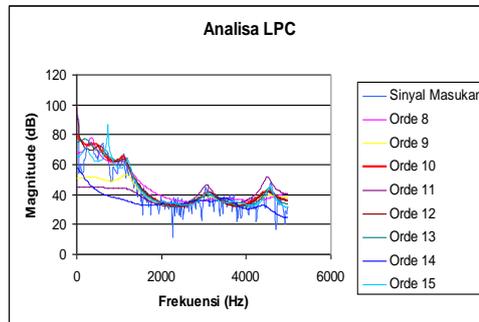
Analisis LPC

Tabel 5. Nilai Koefisien LPC frame ke-1

Nilai Koefisien LPC Frame ke-1
1.000000
-1.774702
0.209568
0.996206
0.033311
-0.228337
-0.714449
0.126185
0.856339
-0.483418

Sumber: Perancangan

Energi dari masing-masing frame ditunjukkan oleh data pertama dari setiap frame pada analisis LPC dan nilai selanjutnya adalah koefisien *LPC* dari suatu frame. Nilai tersebut digunakan untuk menentukan ciri dari sinyal suara yang dihasilkan oleh individu. Gambar 13 menunjukkan sinyal masukan hasil FFT dan hasil analisis LPC dari beberapa macam orde. Pada Gambar 5.7 dapat dilihat bahwa hasil analisis LPC yang paling menyerupai selubung envelope sinyal masukan hasil FFT adalah hasil analisis LPC orde 10. Hal ini sesuai dengan teori pada Tabel 2 yang menggunakan orde 10 untuk proses pengenalan suara dengan $N=300$ dan $M=100$.



Gambar 13. Analisis *LPC* dari Kata “Komputer”

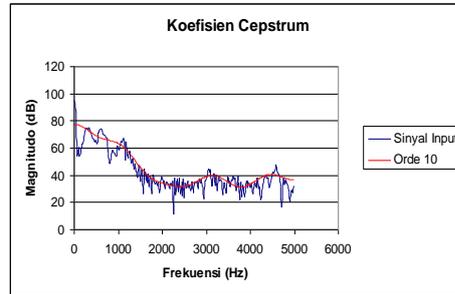
Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack
Parameter LPC ke Koefisien Cepstrum

Tabel 6. Nilai Koefisien Cepstrum frame ke 1

Nilai Koefisien Cepstrum Frame ke-1
10.351636
1.774702
1.365215
0.495053
0.040573
-0.332146
0.023355
0.199097
-0.086470
0.220065

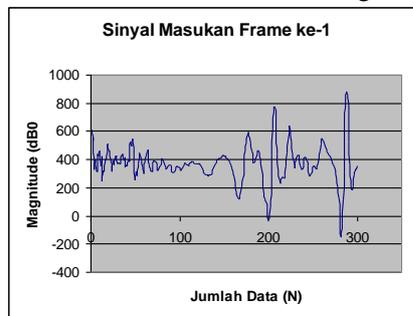
Sumber: Perencanaan

Nilai koefisien *cepstrum* yang bervariasi dapat digunakan untuk menentukan ciri sinyal suara yang dihasilkan oleh individu. Pada Gambar 14 dapat dilihat bahwa hasil koefisien cepstrum dengan orde 10 menyerupai sinyal masukan hasil FFT. Sehingga dapat disimpulkan hasil perencanaan adalah benar.



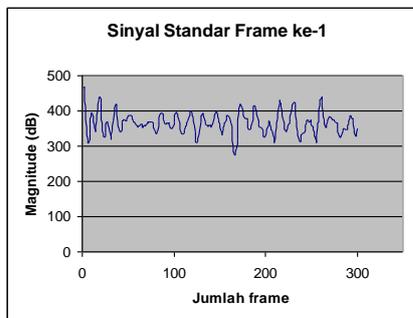
Gambar 14. Koefisien *cepstrum* dari Kata “Komputer”

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack DTW (*Dynamic Time Warping*)
Dynamic Time warping adalah suatu cara untuk membandingkan pola sinyal suara.



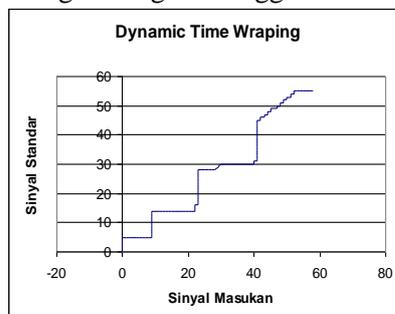
Gambar 15. Sinyal Masukan Frame ke-1 dari Kata “Komputer”

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack



Gambar 16. Sinyal Standar Frame ke-1 dari Kata “Komputer”

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack



Gambar 17. DTW antara Sinyal Masukan dan Sinyal Standar dari Kata “Komputer”

Sumber: Hasil Perancangan dengan Menggunakan Bahasa C++ dan Snack

Pada Gambar 17 dapat diketahui *Dynamic Time Warping* dari keseluruhan frame yang terdapat pada kata "Komputer". *Dynamic Time warping* digunakan untuk mendapatkan panjang jalur sinyal yang sama sehingga jarak antara dua buah sinyal dapat dihitung.

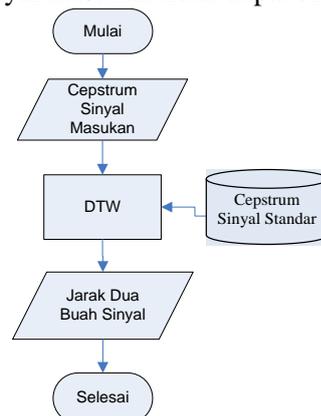
Pengukuran Jarak Dua Buah Sinyal Dengan *Dynamic Time Warping*.

Sinyal masukan diambil sebanyak 10 kali untuk setiap kata dan sinyal standar didapatkan dari rata-rata jumlah sinyal masukan. Kata yang diuji ada 5 jenis kata, yaitu "Komputer", "Cantik", "Malang", "Basket", "Juni". Pengambilan sinyal suara dilakukan dalam berbagai kondisi baik dalam kondisi sepi maupun ramai. Hal ini dilakukan agar sistem dapat bekerja lebih baik pada semua keadaan.

Tabel 7. Penutur Standar

No.	Jenis Kelamin	Umur	Kata Standar yang Diucapkan	Daerah Asal
1.	Perempuan	22 Th	Komputer	Surabaya
2.	Perempuan	21 Th	Cantik	Jember
3.	Laki-Laki	22 Th	Malang	Solo
4.	Laki-Laki	23 Th	Basket	Jember
5.	Perempuan	22 Th	Juni	Malang

Pada pengujian dan analisis sistem pengolahan suara secara keseluruhan, keberhasilan sistem ditentukan berdasarkan nilai batas. Nilai batas diambil dari rata-rata jarak sinyal dan rata-rata error. Jika jarak antara sinyal masukan dan sinyal standar sesuai dengan nilai batas maka sinyal masukan dapat dikenali oleh sistem, namun jika jarak antara sinyal masukan dan sinyal standar diluar nilai batas maka sinyal masukan tidak dapat dikenali oleh sistem.



Gambar 18. Diagram Alir Proses Perbandingan 2 Sinyal.

Sumber: Perencanaan

Tabel 8. Rata-Rata Jarak Sinyal Masukan yang Diucapkan Penutur Standar dengan Sinyal Standar.

Kata ke-	Kondisi	Jarak Sinyal "Komputer"	Jarak Sinyal "Cantik"	Jarak Sinyal "Malang"	Jarak Sinyal "Basket"	Jarak Sinyal "Juni"
1.	Sepi	0.144458	0.087258	0.115630	0.128966	0.105043
2.	Sepi	0.118718	0.062014	0.094162	0.115696	0.083804
3.	Sepi	0.121052	0.059582	0.083296	0.093144	0.114385
4.	Sepi	0.103972	0.060363	0.091878	0.129569	0.087136
5.	Sepi	0.144045	0.065933	0.101962	0.122008	0.093869
6.	Ramai	0.095311	0.075106	0.110137	0.142280	0.102762
7.	Ramai	0.085541	0.062375	0.077333	0.120067	0.111722
8.	Ramai	0.087154	0.059417	0.150256	0.139528	0.112024
9.	Ramai	0.114555	0.060204	0.075568	0.130666	0.133870
10.	Ramai	0.115743	0.090127	0.079233	0.104745	0.097423
Rata-Rata		0.113055	0.068238	0.097946	0.122667	0.104204

Sumber: Pengujian

$$\text{var}_x = | \text{Rata-Rata} - \text{Jarak 2 Sinyal} |$$

$$\text{Rata - Rata} \sum \text{var} = \frac{\sum \text{var}}{\text{Jumlah Data}}$$

Perhitungan rata-rata jarak 2 sinyal dapat dilihat pada Tabel 8. Dari Rumus diatas, maka nilai batas dari kata "Komputer" adalah 0.113055 ± 0.016048 . Sehingga tampilan dari penutur ke-1 dapat diakses jika jarak terpendek adalah jarak antara sinyal masukan dan kata standar "Komputer", dengan nilai batas antara 0.097007 s.d 0.129103. Nilai batas dari kata "Cantik" adalah 0.068238 ± 0.009556 . Sehingga tampilan dari penutur ke-2 dapat diakses jika jarak terpendek adalah jarak antara sinyal masukan dan kata standar "Cantik", dengan nilai batas antara 0.058683 s.d 0.077794. Nilai batas dari kata "Malang" adalah 0.097946 ± 0.017241 . Sehingga tampilan dari penutur ke-3 dapat diakses jika jarak terpendek adalah jarak antara sinyal masukan dan kata standar "Malang", dengan nilai batas antara 0.080705 s.d 0.115187. Nilai batas dari kata "Basket" adalah 0.122667 ± 0.011535 . Sehingga tampilan dari penutur ke-4 dapat diakses jika jarak terpendek adalah jarak antara sinyal masukan dan kata standar "Basket", dengan nilai batas antara 0.111132 s.d 0.134202. Nilai batas dari kata "Juni" adalah 0.104204 ± 0.011205 . Sehingga tampilan dari penutur ke-5 dapat diakses jika jarak terpendek adalah jarak antara sinyal masukan dan kata "Juni", dengan nilai batas antara 0.092999 s.d 0.115409. Berdasarkan data diatas, nilai batas maksimal yang digunakan pada sistem pengenalan suara untuk tiap kata berbeda-beda.

Pengujian dan Analisis Sistem Pengenalan Sinyal Suara.



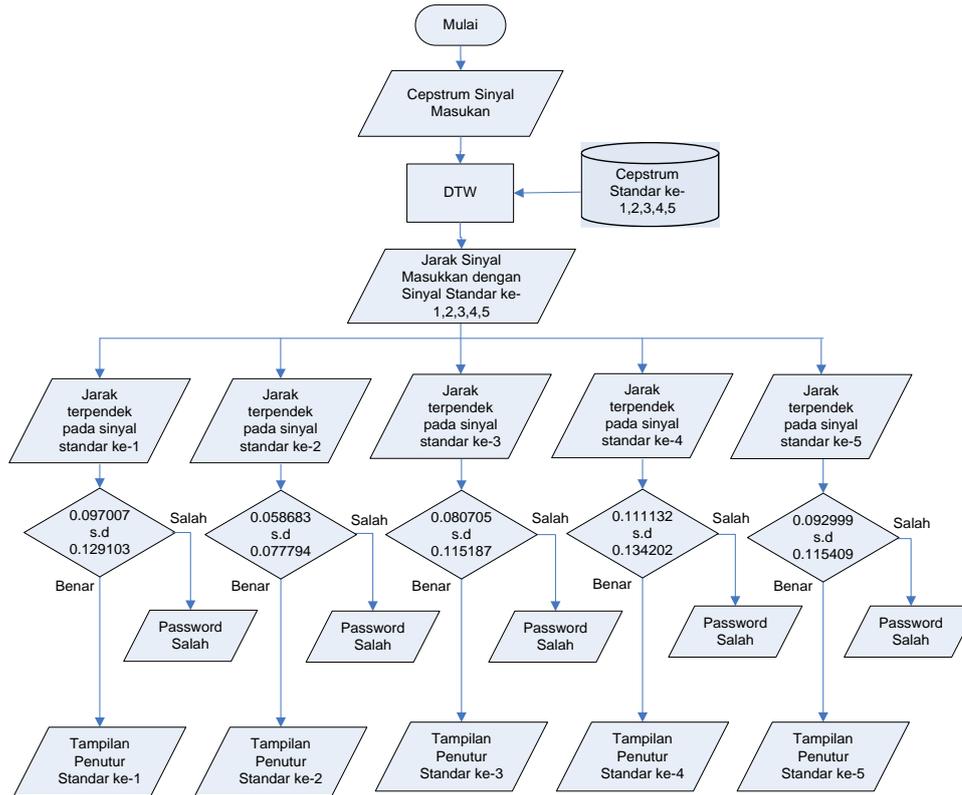
Gambar 19. Tampilan Sistem Pengenalan Suara

Sumber: Hasil Perancangan dengan Menggunakan Tcl/Tk 8.3, Bahasa C++ dan Snack

Tabel 9. Persentase Keberhasilan Sinyal Masukan yang Diucapkan Penutur Standar.

Sinyal Masukan	Kondisi	Penutur ke-	Jumlah Ucapan	Persentase Keberhasilan
Komputer	Sepi	1	5 kali	80%
Komputer	Ramai	1	5 kali	60%
Cantik	Sepi	2	5 kali	80%
Cantik	Ramai	2	5 kali	0%
Malang	Sepi	3	5 kali	60%
Malang	Ramai	3	5 kali	80%
Basket	Sepi	4	5 kali	60%
Basket	Ramai	4	5 kali	60%
Juni	Sepi	5	5 kali	60%
Juni	Ramai	5	5 kali	40%
Rata-Rata Persentase Keberhasilan				58%

Sumber: Pengujian



Gambar 20. Diagram Alir Proses Pengujian Sistem Pengenalan Suara
Sumber: Pengujian

Pengujian dan Analisis Pengenalan Suara dengan Masukan Sinyal Suara yang Diucapkan Oleh Penutur Standar.

Jenis kata yang digunakan pada proses pengujian dan analisis ada 5 jenis, yaitu "Komputer", "Cantik", "Malang", "Basket" dan "Juni". Setiap penutur standar mengucapkan satu jenis kata sebanyak 10 kali dalam kondisi yang bermacam-macam, baik dalam kondisi sepi maupun ramai. Hal ini dilakukan agar sistem dapat bekerja lebih baik pada semua keadaan.

Pada pengujian dan analisis ini, keluaran akan bernilai benar jika daftar nilai hasil studi dapat diakses dan bernilai salah jika daftar nilai hasil studi tidak dapat diakses. Berdasarkan Tabel 9 dapat disimpulkan bahwa persentase keberhasilan sistem pengenalan suara dengan sinyal masukan dari penutur standar adalah 58%.

$$\text{Keberhasilan} = \frac{70\% + 40\% + 70\% + 60\% + 50\%}{5} \\ = 58\%$$

Ketidakberhasilan sistem pengenalan suara dengan masukan sinyal standar yang diucapkan oleh penutur standar disebabkan oleh beberapa hal, yaitu noise lingkungan misalnya dari *Personal Computer* (PC) dan perubahan pola suara penutur yang disebabkan karena perubahan kondisi kesehatan penutur maupun unsur kesengajaan.

Pengujian dan Analisis Pengenalan Suara dengan Masukan Sinyal Suara yang Diucapkan Oleh Penutur Lain.

Jenis kata yang digunakan pada pengujian dan analisis ada 5 jenis, yaitu "Komputer", "Cantik", "Malang", "Basket" dan "Juni". Setiap kata diucapkan sebanyak 5 kali oleh penutur lain yang bukan standar. Pengujian ini dilakukan pada kondisi sepi maupun ramai. Hal ini dilakukan agar sistem dapat bekerja lebih baik pada semua

keadaan. Keluaran bernilai benar jika daftar nilai hasil studi tidak dapat diakses oleh orang lain dan bernilai salah jika daftar nilai hasil studi dapat diakses oleh orang lain.

Tabel 10. Keberhasilan Sistem Pengenalan Suara dengan Masukan berupa kata “Komputer” yang Diucapkan Oleh Penutur Lain

Sinyal Masukan	Penutur ke-	Jumlah Ucapan	Persentase Keberhasilan
Komputer	2	5 kali	100%
Komputer	3	5 kali	100%
Komputer	4	5 kali	100%
Komputer	5	5 kali	100%
Rata-Rata Persentase Keberhasilan			100%

Tabel 11. Keberhasilan Sistem Pengenalan Suara dengan Masukan berupa kata “Cantik” yang Diucapkan Oleh Penutur Lain

Sinyal Masukan	Penutur ke-	Jumlah Ucapan	Persentase Keberhasilan
Cantik	1	5 kali	100%
Cantik	3	5 kali	100%
Cantik	4	5 kali	100%
Cantik	5	5 kali	100%
Rata-Rata Persentase Keberhasilan			100%

Tabel 12. Keberhasilan Sistem Pengenalan Suara dengan Masukan berupa kata “Malang” yang Diucapkan Oleh Penutur Lain

Sinyal Masukan	Penutur ke-	Jumlah Ucapan	Persentase Keberhasilan
Malang	1	5 kali	100%
Malang	2	5 kali	100%
Malang	4	5 kali	80%
Malang	5	5 kali	100%
Rata-Rata Persentase Keberhasilan			95%

Tabel 13. Keberhasilan Sistem Pengenalan Suara dengan Masukan berupa kata “Basket” yang Diucapkan Oleh Penutur Lain

Sinyal Masukan	Penutur ke-	Jumlah Ucapan	Persentase Keberhasilan
Basket	1	5 kali	100%
Basket	2	5 kali	100%
Basket	3	5 kali	100%
Basket	5	5 kali	100%
Rata-Rata Persentase Keberhasilan			100%

Tabel 14. Keberhasilan Sistem Pengenalan Suara dengan Masukan berupa kata “Juni” yang Diucapkan Oleh Penutur Lain

Sinyal Masukan	Penutur ke-	Jumlah Ucapan	Persentase Keberhasilan
Juni	1	5 kali	100%
Juni	2	5 kali	100%
Juni	3	5 kali	100%
Juni	5	5 kali	100%
Rata-Rata Persentase Keberhasilan			100%

Berdasarkan hasil pengujian diatas, tingkat keberhasilan sistem pengenalan suara dengan masukan sinyal suara yang diucapkan oleh penutur lain adalah 99%.

$$\text{Keberhasilan} = \frac{100\% + 100\% + 95\% + 100\% + 100\%}{5}$$

$$= 99\%$$

Ketidakterhasilan sistem pengenalan suara dengan masukan sinyal standar yang diucapkan oleh penutur lain yang bukan standar dapat disebabkan oleh beberapa hal, yaitu noise lingkungan misalnya dari *Personal Computer* (PC), perubahan pola suara penutur yang diakibatkan oleh perubahan kondisi kesehatan penutur maupun unsur kesengajaan, serta kesamaan pola suara yang disebabkan jenis kelamin penutur lain yang mengucapkan sinyal masukan dan penutur standar sama.

KESIMPULAN

1. Sinyal suara diolah dengan metode LPC (*Linier Predictive Coding*) dan pengenalan suara dengan metode DTW (*Dynamic Time Warping*). Peralatan yang digunakan adalah mikropon, *speaker*, *sound card*, PC (*Personal Computer*) serta bahasa pemrograman Tcl/Tk, *snack* dan bahasa C++.
2. Persentase keberhasilan sistem pengenalan suara dengan sinyal masukan yang diucapkan oleh pengucap standar adalah 58%, sedangkan sinyal masukan yang diucapkan oleh pengucap lain yang bukan standar adalah 99%.
3. Tingkat keberhasilan sistem pengenalan suara dipengaruhi oleh beberapa faktor, diantaranya kondisi kesehatan, pola suara, tipe hardware, keadaan lingkungan.

SARAN

Penggunaan algoritma yang lebih baik, pengambilan sampel suara yang lebih banyak dari berbagai macam kondisi, penggunaan hardware yang berkualitas tinggi, otomatisasi sistem, penggunaan pengenalan suara untuk independent speaker dalam aplikasi lain.

DAFTAR PUSTAKA

- Anonim. *Discrete Fourier Transform dan FFT*: Student.ee.itb.ac.id.
- Anonim. *Window Function*:Wikipedia.
- Airlie, David Mel. 1997. *A Voice Controlled Window Manager for the X-Windowing System*. Ireland
- Armand, Arry Akhmad. 2004. *Teknologi Pemrosesan Bahasa Alami sebagai Teknologi Kunci untuk Meningkatkan Cara Interaksi antara Manusia dengan Mesin*. Bandung.
- Armand, Arry Akhmad. *Proses Pembentukan dan Karakteristik Sinyal Ucapan*. Bandung
- Bristow, Geoff.1984. *Electronic Speech Synthesis, Techniques, Technology and Applications*. McGRAW-HILL Book Company. New York, USA.
- Dewantara, Bima Sena Bayu. 2004. *Aplikasi Pengenalan Wicara untuk Perintah Nirkabel Robot Mikro Mouse*. Surabaya.
- Furui, Sadaoki. 1989. *Digital Speech Processing Synthesis and Recognition*. Marcel Dekker, Inc. New York, USA.
- Kondoz, A. M. 2004. *Digital Speech, Coding for Low Bit Rate Communication System*. John Wiley and Sons, Ltd.
- Lawrence Rabiner, Biing Hwang Juang. 1993. *Fundamentals of Speech Recognition*. Prentice Hall International Inc.
- Proakis, John G. 1995. *Pemrosesan Sinyal Digital*. Pearson Education Asia Prentice Hall.

-
- Robinson, Tony. *Practical Application of the Short Term Fourier Transform*.
- Salivahanan, S, Valavaraj, A, Gnanapriya, C. 2001. *Digital Signal Processing*. Mcgraw-Hill International Edition. Singapore.
- Sunarto, Agus. 2003. *Pengenalan Suara Manusia dengan Metode LPC dan Jaringan Saraf Tiruan Propagasi Balik*. Malang.
- Untung, Nyaris. 2000. *Voice dialing Telepon Rumah dengan Komputer Pribadi*. Malang.
- Upperman, Gina. 2004. *Linear Predictive Coding in Voice Conversion*. US/Central.