

Aplikasi Teknik Speech Recognition pada Voice Dial Telephone



Oleh:
Ahmad Irfan
Abdul Rahman
Tri Budi Santoso
Titon Dutono

Laboratorium Sinyal, Politeknik Elektronika Negeri Surabaya (PENS)
Institut Teknologi Sepuluh Nopember (ITS)
E-mail: tribudi@eepis-its.edu



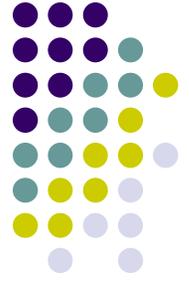
1. Pendahuluan:

Speech Recognition ?

Merupakan salah satu dari aplikasi proses *digital signal processing* (pengolahan sinyal digital)

Aplikasi Speech Recognition

- *voice dialing*
- *banking by telephone*
- *telephone shopping*
- *data base access service*
- *security control*



Penelitian ...

- Oleh Min Do [1] telah dikembangkan sebuah mini project *Automatic Speaker Recognition System*, simulai *off-line* menggunakan perangkat lunak Matlab.
- Darren Ellis [2], sebuah rancangan speaker recognition. Disini simulasi pengkondisian lingkungan disajikan dalam bentuk munculnya noise gaussian.
- Maria Ulfa [3] telah berhasil membangun sebuah aplikasi sederhana pada layanan perbankan berbasis *voice recognition*.
- Paper [4] dan [5] disajikan bentuk aplikasi teknologi digital signal processing secara real time.



Pada paper ini ...

- membangun sebuah *voice dial telephone* menggunakan perintah sinyal wicara.
- sistem yang dibangun masih pada tingkatan *independet speaker*

Tujuan:

Secara keseluruhan tujuan penelitian ini adalah untuk membangun sebuah *Modul Praktikum Speech Processing* di Laboratorium Sinyal, PENS-ITS.

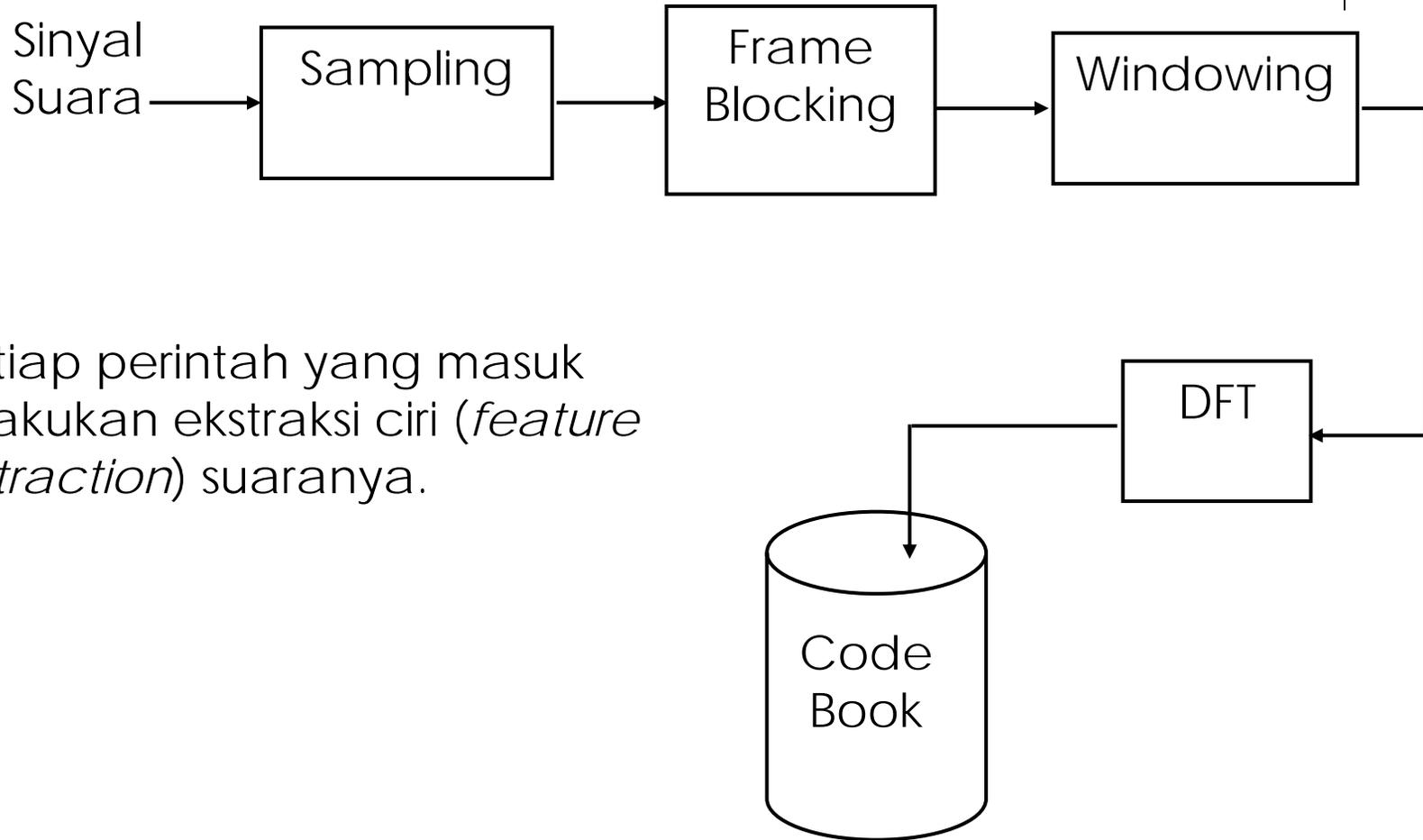
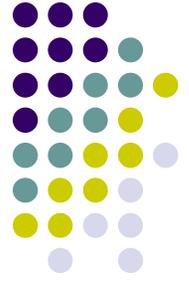
2. Prinsip Pengenal Wicara



Perancangan sistem pengolah wicara:

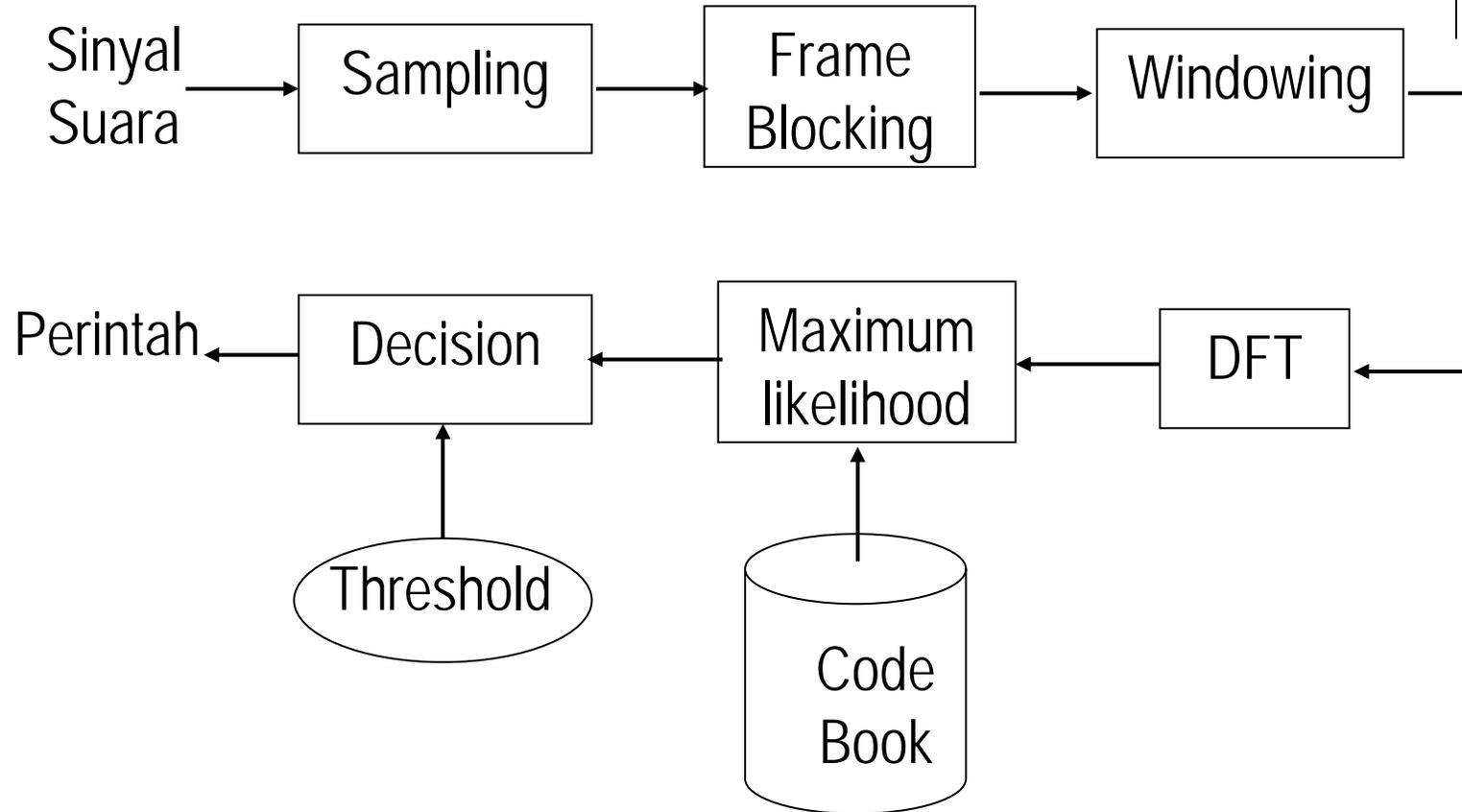
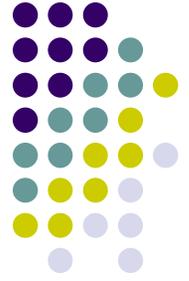
1. Fase pertama adalah pada pembuatan sebuah sistem pengenal wicara adalah *enrollment session* (fase pembelajaran) atau yang dikenal juga sebagai *training phase*.
2. Sedangkan fase kedua adalah *operating session* atau *testing phase* (fase pengujian).

Fase Pembelajaran



Setiap perintah yang masuk dilakukan ekstraksi ciri (*feature extraction*) suaranya.

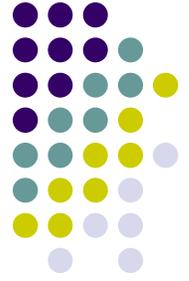
Fase Pengujian





- Fase pengujian diawali dengan ekstraksi ciri suara (perintah) yang masuk.
- Ciri dari suara yang masuk selanjutnya di cocokkan dengan yang ada pada *code book*. Ini juga dikenal sebagai proses *feature matching* (pencocokan ciri).
- Proses matching menggunakan algoritma *maximum likelihood*, yaitu ciri perintah yang masuk dicari yang paling mirip dengan ciri perintah yang ada pada *code book*

Teori Pendukung...



- **Proses Sampling**

Nyquist mempersyaratkan frekuensi sampling minimal dua kali frekuensi sinyal.

$$f_s \geq 2 \times f_{\max}$$

dimana:

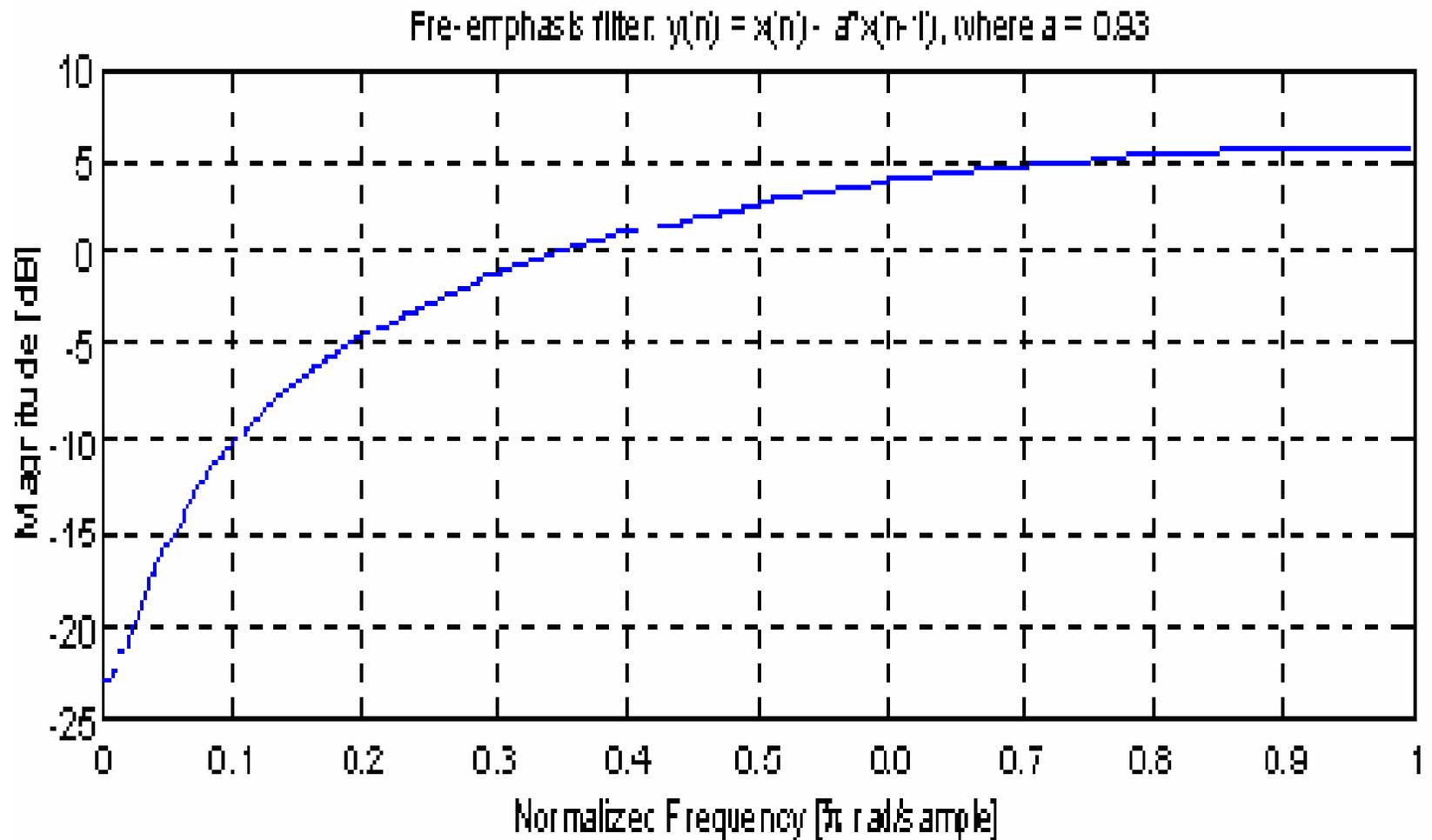
f_s : frekuensi sampling

f_{\max} : frekuensi tertinggi sinyal yang disampel

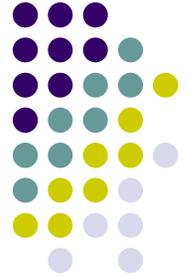


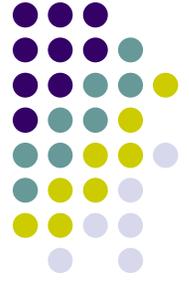
• Pre Emphasis

Spectrum sinyal wicara akan diperhalus dengan filter pre-emphasis $1-az^{-1}$, dengan nilai $a=0,95$



- Frame Blocking dan Windowing





• Discrete Fourier Transform (DFT)

Untuk mendapatkan ciri sinyal dalam domain frekuensi

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-jk\omega_0 n} \quad 0 \leq k \leq N-1$$

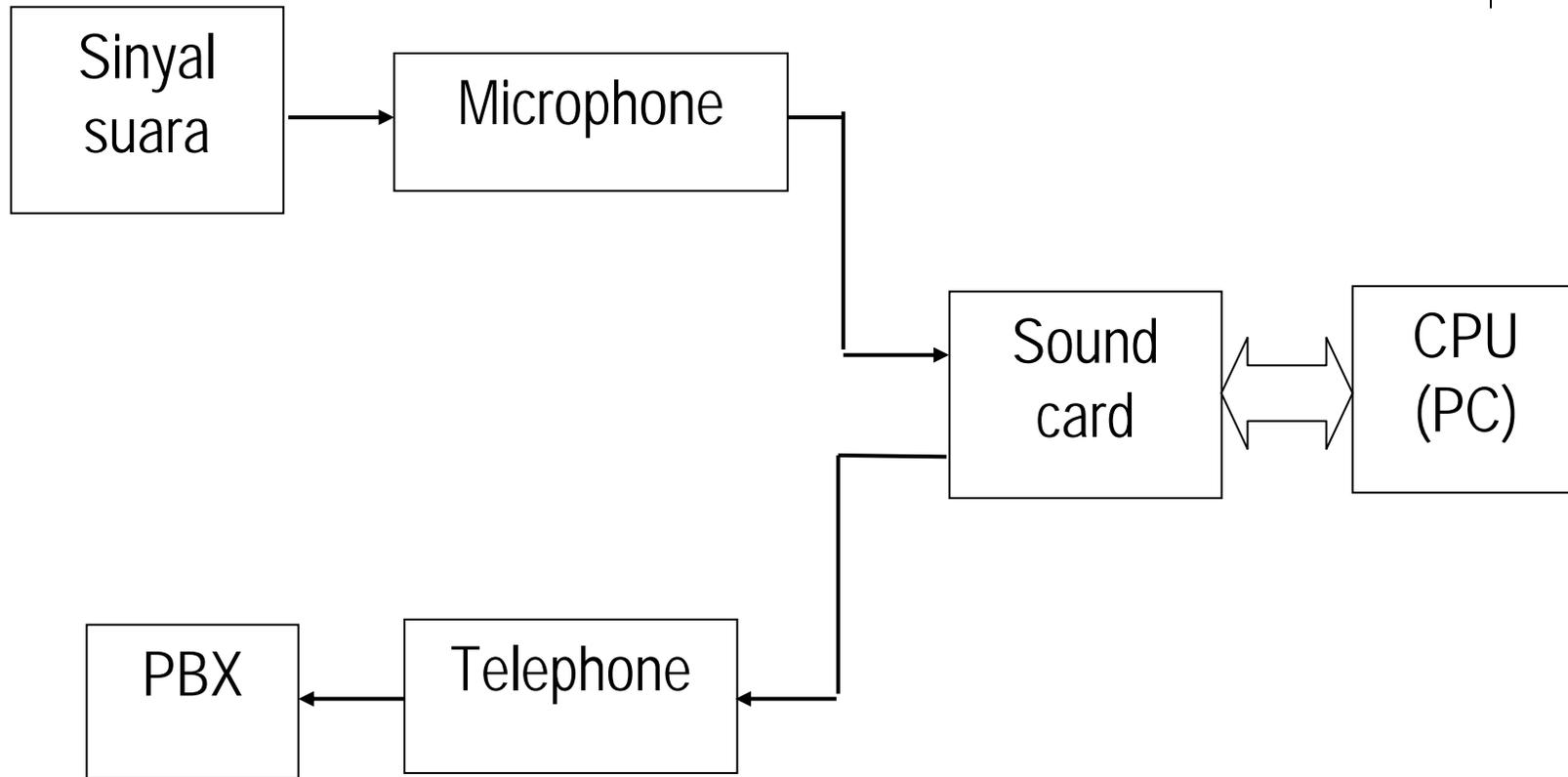
dimana:

$$\begin{aligned} \omega_0 &= \text{frekuensi fundamental} \\ &= 2\pi/\text{sampling rate} \\ &= 2\pi/N \end{aligned}$$

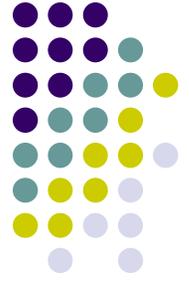
• Maximum Likelihood

Maximum likelihood atau *maximum of posterior* merupakan prosedur untuk mendapatkan nilai atau parameter yang memiliki bentuk terdekat dengan cara membandingkan suatu parameter masukan dengan parameter pilihan tersedia.

3. Perancangan Sistem Voice Dial Telephone



Untuk fase pembelajaran dan pengujian



- **Kerja Sistem Fase Pembelajaran**
- Suara dari pembicara akan ditangkap microphone
- Sinyal elektrik disampling dan dikonversi ke dalam bentuk sinyal digital oleh sound card.
- Dilanjutkan dengan Pre-emphasis, frame blocking, dan windowing untuk menghaluskan spektral sinyal.
- Setiap sinyal dicirikan oleh bentuk spektral frekuensinya dan disipan sebagai code book untuk masing-masing jenis perintah.



• Kerja Sistem Fase Pembelajaran

- fase pengujian proses dari sampling sampai ekstraksi ciri sama dengan yang terjadi pada fase pembelajaran.
- Setelah ciri sebuah sinyal uji didapatkan dilanjutkan dengan proses matching dengan ciri-ciri sinyal perintah yang tersipan dalam code book.
- Dengan menggunakan algoritma maximum likelihood bentuk perintah yang masuk akan dapat diterjemahkan.

4. Analisa Hasil Pengujian



mode_test.tcl

PERANCANGAN VOICE DIAL TELEPHONE MENGGUNAKAN TEKNIK SPEECH RECOGNITION
OLEH
ACHMAD IRFAN DAN ABDUL RAHMAN

Mode Test



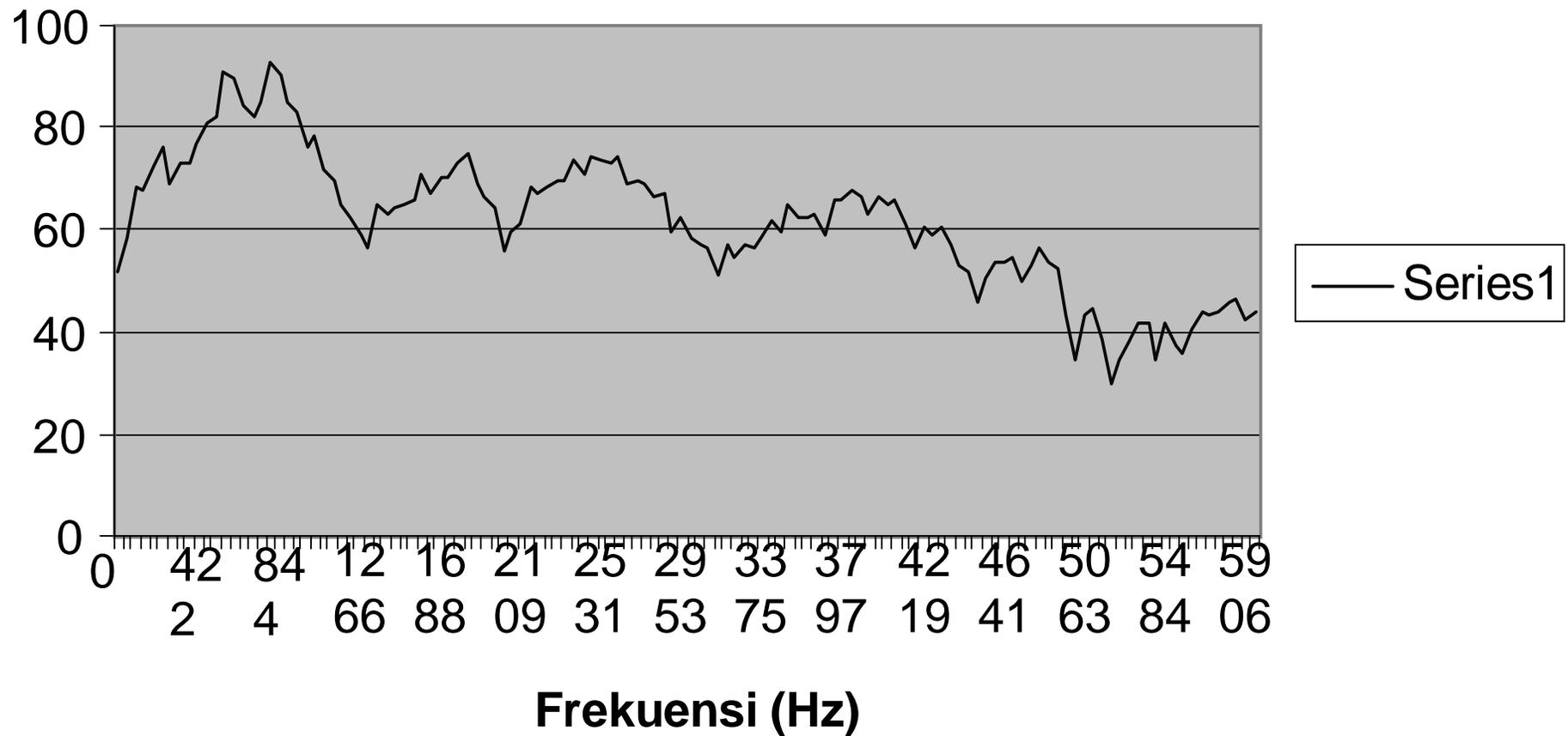
0.000 sec
Status Ready

The image shows a graphical user interface for a voice dial telephone system. It features a central black rectangular area representing a video feed, flanked by two vertical bar graphs. To the left is a microphone icon, and to the right is a rotary telephone icon. The status at the bottom indicates '0.000 sec' and 'Status Ready'. The window title is 'mode_test.tcl' and the title bar contains standard window controls.

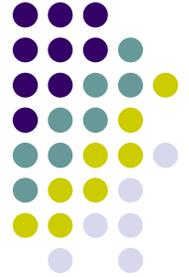
Hasil DFT Sinyal Input



DFT

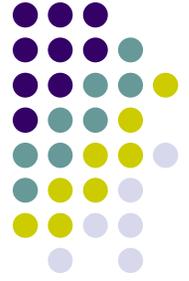


Pentuan Jenis Perintah dengan Maximum Likelihood



Konsep ini dapat direalisasikan dalam bentuk persamaan dasar sederhana yang kita kenal sebagai *least squares error* (LSE).

Secara formal LSE dapat dibentuk dengan mencari nilai *sum squares error* (SSE) dari dua nilai masukan terhadap nilai standar yang digunakan untuk menguji.



Formulasinya:

$$SSE(X) = \sum_{k=1}^m (X_k - c(X))^2$$

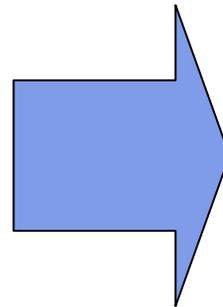
Dimana:

X_k = nilai ciri kata yang masuk dalam domain frekuensi

$c(X)$ = nilai ciri dari kata standar dalam code book

contoh pengujian kata 'optik' yang masuk didapatkan hasil seperti berikut:

- Optik : 0.220744
- Multimedia : 1.076151
- Telkom : 0.710211
- Jarkom : 0.850238
- Digital : 1.049321

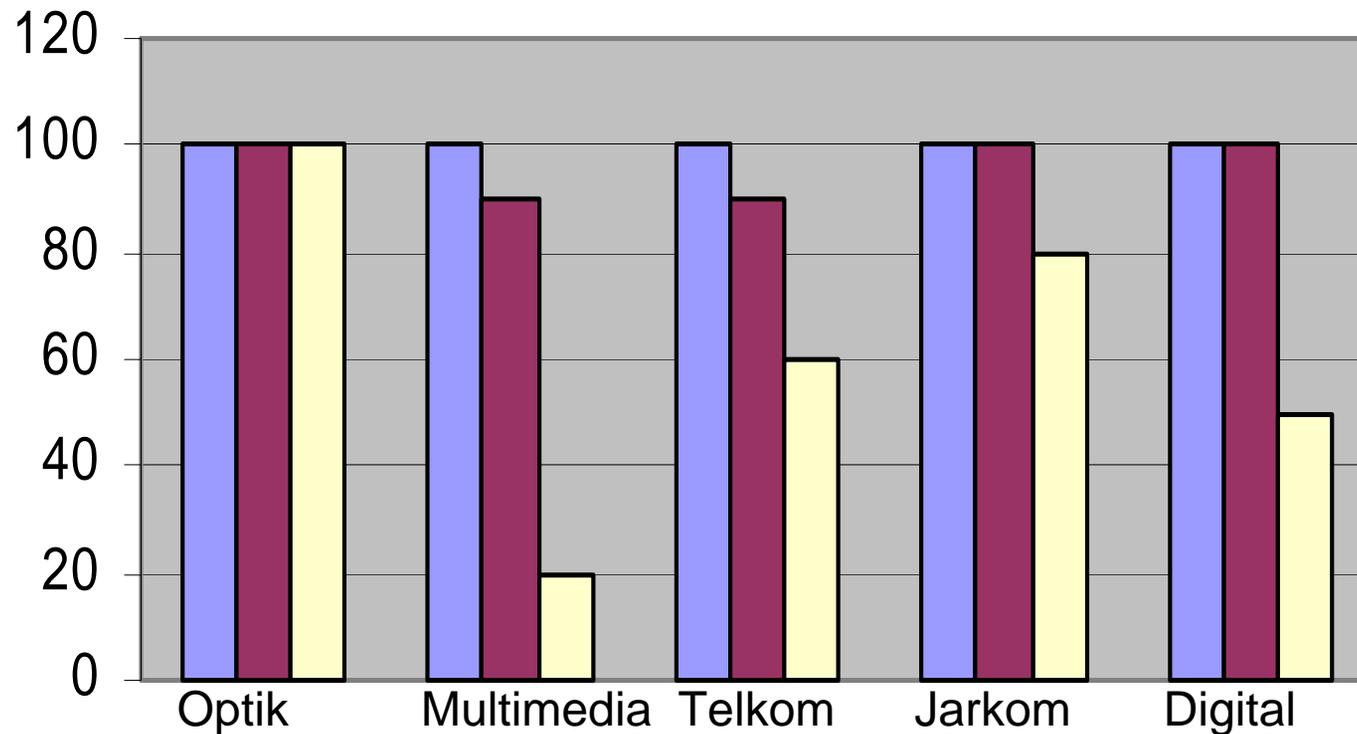
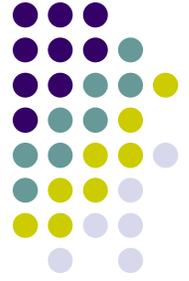


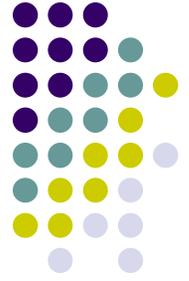
kata "optik" , karena memiliki nilai error terkecil yaitu 0.220744.

Hasil Pengujian Sistem

3 pengujian,

- pertama diambil sampel data dari orang pria yang sama
- kedua orang pria yang berbeda
- ketiga adalah wanita.

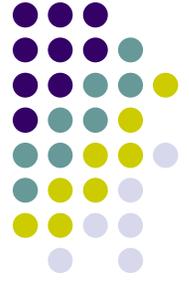




5. Kesimpulan

- sistem yang telah dirancang dengan menggunakan sistem pengucap bebas ini memiliki prosentase keberhasilan yang tinggi
- Diharapkan untuk melakukan penelitian lanjut dengan menggunakan sistem pengucap tak bebas

Referensi:



- [1] Min N. Do, " *An Automatic Speaker Recognition System* ", Audio Visual Communication Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland, 2001.
- [2] Darren Ellis, " *Design of a Speaker Recognition Code using Matlab* ", Department of Computer and Electrical Engineering, University of Tennessee, 2001.
- [3] Maria Ulfa, Prima K, Titon Dutono, " *Aplikasi Speech Recognition sebagai Fungsi Layanan pada Sistem Perbankan* ", Seminar SNTE, Jogjakarta 2003.
- [4] Bima Sena, Linda Indra, Titon Dutono, " *Aplikasi Pengenalan Wicara untuk Perintah pada Micromouse Robot Nirkabel* ", Seminar IES 2004.
- [5] Anas, Wahyudin, Tri Budi, Titon Dutono, " *Pengenalan Nada-Nada Tunggal pada Gitar dengan Menggunakan Teknik Pengolahan Sinyal Digital* ", EEPIS Journal, Volume 10, Juli 2005.
- [6] Titon Dutono, Moh Nuh, " *Dasar-Dasar Pengolahan Sinyal* ", EEPIS Press 2000.
- [7] Rabiner, Juang, " *An Introduction to Speech Recognition* ", Prentice Hall, USA, 1993.