

2018

Identifikasi Voice Menggunakan Jaringan Syaraf Tiruan dan Mel Frequency Cepstrum Coefficients

Simamora, Corianti Gabeanna Maria

Universitas Sumatera Utara

<http://repositori.usu.ac.id/handle/123456789/12169>

Downloaded from Repositori Institusi USU, Universitas Sumatera Utara

**IDENTIFIKASI VOICE MENGGUNAKAN JARINGAN SYARAF
TIRUAN DAN MEL FREQUENCY CEPSTRUM COEFFICIENTS**

TESIS

Oleh :

**CORIAN TI G.M SIMAMORA
147034008/MTE**



**FAKULTAS TEKNIK
UNIVERSITAS SUMATERA UTARA
MEDAN
2018**

**IDENTIFIKASI VOICE MENGGUNAKAN JARINGAN SYARAF
TIRUAN DAN MEL FREQUENCY CEPSTRUM COEFFICIENTS**

TESIS

Untuk Memperoleh Gelar Magister Teknik
Dalam Program Studi Magister Teknik Elektro
Universitas Sumatera Utara

Oleh :

**CORIAN TI G.M SIMAMORA
147034008/MTE**

**FAKULTAS TEKNIK
UNIVERSITAS SUMATERA UTARA
MEDAN
2018**

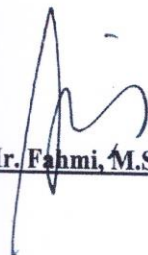
Judul Tesis : Identifikasi *Voice* Menggunakan Jaringan Syaraf Tiruan dan *Mel Frequency Cepstrum Coefficients* 2018


Nama Mahasiswa : Corianti Gabeanna Maria Simamora

Nomor Induk Mahasiswa : 147034008


Program Studi : Magister Teknik Elektro

Menyetujui
Komisi Pembimbing


(Dr. Ir. Fahmi, M.Sc. IPM)

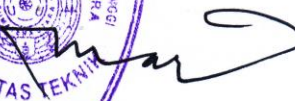

(Dr. Maksum Pinem, ST. MT)

Ketua Program Studi,


(Suherman, Ph.D)



Dekan,


(Ir. Seri Maulina, M.Si., Ph.D)

Telah Lulus : 15 Maret 2018

Telah Lulus : 15 Maret 2018

Diuji dan dinyatakan lulus

Tanggal : 15 Maret 2018

PANITIA PENGUJI TESIS

Ketua : Dr. Ir. Fahmi, M.Sc.IPM

**Anggota : 1. Dr. Maksum Pinem, ST.MT
2. Ir. Surya Hardi, MS. Ph.D
3. Ir. Syafruddin HS, MS. Ph.D**

ABSTRAK

Perkembangan teknologi di bidang pengenalan yang semakin pesat membutuhkan daya pikir, kreatifitas dan inovasi untuk meningkatkan sistem teknologi *modern* pada bidang ilmu pengetahuan khususnya bidang *access*. Ilmu yang membahas bidang pengenalan suara dikenal dengan nama pemrosesan sinyal suara digital dan *Natural Language Processing*. Untuk menyelesaikan permasalahan pengenalan suara *voice identification* menggunakan pemrosesan sinyal informasi yang menggunakan proses *Mel Frequency Cepstrum Coefficients* (MFCC) dan menggunakan Jaringan syaraf tiruan (JST) *backpropagation* dengan proses pembelajaran perubahan bobot. Siklus perubahan bobot atau EPOCH (Exponential Decay) digunakan sebagai inisialisasi pengidentifikasian suara *voice* pria dan wanita yang diperoleh dengan menggunakan beberapa sampel suara pria dan wanita. Hasil pengenalan *voice identification* dari gender wanita lebih baik dibandingkan pria dengan nilai error pada pendeteksian suara *voice* wanita $0.05 < \text{Error} < 1$ lebih kecil dari siklus perubahan bobot jaringan syaraf tiruan *backpropagation* dengan konstanta laju pelatihan ($\alpha=0.2$).

Kata Kunci : Pengenalan suara *voice*, *Mel Frequency Cepstrum Coefficients* (MFCC), Jaringan syaraf tiruan (JST), *Backpropagation*.

ABSTRACT

The development of information technology requires the power of thinking, creativity and innovation of progressing the increasingly varied life systems to improve modern system technology and sciences of access. The development of voice is known speech and natural language processing. Voice becomes communication medium for human and computer (machine) is used for recognition/identification and access system. One of its application is speech to text application. Some of speech to text research already claimed good accuracy for some languages or voice namely for Robotic, furniture, digital signal processing and biometric technology. For managing the recognition, we used several approaches in this research. The researcher performs information signal processing by using Mel Frequency Cepstrum Coefficients (MFCC) and using artificial neural network (ANN) backpropagation with learning process through weight change. The weight change cycles or EPOCH (Exponential Decay) are used as initialization of voice identification of male and female voice obtained by using multiple sound. The result of voice identification of female gender is better than man with error on detection of female voice $0,05 < \text{error} < 1$ smaller than cycle change of artificial neural network backpropagation with $\alpha = 0.2$.

Keywords : Indonesian's voice recognition, Mel Frequency Cepstrum Coefficients (MFCC), Aritificial Neural Network (ANN), Backpropagation

KATA PENGANTAR

Puji syukur penulis ucapkan kehadiran Tuhan Yang Maha Esa atas Berkat-Nya penulis dapat menyelesaikan Tesis ini dengan judul “**Identifikasi Voice Menggunakan Jaringan Syaraf Tiruan dan Mel Frequency Cepstrum Coefficients** ” dalam rangka memenuhi persyaratan menyelesaikan pendidikan Program Studi S-2 Fakultas Teknik Universitas Sumatera Utara.

Penulis menyadari dalam penulisan tesis ini banyak mengalami kesulitan, akan tetapi berkat bantuan dari berbagai pihak akhirnya penulis dapat menyelesaikan Tesis ini, untuk itu dalam kesempatan yang berbahagia ini penulis menyampaikan rasa hormat dan terima kasih kepada bapak/ibu :

1. Prof. Dr. Runtung, S.H., M.Hum, selaku Rektor Universitas Sumatera Utara yang telah memberikan kesempatan kepada penulis untuk mengikuti pendidikan pada Program Studi S2 Teknik Elektro Komputer Fakultas Teknik Elektro Universitas Sumatera Utara.
2. Ir. Seri Maulina, M.Si., Ph.D, selaku Dekan Fakultas Teknik Elektro Universitas Sumatera Utara.
3. Suherman, Ph.D, selaku Ketua Program Studi S2 Teknik Elektro Komputer Fakultas Teknik Universitas Sumatera Utara dan Arman Sani, selaku Sekretaris Program Studi S2 Teknik Elektro Komputer Fakultas Teknik Universitas Sumatera Utara serta jajarannya yang telah memberikan bimbingan dan arahan selama penulis mengikuti pendidikan.

4. Dr.Ir. Fahmi, M.Sc. IPM, selaku pembimbing I saya yang telah banyak memberikan waktunya dalam membimbing saya selama penulisan Tesis ini.
 5. Dr. Maksum Pinem, ST.MT, selaku pembimbing II saya yang telah banyak memberikan waktunya dalam membimbing saya selama penulisan Tesis ini.
 6. Ir. Syafruddin HS, MS. Ph.D, selaku penguji I saya yang juga telah banyak memberikan arahan dan masukan dalam menyelesaikan Tesis ini.
 7. Ir. Surya Hardi, MS. Ph.D, selaku penguji II saya yang telah banyak memberikan waktunya dalam membimbing saya selama penulisan Tesis ini.
 8. Seluruh Staff dan Dosen Program Studi S2 Teknik Elektro Komputer Fakultas Teknik Universitas Sumatera Utara yang secara langsung banyak memberikan ilmu kepada penulis selama menjalani pendidikan.
 9. Kepada semua keluarga tercinta, Orang Tua Drs. H. Simamora dan N.Naibaho dan saudara M.Simamora, K.Simamora, O.Simamora, T.Simamora atas segala doa dan dukungannya dalam memberi materi, semangat dan perhatiannya kepada penulis dalam penyusunan Tesis ini.
- Semoga Tesis ini bermanfaat bagi pengembangan ilmu pengetahuan dalam bidang Teknik Elektro Komputer. Akhir kata penulis ucapkan terima kasih.

Medan, Mei 2018
Penulis

Corianti GM Simamora
147034008

DAFTAR RIWAYAT HIDUP

Corianti Gabeanna Maria Simamora ST. MT lahir pada tanggal 26 Febuari 1986 di Pematang Siantar, anak ke 3 dari 5 bersaudara dari pasangan ayahanda H.Simamora dan ibunda N.Naibaho. Pengalaman kerja PT. Bank PANIN Tbk Jakarta , PT. Indosat Tbk Jakarta dan Dosen MBP DAN ISTP Medan.

Pendidikan formal penulis dimulai dari pendidikan di Sekolah Dasar Cinta Rakyat Pematang Siantar, Sekolah Lanjutan Tingkat Pertama Budi Mulia Pematang Siantar, Sekolah Lanjutan Tingkat Atas Budi Mulia Pematang Siantar, Pendidikan D-III Teknik Elektro Elektronika Politeknik Negeri Medan, Pendidikan S1 Teknik Elektro Telekomunikasi Institut Teknologi Sepuluh November Surabaya. Penulis mengikuti pendidikan lanjutan di Program Studi S2 Teknik Elektro Komputer di Fakultas Teknik Elektro Universitas Sumatera Utara.

DAFTAR ISI

	Halaman
ABSTRAK	ii
ABSTRACT	iii
KATA PENGANTAR.....	iv
DAFTAR RIWAYAT HIDUP	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xiii
 BAB 1. PENDAHULUAN	 1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
 BAB 2. TINJAUAN PUSTAKA	 6
2.1 <i>State of Art</i> Pemrosesan Suara dan Bahasa.....	6
2.2 Landasan Teori	9
2.2.1 Pengenalan Suara	13
2.2.2 Karakteristik Telinga	14
2.2.3 Proses Pendengaran	15
2.3 Pengolahan Sinyal Digital (PSD)	16
2.3.1 Representasi Sinyal Digital.....	18
2.3.2 Pengolahan Sinyal	20
2.4 Sistem Pengenalan Bahasa (<i>Speech Recognition</i> <i>System</i>)	24
2.4.1 Struktur Suku Kata dalam Bahasa Indonesia	25
2.4.2 Transformasi Sinyal menjadi Informasi	26
2.5 Sinyal	26
2.5.1 Sinyal Multichannel dan Sinyal Multidimensional	28
2.5.2 Sinyal Waktu Kontinu dan Sinyal Waktu Diskret	28
2.5.3 Sinyal <i>Continuous Valued</i> dan <i>Discete Valued</i>	29
2.5.4 Sinyal Deterministic dan Sinyal Random	29
2.6 Ukuran Kualitas Sinyal	30
2.6.1 Pembacaan Sinyal	31
2.7 Ekstraksi Ciri	32
2.7.1 <i>Mel Frequency Cepstrum Coefficients</i> (MFCC)	33
2.7.2 <i>Mel Frequency Wrapping</i> (MFW)	43

2.7.3	<i>Dynamic Time Warping (DTW)</i>	44
2.7.4	<i>Discrete Cosine Transform (DCT)</i>	45
2.8	Jaringan Syaraf Tiruan (JST)	46
2.8.1	Sifat Jaringan Syaraf Tiruan (JST)	49
2.8.2	Jaringan Syaraf Biologis Manusia	49
2.8.3	Karakteristik Jaringan Syaraf Tiruan	50
2.8.4	Konsep Dasar Jaringan Syaraf Tiruan	51
2.8.5	Istilah-istilah Jaringan Syaraf Tiruan	52
2.8.6	Aplikasi Jaringan Syaraf Tiruan	54
2.8.7	<i>Multi Layer Perceptron</i>	55
2.9	Backpropagation	56
2.9.1	Arsitektur Jaringan Backpropagation	58
2.9.2	Algoritma	59
BAB 3.	METODE PENELITIAN.....	63
3.1	Langkah-Langkah Penelitian	63
3.1.1	Pemrosesan Pengambilan data dan Pemrosesan Suara	64
3.2	Diagram Alir Pemrosesan Suara	65
3.2.1	Proses <i>Mel Frequency Cepstrum Coefficients</i> (MFCC).....	67
3.2.2	Proses Jaringan Syaraf Tiruan Backpropagation	69
3.3	Evaluasi Penelitian <i>Mel Frequency Cepstrum Coefficients</i> (MFCC).....	72
3.3.1	<i>Pre-Emphasize</i>	72
3.3.2	<i>Frame Blocking</i>	73
3.3.3	<i>Windowing</i>	74
3.3.4	<i>Fast Fourier Transform (FFT)</i>	75
3.3.5	<i>Mel Frequency Wrapping</i>	76
3.3.6	<i>Transformasi Kosinus</i>	77
3.4	Evaluasi Penelitian Jaringan Syaraf Tiruan <i>Backpropagation</i>	78
BAB 4.	HASIL DAN PEMBAHASAN.....	79
4.1	Analisa Proses Ekstraksi Ciri <i>Mel Frequency Cepstrum Coefficients</i>	80
4.1.1	Hasil Koefisien <i>Mel Frequency Cepstrum Coefficients</i> (MFCC).....	87
4.2	Analisa Proses Jaringan Syaraf Tiruan <i>Backpropagation</i>	90
4.2.1	Hasil Pelatihan Jaringan Syaraf Tiruan (<i>Backpropagation</i>)	91
4.3	Analisa pengaruh nilai α terhadap Perubahan Bobot Jaringan Syaraf Tiruan <i>Backpropagation</i>	93

BAB 5. KESIMPULAN DAN SARAN.....	102
6.1 Kesimpulan	102
6.2 Saran	103
DAFTAR PUSTAKA	105
LAMPIRAN	107

DAFTAR TABEL

Nomor	Judul	Halaman
4.1	<i>Vector</i> Ciri suara wanita “Apa kabar”	87
4.2	<i>Vector</i> Ciri Suara Pria dengan jumlah koefisien 6 dan 8	89
4.3	<i>Vector</i> Ciri Suara Wanita dengan jumlah koefisien 6 dan 8	90
4.4	Hasil Perubahan bobot dengan 6 koefisien ciri	93
4.5	Hasil Perubahan bobot dengan 8 koefisien ciri	93
4.6	Inisialisasi Perubahan Bobot dengan Jumlah Koefisien ($C_n=6$).....	95
4.7	Inisialisasi Perubahan Bobot dengan Jumlah Koefisien ($C_n=8$)	96
4.8	Inisialisasi Perubahan Bobot dengan Jumlah Koefisien ($C_n=8$)	99
4.9	Inisialisasi Perubahan Bobot dengan Jumlah Koefisien ($C_n=8$)	100

DAFTAR GAMBAR

Nomor	Judul	Halaman
2.1	Organ-organ produksi suara	10
2.2	Organ Telinga.....	14
2.3	Proses Sampling	18
2.4	Alur Proses Transformasi Sinyal Suara Analog <i>voice identification</i> ..	21
2.5	Proses <i>Frame Blocking</i> dan <i>Windowing</i>	32
2.6	Node sederhana.....	47
2.7	Topologi Jaringan Backpropagation.....	59
3.1	Skema Alur Penelitian Identifikasi suara <i>Voice</i>	64
3.2	Proses Pengambilan data/suara	65
3.3	Skema Alur Penelitian <i>Voice Identification</i>	66
3.4	Diagram alir Penelitian menggunakan metode MFCC	67
3.5	Proses Pelatihan Jaringan Syaraf Tiruan	69
3.6	Model Tiruan Neuron.....	70
3.7	Topologi Jaringan <i>Backpropagation</i>	71
3.8	<i>Script</i> program <i>Preemphasize</i>	73
3.9	<i>Script</i> program <i>Blocking</i>	74
3.10	<i>Script</i> program <i>Hamming Window</i>	75
3.11	<i>Script</i> program <i>FFT</i>	76
3.12	<i>Script</i> program <i>Mel Frequency</i>	77
3.13	<i>Script</i> program <i>Discrete Cosine Transform</i>	78
3.14	<i>Script</i> program Jaringan Syaraf Tiruan	79

4.1	Proses <i>preemphasis</i>	82
4.2	Proses <i>Blocking</i> Sinyal Suara	83
4.3	Proses <i>Hamming Window</i>	84
4.4	<i>Fast Fourier Transform</i> (FFT)	85
4.5	Proses <i>Mel Filterbank</i>	86
4.6	Proses <i>Discrete Cosine Transform</i> (DCT)	87
4.7	Hasil <i>Mel Frequency Cepstrum Coefficients</i> (MFCC).....	89
4.8	Arsitektur Jaringan <i>Backpropagation</i>	92
4.9	<i>Script</i> program Jaringan Syaraf Tiruan	94
4.10	Perubahan bobot suara <i>voice</i> Pria	97
4.11	Perubahan bobot suara <i>voice</i> Wanita	98

DAFTAR LAMPIRAN

Nomor	Judul	Halaman
1.	Lampiran 1	107
2.	Lampiran 2.....	113
3.	Lampiran 3	115
4.	Lampiran 4	123
5.	Lampiran 5	125
6.	Lampiran 6	126

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang semakin pesat membutuhkan daya pikir, kreatifitas dan inovasi dalam perkembangan kemajuan untuk mempermudah sistem kehidupan yang semakin bervariasi. Salah satu hal yang melatar belakangi daya pikir manusia dalam perkembangan teknologi adalah komputer yang mampu berkomunikasi dengan manusia secara alami. Ilmu yang membahas bidang ini dikenal dengan nama pemrosesan suara dan bahasa secara alami (*speech and natural language processing*). Secara umum ada dua paradigma dalam bidang keilmuan *speech* yaitu paradigma stokastik dan paradigma logika formal. Paradigma logika formal difokuskan pada pengembangan *grammar* (metamorfosis grammar, definite clause grammar, functional grammar) dan juga struktur unifikasi. Sedangkan paradigma stokastik terletak pada pemrosesan data suara baik pada tahap praproses maupun pengenalan pola [1]. Pada perkembangan sekarang ini, bidang tersebut sudah meluas sehingga tidak hanya masalah pemrosesan (*processing*) tetapi juga mencakup *speech generation* baik *speech recognition* (pengenalan ucapan), *speech understanding* (pemahaman ucapan), *speaker identification* (identifikasi pembicara), *speaker verification* (verifikasi pembicara).

Hal ini yang menyebabkan perkembangan paradigma dalam bidang pengolahan sinyal digital khususnya *Voice Identification*. Oleh karena itu aplikasi dalam bidang *speech processing and understanding* menjadi luas mulai dari *document summarization* (coding), *transmission*, *text parsing* (analysis), *spelling/grammar correction* (enhancement), *natural language generation* (synthesis), *natural language understanding* (understanding), *web search* (retrieval/mining), serta banyak lagi aspek *recognition* seperti mesin pendikte, identifikasi maupun verifikasi pembicara, mesin penjawab otomatis, serta interaksi manusia dengan komputer lainnya melalui suara. Selain aspek terapan yang begitu luas pada bidang *speech and language processing*, investasi yang diperlukan lebih pada aspek software [2]. Suara adalah salah satu alat komunikasi antar manusia yang efektif dan digemari karena merupakan fenomena perpaduan multidimensi mulai dari dimensi linguistic, semantik, artikulasi dan akustik. Dimensi linguistic dan semantik bersifat *linguistic dependent*, sedangkan artikularis dan akustik bersifat *Linguistic Independent*. Dimensi artikularis secara detail menjadi kajian dari bidang fonologi yang mempelajari jenis-jenis suara yang dihasilkan. Sedangkan akustik terbagi menjadi dua bagian yaitu yang mengkaji sinyal suara secara fisik dan bagian lainnya yang melakukan analisis terhadap sinyal suara. Suara merupakan fungsi kompleks dari beberapa aspek, seperti karakteristik pembicara (dimensi titik artikulasi, emosi, kesehatan, umur, jenis kelamin, dialek) bahasa dan lingkungan (background dan media transmisi). Maka pemodelan sinyal membutuhkan proses yang kompleks dalam proses analisa data.

Dengan latar belakang perkembangan teknologi dalam bidang *processing* maka dilakukan penelitian ini untuk proses pemrosesan suara dalam proses pengidentifikasian (voice identification). Sistem pembicara mengenali pembicara berdasarkan suara yang direkam dan disimpan dalam bentuk format audio wav (waveform audio format). Pengidentifikasian juga dipengaruhi adanya noise yang disebabkan oleh lingkungan, distorsi alat perekam yang bisa menjadi sumber error yang dapat mengakibatkan pengidentifikasian tidak akurat. Namun penelitian ini telah banyak dilakukan dalam pengaplikasian teknologi komunikasi baik sebagai biometrik sistem maupun *security system*[3]. Hal ini dapat diamati melalui aplikasi-aplikasi yang telah menerapkan suara sebagai alat komunikasi seperti pada perintah-perintah ke robot, perabotan rumah tangga, aplikasi konversi suara menjadi teks digital dalam bahasa tertentu, mendeteksi penyakit, penyelidikan kasus kriminal dan lain sebagainya. Peneliti membagi sistem dalam 4 bagian besar, yakni proses perekaman, *pre-processing*, ekstraksi ciri, dan pengenalan. Pada *pre-processing* digunakan proses bantuan seperti *pre-emphasis*, *segmentasi*, *framing*, dan *windowing*. Proses ekstraksi ciri dan pengenalan digunakan ciri MFCC (Mel Frequency Cepstrum Coefficients) dan *identifier* jaringan syaraf tiruan (JST) *backpropagation*[4][5]. Berdasarkan penelitian sebelumnya teknik ekstraksi ciri sinyal suara yang digunakan yaitu MFCC telah menghasilkan ciri vektor yang efektif dan pencapaian keakuratan hasil sebesar 99% sedangkan pada tahapan pengenalan pola yang digunakan sebagai *classifier* menggunakan jaringan syaraf tiruan. Metode ini menggunakan rata-rata dari seluruh unit waktu untuk memperoleh informasi.

1.2 Rumusan Masalah

Dalam Penelitian ini dapat dirumuskan masalah sebagai berikut :

- a. Bagaimana proses ekstraksi ciri suara menggunakan metode MFCC.
- b. Bagaimana proses pengenalan suara pria dan wanita terhadap inisialisasi perubahan bobot menggunakan jaringan syaraf tiruan dengan metode *backpropagation*.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah :

- a. Mengidentifikasi suara menggunakan metode MFCC yang merepresentasikan sinyal suara (recorded voice) dalam bentuk data.
- b. Pengimplementasian metode *neural network* yaitu jaringan syaraf tiruan dengan metode *backpropagation*.
- c. Menemukan suatu sistem dari program aplikasi pengenalan suara *voice* menggunakan jaringan syaraf tiruan (JST).

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

- a. Untuk memberikan kontribusi ilmu pada bidang teknik elektro khususnya dibidang pengolahan sinyal digital.
- b. Mendeteksi suara menggunakan metode ekstraksi ciri MFCC sebagai proses pembelajaran dan pengenalan pola suara setelah proses perekaman yang mempengaruhi keakuratan *voice identification*.

- c. Untuk menemukan sistem yang bisa mengidentifikasi dan memverifikasi suara di berbagai aspek bidang baik teknologi, perbankan, pendidikan, ilmu medis dan lain sebagainya.
- d. Untuk mengetahui tingkat keakuratan metode yang digunakan untuk mendeteksi pernyataan dari pengidentifikasian keaslian sumber suara.

1.5 Batasan Masalah

Dari rumusan masalah yang ada dan untuk mengurangi pembahasan yang meluas maka akan dilakukan pembatasan masalah sebagai berikut :

- a. Implementasi sistem identifikasi suara menggunakan algoritma Matlab
- b. Data masukan berupa sinyal suara yang diambil dari 2 orang responden (1 pria dan 1 wanita).
- c. Sistem identifikasi suara *voice* menggunakan metode MFCC sebagai ekstraksi ciri suara yang mengalami distorsi akibat kondisi lingkungan.
- d. Sistem pengenalan pola menggunakan jaringan syaraf tiruan dengan metode *backpropagation*.
- e. Penelitian ini dibatasi pada pengidentifikasian suara yang direkam dan disimpan dalam .wav dengan penanganan gangguan noise yang ditemukan pada saat proses perekaman.
- f. Penelitian ini mendeteksi, mengidentifikasi dan memverifikasi sumber suara asli.

BAB 2

TINJAUAN PUSTAKA

Suara adalah salah satu alat komunikasi antar manusia yang efektif dan digemari. Selain sebagai alat komunikasi antar manusia, suara manusia telah digunakan pada taraf yang lebih besar saat sekarang. Suara telah dijadikan sebagai salah satu alat komunikasi antara manusia dan komputer (mesin). Hal ini dapat diamati melalui aplikasi-aplikasi yang telah menerapkan suara sebagai alat komunikasi tersebut seperti pada perintah-perintah ke robot, perabot rumah tangga, aplikasi konversi suara menjadi teks digital dalam bahasa tertentu dan lain sebagainya. Pada penelitian ini, peneliti memfokuskan diri pada pengenalan suara ucapan kata dalam Bahasa Indonesia. Penelitian mengenai pengenalan bahasa sudah banyak dilakukan oleh para peneliti sebelumnya dalam berbagai bahasa. Metode yang digunakan dalam penelitian-penelitian dalam proses ekstraksi ciri dan pengenalan pola pun sangat beragam.

2.1 *State of Art* Pemrosesan Suara dan Bahasa

Algoritma *endpoint detection* (EPD) digunakan untuk memotong durasi diam dalam sampel bahasa yang mendeteksi keberadaan bahasa/ kata yang terdeteksi sebagai noise[6]. Teknik *Mel Frequency Cepstral Coefficient* (MFCC) digunakan

untuk mengestrak ciri dari bahasa, algoritma *subtractive clustering* diterapkan untuk mengidentifikasi sistem interferensi *fuzzy*, dan *adaptive Neuro Fuzzy Inerference System* (ANFIS) digunakan sebagai teknik klasifikasi modern untuk melatih dalam mengidentifikasi ciri dari bahasa. Unjuk kerja sistem dievaluasi dengan menggunakan 630 sampel ucapan untuk pelatihan, pengujian, dan hasil-hasil eksperimen menunjukkan bahwa pengembangan sistem ini telah mencapai akurasi yang tinggi untuk fase pelatihan dan pengenalan berturut-berturut 100% dan 85,24%. Penelitian dengan judul “*Isolated Digit Speech Recognition in Malay Language using Neuro-Fuzzy Approach*”. Dalam penelitian ini dilakukan pengembangan dan implementasi sebuah sistem pengenalan bahasa berupa angka (0 sampai 9) dalam bahasa melayu. Sistem ini dikembangkan dengan menggunakan pendekatan neuro-fuzzy yang menggabungkan gaya penalaran dan sistem fuzzy dengan pembelajaran dan sruktur *connection* dari *neural network*[7].

Penelitian pada tahun 2008 oleh Vimal Krishnan V.R, V. Jayakumar A, dan Anto P.B dengan judul “*Speech Recognition of Isolated Malayalam Words Using Wavelet Features dan Artificial Neural Network*”, menggunakan daubechies 4 type dari wavelet untuk ekstraksi ciri dan jaringan syaraf tiruan (Artificial Neural Network) untuk tujuan klasifikasi dan pengenalan [8]. Lima kata bahasa Malayalam (salah satu bahasa di India selatan) digunakan untuk eksperimen. Sebanyak 150 sampel dikumpulkan, dikategorikan diberi label dan disimpan dalam sebuah basis data. vektor ciri (feature vector) dihasilkan untuk semua kata dan membentuk sebuah himpunan pelatihan untuk tujuan klasifikasi dan pengenalan. Sebuah runtun level –

level dekomposisi dihasilkan untuk mencapai sebuah vektor ciri yang baik. Sebuah vektor ciri dari elemen berukuran dua belas dikumpulkan untuk semua kata pada level kedelapan dekomposisi. Dengan menggunakan metode ini dapat dicapai tingkat pengenalan 89%. Menurut Guang Li, Zhang, J da dan Freeman, W.J pada tahun 2007 dalam penelitian berjudul “*Mandarin Digital Speech Recognition Based on a Chaotic Neural Network and Fuzzy C-means Clustering*” menggunakan pemodelan sistem syaraf penciuman model KIII (Chaotic Neural Network) dalam pengenalan tutur bahasa mandarin. Namun suara adalah perpaduan multidimensi yang dipengaruhi berbagai aspek baik internal maupun eksternal. Pengaruh aspek internal berupa artikulasi, emosi, kesehatan, umur, jenis kelamin dan perbedaan dialek dan aspek eksternal dapat berupa pengaruh lingkungan dan media transmisi yang digunakan dalam pengidentifikasian voice. Hal inilah yang menjadi latar belakang dari penelitian *voice identification* [9][10].

Setelah memahami bidang kajian melalui studi literature maka dilakukan identifikasi permasalahan utama yang muncul. Dengan memfokuskan pada satu permasalahan utama, dilakukan percobaan menggunakan metode yang umum dipakai. Hal ini dilakukan untuk mengidentifikasi tingkat keakuratan metode tersebut dalam mengidentifikasi suara. Pemahaman terhadap bidang kajian yang disajikan memberikan deskripsi mengenai transformasi sinyal suara menjadi informasi. Proses transformasi ini secara garis besar terdiri dari tiga bagian yaitu proses transformasi sinyal, proses pengenalan suara, dan ekstraksi ciri. Pembahasan dimulai dengan bidang-bidang utama pengenalan suara hingga ke identifikasi pembicara. Selain

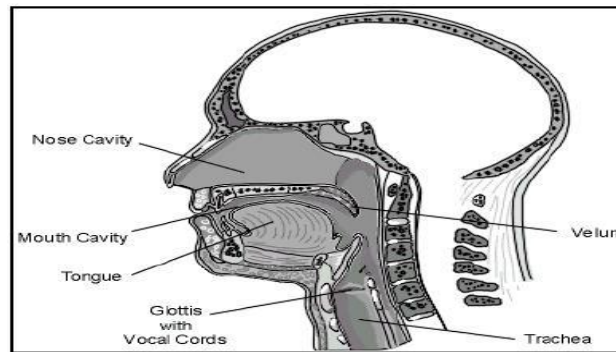
penyajian aspek teknis yang dinyatakan secara skematis juga perkembangan metodologi yang ada serta pencapaian hasilnya. Hal ini dimaksudkan untuk mengetahui sejauh mana perkembangan riset yang ada serta identifikasi permasalahan yang muncul. Studi empiris berbasis teknik yang sudah umum dipergunakan pada pemrosesan sinyal disajikan pada bab-bab berikut terdiri dari bagian utama yaitu identifikasi menggunakan *Mel Frequency Cepstrum Coefficients* (MFCC) sebagai ekstraksi ciri sebagai pengenalan pola [11][12].

Metode ini menggunakan data sinyal yang asli dengan adanya pengidentifikasian noise. Bagian kedua adalah pembelajaran pola yang berhubungan dengan bunyi suara dari kelas yang sama yang digunakan untuk membuat pola representative. Hasilnya digunakan sebagai pola referensi yang menjadi sebuah model yang mempunyai karakteristik dari ciri-ciri pola referensi [13].

2.2 Landasan Teori

Suara manusia terbentuk oleh tiga komponen utama yaitu sumber tenaga yang berasal dari paru-paru, sumber getar yang berasal pangkal tenggorokan serta *resonator* yang diperankan oleh tenggorokan, hidung, mulut dan sinus. Tenaga suara berasal dari udara yang dikeluarkan. Pada saat bernafas, diafragma bagian bawa tulang rusuk mengembang, membawa udara ke paru-paru dan proses sebaliknya terjadi saat membuang nafas dimana udara keluar dari paru-paru dan menciptakan aliran udara menuju trakea. Aliran udara ini menghasilkan energi untuk pita suara

pada pangkal tenggorokan yang berfungsi untuk memproduksi suara. Semakin besar aliran udara maka semakin besar pula suara yang dihasilkan [14].



Gambar 2.1 Organ-organ produksi suara

Sumber getar berasal dari larik atau kotak suara yang berada di ujung batang tenggorokan. Terdiri dari dua pita suara yang terbuka ketika bernafas dan tertutup ketika menelan dan memproduksi suara. Saat kita memproduksi suara, aliran udara melewati dua pita suara secara bersamaan. Pita suara ini lunak dan akan bergetar ketika dilewati aliran udara. Pita suara bergetar sangat cepat antara 100 sampai 1000 kali perdetik, tergantung nada suara yang dihasilkan. Nada suara ditentukan oleh panjang dan ketegangan pita suara, yang dikontrol oleh otot di larik. Pita suara menghasilkan suara yang mirip dengan dengungan. Struktur pada pita suara termasuk tenggorokan, hidung, mulut adalah bagian dari sistem resonator. Struktur ini dapat dibandingkan dengan sebuah terompet. Suara dengungan yang dihasilkan oleh getaran pita suara diubah oleh bentuk dari sistem resonator untuk menghasilkan suara manusia yang unik.

Organ-organ pembentuk sinyal suara dapat diperlihatkan pada Gambar 2.1 meliputi :

1. Bagian *pulmonary tract*; terdiri dari paru-paru dan batang tenggorokan.
2. Bagian *larynx*; terdiri dari tali-tali vokal. Ruang antara pita suara disebut glottis.
3. Bagian *vocal tract*; terdiri dari pharynx, rongga mulut, dan rongga hidung.

Berdasarkan cara eksitasinya, suara dibagi menjadi 3, yaitu

a. Suara *Voiced* (Voiced Sound)

Tekanan udara yang dibangkitkan dilewatkan pada *glottis* dengan pita suara yang tegang. Getaran pita suara akan mengubah aliran udara menjadi pulsa-pulsa yang *quasi-periodic*.

b. Suara Desah (Fricative/Unvoiced Sound)

Suara desah diakibatkan penyempitan pada beberapa bagian *vocal tract*. Aliran udara yang berasal dari tenggorokan dipaksa melewati daerah penyempitan tersebut dengan kecepatan yang cukup tinggi.

c. Suara letupan (Plosive/Stop Sound)

Adanya penutupan pada bagian *vocal tract* (biasanya di bagian depan) mengakibatkan aliran udara berkumpul di bagian belakang sehingga menimbulkan tekanan yang lebih besar.

Bagian *vocal tract* yang tertutup ini dibuka secara tiba-tiba untuk menghasilkan eksitasi letupan vokal (vowel) adalah bunyi bahasa yang arus udaranya tidak mengalami rintangan dan kualitasnya ditentukan oleh 3 faktor: tinggi-rendahnya posisi lidah, bagian lidah yang dinaikkan, dan bentuk bibir pada pembentukan vokal tersebut. Pada saat vokal diucapkan, bagian depan, tengah atau belakang lidah dapat

dinaikkan atau diturunkan bersama rahang. Berdasarkan frekuensinya, suara dibagi menjadi 4 bagian :

- a. Infrasonic = 0 Hz – 20 Hz
- b. Pendengaran manusia = 20 Hz – 20 KHz
- c. Ultrasound = 20 KHz – 1 GHz
- d. Hypersonic = 1 GHz – 10 THz

Suara merupakan fenomena fisik yang dihasilkan oleh getaran benda atau getaran suatu benda yang berupa sinyal analog dengan amplitudo yang berubah secara kontinyu terhadap waktu. Suara dihasilkan oleh getaran suatu benda. Selama bergetar, perbedaan tekanan terjadi di udara sekitarnya. Pola osilasi yang terjadi dinamakan sebagai Gelombang. Gelombang mempunyai pola sama yang berulang pada interval tertentu, yang disebut sebagai “Periode”. Hal-hal yang berkaitan dengan suara antara lain adalah frekuensi. Frekuensi merupakan banyaknya getaran dalam satu detik, satuannya adalah Hertz (Hz) atau *cycles per second* (CPS), sedangkan panjang gelombang suara (Wavelength) dirumuskan dengan c/f , dimana c = Kecepatan rambat bunyi , f = Frekuensi.

Manusia membuat suara dengan frekuensi berkisar antara 50Hz – 10KHz. Suara merupakan fenomena fisik yang dihasilkan oleh getaran benda atau getaran suatu benda yang berupa sinyal analog dengan amplitudo yang berubah secara kontinyu terhadap waktu. Sinyal suara musik memiliki frekuensi 20Hz – 20KHz. Oleh karena itu sistem *multimedia* menggunakan suara yang berada dalam rentang pendengaran manusia. Suara diluar rentang pendengaran manusia dapat dikatakan

sebagai “*Noise*”, merupakan getaran yang tidak teratur dan tidak berurutan dalam berbagai frekuensi.

Amplitudo merupakan keras lemahnya bunyi atau tinggi rendahnya gelombang yang memiliki satuan *decibel* (dB). Bunyi yang berada pada tingkat volume lebih besar dari 85 dB akan mampu membuat hancur gendang telinga. Sedangkan *velocity* merupakan kecepatan perambatan gelombang bunyi sampai ke telinga pendengar dan memiliki satuan m/s. Pada udara kering dengan suhu 20 °C (68 °F) kecepatan rambat suara sekitar 343 m/s.

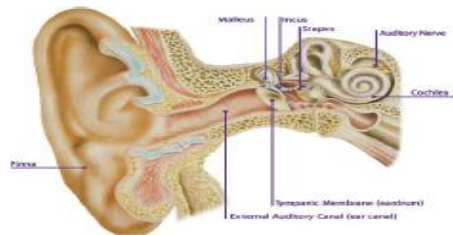
2.2.1 Pengenalan suara

Pengenalan suara (speech recognition) merupakan salah satu kajian utama dalam kecerdasan buatan. Pengenalan suara pada komputer ini bertujuan agar komputer dapat mengerti yang diucapkan oleh manusia. Aplikasi yang digunakan adalah *Voice Command*. Dengan aplikasi ini diharapkan komputer dapat mengerti perintah yang diberikan manusia. Sehingga komputer dapat mengerti setiap perintah yang diberikan manusia melalui ucapan [15].

Pada sistem pengenalan suara pada manusia memiliki tiga organ penting yang digunakan dan saling berhubungan yaitu telinga yang berperan sebagai transduser dengan menerima sinyal masukan suara dan mengubahnya menjadi sinyal syaraf. Jaringan syaraf ini berfungsi untuk mentransmisikan sinyal ke otak dan otak akan mengidentifikasi dan mengklasifikasikan informasi yang terkandung dalam sinyal masukan [16].

2.2.2 Karakteristik telinga

Telinga terbagi menjadi tiga bagian, yaitu bagian luar, tengah dan dalam. Bagian dari Organ Telinga dapat diperlihatkan pada Gambar 2.2.



Gambar 2.2 Organ Telinga

Beberapa Bagian telinga yang berfungsi sebagai penerima suara meliputi :

Pinna, merupakan bagian luar telinga yang berfungsi sebagai corong, untuk mengumpulkan sinyal suara menuju auditory canal sehingga dapat memberikan kesan arah sinyal suara yang diterima. *Auditory canal* adalah struktur berbentuk pipa lurus sepanjang 2,7 cm, dengan diameter sekitar 0,7 cm yang pada bagian ujungnya terdapat selaput membranc yaitu gendang telinga. Membran ini merupakan pintu masuk telinga bagian tengah yaitu ruangan berisi udara dengan volume sebesar 2 cm³, yang terdiri dari tiga buah tulang, yaitu *malleus* (martil), *incus* (landasan), dan *stapes* (sanggurdi). Bagian ini terhubung dengan tenggorokan melalui *Eustachian tube*. Getaran pada gendang telinga ditransmisikan ke *malleus* melalui *incus* dan *stapes* yaitu membentuk *oval window*. Telinga bagian dalam (labyrinth) memiliki tiga

bagian yaitu *vestibule* (ruang pintu masuk), *semicircular canal* dan *cochlea*. *Vestibule* terhubung dengan telinga bagian tengah melalui dua jalur yaitu *oval window* dan *round window*. Keduanya tertutup untuk mencegah keluarnya cairan yang mengisi telinga bagian dalam. Pada *cochlea*, yang berstruktur seperti rumah siput, terdapat syaraf pendengaran. Syaraf ini memanjang sampai ke basilar membrane.

Pada bagian atas basilar membrane terdapat *organ of corty* yang memiliki empat baris sel rambut (sekitar 3×10^4 sel seluruhnya). Telinga bagian dalam (labyrinth) memiliki tiga bagian yaitu *vestibule* (ruang pintu masuk), *semicircular canal* dan *cochlea*. *Vestibule* terhubung dengan telinga bagian tengah melalui dua jalur yaitu *oval window* dan *round window*. Keduanya tertutup untuk mencegah keluarnya cairan yang mengisi telinga bagian dalam. Pada *cochlea*, yang berstruktur seperti rumah siput, terdapat syaraf pendengaran. Syaraf ini memanjang sampai ke basilar membrane. Pada bagian atas basilar membrane terdapat *organ of corty* yang memiliki empat baris sel rambut (sekitar 3×10^4 sel seluruhnya).

2.2.3 Proses pendengaran

Proses Pendengaran pada telinga manusia dijelaskan sebaga berikut :

1. Sinyal suara memasuki saluran telinga dan variasi tekanan yang dihasilkannya menekan gendang telinga. Karena sisi bagian dari gendang telinga mempunyai tekanan yang nilainya dijaga konstan maka gendang telinga akan bergetar.
2. Getaran dari gendang telinga disalurkan pada tiga rangkaian tulang yaitu; *martil*, *incus*, dan *stapes*. Mekanisme ini dirancang untuk mengkopel variasi suara dari udara ke telinga bagian dalam. Karena luas permukaan penampang yang ditekan

stapes lebih kecil dari luas penampang gendang telinga maka tekanan suara yang sampai ke telinga bagian dalam bertambah besar.

3. Cairan pada *cochlea* bergetar dengan frekuensi yang sama dengan gelombang yang datang. *Basilar membrane* kemudian memisahkan sinyal berdasarkan frekuensinya. *Basilar membrane* berstruktur kuat dan panjang di daerah sekitar *oval window* namun bersifat lentur pada bagian ujungnya. Frekuensi resonansi yang dihasilkan membrane tersebut berbeda sepanjang dimensi.

Dimana resonansi frekuensi tinggi terjadi pada bagian - bagian *basilar membrane* yang berada dekat dengan *oval window*, sedangkan resonansi terjadi pada frekuensi rendah terjadi pada daerah ujung lainnya. Syaraf yang berada pada membran kemudian mendeteksi posisi terjadinya resonansi yang juga akan menentukan frekuensi suara yang datang. Ukuran dari basilar membrane rata-rata sekitar 35 mm. Dari ukuran panjang tersebut dapat dihasilkan 10 resolusi frekuensi, sehingga pada setiap 3.5 mm panjang membrane terdapat 1 oktaf frekuensi resonansi.

2.3 Pengolahan sinyal digital (PSD)

Penerapan pengolahan sinyal secara digital telah diterapkan begitu luas. Dari peralatan instrumentasi, kontrol, peralatan musik, peralatan kesehatan dan peralatan lainnya. Istilah pengolahan sinyal digital sebenarnya kurang begitu tepat, yang lebih tepat adalah pengolahan sinyal diskrit. Tetapi karena istilah ini sudah luas digunakan, maka istilah pengolahan sinyal digital tetap digunakan. Proses pengolahan sinyal digital, diawali dengan proses pencuplikan sinyal masukan yang berupa sinyal kontinyu. Proses ini mengubah representasi sinyal yang tadinya berupa sinyal

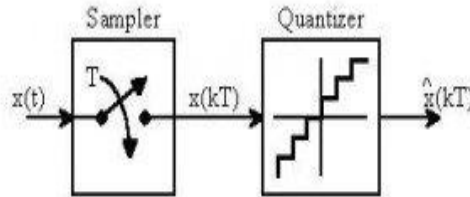
kontinyu menjadi sinyal diskrit. Proses ini dilakukan oleh suatu unit ADC (Analog to Digital Converter). Unit ADC ini terdiri dari sebuah bagian sample/hold dan sebuah bagian *quantiser*. Unit sample/hold merupakan bagian yang melakukan pencuplikan orde ke-0, yang berarti nilai masukan selama kurun waktu T dianggap memiliki nilai yang sama. Pencuplikan dilakukan setiap satu satuan waktu yang lazim disebut sebagai waktu cuplik (sampling time). Bagian *quantiser* akan merubah menjadi beberapa level nilai, pembagian level nilai ini bisa secara *uniform* ataupun secara *non-uniform* misal pada *Gaussian quantiser*. Unjuk kerja dari suatu ADC bergantung pada beberapa parameter.

Parameter utama yang menjadi pertimbangan adalah sebagai berikut :

- a. Kecepatan maksimum dari waktu cuplik.
- b. Kecepatan ADC melakukan konversi.
- c. Resolusi dari quantiser, misal 8 bit akan mengubah menjadi 256 tingkatan nilai.
- d. Metoda kuantisasi akan mempengaruhi terhadap kekebalan noise.

Sinyal input asli yang tadinya berupa sinyal kontinyu, $x(t)$ akan dicuplik dan *diquantise* sehingga berubah menjadi sinyal discrete $x(kT)$. Dalam representasi yang baru inilah sinyal diolah. Keuntungan dari metoda ini adalah pengolahan menjadi mudah dan dapat memanfaatkan program sebagai pengolahnya. Dalam proses sampling ini diasumsikan menggunakan waktu cuplik yang sama dan konstan, yaitu T_s . Parameter cuplik ini menentukan dari frekuensi harmonis tertinggi dari sinyal yang masih dapat ditangkap oleh proses cuplik. Frekuensi sampling minimal adalah 2 kali dari frekuensi harmonis dari sinyal.

Proses sampling dapat diperlihatkan pada Gambar 2.3.



Gambar 2.3 Proses sampling

2.3.1 Representasi sinyal digital

Menurut Representasinya, sinyal dalam Pengolahan sinyal digital dapat dibagi menjadi dua jenis, yaitu:

a. *Sinyal Analog*

Sinyal Analog adalah sinyal data dalam bentuk gelombang kontinyu yang membawa informasi dengan mengubah karakteristik gelombang. Dua parameter/karakteristik terpenting yang dimiliki oleh isyarat analog adalah amplitudo dan frekuensi. Gelombang pada sinyal analog yang umumnya berbentuk gelombang sinus memiliki tiga variable dasar, yaitu amplitudo, frekuensi dan phase. Amplitudo merupakan ukuran tinggi rendahnya tegangan dari sinyal analog. Frekuensi adalah jumlah gelombang sinyal analog dalam satuan detik. Phase adalah besar sudut dari sinyal analog pada saat tertentu. Analog disebarluaskan melalui gelombang elektromagnetik (gelombang radio) secara terus menerus, yang banyak dipengaruhi oleh faktor "pengganggu". Analog merupakan bentuk komunikasi elektromagnetik yang merupakan proses pengiriman sinyal pada gelombang elektromagnetik dan bersifat variable yang berurutan. Jadi sistem analog merupakan suatu bentuk sistem

komunikasi elektromagnetik yang menggabungkan proses pengiriman sinyalnya pada gelombang elektromagnetik.

b. *Sinyal Digital*

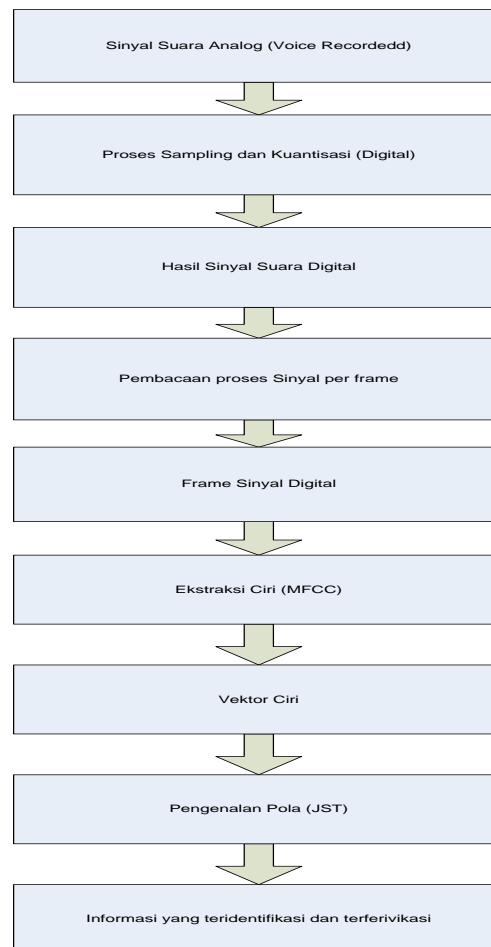
Sinyal Digital adalah sinyal data dalam bentuk pulsa yang dapat mengalami perubahan yang tiba-tiba dan mempunyai besaran 0 dan 1. Sinyal digital hanya memiliki dua keadaan yaitu 0 dan 1 sehingga tidak mudah terpengaruh oleh derau, tetapi transmisi dengan sinyal digital hanya mencapai jarak jangkauan pengiriman data yang relatif dekat. Biasanya sinyal ini juga dikenal dengan sinyal diskret. Sinyal yang mempunyai dua keadaan ini biasa disebut dengan bit. Bit merupakan istilah khas pada sinyal digital. Sebuah bit dapat berupa nol (0) atau satu (1). Kemungkinan nilai untuk sebuah bit adalah 2 buah. Kemungkinan nilai untuk 2 bit adalah sebanyak 4, berupa 00, 01, 10, dan 11. Secara umum, jumlah kemungkinan nilai yang terbentuk oleh kombinasi n bit adalah sebesar 2^n buah. Teknologi digital memiliki beberapa keistimewaan unik yang tidak dapat ditemukan pada teknologi analog, yaitu mampu mengirimkan informasi dengan kecepatan cahaya yang mengakibatkan informasi cepat diterima. Penggunaan yang berulang-ulang terhadap informasi tidak mempengaruhi kualitas dari informasi. Pada saat ini semakin banyak penggunaan teknik analog dan digital dalam suatu sistem untuk memanfaatkan keunggulan masing-masing. Tahapan terpenting adalah menentukan bagian yang menggunakan teknik analog dan bagian yang menggunakan teknik digital. Dan dapat diramalkan di masa depan bahwa teknik digital akan menjadi lebih murah dan berkualitas.

4.3.2 Pengolahan sinyal

- a. Penapisan Signal, Penapisan signal bertujuan untuk memisahkan suatu sinyal yang tercampur dengan derau atau sinyal lain yang tidak diperlukan. Hal ini dapat diperlihatkan pada pengukuran gelombang laut dengan alat yang bernama waverecorder, maka sinyal yang didapatkan sebenarnya adalah kombinasi ‘sinyal’ gelombang dan ‘sinyal’ pasang surut. Penapisan signal dapat dilakukan untuk memisahkan kedua sinyal tersebut.
- b. Pendeteksian Signal, Pendeteksian Signal untuk mengetahui keberadaan suatu sinyal dalam sinyal kompleks yang diolah. Contoh, dalam sinyal EKG.
- c. Kompresi Signal, kompresi signal bertujuan untuk memperkecil ukuran sinyal tanpa harus kehilangan informasi yang terdapat pada sinyal. Contoh, untuk dapat menggunakan bandwidth yang tersedia, maka sinyal yang akan ditranfer lewat internet biasanya akan dikompres terlebih dahulu.
- d. Pengenalan Pola, serta
- e. Restorasi signal dan Rekonstruksi signal

Sinyal merupakan gelombang longitudinal yang tercipta dari tekanan udara yang berasal dari paru-paru yang berjalan melewati lintasan suara menuju mulut dan rongga hidung dengan bentuk *articulator* yang senantiasa berubah. Pemrosesan suara merupakan teknik mentransformasi gelombang longitudinal menjadi informasi yang sesuai dengan yang diinginkan. Secara umum proses transformasi tersebut terdiri dari digitalisasi sinyal analog, ekstraksi ciri, dan pengenalan pola menggunakan jaringan syaraf tiruan untuk klasifikasi. Alur proses

transformasi sinyal suara analog menjadi Informasi dapat diperlihatkan pada Gambar 2.4. Proses tranformasi sinyal suara manjadi informasi dapat disajikan dalam proses mengenai pengenalan sinyal, ekstraksi ciri dan pengenalan pola. Pada bagian sinyal, pembahasan dimulai dengan terminology sinyal, ukuran kualitas sinyal, sampling dan kuantisasi serta pembacaan sinyal untuk pemrosesan. Untuk proses ekstraksi ciri, pembahasan difokuskan pada teknik MFCC. Untuk pengenalan pola, akan difokuskan pada pemodelan jaringan syaraf tiruan dengan metode *backpropagation* [17].



Gambar 2.4 Alur Proses Transformasi Sinyal Suara Analog
(voice identification)

Sampling

Sampling merupakan pengamatan nilai sinyal waktu kontinyu (sinyal analog) pada suatu waktu tertentu, sehingga diperoleh sinyal waktu diskrit. Beberapa cara untuk melakukan sampling pada sinyal analog. Salah satu yang sering digunakan adalah *periodic* atau *uniform sampling*. Dalam hal ini sampling dilakukan pada setiap selang waktu yang tetap yaitu pada setiap selang waktu T . Hubungan antara sinyal diskrit hasil sampling dengan sinyal analog pada persamaan 2.1.

$$X(n) = X_a(nT) \text{ dengan } -\alpha < n < \alpha \dots\dots\dots (2.1)$$

$X(n)$ adalah sinyal waktu diskret yang diperoleh dari sampling terhadap sinyal waktu nyata $X_a(T)$ setiap T detik. Dalam hal ini T sebagai periode sampling dan $F_s = 1/T$ adalah sampling rate (Hertz). Permasalahan yang dihadapi saat proses sampling adalah dalam hal penentuan T atau F_s dari suatu sinyal analog dengan frekuensi F .

Sinyal analog dapat direpresentasikan sebagai penjumlahan dari gelombang sinus dengan amplitud, frekuensi dan fase yang berbeda, yang dalam hal ini sebanyak N komponen. Nilai sampling rate yang dapat menangkap semua komponen sinyal minimal dua kali dari frekuensi maksimum yang ada pada sinyal. Nilai sampling rate sebesar $F_s = 2F_{\max}$ yang disebut sebagai *Nyquist rate*.

Kuantisasi dan Coding

Kuantisasi merupakan proses mengkonversi nilai amplitud yang bersifat kontinyu pada suatu titik waktu tertentu menjadi sinyal digital dengan mengekspresikannya menggunakan sejumlah digit tertentu. Konversi nilai kontinyu menggunakan sejumlah digit ini akan menghasilkan error yang disebut *quantization*

error atau *quantization noise*. Proses kuantisasi dilakukan dengan pembulatan ke nilai terdekat (rounding) atau bisa dengan pemotongan bagian sisa (*truncating*). Error karena kuantisasi dengan metode pembulatan, $e_q(n)$ pada persamaan 2.2.

$$\frac{-\Delta}{2} \leq e_q(n) \leq \frac{\Delta}{2} \text{ dengan } \Delta = \frac{X_{\max} - X_{\min}}{L-1} \dots\dots\dots(2.2)$$

L adalah banyaknya level kuantisasi, X_{\max} dan X_{\min} adalah nilai maksimum dan minimum yang akan dikuantisasi. Dalam hal ini *error* kuantisasi merupakan selisih antara nilai sinyal analog dengan nilai hasil kuantisasinya pada persamaan 2.3.

$$e_q(t) = x_a(t) - x_q(t) \dots\dots\dots(2.3)$$

Ukuran kualitas output dari suatu mesin konversi analog ke digital (A/D converter) biasanya diukur dengan *signal-to-quantization noise ratio* (SQNR) yang dinyatakan sebagai rasio energi signal terhadap energi noise pada persamaan 2.4.

$$\text{SQNR} = \frac{P_x}{P_y} = \frac{3}{2} 2^{2b} \dots\dots\dots(2.4)$$

Dengan satuan decibel (dB), maka SQNR dirumuskan pada persamaan 2.5.

$$\text{SQNR (dB)} = 10\log_{10}(\text{SQNR}) = 1.76 + 6.02b \dots\dots\dots(2.5)$$

Coding merupakan pemberian bilangan biner pada setiap level kuantisasi. Jika kuantisasi yang diterapkan mempunyai level sebanyak L, maka setidaknya harus tersedia L bilangan biner yang berbeda. Sedangkan kode biner dengan panjang b akan dapat menghasilkan kode berbeda sebanyak 2^b . Oleh karena itu untuk kuantisasi dengan L level diperlukan bilangan biner dengan panjang $b \geq \log_2 L$.

2.4 Sistem Pengenalan Bahasa (Speech Recognition System)

Sistem pengenalan bahasa dimodelkan sebagai fungsi energi dari frekuensi. Bahasa dapat dianggap sebagai campuran dari runtun nada yang berbeda. Pengenalan bahasa melibatkan transformasi isyarat bahasa akustik kepada nilai-nilai numeris untuk pembelajaran mesin. Bahasa yang dimodelkan sebagai fungsi energi dari frekuensi. Dengan suatu kombinasi frekuensi dan energi yang diberikan sebuah nada yang dapat dihasilkan. Suara dapat dianggap sebagai campuran dari runtun nada yang berbeda. Pengenalan bahasa otomatis (ASR) telah mencapai keberhasilan besar dalam beberapa dekade terakhir namun studi lanjut diperlukan karena tidak ada metode saat ini yang cukup cepat dan tepat sebanding dengan pengenalan suara manusia. ASR dapat diklasifikasikan kedalam teknologi pemrosesan tutur digital yang juga melibatkan sintesis bahasa dan biometric suara (identifikasi pembicara, verifikasi pembicara). Secara umum tujuannya adalah untuk memungkinkan mesin dapat meniru kemampuan manusia untuk mendengar, mengidentifikasi dan mengucapkan bahasa lisan alami manusia. Pikiran pembicara memutuskan runtun kata sumber voice (W) yang disampaikan melalui generator teksnya. Sumber tersebut dilewatkan melalui saluran komunikasi yang terdiri dari alat vokal pembicara untuk menghasilkan gelombang bahasa dan komponen pengolahan isyarat bahasa dari *speech recognizer* [18]. Akhirnya *speech decoder* mendecode isyarat akustik menjadi sebuah runtun kata/bahasa yang harus paling mirip dengan runtun kata aslinya (sumber voice).

2.4.1 Struktur suku kata dalam bahasa Indonesia

Suku kata adalah pembagian kata sesuai dengan pembangian lafal atau bunyinya. Kata dalam bahasa Indonesia bisa tersusun dari satu hingga banyak suku kata dengan bunyi vokal (V) dan kosonan (K). Kebanyakan kata dalam bahasa Indonesia tersusun dari dua suku kata dan tiga suku kata. Oleh karena itu, sebagai sampel bahasa yang digunakan dalam penelitian ini dipilih lima kata yang tersusun dari dua suku kata dan tiga suku kata, dengan penjelasan sebagai berikut:

1. Tersusun dari berbagai struktur suku kata (V,KV,KVK,KKV,KKVK)
2. Kata yang diawali vokal dan kata yang diawali konsonan
3. Kata yang diakhiri vokal dan kata yang diakhiri konsonan.

Menunjukkan lafal, struktur fonem, struktur suku kata dan jumlah suku kata untuk setiap bahasa Indonesia yang digunakan dalam penelitian ini. Pengenalan ucapan kata dalam bahasa Indonesia masih minim dilakukan, selain itu cakupan jumlah katapun masih kecil. Pada penelitian ini, peneliti melakukan pengenalan ucapan kata Bahasa Indonesia dengan memecah kata bahasa Indonesia kedalam bentuk suku kata bahasa lisan. Pemecahan kedalam bentuk lafal kata diharapkan mampu mengurangi jumlah kata yang sangat besar, namun tetap mengakomodasi seluruh kata dalam bahasa Indonesia. Peneliti membagi sistem dalam 4 bagian besar yakni proses perekaman, *pre-processing*, ekstraksi ciri, dan pengenalan pola. Penelitian ini proses ekstraksi ciri menggunakan MFCC dan untuk pengenalan pola digunakan *identifier* menggunakan jaringan syaraf tiruan dengan metode *backpropagatin* [19].

2.4.2 Transformasi sinyal menjadi informasi

Sinyal suara merupakan gelombang longitudinal yang tercipta dari tekanan udara yang berasal dari paru-paru yang berjalan melewati lintasan suara menuju mulut dan rongga hidung dengan bentuk articulator yang senantiasa berubah. Pemrosesan suara merupakan teknik mentransformasi gelombang longitudinal tersebut menjadi informasi sesuai yang diinginkan. Secara umum proses transformasi tersebut terdiri dari digitalisasi sinyal analog, ekstraksi ciri dan diakhiri dengan pengenalan pola untuk klasifikasi. Proses transformasi sinyal suara menjadi informasi yang akan dijelaskan pada sub bab disajikan dalam tiga konsep yaitu sinyal mengenai sinyal, ekstraksi ciri dan pengenalan pola. Pada bagian sinyal, pembahasan dimulai dengan terminology sinyal, ukuran kualitas sinyal serta pembacaan sinyal untuk pemrosesan. Untuk ekstraksi ciri pembahasan difokuskan pada teknik *mel frequency cepstrum coefficients* (MFCC). Hal ini dengan pertimbangan teknik tersebut relatif lebih baik dibandingkan teknik lain yang sudah ada.

2.5 Sinyal

Sinyal adalah suatu kuantitas fisik yang bervariasi dengan waktu, ruang maupun sembarang satu atau lebih peubah bebas lainnya. Secara matematika suatu sinyal dirumuskan sebagai fungsi dari satu atau lebih peubah bebas. Sebagai contoh berikut adalah dua sinyal yang pertama sebagai fungsi dari waktu dan kedua sebagai fungsi dari bidang spasial [20].

$$S_1(t) = 20t^2$$

t adalah waktu

$$S_2(x, y) = 3x + 2xy + 10y^2 \quad x \text{ dan } y \text{ menunjukkan posisi koordinat}$$

Dalam banyak kasus, fungsi yang menghubungkan antara waktu dengan kuantitas sinyal tidak diketahui dan sangat kompleks, sehingga tidak praktis dalam penerapannya seperti pada sinyal suara. Suatu segmen dari suatu sinyal suara direpresentasikan dengan akurasi tinggi yang merupakan jumlah dari beberapa fungsi sinus yang berbeda amplitudo dan frekuensi. Hal ini dapat dituliskan pada persamaan 2.6.

$$X(t) = \sum_{t=1}^N A_t(t) \sin[2\pi F_1(t) + \theta(t)] \dots\dots\dots(2.6)$$

$\{A_t(t)\}$ A adalah amplitudo dari gelombang sinus untuk setiap fungsi

$\{F_t(t)\}$ F adalah frekuensi dari gelombang sinus untuk setiap fungsi

$\{\theta, (t)\}$ θ adalah fase dari gelombang sinus untuk setiap fungsi t

Untuk menangkap pesan dari satu segmen sinyal suara yang dilakukan adalah dengan mengukur amplitudo, frekuensi dan fase pada segmen tersebut. Sinyal image adalah contoh sinyal sebagai fungsi dari dua peubah bebas yaitu koordinat spasial. Sinyal dibangkitkan oleh suatu sistem yang merupakan respon dari suatu stimulus atau gaya. Misalkan sistem sinyal suara terdiri dari *vocal cords* dan *vocal tract* yang disebut sebagai *vocal cavity*. Sedangkan sebagai stimulus adalah udara yang dihembuskan dari paru-paru melalui batang tenggorokan. Stimulus dan sistem secara bersama-sama disebut sebagai sumber sinyal (signal source). Secara umum sinyal dapat dikategorikan sesuai dengan sumber, jenis peubah bebas, skala nilainya serta kemunculannya.

2.5.1 Sinyal multichannel dan Sinyal multidimensional

Sinyal seringkali dibangkitkan oleh beberapa sumber (multiple sensor). Sebagai contoh sinyal pada *electrocardiogram* (ECG) untuk mendeteksi kondisi hati dari tiga sudut pandang yang berbeda dan merupakan fungsi dari waktu t . Oleh karena itu sinyal ditangkap sebagai sinyal multichannel dan representasikan dalam vektor 3 elemen.

Pada vektor diatas ada satu peubah bebas yaitu t . Jika peubah bebas lebih dari satu misal ditambah x dan y menunjukkan koordinat, maka sinyal tersebut menjadi *multidimensional*.

2.5.2 Sinyal waktu kontinyu dan Sinyal waktu diskret

Berdasarkan peubah bebas waktu (t), sinyal dapat dibedakan menjadi dua, yaitu *Sinyal Waktu Kontinyu* (Sinyal Analog); kuantitas sinyal terdefinisi pada setiap waktu dalam selang kontinyu (a,b) pada persamaan 2.7.

Misalkan $x(t) = (\cos \pi t) \quad 0 < t < \infty$ (2.7)

Sinyal Waktu Diskret ; kuantitas sinyal terdefinisi pada waktu diskret tertentu yang dalam hal ini jarak antar waktu tidak harus sama. Namun untuk kemudahan penurunan sifat matematikanya, maka biasanya jarak antar waktu adalah sama pada persamaan 2.8.

Misalkan $x[t_n] = \cos[\pi t_n] \quad n = 0, \pm 1, \pm 2, \pm 3$ (2.8)

Sinyal diskret ini biasanya dibangkitkan dengan cara sampling ataupun dengan cara mengkumulatikan dalam suatu selang waktu.

Sinyal diskret ini biasanya dibangkitkan dengan cara sampling ataupun dengan cara mengkumulatitkan dalam suatu selang waktu.

2.5.3 Sinyal *continuous valued* dan *discete valued*

Jika nilai dari sinyal waktu kontinyu maupun waktu diskret adalah sembarang nilai dari suatu selang yang terbatas maupun tidak terbatas maka disebut sinyal nilai kontinyu (*Continuous Valued*). Sedangkan jika nilai sinyal tersebut adalah satu dari beberapa kemungkinan nilai yang ada dalam gugus yang terbatas, maka disebut sebagai sinyal nilai diskret (*Discrete valued*). Sinyal waktu diskret yang mempunyai nilai bersifat diskrit disebut sebagai sinyal digital. Untuk tujuan pemerosesan sinyal secara digital, maka harus bentuk diskrit dalam waktu dan nilai kuantitas sinyal adalah diskrit (sinyal digital).

2.5.4 Sinyal *deterministic* dan Sinyal *random*

Jika nilai sinyal dari waktu ke waktu dapat dideskripsikan secara langsung menggunakan suatu fungsi secara tepat tanpa ada ketidakpastian, maka sinyal seperti ini disebut sebagai sinyal *deterministic*. Persamaan matematika yang menghubungkan peubah bebas dengan kuantitas sinyal disebut sebagai model sinyal. Sedangkan jika nilai sinyal pada setiap periode waktu tidak dapat diprediksi sebelumnya dengan tepat tanpa kesalahan, maka sinyal seperti ini dikenal dengan sinyal *random*.

2.6 Ukuran kualitas sinyal

Kualitas sinyal merujuk pada sejauh mana sinyal asli telah terdistorsi oleh suatu sebab tertentu. Makin banyak distorsi yang terjadi maka kualitas sinyal semakin buruk. Ada beberapa ukuran kualitas, yaitu SNR (*Signal to Noise Rasio*), Segmental-

mean SNR dan Segmental-median SNR. Andaikan $s(n)$ menotasikan sinyal bebas noise (noise free signal) pada periode n dan $\hat{s}(n)$ adalah sinyal $s(n)$ yang telah terdistorsi, maka sinyal error dapat dituliskan pada persamaan 2.9.

$$\varepsilon(n) = s(n) - \hat{s}(n) \quad \dots\dots\dots(2.9)$$

Oleh karena itu energi sinyal error ini pada persamaan 2.10.

$$E_{\varepsilon} = \sum_{n=-\infty}^{\infty} \varepsilon^2(n) \quad \dots\dots\dots(2.10)$$

Sedangkan energi sinyal asli yang tidak terdistorsi pada persamaan 2.11.

$$E_s = \sum_{n=-\infty}^{\infty} s^2(n) \quad \dots\dots\dots(2.11)$$

Signal to Noise Rasio (SNR) dirumuskan pada persamaan 2.12.

$$SNR = 10 \log 10 \frac{E_s}{E_{\varepsilon}} = 10 \log 10 \frac{\sum_n s^2(n)}{\sum_n [s(n) - \hat{s}(n)]^2} \quad \dots\dots\dots(2.12)$$

Sedangkan ukuran lainnya pada persamaan 2.13 dan 2.14.

$$SNR_{seg-mean} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log 10 \left[\frac{\sum_{n=m_1-N+1}^{m_1} s^2(n)}{\sum_{n=m_1-N+1}^{m_1} [s(n) - \hat{s}(n)]^2} \right] \quad \dots\dots\dots(2.13)$$

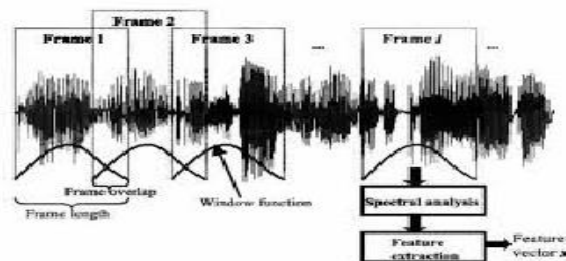
$$SNR_{seg-mean} = median \left\{ 10 \log 10 \left[\frac{\sum_{n=m_1-N+1}^{m_1} s^2(n)}{\sum_{n=m_1-N+1}^{m_1} [s(n) - \hat{s}(n)]^2} \right] \right\} \quad \dots\dots\dots(2.14)$$

Dalam hal ini $m_0, m_1, m_3, \dots, m_{M-1}$ adalah periode akhir dari frame ke 1, frame ke 2, dan seterusnya hingga frame ke M dengan panjang setiap frame adalah N.

2.6.1 Pembacaan sinyal

Untuk keperluan pemrosesan, sinyal analog yang sudah digitalkan (sampling dan kuantisasi) dibaca dari frame demi frame dengan lebar tertentu yang saling tumpang tindih. Panjang frame ini biasanya 5 hingga 100 millisecond dengan overlape antar frame yang berurutan adalah 25%, 50%, 75%. Proses ini dikenal dengan *frame blocking*. Satu frame tersebut sebagai satu unit terkecil yang mengandung satu unit informasi, sehingga barisan frame akan menyimpan suatu informasi yang lengkap dari sebuah sinyal suara.

Untuk itu distorsi antar frame harus diperkecil atau diminimalisasi. Satu teknik untuk meminimalkan distorsi antar frame adalah dengan melakukan proses filtering pada setiap frame. Secara umum fungsi *Filtering* berfungsi untuk memisahkan sinyal dari berbagai sumber lain yang “mengotori” serta “menjernihkan” sinyal dari adanya distorsi. Proses *blocking* dan *windowing* dapat diilustrasikan pada Gambar 2.5.



Gambar 2.5 Proses *Frame Blocking* dan *Windowing*

Pemrosesan dalam penelitian ini dilakukan dalam domain frekuensi, maka proses filtering yang dipilih akan menggunakan *windowing*. Proses *windowing*

dilakukan pada setiap frame dimana sinyal digital tersebut dikalikan dengan fungsi window tertentu yang berukuran sama dengan ukuran frame. Terlihat bahwa sinyal analog yang sudah diubah menjadi digital dibaca frame demi frame dan pada setiap frame dilakukan *windowing* dengan fungsi window tertentu.

2.7 Ekstraksi Ciri

Ekstraksi ciri merupakan proses untuk menentukan satu nilai atau *vector* ciri yang dapat dipergunakan sebagai penciri objek atau individu. Dalam pemrosesan suara, ciri yang biasa dipergunakan adalah nilai koefisien cepstral dari sebuah frame. Satu teknik ekstraksi ciri sinyal suara yang umum dan menunjukkan kinerja yang baik adalah teknik *Mel Frequency Cepstrum Coefficients* (MFCC) yang menghitung koefisien cepstral dengan mempertimbangkan persepsi sistem pendengaran manusia terhadap frekuensi suara. Davis dan Mermelstein memperlihatkan bahwa MFCC sebagai teknik ekstraksi ciri memberikan hasil pengenalan yang tinggi terutama dalam hal jumlah, bentuk, dan lebar filter serta cara membentuk intervalnya.

MFCC sebagai pengekstraksi ciri dan teknik untuk parameterisasi sinyal suara telah banyak digunakan pada berbagai bidang area pemrosesan suara, terutama pada sistem identifikasi suara pembicara. Sinyal dibaca frame demi frame dan dilakukan *windowing* untuk setiap frame untuk berikutnya dilakukan transformasi *fourier*. Hasil dari transformasi *fourier* selanjutnya dihitung spektrum mel menggunakan sejumlah filter yang dibentuk sehingga jarak antara pusat filter konstan pada ruang frekuensi mel. Dari literature yang ada, skala mel dibentuk untuk

mengikuti persepsi sistem pendengaran manusia yang bersifat linear untuk frekuensi rendah dan logaritmik untuk frekuensi tinggi dengan batas ada nilai frekuensi akustik sebesar 1000Hz. Proses ini dikenal dengan nama *Mel-Frequency Wrapping*. Koefisien MFCC merupakan hasil transformasi *Cosinus* dari spectrum mel tersebut dan dipilih koefisien. Transformasi kosinus berfungsi untuk mengembalikan domain, dari frekuensi ke domain waktu. Berikut disajikan detail proses setiap tahapan serta teori yang melandasinya yang dimulai dengan transformasi fourier.

2.7.1 *Mel Frequency Cepstrum Coefficients (MFCC)*

MFCC merupakan salah satu metode yang banyak digunakan dalam bidang *speech technology*, baik *speaker recognition* maupun *speech recognition* untuk melakukan *feature extraction*. Proses ini digunakan untuk mengkonversikan sinyal suara menjadi beberapa parameter. Beberapa keunggulan dari metode ini :

- a. Mampu menangkap karakteristik suara yang sangat penting bagi pengenalan suara. Atau dengan kata lain mampu menangkap informasi-informasi penting yang terkandung dalam sinyal suara.
- b. Menghasilkan data seminimal mungkin tanpa menghilangkan informasi-informasi penting yang ada.
- c. Mereplika organ pendengaran manusia dalam melakukan persepsi sinyal suara.

Perhitungan yang dilakukan dalam MFCC menggunakan dasar-dasar perhitungan *short-term analysis*. Hal ini dilakukan mengingat sinyal suara yang bersifat *quasi stationary*. Karakteristik sinyal suara pada metode MFCC akan menunjukkan sinyal suara yang *stationary* pada periode waktu yang panjang dan

sesuai dengan perubahan kata yang diucapkan. MFCC, *feature extraction* merupakan adaptasi dari sistem pendengaran manusia, dimana sinyal suara akan di filter secara linier untuk frekuensi rendah (dibawah 1000Hz) dan secara logaritmik untuk frekuensi tinggi (diatas 100Hz). Proses pengenalan suara yang diterapkan pada metode ini memiliki beberapa tahapan yang terdiri dari *Pre Emphasize*, *Frame Blocking*, *Windowing*, *Fast Fourier Transform*, *Mel Frequency Warping*, dan *Discrete Cosine Transform* [21] .

Pre-emphasize Fiter

Pre-emphasize Filter merupakan satu jenis filter yang sering digunakan sebelum sebuah signal diproses lebih lanjut. Filter ini mempertahankan frekuensi-frekuensi tinggi pada sebuah spektrum, yang tereliminasi pada saat proses produksi suara. Filter ini juga digunakan untuk memperbaiki signal dari gangguan noise, sehingga dapat meningkatkan tingkat akurasi dari proses *feature extraction*.

Tujuan dari *Pre-emphasize Filtering* adalah

- a. Mengurangi *noise ratio* pada signal sehingga dapat meningkatkan kualitas signal.
- b. Menyeimbangkan spectrum dari *voiced sound*. Pada saat memproduksi *voiced sound* glottis manusia menghasilkan sekitar -12Db octave slope. Namun ketika energy akustik tersebut dikeluarkan melalui bibir, terjadi peningkatan sebesar +6. Sehingga signal yang terekam oleh microphone adalah sekitar -6 Db octave slope. Bentuk umum dalam persamaan 2.15 yang digunakan dalam *Pre-emphasize filter* adalah sebagai berikut :

$$y[n] = s[n] - \alpha s[n - 1], 0 \leq \alpha \leq 1.0 \dots \dots \dots (2.15)$$

Dimana :

$y[n]$ = *signal* hasil *pre-emphasize filter*

$s[n]$ = *signal* sebelum *pre-emphasize filter*

α = konstanta *pre-emphasize filter*

Frame blocking

Hasil Perekaman suara merupakan sinyal analog yang berada dalam domain waktu yang bersifat varian time (Fungsi yang tergantung oleh waktu). Karena signal suara terus mengalami perubahan akibat adanya pergeseran artikulasi dari organ produksi vocal, signal harus diproses secara short segments (short frame). Panjang *frame* yang biasanya digunakan untuk pemrosesan *signal* adalah antara 10-30 milidetik. Panjang *frame* yang digunakan sangat mempengaruhi keberhasilan dalam analisa spektral. Di satu sisi, ukuran dari *frame* harus sepanjang mungkin untuk dapat menunjukkan resolusi frekuensi yang baik. Tetapi di lain sisi, ukuran *frame* juga harus cukup pendek untuk dapat menunjukkan resolusi waktu yang baik. Proses *frame* ini dilakukan terus sampai seluruh *signal* dapat diproses sehingga terdapat sinyal suara. Hal ini dapat diperlihatkan pada persamaan 2.16. Selain itu, proses ini umumnya dilakukan secara *overlapping* untuk setiap *frame*-nya. Panjang daerah *overlap* yang umum digunakan adalah kurang lebih 30% sampai 50% dari panjang *frame*. *Overlapping* dilakukan untuk menghindari hilangnya ciri atau karakteristik suara pada perbatasan perpotongan setiap *frame*.

$$x_1 = x(Ml + 1), x(Ml+2), \dots, x(Ml + N) \dots \dots \dots (2.16)$$

Dengan

x_l adalah isyarat suara pada frame ke- l

x adalah isyarat suara

M adalah jarak antar *frame* yang berdekatan

N adalah jumlah cuplikan isyarat suara pada sebuah *frame*

$l = 1, 2, \dots, L$; dengan L adalah jumlah *frame*

Windowing

Proses *framing* dapat menyebabkan terjadinya kebocoran spektral (*spectral leakage*) atau *aliasing*. *Aliasing* adalah *signal* baru dimana memiliki frekuensi yang berbeda dengan *signal* aslinya. Efek ini dapat terjadi karena rendahnya jumlah *sampling rate*, ataupun karena proses *frame blocking* yang menyebabkan *signal* menjadi *discontinue*. Untuk mengurangi kemungkinan terjadinya kebocoran spektral, maka hasil dari proses *framing* harus melewati proses *window*. Sebuah fungsi *window* yang baik harus menyempit pada bagian *main lobe* dan melebar pada bagian *side lobe*-nya. Berikut ini adalah representasi dari fungsi *window* terhadap *signal* suara yang diinputkan pada persamaan 2.17.

$$Y(n) = X(n) \times W(n) \dots\dots\dots(2.17)$$

$Y(n)$ = nilai sampel *signal* hasil *windowing*

$X(n)$ = nilai sampel dari *frame signal* ke i

$w(n)$ = fungsi *window*

N = *frame size*, merupakan kelipatan 2

Ada banyak fungsi *window*, namun yang paling sering digunakan dalam aplikasi *speaker recognition* adalah *hamming window*. Fungsi *window* ini menghasilkan *sidelobe level* yang tidak terlalu tinggi (kurang lebih -43 dB), selain itu noise yang dihasilkanpun tidak terlalu besar.

Fungsi Hamming window dapat diperlihatkan pada persamaan 2.18.

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N-1 \dots\dots\dots(2.18)$$

Dengan :

$w(n)$ adalah Hamming window

$n = 1, 2, \dots, N$; N adalah jumlah isyarat suara dalam sebuah frame

Transformasi Fourier

Transformasi *Fourier* dilakukan untuk setiap *frame* yang telah dibentuk. Tujuannya adalah untuk merubah isyarat dari domain waktu menjadi domain frekuensi. Karena hasil dari transformasi *fourier* ini adalah simetris, maka hanya diambil setengah dari hasil transformasinya saja. Setelah itu hasil ini dikalikan dengan konjugatannya untuk mendapatkan spektrum isyarat suara. Adapun untuk menghitung transformasi *fourier* untuk setiap frame digunakan persamaan 2.19 dibawah ini.

$$Xl(k) = \sum_{n=0}^{N-1} \widehat{x_l}(n) e^{-j2\pi nk/N} \dots\dots\dots(2.19)$$

Dengan :

Xl adalah hasil LFT dari isyarat suara pada *frame* l

$\widehat{x_l}$ adalah hasil windowing isyarat suara pada frame ke-l

N = jumlah cuplikan isyarat dalam sebuah *frame*

$$k = 0, 1, \dots, N-1, \quad j = \sqrt{-1}$$

Dengan $Mag(Xl(k))$ adalah magnitude kompleks isyarat tutur pada *frame* 1, $real(Xl(k))$ adalah bagian real hasil transformasi *fourier frame* 1, $imag(Xl(k))$ adalah bagian imajiner hasil transformasi *fourier frame* 1, dan $k = 0, 1, \dots, N/2$; dan N adalah jumlah titik frekuensi hasil dari transformasi *fourier*.

Analisa *fourier*

Analisa *fourier* adalah sebuah metode yang memungkinkan untuk melakukan analisa terhadap *spectral properties* dari *signal* yang diinputkan. Representasi dari *spectral properties* sering disebut sebagai *spectrogram*. Dalam *spectrogram* terdapat hubungan yang sangat erat antara waktu dan frekuensi. Hubungan antara frekuensi dan waktu adalah hubungan berbanding terbalik. Bila resolusi waktu yang digunakan tinggi, maka resolusi frekuensi yang dihasilkan akan semakin rendah.

Discrete Fourier Transform (DFT)

DFT merupakan perluasan dari transformasi *fourier* yang berlaku untuk *signal-signal* diskrit dengan panjang yang terhingga. Semua *signal* periodik terbentuk dari gabungan *signal-signal* sinusoidal yang menjadi satu yang dapat dirumuskan pada persamaan 2.20.

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt \dots\dots\dots (2.20)$$

Dengan :

$X(\omega)$ adalah hasil DFT yaitu isyarat dalam domain frekuensi

ω adalah frekuensi yang telah dinormalisasi antara 0 dan 2π

$x(t)$ adalah isyarat analog dalam domain waktu

Pada umumnya, kedua fungsi $x(t)$ dan $x(\omega)$ merupakan fungsi kompleks dari variable t dan variable ω . Isyarat kontinyu $x(t)$ diubah menjadi isyarat diskrit $x(nT)$ dengan cara mengambil sampling setiap T detik. Jika T tidak dituliskan maka isyarat diskrit dapat ditulis dengan notasi $x(n)$ sehingga persamaan transformasi *Fourier* untuk isyarat diskrit dapat ditulis pada persamaan 2.21.

$$X(\omega) = \int_{-\infty}^{\infty} x(n) \cdot e^{-j\omega n} dt \dots\dots\dots(2.21)$$

Dengan :

ω melambangkan frekuensi yang telah dinormalisasi antara 0 dan 2π

$X(\omega)$ adalah hasil DFT yaitu isyarat dalam domain frekuensi

$x(n)$ adalah isyarat diskrit dalam domain waktu

$X(\omega)$ menjadi periodik akibat langsung dari $x(n)$ yang disampling secara periodik pula. Asumsikan bahwa isyarat $x(n)$ terdiri dari N sampel, karena tidak ada batasan interval dari N maka diasumsikan isyarat tersebut merupakan isyarat yang periodik (berulang setiap periode waktu tertentu). Karena asumsi inilah maka transformasi fourier dalam sistem waktu diskrit dapat ditulis pada persamaan 2.22 dan persamaan 2.23.

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn} \dots\dots\dots(2.22)$$

Dengan :

$X(k)$ adalah hasil DFT isyarat dalam domain frekuensi

$x(n)$ adalah isyarat diskrit dalam domain waktu

$$W_N^{kn} = e^{-j2\pi nk/N}, \quad k = 0, 1, \dots, N-1 \dots\dots\dots (2.23)$$

N adalah jumlah sampel dari isyarat diskrit

DFT menghasilkan serangkaian N buah nilai yang berindeks k di dalam domain frekuensi yang merupakan transformasi dari isyarat domain waktu yang berindeks n .

Dari hasil tersebut, $X(K)$ dan $X(N-k)$ merupakan konjugasi kompleks yang dapat diperlihatkan pada persamaan 2.24 yang disebut sebagai DFT sebanyak N buah sample.

Karena magnitude dari konjugasi kompleks adalah sama, maka didapatkan untuk k bernilai 0 sampai $N/2$.

$$X(k) = |X(N-k)| \dots\dots\dots (2.24)$$

Dengan demikian, nilai hasil transformasi dalam domain frekuensi yang digunakan untuk analisis isyarat hanya nilai yang berindeks 0 sampai $N/2$ saja. W_N biasa disebut *twiddle factor*. Karena perhitungan menggunakan bilangan kompleks maka operasi penjumlahan maupun perkalian membutuhkan kira-kira N^2 untuk N point, maka total operasi aritmatika yang dibutuhkan untuk nilai N tertentu semakin lama semakin besar seiring bertambahnya nilai N . Oleh karena itu dibutuhkan metode alternatif untuk dapat menghitung DFT secara efisien. Metode untuk mengefisienkan perhitungan DFT ini biasanya berdasarkan sifat simetris dan periodic *twiddle factor*. Dengan rumus diatas, suatu *signal* suara dalam domain waktu dapat dicari frekuensi

pembentuknya. Tujuan penggunaan analisa *fourier* pada data suara yaitu untuk merubah data dari domain waktu menjadi data spektrum di domain frekuensi. Untuk pemrosesan *signal* suara, hal ini menguntungkan karena data pada domain frekuensi dapat diproses dengan lebih mudah dibandingkan data pada domain waktu, karena pada domain frekuensi, keras lemahnya suara tidak seberapa berpengaruh. Untuk mendapatkan spektrum dari sebuah *signal* dengan DFT diperlukan N buah sampel data berurutan pada domain waktu, yaitu $x[m]$ sampai $x[m+N-1]$. Data tersebut dimasukkan dalam fungsi DFT maka akan menghasilkan N buah data. Namun karena hasil dari DFT adalah simetris, maka hanya $N/2$ data yang diambil sebagai spektrum.

Fast Fourier Transform (FFT)

Perhitungan DFT secara langsung dalam komputerisasi dapat menyebabkan proses perhitungan yang sangat lama. Hal itu disebabkan karena dengan DFT, dibutuhkan N^2 perkalian bilangan kompleks. Karena itu dibutuhkan cara lain untuk menghitung DFT dengan cepat. Hal itu dapat dilakukan dengan menggunakan algoritma *fast fourier transform* (FFT) dimana FFT menghilangkan proses perhitungan yang kembar dalam DFT. *Fast Fourier Transform* (FFT) dikembangkan oleh *Cooley* dan *Tukey* pada tahun 1965. Algoritma FFT merupakan penyederhanaan dari DFT yang memiliki persyaratan jumlah data harus merupakan bilangan 2^n untuk $n=0,1,2,\dots$. Waktu komputasi DFT memiliki kompleksitas N^2 sedangkan FFT memiliki kompleksitas $Np/2$ dengan $p = \log N$. Sehingga FFT lebih cepat daripada DFT dengan rasio FFT terhadap DFT adalah seperti pada persamaan 2.25.

$$\frac{N^2}{Np/2} = \frac{2N}{p} \dots\dots\dots (2.25)$$

Dilihat dari metode yang digunakan, FFT dibagi menjadi 2, yaitu metode DIT (*Decimation In Time*) dan metode DIF (*Decimation In Frequency*), namun keduanya memiliki fungsi yang sama, yaitu untuk mentransformasikan isyarat menjadi frekuensi dasarnya. *Decimation* adalah proses pembagian isyarat menjadi beberapa bagian yang lebih kecil yang bertujuan untuk memperoleh waktu proses yang lebih cepat. Jika input isyarat pada *time domain* dari *N-points* adalah $x(n)$, langkah awal yang dilakukan adalah dengan memisahkan menjadi 2 bagian yang sama ($N/2$ *points*). Proses FFT diperlukan untuk persiapan ekstraksi ciri MFCC dapat diperlihatkan pada persamaan 2.26.

$$S[k] = \sum_{n=0}^{N-1} s[n] e^{-j2\pi nk/N}, 0 \leq k \leq N-1 \dots\dots\dots (2.26)$$

N = jumlah sampel yang akan diproses ($N \in \mathbb{N}$)

$s[n]$ = nilai sampel *signal* bernilai $k = N/2, k \in \mathbb{N}$

K = variable frekuensi discrete

Dengan rumus diatas, suatu *signal* suara dalam domain waktu dapat dicari frekuensi pembentuknya. Tujuan penggunaan analisa *fourier* pada data suara yaitu untuk merubah data dari domain waktu menjadi data spektrum di domain frekuensi. Untuk pemrosesan *signal* suara, hal ini sangatlah menguntungkan karena data pada domain frekuensi dapat diproses dengan lebih mudah dibandingkan data pada domain waktu, karena pada domain frekuensi, keras lemahnya suara tidak seberapa

berpengaruh. Untuk mendapatkan spektrum dari sebuah *signal* dengan DFT diperlukan N buah sampel data berurutan pada domain waktu, yaitu $x[m]$ sampai $x[m+N-1]$. Data tersebut dimasukkan dalam fungsi DFT maka akan menghasilkan N buah data. Namun karena hasil dari DFT adalah simetris, maka hanya N/2 data yang diambil sebagai spektrum.

2.7.2 *Mel Frequency Wrapping* (MFW)

Mel Frequency Wrapping umumnya dilakukan dengan menggunakan *Filterbank*. *Filterbank* adalah salah satu bentuk dari *filter* yang dilakukan dengan tujuan untuk mengetahui ukuran energi dari *frequency band* tertentu dalam *signal* suara. *Filterbank* dapat diterapkan baik pada domain waktu maupun pada domain frekuensi, tetapi untuk keperluan MFCC, *filterbank* harus diterapkan dalam domain frekuensi. *Filterbank* menggunakan representasi konvolusi dalam melakukan *filter* terhadap *signal*. *Mel-Frequency* adalah frekuensi yang linier di bawah 1 kHz dan logaritmik di atas 1 kHz. Relasi pendekatan empiris untuk menghitung skala Mel untuk frekuensi yang diberikan dalam Hz ditunjukkan dengan persamaan 2.27.

$$\hat{f}_{\text{mel}} = 2595 \cdot \log_{10} \left(1 + \frac{f_{\text{lin}}}{700} \right) \dots\dots\dots (2.27)$$

Persepsi manusia terhadap frekuensi dari *signal* suara tidak mengikuti *linear scale*. Frekuensi yang sebenarnya dalam sebuah *signal* akan diukur manusia secara subyektif dengan menggunakan *mel scale*. *Mel frequency scale* adalah *linear frequency scale* pada frekuensi dibawah 1000 Hz, dan merupakan *logarithmic scale* pada frekuensi diatas 1000 Hz.

2.7.3 *Dynamic Time Warping (DTW)*

DTW adalah metode yang digunakan untuk menghitung jarak antara dua data *time series*. Keunggulan DTW dari metode jarak yang lainnya adalah mampu menghitung jarak dari dua vektor data dengan panjang berbeda. Satu masalah yang cukup rumit dalam *speech recognition* (pengenalan wicara) adalah proses perekaman yang terjadi seringkali berbeda durasinya biarpun kata atau kalimat yang diucapkan sama bahkan untuk satu suku kata yang sama atau vocal yang sama seringkali proses perekaman terjadi dalam durasi yang berbeda. Sebagai akibatnya proses matching antara sinyal uji dengan sinyal referensi (*template*) seringkali tidak menghasilkan nilai yang optimal [22].

Sebuah teknik yang cukup populer di awal perkembangan teknologi pengolahan sinyal wicara adalah dengan memanfaatkan sebuah teknik *dynamic-programming* yang juga lebih dikenal sebagai DTW. Teknik ini ditujukan untuk mengakomodasi perbedaan waktu antara proses perekaman saat pengujian dengan yang tersedia pada *template* sinyal referensi. Prinsip dasarnya adalah dengan memberikan sebuah rentang '*steps*' dalam ruang (dalam hal ini sebuah frame-frame waktu dalam sample, frame-frame waktu dalam template) dan digunakan untuk mempertemukan lintasan yang menunjukkan *local match* terbesar (kemiripan) antara time frame yang lurus. Total '*similarity cost*' yang diperoleh dengan algorithm ini merupakan sebuah indikasi seberapa bagus *sample* dan *template* ini memiliki kesamaan yang selanjutnya akan dipilih *best-matching template*. DCT merupakan langkah terakhir dari proses utama MFCC *feature extraction*.

2.7.4 Discrete Cosine Transform (DCT)

Konsep dasar dari DCT adalah mendekorelasikan *mel spectrum* sehingga menghasilkan representasi yang baik dari property spektral local. Pada dasarnya konsep dari DCT sama dengan *inverse fourier transform*. Namun hasil dari DCT mendekati PCA (*principle component analysis*). PCA adalah metode static klasik yang digunakan secara luas dalam analisa data dan kompresi. Hal inilah yang menyebabkan seringkali DCT menggantikan *inverse fourier transform* dalam proses MFCC *feature extraction*. Formula DCT pada persamaan 2.28 dapat diperlihatkan sebagai berikut.

$$C_n = \sum_{k=1}^K (\log s_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right], n = 1, 2, \dots, K \quad \dots\dots\dots(2.28)$$

S_k = keluaran dari proses *filterbank* pada *index* k

K = jumlah koefisien yang diharapkan

Koefisien ke nol dari DCT pada umumnya akan dihilangkan, walaupun sebenarnya mengindikasikan energi dari frame *signal* tersebut. Hal ini dilakukan, berdasarkan penelitian-penelitian yang pernah dilakukan.

2.8 Jaringan Syaraf Tiruan (JST)

Jaringan syaraf tiruan sederhana pertama kali diperkenalkan oleh McCulloch dan Pitts di tahun 1943. McCulloch dan Pitts menyimpulkan bahwa kombinasi beberapa neuron sederhana menjadi sebuah sistem neural akan meningkatkan kemampuan komputasinya. Bobot dalam jaringan yang diusulkan oleh McCulloch

dan Pitts diatur untuk melakukan fungsi logika sederhana. Fungsi aktivasi yang dipakai adalah fungsi threshold [23]. Jaringan syaraf tiruan adalah sistem pemrosesan informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologi. JST dibentuk sebagai generalisasi model matematika dari jaringan syaraf biologi, dengan asumsi bahwa:

- a. Pemrosesan informasi terjadi pada banyak elemen sederhana (neuron).
- b. Sinyal dikirimkan diantara neuron-neuron melalui penghubung-penghubung.
- c. Penghubung antar neuron memiliki bobot yang akan memperkuat atau memperlemah sinyal
- d. Untuk menentukan output, setiap neuron menggunakan fungsi aktivasi (biasanya bukan fungsi linear) yang dikenakan pada jumlah input yang diterima. Besarnya output ini selanjutnya dibandingkan dengan suatu batas ambang.

Jaringan syaraf tiruan merupakan salah satu representasi buatan dari otak manusia yang mencoba untuk mensimulasikan proses pembelajaran pada otak manusia dengan menggunakan program komputer yang mampu menyelesaikan sejumlah proses perhitungan selama proses pembelajaran. Adapun beberapa aplikasi Jaringan syaraf tiruan (ANN) adalah sebagai berikut [24] :

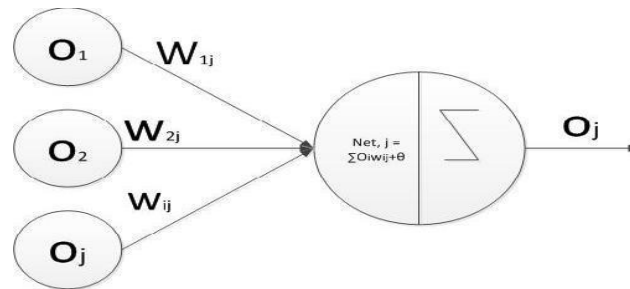
1. Pengenalan pola (Pattern Recognition)

Jaringan syaraf tiruan sangat cocok digunakan untuk mengenali pola, misalnya mengenali pola huruf, angka, sidik jari atau bahkan penyakit. Hal ini mirip dengan

cara kerja otak manusia ketika membedakan dan mengenali perubahan wajah setiap orang yang pernah dilihatnya.

2. Signal Processing

Mengenali pola suatu signal atau noise dalam suatu perubahan nada suara misalnya. Secara sederhana Jaringan syaraf tiruan terdiri dari module yang saling keterkaitan secara sederhana yang disebut juga dengan *Neuron* dimana analogi dari syaraf biologi pada otak.



Gambar 2.6 Node sederhana

Node sederhana seperti pada Gambar 2.6 memperlihatkan bahwa *Neural Network* terdiri dari satuan-satuan pemroses berupa neuron. Y sebagai output menerima input dari $x_1, x_2, x_3, \dots, x_n$ dengan bobot $W_1, W_2, W_3, \dots, W_n$. Hasil penjumlahan seluruh impuls neuron dibandingkan dengan nilai ambang tertentu melalui fungsi aktivasi f setiap neuron.

3. Peramalan

Jaringan syaraf tiruan juga dapat dipakai untuk meramalkan apa yang akan terjadi di masa yang akan datang berdasarkan pola kejadian yang ada di masa lampau. Ini dapat dilakukan mengingat kemampuan jaringan syaraf tiruan untuk mengingat dan membuat generalisasi dari apa yang sudah ada sebelumnya.

Disamping area-area tersebut, jaringan syaraf tiruan juga dilaporkan dapat menyelesaikan masalah dalam bidang control, kedokteran dan lain-lain. Meskipun banyak aplikasi menjanjikan yang dapat dilakukan oleh jaringan syaraf tiruan, namun jaringan syaraf tiruan juga memiliki beberapa keterbatasan umum. Pertama adalah ketidakakuratan hasil yang diperoleh. Jaringan syaraf tiruan bekerja berdasarkan pola yang terbentuk pada inputnya. Jaringan syaraf tiruan atau *neural network* merupakan suatu metode komputasi yang meniru sistem jaringan syaraf biologis manusia. Metode ini menggunakan elemen perhitungan *non-linier* dasar yang disebut *neuron* yang diorganisasikan sebagai jaringan yang saling berhubungan, sehingga mirip dengan jaringan syaraf manusia. Jaringan syaraf tiruan dibentuk untuk memecahkan suatu masalah tertentu seperti pengenalan pola atau klasifikasi karena proses pembelajaran [25].

2.8.1 Sifat Jaringan Syaraf Tiruan (JST)

Jaringan syaraf tiruan merupakan sistem yang bersifat “*fault tolerant*” layaknya *neuron* biologi. Dikatakan fault toleran jika sinyal *input* yang berbeda dari yang pernah diterima sebelumnya dan mampu bekerja meskipun beberapa *neuronnya* tidak mampu bekerja dengan baik. Jika sebuah *neuron* rusak maka *neuron* lain dapat dilatih untuk menggantikan fungsi *neuron* yang rusak tersebut. JST juga memiliki sifat yang adaptif, yaitu dapat belajar dari data-data sebelumnya dan mengenal pola data yang selalu berubah. Sistem JST tidak terprogram disebabkan semua keluaran atau kesimpulan yang ditarik oleh jaringan didasarkan pada pengalamannya selama

mengikuti proses pembelajaran/pelatihan. Sistem pelatihan pada JST digunakan untuk mencapai kemampuan keseimbangan baik kemampuan memorisasi maupun kemampuan generalisasi. Kemampuan memorisasi merupakan suatu kemampuan untuk mengambil kembali secara sempurna sebuah pola yang telah dipelajari dan Kemampuan generalisasi merupakan kemampuan untuk menghasilkan respon yang bisa diterima terhadap pola-pola masukan yang serupa dengan pola-pola sebelumnya telah. Hal ini sangat bermanfaat bila JST memiliki informasi baru yang belum pernah dipelajari, namun tetap dapat memberikan tanggapan yang baik dan memberikan keluaran yang paling mendekati.

2.8.2 Jaringan Syaraf Biologis Manusia

Jaringan syaraf biologis manusia terdiri dari kumpulan sel-sel syaraf (neuron) yang mempunyai tugas untuk mengolah informasi. Neuron terdiri dari 3 komponen utama, yaitu *dendrit* sebagai penerima informasi, badan sel (*soma*) sebagai tempat pengolahan informasi, dan akson (*neurit*) untuk mengirimkan impuls-impuls ke sel syaraf lainnya. Ketika *dendrit* menerima sinyal dari neuron lain, sinyal tersebut berupa implus elektrik yang dikirim melalui celah sinaptik melalui proses kimiawi. Sinyal tersebut di modifikasi (diperkuat/diperlemah) di celah sinaptik. Berikutnya, *Soma* menjumlahkan sinyal-sinyal yang masuk. Jika jumlah tersebut cukup kuat dan melebihi batas ambang, maka sinyal tersebut akan diteruskan ke sel lain melalui axon.

Frekuensi penerusan sinyal berbeda antara satu sel dengan yang lain. Sistem Neuron biologi merupakan sistem yang tolerant terhadap dua hal. Pertama, manusia dapat mengenali sinyal input yang berbeda dari yang pernah diterima sebelumnya.

Kedua, otak manusia tetap mampu belajar meskipun beberapa neuronnya tidak mampu bekerja dengan baik. Jika sebuah neuron rusak maka neuron lain dapat menggantikan fungsi sel yang rusak.

2.8.3 Karakteristik Jaringan Syaraf Tiruan

Sistem Jaringan syaraf tiruan memiliki tiga karakteristik utama , yaitu :

a. **Arsitektur Jaringan**

Merupakan pola keterhubungan antara neuron. Keterhubungan neuron-neuron inilah yang membentuk suatu jaringan.

b. **Algoritma Jaringan**

Merupakan metode untuk menentukan nilai bobot hubungan. Ada dua metode pada algoritma jaringan saraf tiruan, yaitu metode pelatihan dan metode pengenalan aplikasi.

c. **Fungsi Aktivasi**

Merupakan fungsi untuk menentukan nilai keluaran berdasarkan nilai total masukan pada neuron. Fungsi aktivasi suatu algoritma jaringan dapat berbeda dengan fungsi aktivasi algoritma jaringan lain.

2.8.4 Konsep Dasar Jaringan Syaraf Tiruan

Pembagian arsitektur jaringan syaraf tiruan bisa dilihat dari kerangka kerja dan skema interkoneksi. Kerangka kerja jaringan syaraf tiruan bisa dilihat dari jumlah lapisan (layer) dan jumlah node pada setiap lapisan.

Lapisan-lapisan penyusun jaringan saraf tiruan dapat dibagi menjadi tiga, yaitu:

a. Lapisan Input

Node-node didalam lapisan input disebut unit-unit input. Unit-unit input menerima input dari dunia luar. Input yang dimasukkan merupakan penggambaran dari suatu masalah.

b. Lapisan Tersembunyi

Node-node di dalam lapisan tersembunyi disebut unit-unit tersembunyi. Output dari lapisan ini tidak secara langsung dapat diamati.

c. Lapisan Output

Node-node pada lapisan output disebut unit-unit output. Keluaran atau output dari lapisan ini merupakan output jaringan saraf tiruan terhadap suatu permasalahan.

Jaringan saraf tiruan dibagi kedalam tiga arsitektur, yaitu:

a. Jaringan lapis tunggal

Jaringan yang memiliki arsitektur jenis ini hanya memiliki satu buah lapisan bobot koneksi. Jaringan lapisan-tunggal terdiri dari unit-unit input yang menerima sinyal dari dunia luar dan unit-unit output bisa membaca respons dari jaringan saraf tiruan tersebut.

b. Jaringan Multilapis

Merupakan jaringan dengan satu atau lebih lapisan tersembunyi. Multilayer net ini memiliki kemampuan lebih dalam memecahkan masalah bila dibandingkan dengan single-layer net, namun pelatihannya mungkin lebih rumit.

c. Jaringan Kompetitif

Pada jaringan ini sekumpulan neuron bersaing untuk mendapatkan hak menjadi aktif.

2.8.4 Istilah-istilah Jaringan Syaraf Tiruan

Berikut ini merupakan beberapa istilah jaringan saraf tiruan yang sering ditemui;

a. Neuron atau Node atau Unit

Sel syaraf tiruan yang merupakan elemen pengolahan jaringan syaraf tiruan. Setiap neuron menerima data input, memproses input tersebut (melakukan sejumlah perkalian dengan melibatkan *summation function* dan fungsi aktivasi dan mengirimkan hasilnya berupa output.

b. Jaringan

Kumpulan neuron yang saling terhubung dan membentuk lapisan.

c. Input atau masukan

Berkorespon dengan sebuah artikel tunggal dari sebuah pola atau data lain dari dunia luar. Sinyal-sinyal input ini kemudian diteruskan kelapisan selanjutnya.

d. Output atau keluaran

Solusi atau hasil pemahaman jaringan terhadap data input. Tujuan pembangunan jaringan saraf tiruan dalam menghadapi masalah-masalah yang kompleks.

e. Bobot

Bobot dalam JST merupakan nilai matematis dari koneksi, yang mentransfer data dari satu lapisan ke lapisan lainnya.

f. Lapisan tersembunyi (hidden layer)

Lapisan yang tidak langsung berinteraksi dengan dunia luar. Lapisan ini memperluas jaringan saraf tiruan untuk menghadapi masalah-masalah yang kompleks.

g. *Summation function*

Fungsi yang digunakan untuk mencari rata-rata bobot dari semua elemen input. Yang sederhana adalah dengan mengalikan setiap nilai input (X_j) dengan bobotnya (W_{ij}) dan menjumlahkannya (penjumlahan berbobot atau S_j).

h. Fungsi aktivasi atau fungsi transfer

Fungsi yang menggambarkan hubungan antara tingkat aktivasi internal (*summation function*) yang mungkin berbentuk linear atau non linear. Yang populer digunakan ialah fungsi sigmoid yang memiliki beberapa varian yaitu sigmoid logaritma, sigmoid biner, sigmoid bipolar, dan sigmoid tangent.

i. Paradigma pembelajaran

Cara berlangsungnya pembelajaran atau pelatihan jaringan syaraf tiruan yang terawasi (*supervised learning*), tidak terawasi (*unsupervised learning*) atau merupakan gabungan keduanya (*hybrid*).

2.8.6 Aplikasi Jaringan Syaraf Tiruan

Aplikasi yang sering menggunakan jaringan syaraf tiruan diantaranya :

a. Pengenalan Pola

Jaringan syaraf tiruan sering digunakan untuk pengenalan pola yang sudah sedikit berubah. Misalnya pengenalan huruf, angka, suara atau tanda tangan. Mirip dengan otak manusia yang masih mampu mengenali orang yang sudah beberapa waktu tidak dijumpai.

b. *Signal Processing*

Jaringan syaraf tiruan (ADALINE) dapat digunakan untuk menekan noise dalam saluran telepon.

c. Peramalan

Jaringan syaraf tiruan dapat meramalkan apa yang akan terjadi di masa akan datang berdasarkan pembelajaran pola kejadian yang ada di masa lalu.

2.8.7 *Multi Layer Perceptron*

MLP juga dikenal setara *Multilayer Feedforward Neural Network* (MLFNN) adalah salah satu yang paling terkenal dan banyak digunakan sebagai type model ANN oleh karena arsitektur dan perbandingan algoritma yang sederhana. Dapat juga digunakan sebagai fungsi pembangkit secara menyeluruh, bahkan cocok digunakan untuk jenis aplikasi yang besar. *Multi layer perceptron* tersusun oleh seperangkat sensor modern yang dikelompokkan dalam tiga tingkatan lapisan yang terdiri dari lapisan modul masukan, satu atau lebih perantara atau lapisan tersembunyi dari modul perhitungan dan lapisan keluaran modul yang mengkalkulasikan keluaran dari jaringan.

Semua lapisan berurutan terhubung secara lengkap. hubungan antara modul berbatasan lapisan relay sinyal keluaran dari satu lapisan ke berikutnya. Sebagai contoh, gambar diatas mempunyai 4 vektor dimensi, diikuti oleh 3 lapisan tersembunyi dan yang terakhir lapisan keluaran dimana terdiri dari 1 modul. Jaringan ANN disebut dengan sebagai jaringan 4-3-1. Jaringan syaraf tiruan juga memiliki sejumlah besar kelebihan dibandingkan dengan metoda perhitungan lainnya (sistem pakar, statistik, dll), yaitu

- a. Kemampuan mengakuisisi pengetahuan walaupun dalam kondisi adanya gangguan dan ketidakpastian. Hal ini dikarenakan JST mampu melakukan generalisasi, abstraksi, dan ekstraksi terhadap properti statistik dari data.
- b. Kemampuan merepresentasikan pengetahuan secara fleksibel. JST dapat menciptakan sendiri representasi melalui pengaturan diri sendiri atau kemampuan belajar (*self organizing*).
- c. Kemampuan mentolerir suatu distorsi (*error/fault*). Dimana gangguan kecil pada data dapat dianggap hanya *noise* (guncangan) belaka. Kemampuan memproses pengetahuan secara efisien karena memakai sistem paralel, maka waktu yang diperlukan untuk mengoperasikannya menjadi lebih singkat.
- d. Kemampuan untuk memperoleh pengetahuan melalui pembelajaran dari pengalaman.

Walaupun dengan segudang kelebihan yang dimiliki, jaringan syaraf tiruan tetap mempunyai sejumlah batasan. Misal : Kekurangmampuannya dalam melakukan operasi-operasi numerik dengan presisi tinggi, operasi algoritma aritmatik, operasi

logika, dan operasi simbolis serta lamanya proses pelatihan yang kadang-kadang membutuhkan waktu sehari-hari untuk jumlah data yang besar. Hal itu terjadi karena sulitnya mengukur performansi sebenarnya dari jaringan syaraf tiruan itu sendiri. Saat ini implementasi jaringan syaraf tiruan sudah cukup luas digunakan mulai dari teknologi militer, satelit ruang angkasa, bisnis dan kesehatan.

2.9 Backpropagation

Jaringan syaraf tiruan *backpropagation* pertama kali diperkenalkan oleh Rumelhart, Hinton dan William pada tahun 1986, kemudian Rumelhart dan McClelland mengembangkannya pada tahun 1988. Algoritma *backpropagation* untuk *neural network* umumnya diterapkan pada perceptron berlapis banyak (*multilayer perceptrons*). Perceptron paling tidak mempunyai bagian input, bagian output dan beberapa lapis yang berada di antara input dan output. Lapis di tengah ini, yang juga dikenal dengan lapis tersembunyi (*hidden layers*), bisa satu, dua, tiga dst. Dalam praktek, banyaknya *hidden layer* paling banyak adalah tiga lapis. Dengan tiga lapis ini hampir semua permasalahan dalam dunia industri telah bisa diselesaikan. Output lapis terakhir dari *hidden layer* langsung dipakai sebagai output dari *neural network*.

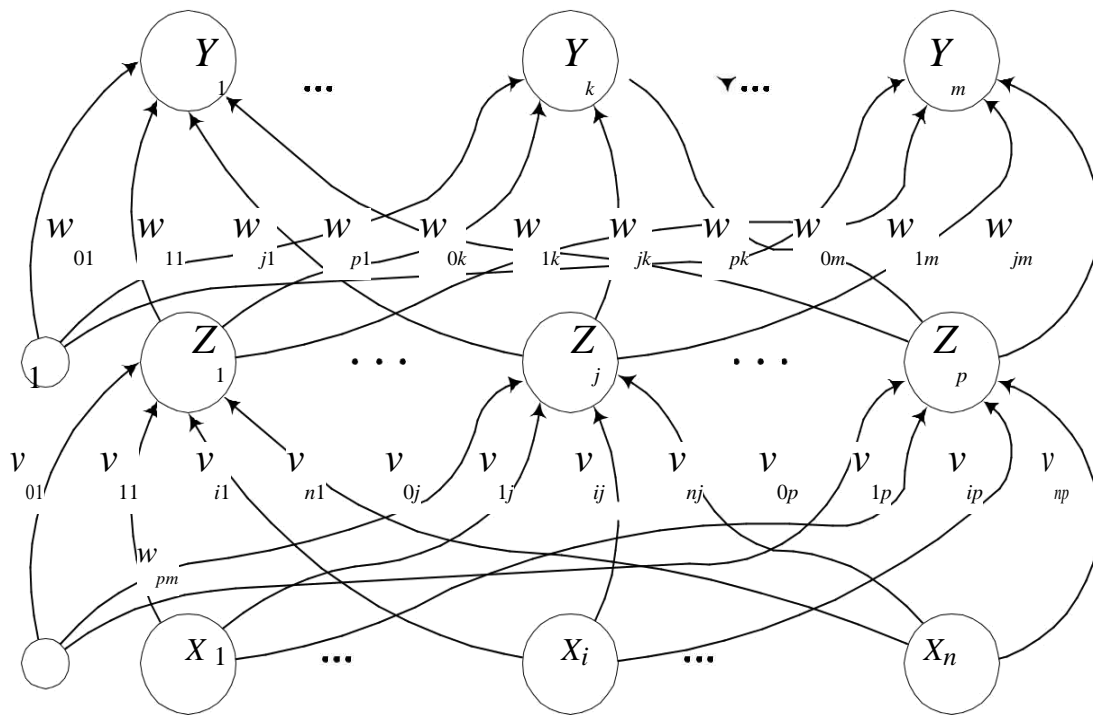
Training pada metode *backpropagation* melibatkan 3 tahapan: pola training *feedforward*, penghitungan *error* dan penyesuaian bobot. Setelah training aplikasi jaringan hanya menggunakan komputasi tahapan pertama yaitu *feedforward*. Walaupun tahap *training* sangat lambat, namun jaringan dapat menghasilkan output dengan sangat cepat. Metode *backpropagation* telah divariasikan dan dikembangkan untuk meningkatkan kecepatan proses *training*. Walaupun satu lapisan jaringan

sangat terbatas dalam pembelajarannya, jaringan dengan lapis banyak dapat mempelajari lebih banyak lagi. Lebih dari satu lapisan tersembunyi mungkin bermanfaat untuk beberapa aplikasi, namun satu lapisan tersembunyi adalah cukup.

Metode *backpropagation* dapat digunakan untuk melakukan pendeteksian suatu jenis penyakit, gangguan, maupun kasus yang memiliki data masa lalu dan dengan metode *backpropagation* target output yang diinginkan lebih mendekati ketepatan dalam melakukan pengujian karena terjadi penyesuaian nilai bobot dan bias yang semakin baik pada proses pelatihan. Bahkan, jaringan syaraf tiruan telah digunakan untuk menganalisa perilaku panas pada kultur jaringan. Pada tahun 1997, prinsip *plant speaking* digunakan untuk memonitor perkembangan tanaman selada di *phytotron* dimana jaringan saraf tiruan digunakan sebagai otak untuk mengendalikan pemberian air dan unsur hara untuk pertumbuhannya. Selain itu jaringan syaraf tiruan dengan metode *backpropagation*, juga diterapkan dalam lokalisasi target *through-the-wall* dengan metode *time reversal music* dan terbukti sangat efisien secara komputasi.

2.9.1 Arsitektur Jaringan Backpropagation

Jaringan saraf tiruan banyak lapis dengan satu lapisan tersembunyi (unit Z), unit output (unit Y) dan unit tersembunyi yang juga mempunyai bias. Bias pada unit output Y_k ditandai dengan w_{0k} , bias pada unit tersembunyi Z_j ditandai dengan v_{0j} . Bias bertindak sebagaimana halnya bobot dimana outputnya selalu 1. Selama proses pembelajaran *backpropagation*, sinyal dikirim dalam arah yang berlawanan. Simbol-simbol yang digunakan ini tidaklah mutlak, bisa saja berganti dengan simbol-simbol yang lainnya asalkan fungsi logika yang dimaksudkannya tetap sama.



Gambar 2.7 Topologi Jaringan *Backpropagation*

2.9.2 Algoritma

Algoritma pelatihan jaringan *backpropagation*:

1. Langkah 0 : Inisialisasi bobot.
2. Langkah 1 : Selama kondisi berhenti bernilai salah,

Untuk masing-masing pasangan pelatihan, lakukan:

1. Masing-masing unit input (X_i , $i = 1, 2, 3, \dots, n$) menerima sinyal x_i dan meneruskan sinyal tersebut ke semua unit pada lapisan yang ada di atasnya (lapisan tersembunyi).
2. Masing-masing unit tersembunyi (Z_j , $j = 1, 2, 3, \dots, p$) menjumlahkan sinyal-sinyal input berbobot:

$$Z_inj = v0j + \sum_{i=1}^n Xivij \quad v0 = \text{bias} ; v = \text{bobot}$$

3. Gunakan fungsi aktivasi untuk menghitung sinyal outputnya:

$$Zj = f_.(Z_inj)$$

Kirimkan sinyal tersebut ke semua unit lapisan atasnya (unit-unit output).

4. Gunakan fungsi aktivasi untuk menghitung sinyal outputnya:

$$Zj = f_.(Z_inj)$$

Kirimkan sinyal tersebut ke semua unit lapisan atasnya (unit-unit output).

5. Masing-masing unit output (Y_k , $k=1,2,3,\dots,m$) menjumlah sinyal-sinyal input berbobot:

$$Y_ink = w0k + \sum_{j=1}^p Zjwjk \quad w0 = \text{bias} ; v = \text{bobot}$$

6. Gunakan fungsi aktivasi untuk menghitung sinyal outputnya:

$$Yk = f_.(Y_ink)$$

Dan kirimkan sinyal tersebut ke semua unit di lapisan sisanya (unit-unit output).

Untuk masing-masing pasangan pelatihan, lakukan: *Backpropagation* (perambatan mundur):

1. Masing-masing unit output (Y_k , $k = 1,2,3,\dots,m$) menerima target pola yang berhubungan dengan pola input pembelajaran, hitung informasi errornya:

$$k = (tk - Yk) f_.'(Y_ink) \quad t = \text{target output}$$

Kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperoleh nilai w_{jk}): $w_{jk} = \alpha \delta k Zj$ $\alpha = \text{learning rate}$

Hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki

nilai w_{0k}): $w_{0k} = \alpha \delta_k$

Kirimkan ini ke unit-unit yang ada di lapisan bawahnya.

2. Masing-masing unit tersembunyi (Z_j , $j=1,2,3,\dots,p$) menjumlahkan delta inputnya (dari unit-unit yang berada pada lapisan di atasnya):

$$\delta_{inj} = \sum_{k=1}^m \delta_k w_{jk}$$

Kalikan nilai ini dengan turunan dari fungsi aktivasinya untuk menghitung informasi error: $\delta_j = \delta_{inj} f'(Z_{inj})$

Kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai v_{ij}): $v_{ij} = \alpha \delta_j X_i$

Hitung koreksi bias (yang nantinya akan digunakan untuk memperbaiki nilai v_{0j}): $v_{0j} = \alpha \delta_j$

3. Perbaharui bobot dan biasnya:

- a. Masing-masing unit output (Y_k , $k=1,2,3,\dots,m$) memperbaiki bias dan bobotnya ($j = 1,2,3,\dots,p$)

$$j_k(\text{baru}) = w_{jk}(\text{lama}) + \delta_k w_{jk} \quad (\text{bobot})$$

$$w_{0k}(\text{baru}) = w_{0k}(\text{lama}) + \delta_k w_{0k} \quad (\text{bias})$$

- b. Masing-masing unit tersembunyi (Z_j , $j=1,2,3,\dots,p$) memperbaiki bias dan bobotnya ($i = 0,1,2,3,\dots,n$)

$$v_{ij}(\text{baru}) = v_{ij}(\text{lama}) + \delta_j v_{ij} \quad (\text{bobot})$$

$$v_{0j}(\text{baru}) = v_{0j}(\text{lama}) + \delta_j v_{0j} \quad (\text{bias})$$

4. Tes kondisi berhenti.

Setelah algoritma pelatihan di atas dilakukan, dan didapat output yang paling mendekati target, maka bobot dan bias akhir dari pelatihan disimpan dan dilakukan proses pengujian. Algoritma pengujian jaringan *backpropagation*:

1. Langkah 0 : inisialisasi bobot (dari algoritma pelatihan)
2. Langkah 1 : untuk setiap vektor masukan, lakukan langkah 2-4.
3. Langkah 2 : for $i=1...n$: atur aktivasi unit masukan x
4. Langkah 3 : for $j=1...p$

$$Z_inj = v0j + \sum_{i=1}^n Xivij$$

$$Zj = f.(Z_inj)$$

5. Langkah 4 : for $k=1...m$:

$$Y_ink = w0k + \sum_{j=1}^p Zjvjk$$

$$Yk = f.(Y_ink)$$

6. Langkah 5 : jika $Yk \geq 0.5$ maka $Yk = 1$, else $Yk = 0$.

Sekali pelatihan subset dengan pembaharuan bobot jaringan disebut epoch. Banyaknya contoh pada subset disebut epoch size/ukuran epoch. Beberapa peneliti menggunakan sebuah epoch bernilai satu yang berarti bobot diperbaharui setelah setiap kasus pelatihan ditampilkan. Ketika ukuran epoch lebih kecil dari seluruh pelatihan, maka subset dipilih secara acak setiap waktunya atau gangguan bisa saja terjadi. Epoch pelatihan diulang sampai hasil dari jaringan berakhir sesuai waktunya.

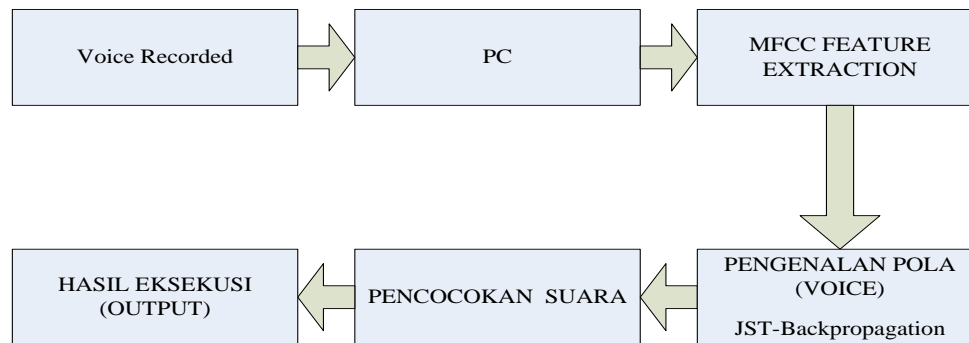
BAB 3

METODE PENELITIAN

Sistem Identifikasi Pembicara (SIP) memiliki tujuan untuk memberikan gambaran mengenai posisi pada ranah pemrosesan suara, komponen SIP serta riset-riset yang sudah ada. Agar diperoleh gambaran yang lebih menyeluruh maka pembahasan diawali dengan menyajikan sekilas mengenai pemrosesan suara dan bahasa. Sistem ini ditujukan untuk mengenali pembicara melalui input suara dari seorang pembicara dan bahasa. Identifikasi pembicara (*speaker Identification*) berbeda dari pengenalan ucapan (*speech recognition*), di mana pengenalan ucapan lebih dikonsentrasikan pada ekstraksi dari beberapa bagian informasi pesan yang didalamnya berisikan teks yang diucapkan. Teks yang diucapkan ini mengandung unit-unit linguistik yang dikenal sebagai *phonem*, kata-kata kunci suatu pesan ataupun pesan yang lengkap. Sedangkan identifikasi pembicara (*speaker identification*) menekankan pada analisis suara yang diucapkan oleh tiap pembicara sehingga dapat dibedakan karakteristik unik suara dari tiap pembicara tersebut.

3.1 Langkah-Langkah Penelitian

Secara umum alur penelitian Identifikasi suara *voice* dapat diperlihatkan pada Gambar 3.1.



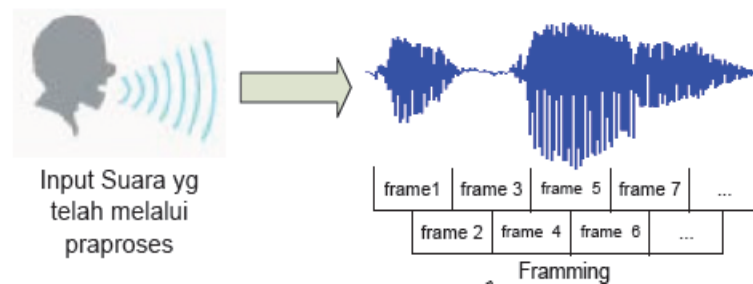
Gambar 3.1 Skema Alur Penelitian Identifikasi suara *Voice*

Studi *literature* merupakan studi awal mempelajari subjek penelitian sistem *voice recognition*. Sistem ini digunakan untuk pendeteksian atau pengenalan suara yang direkam dan disimpan dalam file suara dengan format .wav yang dianalisa menggunakan metode MFCC (Mel Frequency Cepstrum Coefficients). Proses pengklasifikasian yang dilakukan jaringan syaraf tiruan dilakukan setelah mendapatkan analisa koefisien MFCC. Selanjutnya menjadi vektor masukan JST untuk melatih jaringan. Hasil pelatihan berupa perubahan bobot jaringan sehingga diperoleh jaringan yang memiliki kemampuan pengklasifikasian. Jaringan tersebut kemudian diuji dengan mensimulasikannya pada data latihan dan data baru untuk menghasilkan persentase keberhasilan pengenalan. Eksperimen dilakukan dengan beberapa perubahan nilai parameter untuk memperoleh persentase pengenalan tertinggi.

3.1.1 Pemrosesan Pengambilan data dan Pemrosesan Suara

Pengambilan dan pengumpulan data adalah langkah yang paling penting dalam pengenalan bahasa/suara. Hanya sebuah data yang efisien yang dapat

menghasilkan sebuah sistem pengenalan bahasa yang baik. Proses perekaman suara menggunakan mikropon dan laptop. Sumber suara yang diambil adalah satu suara pria dan satu sumber suara wanita. Data yang direkam berupa kata “selamat pagi”, “selamat malam”, “apa kabar”, dan “hallo”. Proses pengambilan data dapat dilihat pada Gambar 3.2. Proses pemrosesan data ini menggunakan program MATLAB dan disimpan dalam format wav kemudian dilakukan pengekstraksian untuk penyimpanan data masukan. Tiap pembicara mengucapkan kata sebanyak tiga kali sehingga diperoleh total keseluruhan bahasa ada 24 buah bahasa.



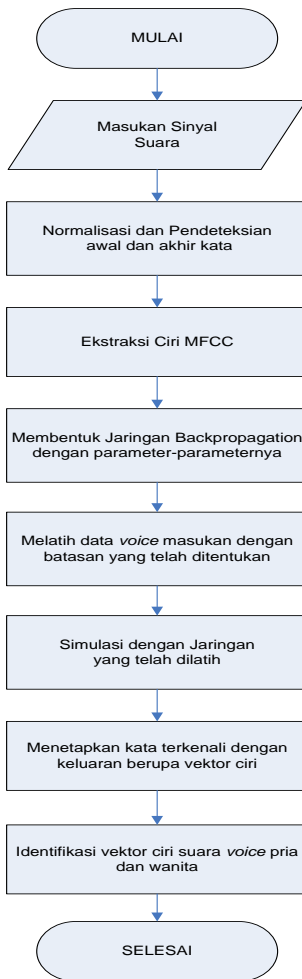
Gambar 3.2 Proses Pengambilan data/suara

3.2 Diagram Alir Pemrosesan Suara

Pada penelitian sebelumnya disajikan paradigma dari perkembangan riset bidang pemrosesan suara dan bahasa. Bidang pemrosesan suara “*voice recognition*” pada penelitian ini diarahkan pada pemrosesan suara untuk klasifikasi secara umum pada subsistem ekstraksi ciri dan pengenalan pola.

Skema alur penelitian *voice identification* dapat diperlihatkan pada Gambar

3.3.



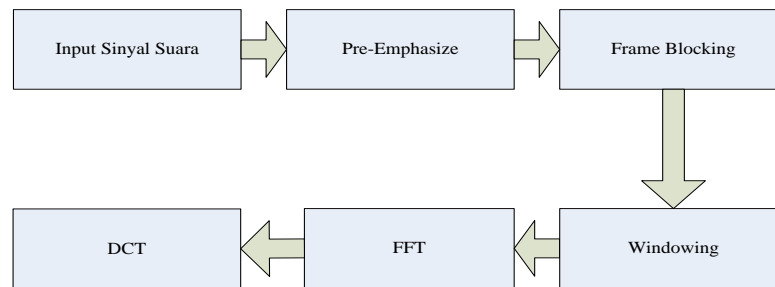
Gambar 3.3 Skema Alur Penelitian *Voice Identification*

3.2.1 Proses *Mel Frequency Cepstrum Coefficients* (MFCC)

Identifikasi pembicara merupakan proses untuk menentukan secara otomatis siapa pemilik dari suara yang diberikan. Proses pengenalan pola dan pengolahan data dengan menggunakan Matlab merupakan kajian bidang recognition yang diarahkan dengan menggunakan metode MFCC sebagai ekstraksi ciri dan pengenalan pola

menggunakan jaringan syaraf tiruan (JST). Input dari sistem pengenalan suara berupa sinyal suara yang berupa gelombang. Pada bagian ini berupa gelombang analog untuk menghasilkan sinyal suara yang dinormalisasi dengan tujuan untuk meyamakan amplitudo dari setiap suara nyanyian yang direkam oleh sistem sehingga berada dalam rentang -1 dan 1.

Data dinormalisasikan amplitudonya dengan cara membagi sampel data dengan nilai tertinggi $x_{norm} = \frac{x}{\max(|x|)}$. Beberapa parameter yang berperan penting dalam menentukan tingkat pengidentifikasian signal suara (MFCC) adalah proses *pre-Emphasis*, *Frame Blocking*, *Windowing*, *Fast Fourier Transform* (FFT), dan *Discrete Cosine Transform* (DCT). Proses Ekstraksi ciri MFCC dapat diperlihatkan pada Skema alur penelitian 3.4 dibawah ini.



Gambar 3.4 Diagram alir Penelitian menggunakan metode MFCC

Setelah melalui tahap perekaman suara dan *pre-processing*, data suara selanjutnya memasuki tahap proses ekstraksi ciri menggunakan metode MFCC untuk mendapatkan vektor ciri adalah proses *pre-emphasis* terhadap sinyal hasil keluaran *pre-processing* dengan koefisien α yang digunakan pada penelitian adalah 0.95. Setelah itu sinyal suara dipotong dengan ukuran *frame* tertentu. Panjang *frame* yang

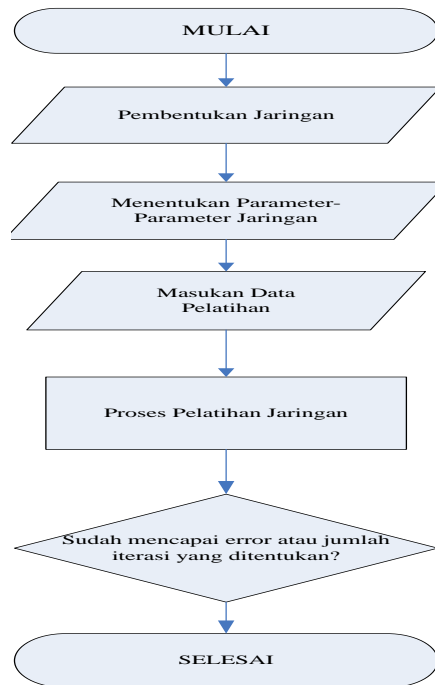
disajikan adalah 512 dengan besar overlape yang diujikan sebesar 50% sehingga didapatkan sejumlah *frame* tertentu. Kemudian setiap *frame* yang telah didapat dikalikan dengan *window hamming* untuk mengurangi kebocoran spectral. Selanjutnya *frame* dari domain waktu diubah ke domain frekuensi dengan menggunakan FFT.

Spektum frekuensi yang didapat dari proses FFT dilewatkan ke *Mel-Filter Bank* yang memiliki kesamaan persepsi pendengaran manusia. Jumlah *Mel-filter Bank* yang digunakan menentukan hasil ciri dari suatu sinyal suara. Pada penelitian ini dilakukan percobaan dengan menggunakan beberapa jumlah *mel filter bank* sebanyak 40 buah filter untuk mengetahui kualitas filter yang menghasilkan ciri paling optimal. Pemilihan jumlah filter tersebut didasarkan pada penelitian yang sudah ada mengenai MFCC. Pada DCT ini ditentukan beberapa jumlah koefisien keluaran MFCC yang merupakan hasil akhir dari proses ekstraksi ciri menggunakan MFCC. Pada penelitian ini dilakukan pengambilan jumlah koefisien suara sebanyak 6, 8, 10, 20 buah koefisien. Hasil akhir ini yang selanjutnya digunakan sebagai vektor ciri untuk masuk ke tahap klasifikasi.

3.2.2 Proses Jaringan Syaraf Tiruan Backpropagation

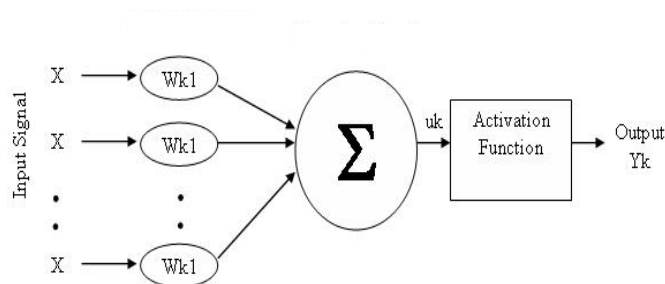
Pada umumnya tujuan jaringan syaraf tiruan melakukan proses pelatihan adalah untuk mendapatkan balancing antara kemampuan jaringan untuk menanggapi secara benar pola-pola input pada saat pelatihan dan kemampuan untuk memberikan penilaian yang layak dari suatu pola masukkan lain yang serupa.

Proses Pelatihan Jaringan Syaraf Tiruan dapat diperlihatkan pada Gambar 3.5.



Gambar 3.5 Proses Pelatihan Jaringan Syaraf Tiruan

Pada proses pelatihan tersebut akan dibentuk suatu harga pembobot yang akan digunakan sebagai faktor pengkali dari pola masukan yang lain. Model Tiruan Neuron dapat diperlihatkan pada Gambar 3.6.

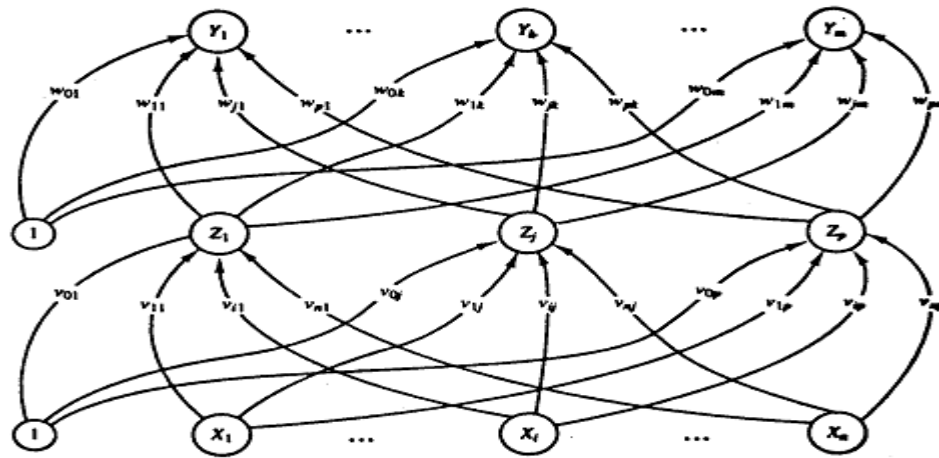


Gambar 3.6 Model Tiruan Neuron

Pada metode ini terdapat tiga tahap dalam proses pelatihan, yaitu: proses umpan maju dari pola input pelatihan, perhitungan dan propagasi balik dari error yang terjadi dan penyesuaian nilai bobot. Pada tahap pelatihan ini merupakan langkah bagaimana suatu jaringan syaraf itu berlatih yaitu proses umpan maju dari pola input pelatihan, perhitungan, dan propagasi balik dari error yang terjadi dari penyesuaian nilai pembobot. Pada tahap pelatihan ini merupakan langkah bagaimana suatu jaringan syaraf itu berlatih, yaitu dengan cara melakukan perubahan bobot sambungan, baik bobot sambungan antar *input layer* dan *hidden layer* maupun antara *hidden layer* dan *output layer*, bila terdapat lebih dari satu *hidden layer* maka juga terdapat pembobot antar *hidden layer* itu sendiri. Sedangkan penyelesaian masalah baru akan dilakukan jika proses pelatihan tersebut selesai, fase tersebut adalah proses pemakaian/testing tentunya dengan menggunakan pembobot yang telah dihasilkan dari proses pelatihan yang telah dilakukan.

Fase Pelatihan Umpan Mundur (Backpropagation)

Jaringan *BackPropagation* terdiri atas beberapa layer yang masing-masing unit pada satu layer terhubung penuh dengan masing-masing unit pada lapisan di atasnya atau dibawahnya, kecuali pada bias hanya terkoneksi penuh dengan unit layer di atasnya yang ditunjukkan pada gambar 3.7. Pada gambar 3.7 tersebut menunjukkan jaringan yang memiliki satu hidden layer dengan input layer X, hidden layer Z dan output layer Y, serta pemberian nilai bias, yaitu suatu masukkan dengan nilai tetap sama yaitu 1.



Gambar 3.7 Topologi Jaringan *Backpropagation*

Algoritma belajar *Backpropagation* terdiri dari dua proses, *feed forward* dan *backpropagation* dari errornya. Selama *feed forward* masing-masing unit masukkan menerima (X) atau sinyal masukkan dari luar, kemudian sinyal tersebut disebarkan masing-masing unit pada *hidden layer* (Z), masing-masing *hidden* unit menghitung sesuai dengan *fungsi aktifasinya*. Dan kemudian mengirim sinyal itu kemasing-masing unit pada *output layer* akan menghitung sesuai dengan fungsi aktifasinya juga, yang akan menghasilkan sinyal keluaran sebagai respon jaringan dengan adanya pemberian pola *input* tersebut. Pada propagasi baliknya, masing-masing output unit dibandingkan dengan hasil perhitungan aktivasi Y dengan nilai target t untuk mendapatkan *error*, berdasarkan *error* inilah akan dihitung nilai δk , selanjutnya harga *error* pada *output* unit akan disebarkan mundur ke masing-masing unit pada *hidden layer*. Selanjutnya *error* tersebut digunakan untuk memperbaiki faktor pembobot antara unit *output* dengan unit *hidden* demikian selanjutnya dicari *error* dari keluaran *hidden* untuk memperbaiki faktor pembobot antara unit *input*.

Fase Pemakaian Neural Network

Pada fase pelatihan akan didapatkan harga pembobot pada masing-masing layer. Pada fase pemakaian ini, pembobot yang dipakai adalah pembobot yang dibentuk pada saat pelatihan saat mencapai harga terbaik tentunya dengan pola jaringan yang sama seperti pada fase pelatihan, jika pola jaringan tidak sama, maka dapat dipastikan jaringan akan memberikan *output* yang tidak diinginkan. Algoritma yang dipakai pada fase ini hanya menggunakan bagian *feedforward* dari pelatihan.

3.3 Evaluasi Penelitian *Mel Frequency Cepstrum Coefficients* (MFCC)

Pada proses Ekstraksi Ciri MFCC ini dilakukan evaluasi proses *pre-emphasize*, *blocking*, *Hamming window*, FFT, dan *Discrete Cosine Transform* (DCT).

3.3.1 *Pre-Emphasize*

Pre-emphasize dilakukan untuk memperbaiki signal dari gangguan *noise*, sehingga dapat meningkatkan tingkat akurasi dari proses *feature extraction*. Default nilai *alpha* yang digunakan dalam proses ini adalah 0.95.

Proses perhitungan menggunakan *script* program MATLAB pada Gambar 3.8.

```
static private double[] Preemphasize()
{
    Try
    {
        double[] sigDataTemp=new double[length];
        for (int i = (int)length - 1; i > 0; i--)
            sigDataTemp[i] = sample[i] - 0.95 *
            sample [i - 1];

        sigDataTemp[0] = sample[0]-0.95 *
            sample[0];
        return sigDataTemp;
    }
    catch (Exception err)
    {
        MessageBox.Show(err.Message);
        return null;
    }
}
```

Gambar 3.8 *Script* program *Preemphasize*

3.3.2 *Frame Blocking*

Untuk keperluan pemrosesan, sinyal analog yang sudah melalui proses *sampling* dan kuantisasi (digitasi suara) dibaca dari *frame* demi *frame* dengan lebar tertentu yang saling tumpang tindih (*overlape*). Proses ini dikenal dengan *frame blocking*. Proses ini bertujuan untuk membagi sampel suara menjadi beberapa *frame* atau slot dengan panjang tertentu. Pada penelitian yang dilakukan menggunakan persamaan (2.16) dimana sinyal suara akan dipotong sepanjang 20 ms disetiap

pergeseran sepanjang 10 ms. Proses *blocking* menggunakan *script* program MATLAB pada Gambar 3.9.

```
//frame Blocking

int N = samplingrate * panjangFrame / 1000; //
=>banyaknya sample dalam 1 frame kurang lebih 20-
30ms dan merupakan kelipatan 2

int M = samplingrate * pergeseranFrame /
1000; /
int nFrame = 0;
nFrame = (int)length / M - 1;
double[,] arrFrame = new double[nFrame, N];
double[,] arrWindow = new double[nFrame, N];
for (int i = 0; i < nFrame; i++)
for (int j = 0; j < N; j++)
{
arrFrame[i, j] = sample[M * i + j];
//frame blocking
}
```

Gambar 3.9 Script program *Blocking*

3.3.3 Windowing

Setiap *frame* mengandung satu unit informasi, sehingga barisan *frame* akan menyimpan suatu informasi yang lengkap dari sebuah sinyal suara. Untuk itu, distorsi antar *frame* harus diminimalisasi. Salah satu teknik untuk meminimalkan distorsi antar *frame* adalah dengan melakukan proses *filtering* pada setiap *frame*. Pada penelitian kali ini jenis *filter* yang digunakan yaitu *windowing* karena pemrosesan sinyal yang akan dilakukan dalam domain frekuensi. Proses *windowing* dilakukan

pada setiap *frame*. Fungsi *window* yang digunakan pada penelitian ini adalah *Hamming* karena memiliki ekspresi matematika yang cukup sederhana menggunakan persamaan (2.18). Proses ini *windowing* berfungsi untuk mengurangi efek diskontinuitas pada ujung-ujung *frame* yang dihasilkan oleh proses *frame blocking*. Proses *Hamming window* menggunakan *script* program MATLAB pada Gambar 3.10.

```
//frame Windowing

    int N = samplingrate * panjangFrame / 1000;//
    =>banyaknya sample dalam 1 frame kuranglebih 20-
    30ms dan merupakan kelipatan 2
    int M = samplingrate * pergeseranFrame /
    1000;//
    int nFrame = 0;
    nFrame = (int)length / M - 1
    double[,] arrFrame = new double[nFrame, N];

    for (int j = 0; j < N; j++)
    {
        arrWindow[i, j] = arrFrame[i, j] * (0.54 - 0.46 *
        Math.Cos(2 * Math.PI * j / (N - 1)));
        //windowing
    }
```

Gambar 3.10 *Script* program *Hamming Window*

3.3.4 Fast Fourier Transform (FFT)

Analisis *Fourier* merupakan suatu teknik matematika untuk mendekomposisi sinyal menjadi sinyal-sinyal sinusoidal. Untuk dapat melihat perbedaan sinyal suara yang berbeda-beda maka harus dilihat dari domain frekuensi karena jika dilihat dari domain waktu sulit terlihat perbedaannya. Untuk itu dari sinyal suara yang berada pada domain waktu diubah ke domain frekuensi dengan *Fast Fourier Transform* (FFT). FFT merupakan suatu algoritma untuk mengimplementasikan *Discrete Fourier Transform* (DFT). Analisa berdasarkan fourier transform sama artinya dengan analisa spektrum, karena fourier transform merubah signal digital dari time domain ke frequency domain yang secara matematis dapat diperoleh dengan menggunakan persamaan (2.25) dan (2.26). Proses FFT menggunakan *script* program MATLAB pada Gambar 3.11.

```
static public double[] FFT(ref double[] x)
{
    // Assume n is a power of 2
    n = x.Length;
    nu = (int)(Math.Log(n) / Math.Log(2));
    int n2 = n / 2;
    int nu1 = nu - 1;
    double[] xre = new double[n];
    double[] xim = new double[n];
    double[] magnitude = new double[n2];
    double[] decibel = new double[n2];
    double tr, ti, p, arg, c, s;
    for (int i = 0; i < n; i++)
```

Gambar 3.11 Script program *Hamming Window*

3.3.5 Mel Frequency Wrapping

Persepsi sistem pendengaran manusia terhadap frekuensi sinyal suara ternyata tidak hanya bersifat linear. Oleh karena itu, untuk setiap nada dengan frekuensi aktual (dalam Hertz) tinggi subjektifnya diukur dengan skala mel. Skala *mel-frequency* adalah selang frekuensi di bawah 1000Hz bersifat linear dan untuk selang frekuensi di atas 1000Hz bersifat logaritmik. *Magnitude* hasil dari proses FFT selanjutnya akan melalui tahap *filterbank*. Untuk dapat melakukan tahap *filterbank* ini, maka terlebih dahulu harus dicari nilai-nilai koefisien dari *filterbank*. *Filterbank* bertujuan untuk mengetahui ukuran energi pada *frequency band* tertentu dalam *signal* suara. Relasi pendekatan empiris untuk menghitung skala Mel untuk frekuensi yang diberikan dalam Hz menggunakan persamaan (2.27). Proses *Mel Frequency* menggunakan *script* program MATLAB pada Gambar 3.12.

```
static private melFB[] initFB(int fftLength)
{
    try
    {
        //-- constant for calculations --/
        m_melWeight = new double[fftLength +
1]; //fft/2+1
        samplingrate)); freq[numChanel + 1] =
(double)(samplingrate / 2.0);
        mel = (double)(2595.0 * Math.Log10(1 +
(samplingrate / 2.0) / 700));
        bin[numChanel + 1] =
(int)(Math.Round(fftLength *
freq[numChanel + 1] /
samplingrate));
    }
}
```

Gambar 3.12 Script program *Mel Frequency Cepstrum Coefficients*

3.3.6 Transformasi kosinus

Langkah terakhir yaitu mengkonversikan *log mel spectrum* ke domain waktu. Hasilnya disebut *mel frequency cepstrum coefficients*. Cara untuk mengkonversikan *log mel spectrum* ke bentuk domain waktu yaitu dengan menggunakan *Discrete Cosine Transform* (DCT). Hasil dari DCT ini adalah fitur-fitur yang dibutuhkan oleh penulis untuk melakukan proses analisa terhadap pengenalan suara tersebut. DCT berfungsi untuk mendekorelasi *mel spectrum* sehingga didapatkan representasi yang baik dari spektral local. Proses *Cosine Transform* ini menggunakan persamaan (2.28) yang menggunakan *script* program MATLAB pada Gambar 3.13.

```
//DCT

double[,] arrDCT = new double[nFrame, koef];
for (int i = 0; i < nFrame; i++)
    for (int j = 0; j < koef; j++)
    {
        for (int k = 0; k < koef; k++)
        {
            arrDCT[i, j] += Math.Log10(H2mel[i,
                k]) * Math.Cos(j * (k - 0.5) * (Math.PI /
                koef));
        }
    }
```

Gambar 3.13 Script program *Discrete Cosine Transform*

3.4 Evaluasi Penelitian Jaringan Syaraf Tiruan *Backpropagation*

Jaringan syaraf tiruan *backpropagation* adalah suatu metode untuk megklasifikasikan dan mengidentifikasi suatu pola masukan tertentu dengan melakukan perbaikan bobot penghantar antar lapisan. Secara definitif, Fauzet mendefinisikan jaringan syaraf tiruan sebagai sistem pemrosesan informasi yang mempunyai karakteristik kinerja tertentu seperti pada jaringan syaraf biologi manusia. Ada beberapa parameter yang digunakan pada jaringan syaraf tiruan *backpropagation* berupa Jumlah Neuron input, Jumlah Hidden Layer, *learning rate*, Perbaikan Bobot, Error Minimum, dan Maksimal Epoch. Proses Jaringan Syaraf Tiruan *Backpropagation* menggunakan *script* program MATLAB pada Gambar 3.14.

```
// JST Enter the Architecture detail

disp('Enter the Architecture detail');
e = 0.05; % error limit
n = 6; %INPUT
p = 3;
m = 1; % OUTPUT
%%% Initialization of the Input / Target
disp('Loading the input vector x and target vectors')
x = [1.5252  1.8376  1.5144  0.6666  -0.0862
- trainRecords = size(x,1);
t = [0; 1; 0];
%% Backpropagation of error (Training Starter
while er==0
    disp(sprintf('Epoch : %4g, max err : %d',iteration,
errorMax(iteration)));
    ==Step_2==For_Each_Training_pair=do_Steps_3-
```

Gambar 3.14 *Script* program Jaringan Syaraf Tiruan

BAB 4

HASIL DAN PEMBAHASAN

Telah ditunjukkan pada bab 3 bahwa penelitian pemrosesan suara difokuskan pada sub sistem pengenalan menggunakan jaringan syaraf tiruan dan telah dikenal luas pemakaiannya adalah *Mel Frequency Cepstrum Coefficients* (MFCC) sebagai ekstraksi ciri seperti yang disajikan pada bagian sebelumnya. Sistem pengenalan pembicara yang telah dibahas adalah yang berbasis pada power spectrum dalam domain frekuensi sebagai penentu ciri sinyal suara. Sistem yang dihasilkan mampu mengenali dengan baik untuk data suara tanpa penambahan *noise*. Untuk data suara yang memiliki *noise* lebih besar, terlihat bahwa akurasi turun secara drastis.

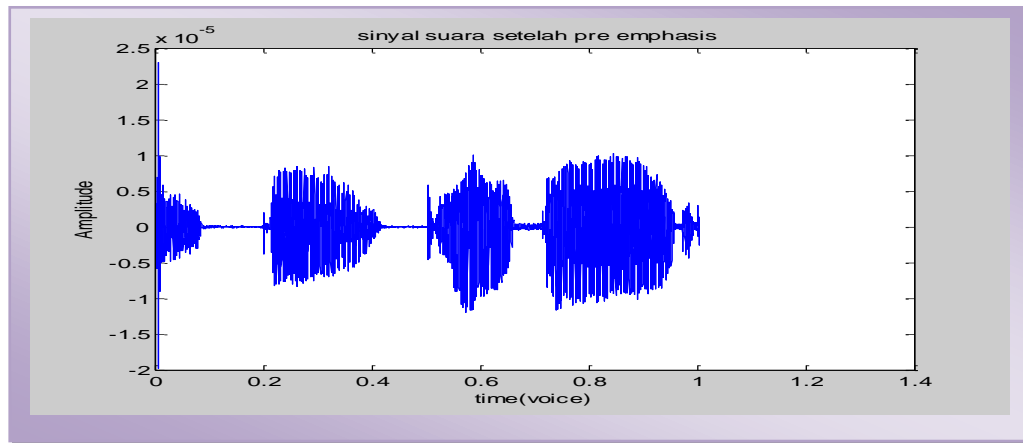
Oleh karena penelitian yang dilakukan pada Tesis ini difokuskan pada sub pemrosesan, sistem ekstraksi ciri dan sub sistem pengenalan hanya sebagai verifikasi hasil, maka dipilih jaringan syaraf tiruan sebagai pengenalan suara. Hal ini dengan pertimbangan peningkatan akurasi. Bukti empiris dari keterbatasan atau kelemahan teknik ekstraksi ciri konvensional dalam melakukan pengenalan untuk situasi ber-*noise*. Sedangkan ekstraksi ciri yang digunakan adalah MFCC dengan suara *voice* sebagai masukannya yang secara detail telah disajikan pada Bab 2.

4.1 Analisa Proses Ekstraksi Ciri *Mel Frequency Cepstrum Coefficients*

Proses pengestraksian ciri merupakan proses untuk mendapatkan parameter-parameter sinyal suara. Parameter-parameter inilah yang nantinya digunakan untuk membedakan satu suara dengan suara yang lain. Proses awal ekstraksi ciri yaitu dengan normalisasi amplitudo terlebih dahulu, setelah itu dilakukan pendeteksian awal dan akhir kata. Normalisasi amplitudo dilakukan dengan cara membagi semua nilai input dengan nilai maksimum dari input itu sendiri. Sehingga untuk semua sinyal masukan memiliki nilai maksimum yang sama yaitu 1. Sedangkan deteksi awal-akhir digunakan pada proses untuk mendeteksi mulai sinyal ucapan awal dan berakhir ketika sudah tidak diucapkan.

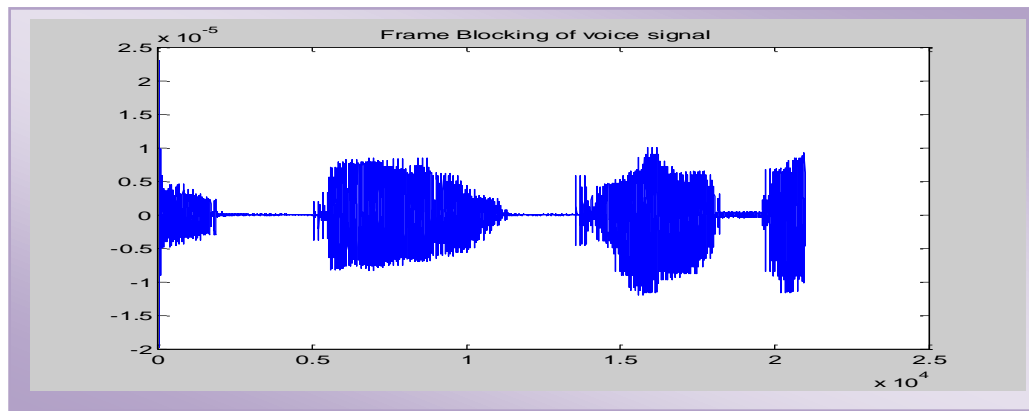
Selanjutnya dilakukan proses analisa *Mel Frequency Cepstrum Coefficients* (MFCC) bisa didapat jumlah frame dan panjang masing-masing frame yang sama. Walaupun tahap ini akan memperpanjang data dan proses pada jaringan syaraf tiruan, tahap ini mampu memperkecil kesalahan pengenalan kata. Terutama untuk kata yang memiliki jumlah suku kata berbeda. Dari hasil normalisasinya selanjutnya melalui proses *preemphasis*. Hal ini dilakukan untuk memperbaiki signal dari gangguan noise sehingga meningkatkan keakurasian dari proses *feature extraction*. Filter ini mempertahankan frekuensi-frekuensi tinggi pada spektrum yang tereliminasi saat proses produksi suara. Setelah melalui proses *preemphasis*, sinyal suara dibagi dalam beberapa *frame* sepanjang N sample suara dengan jarak antar frame yang berdekatan dipisahkan oleh M sample.

Hasil perekaman suara merupakan sinyal analog yang berada dalam domain waktu yang bersifat *varian time*, yaitu suatu fungsi yang tergantung waktu. Proses *preemphasis* dapat diperlihatkan pada Gambar 4.1.

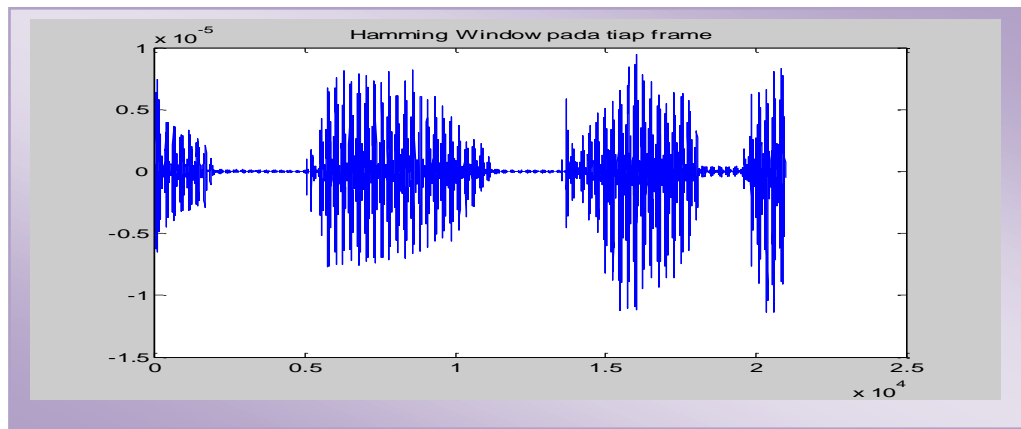


Gambar 4.1 Proses *preemphasis*

Setiap proses *blocking* yang dilakukan pada penelitian ini memiliki slot slot waktu tertentu sekitar 10ms-20ms dari sinyal suara. Setiap potongan yang berupa slot-slot dalam domain waktu ini disebut sebagai *frame*. Proses *Frame Blocking* yang dilakukan pada penelitian *voice recognition* ini ditetapkan 20 ms dengan jarak antar frame 10 ms. Jadi dengan frekuensi sampling 11000 Hz akan menghasilkan jarak antar *frame* (*frame duration*) $11000/(256-156)=110\text{frame/sec}$. Perhitungan jumlah *frame* dengan sample rate 11000 Hz dan sample point sebanyak 256 akan menghasilkan $((I-N)/M))+1=((11000-256)/156))+1=70\text{ frame}$ (Hasil pembulatan dari 69.8). Proses *blocking* dari sinyal suara dapat diperlihatkan pada Gambar 4.2.

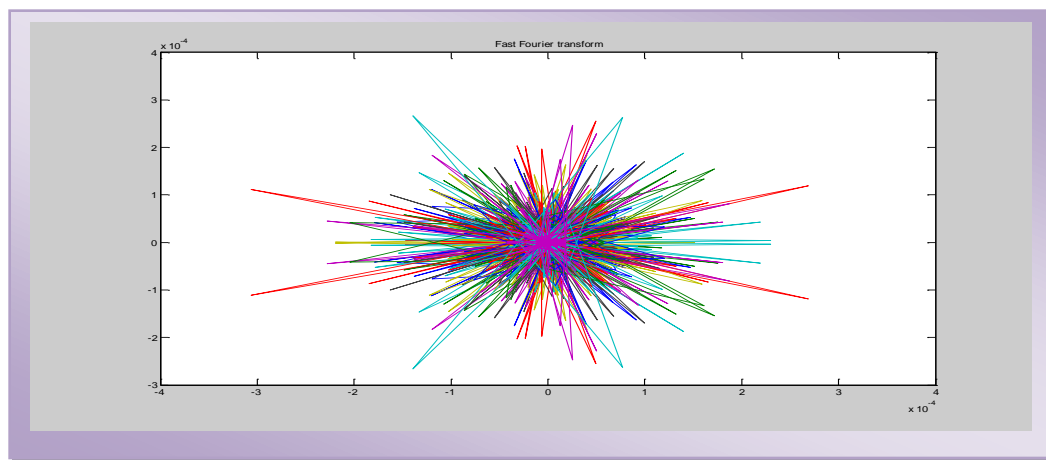
Gambar 4.2 Proses *Blocking*

Setiap frame kemudian dijenelakan menggunakan *Hamming window*. Fungsi ini menghasilkan sidelobe level yang tidak terlalu tinggi (kurang lebih -43dB) selain itu noise yang dihasilkanpun tidak terlalu besar. Proses ini dilakukan untuk mengurangi efek aliasing atau sinyal tak kontinyu pada awal dan akhir masing-masing *frame* yang dapat terjadi akibat proses *blocking* yang dieperlihatkan pada Gambar 4.3.

Gambar 4.3 Proses *Hamming Window*

Selanjutnya proses yang dilakukan untuk mengkonversi masing-masing frame sinyal suara dari domain waktu ke domain frekuensi. Hal ini dilakukan karena dalam sinyal bicara, sistem pendengaran sangat sensitif terhadap karakteristik frekuensi sehingga

sinyal bicara lebih mudah dianalisis pada domain frekuensi. Sehingga perbedaan antar pola kata yang satu dengan yang lain terlihat lebih jelas sehingga ekstraksi parameter sinyal memberikan hasil yang lebih baik. Proses FFT yang dilakukan 512 point dan karena FFT simetris maka keluaran FFT tersebut hanya diambil setengahnya saja seperti pada Gambar 4.4.

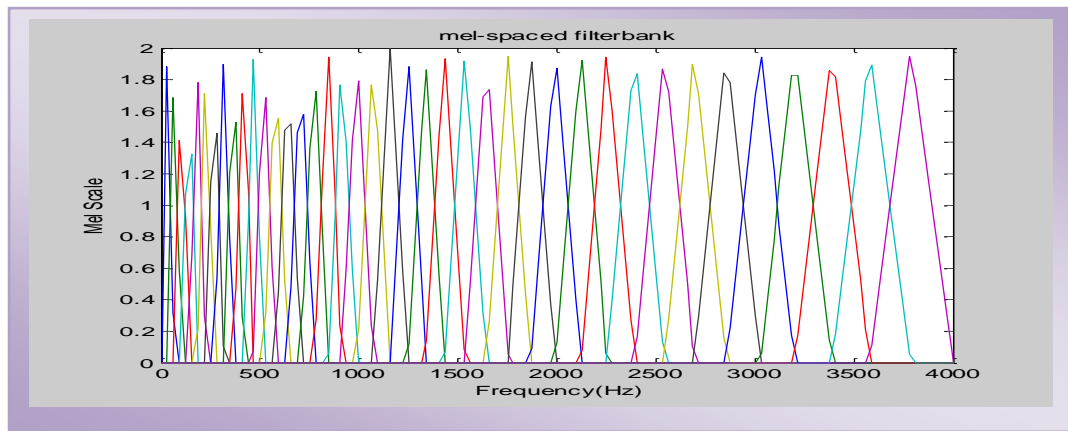


Gambar 4.4 *Fast Fourier Transform (FFT)*

Beberapa parameter yang digunakan pada proses ekstraksi ciri diatas digunakan sebagai analisis keakurasian proses MFCC. Adanya pengaruh panjang frame dari sumber sinyal suara, pengaruh panjang overlape yang menentukan banyaknya sampel yang diproses kembali pada frame berikutnya untuk mengantisipasi apabila ada informasi yang terproses. Dengan jumlah sampel yang lebih banyak akan mempengaruhi panjangnya overlape dan *frame* yang diproses pada sumber sinyal suara sehingga ciri keseluruhan sinyal suara dapat lebih terlihat. Pengaruh dari penggunaan *Mel Filterbank* dalam domain frekuensi. Sedikit *Mel Filterbank* yang

digunakan, maka sedikit pula informasi yang didapat dari setiap frekuensi karena tiap frekuensi memiliki karakteristik yang berbeda.

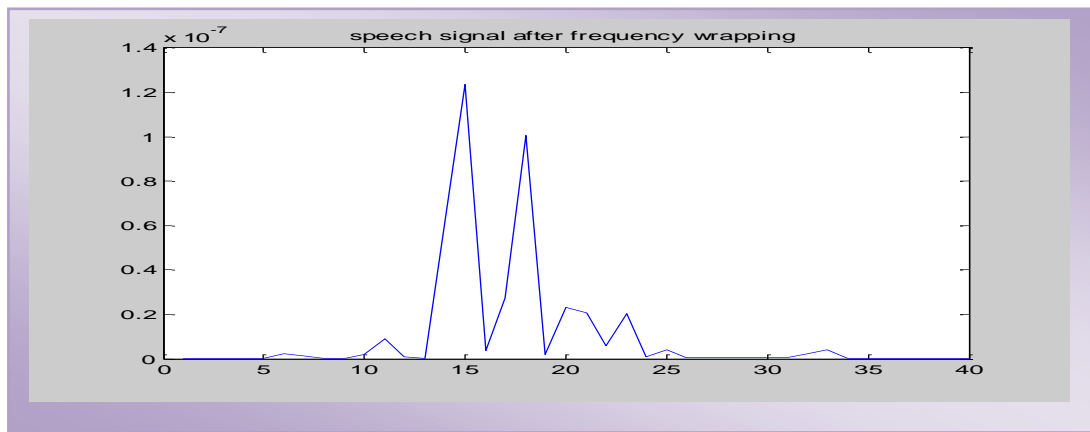
Disisi lain penggunaan *Mel Filterbank* yang banyak membuat filter yang ada semakin rapat sehingga semakin teliti dalam memfilter sinyal suara yang masuk dan memprosesnya. Dalam penggunaanya, terlalu sedikit atau terlalu banyak jumlah *Mel filterbank* dapat memberikan hasil yang kurang baik sehingga perlu dicari jumlah yang memberikan hasil terbaik. Pada penelitian ini digunakan 40 buah *Mel filterbank* untuk mengetahui kualitas filter yang menghasilkan ciri paling optimal. Hasil dari pemrosesan sinyal menggunakan *mel filterbank* dapat diperlihatkan pada Gambar 4.5.



Gambar 4.5 Proses *Mel Filterbank*

Sinyal bicara terdiri dari nada dengan frekuensi yang berbeda-beda. Untuk masing-masing nada dengan frekuensi aktual. F diukur dalam Hz, pitch subjektif diukur dalam skala 'mel'. Skala *mel-frequency* bersifat linier untuk frekuensi dibawah 1000Hz. Dan langkah terakhir dari proses *feature extraction* berupa *spektrum log mel*

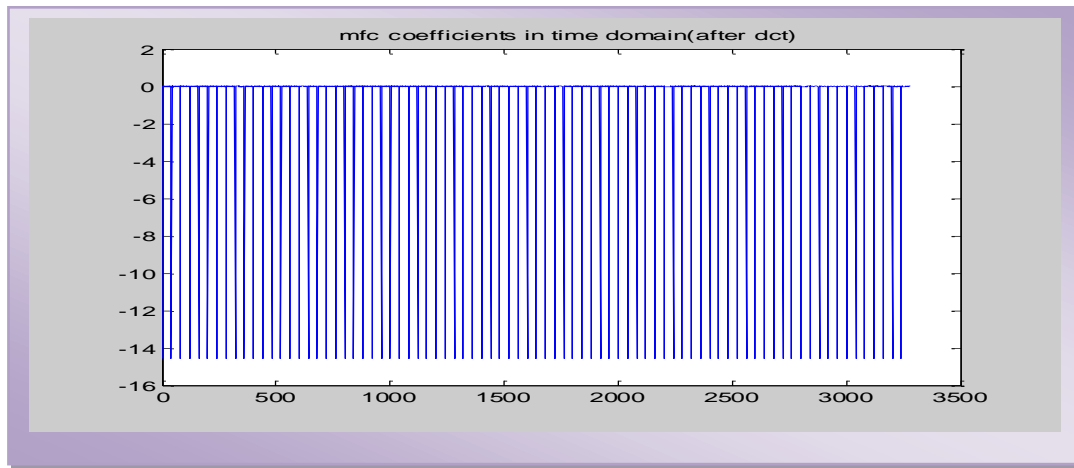
yang dikonversikan kembali dalam waktu menggunakan *Discrete Cosine Transform* (DCT). Fitur-fitur ini yang digunakan sebagai analisa terhadap proses pengenalan suara. Transformasi kosinus berfungsi untuk mengembalikan domain, dari frekuensi ke domain waktu. Proses *Discrete Cosine Transform* diperlihatkan pada Gambar 4.6.



Gambar 4.6 Proses *Discrete Cosine Transform* (DCT)

Dalam pemrosesan suara, ciri yang biasa dipergunakan adalah nilai koefisien cepstral dari sebuah frame. Satu teknik ekstraksi ciri sinyal suara yang umum dan menunjukkan kinerja yang baik adalah teknik *Mel Frequency Cepstrum Coefficients* (MFCC) yang menghitung koefisien cepstral dengan mempertimbangkan persepsi sistem pendengaran manusia terhadap frekuensi suara.

Setiap perubahan jumlah koefisien MFCC mempengaruhi banyaknya koefisien ciri yang digunakan. Semakin banyak koefisien yang digunakan akan memberikan hasil yang semakin baik karena informasi ciri yang dimiliki semakin banyak. Hasil proses MFCC ini dapat diperlihatkan pada Gambar 4.7.



Gambar 4.7 Hasil *Mel Frequency Cepstrum Coefficients* (MFCC)

Vektor ciri yang diperoleh pada proses MFCC ini dapat diperlihatkan pada Tabel 4.1 dengan menggunakan *voice recorded* “apa kabar” yang merupakan dari suara pria dan wanita dengan jumlah koefisien ciri suara yang diambil sebanyak 8 dari sumber sinyal suara asli.

Tabel 4.1 Vector Ciri suara pria dan wanita “Apa kabar”

Suara Voice	Vector Ciri ($C_n = 8$)				Gender
Apa kabar	6.5577	-1.1168	-3.2686	3.4951	Pria
	2.1963	1.2649	-0.1293	0.1714	
	3.0476	2.0393	-4.4242	-4.4484	Wanita
	2.1844	0.5254	-0.1185	0.5810	

Salah satu bentuk pendekatan pengenalan suara adalah pendekatan pengenalan pola yang terdiri dari pembelajaran pola suara dan pengenalan suara melalui perbandingan pola. Tahap perbandingan pola adalah tahap saat suara yang akan dikenali dibandingkan polanya dengan setiap kemungkinan pola yang telah dipelajari dalam

fase pembelajaran untuk diklasifikasikan dengan pola terbaik yang cocok. Pada tahap pengenalan suara secara umum dapat dibagi menjadi tiga. Hasil koefisien keluaran MFCC yang diharapkan merupakan hasil akhir dari proses ekstraksi ciri yang digunakan sebagai vektor ciri untuk masuk ke tahap pengenalan dari pola suara menggunakan jaringan syaraf tiruan.

4.1.1 Hasil Koefisien *Mel Frequency Cepstrum Coefficients* (MFCC)

Vector ciri MFCC yang diperoleh dengan pengambilan beberapa jumlah koefisien yaitu sebanyak 6, 8, 10, 20 dengan jumlah sampel suara sebanyak 3 suara wanita dan 3 suara pria. Proses pre-emphasis terhadap sinyal hasil keluaran *pre-processing* dengan α yang digunakan pada penelitian ini adalah 0.95. Setelah itu sinyal suara dipotong dengan ukuran frame tertentu. Panjang *frame* yang diujikan adalah 256 sehingga didapatkan sejumlah frame tertentu. Kemudian setiap *frame* yang telah didapat dikalikan dengan *window hamming*. Dan dilakukan proses FFT untuk mengubah masing-masing frame dari domain waktu ke domain frekuensi yang menghasilkan spektrum frekuensi. Hasil koefisien keluaran MFCC yang diharapkan yang merupakan hasil akhir dari proses ekstraksi ciri dapat diperlihatkan pada Tabel 4.2, Tabel 4.3 dan Lampiran1. Hasil perubahan spektrum yang merupakan representasi sinyal suara dalam bentuk data dapat diperlihatkan pada Lampiran 3.

Tabel 4.2 Vector Ciri Suara Pria dengan jumlah koefisien 6 dan 8

<i>Suara Voice</i>	Vector Ciri ($C_n=6$)				Vector Ciri ($C_n=8$)			
Apa kabar	-0.4797	-0.2424	-5.1478		-0.4856	-0.2580	-6.0429	-0.2892
	-0.2081	-0.8504	0.0351		-1.5972	0.1343	0.6174	-0.254
	2.8096	-0.4416	-8.8774	-	2.8441	-0.4699	-10.4210	-2.0463
Selamat Malam	1.4726	-2.1110	0.6250		3.9650	2.3927	-0.7388	0.5470
	3.5699	-2.5954	-10.0991		3.6138	-2.7616	-11.8552	-2.5799
	-1.8565	-1.4437	0.6425		2.7117	2.4599	0.0637	0.4953
Selamat Pagi	4.1233	-2.4407	-8.1262	-	4.1740	-2.5970	-9.5392	-2.9577
	2.1284	-3.0554	1.4346		5.7386	5.4921	-2.1221	0.1443
	2.2063	-3.8183	-2.7967		2.2334	-4.0627	-3.2830	1.3270
Hallo	0.9550	-0.3429	1.0123		0.6440	3.8757	-3.8071	-0.4591
	1.0431	-3.5865	-2.2493		1.0559	-3.8161	-2.6404	1.2761
	0.9183	-0.8203	1.1422		1.5407	4.3728	-2.7898	-0.5396
Hallo	3.2366	-4.2290	4.0897		3.2764	-4.4997	4.8008	0.3011
	0.2167	-5.2322	-0.3807		9.8272	-1.4573	-0.6246	0.0687
	3.6363	-5.7737	6.6668		3.6810	-6.1433	7.8260	-0.2127
Hallo	-0.1531	-5.0774	-0.2149		9.5365	-0.8229	-1.7773	-0.1168
	1.7715	-6.9915	5.3845		1.7933	-7.4391	6.3208	3.0253
	2.1771	-4.3714	0.3047		8.2103	1.1666	-2.3823	0.0432
Hallo	3.8303	6.0887	-3.4665		1.7933	-7.4391	6.3208	3.0253
	-5.9714	-2.3607	0.5788		8.2103	1.1666	-2.3823	0.0432
	1.7650	3.7312	-5.0402		1.7867	3.9700	-5.9165	-8.5696
Hallo	-6.1668	-1.2545	0.6083		2.3563	2.3290	0.4173	-0.1315
	1.4463	0	0	0	1.4641	0	0	0
	0	0			0	0	0	

Tabel 4.3 Vector Ciri Suara Wanita dengan jumlah koefisien 6 dan 8

Suara Voice	Vector Ciri ($C_n=6$)				Vector Ciri ($C_n=8$)			
Apa kabar	3.1367	2.7135	4.8906	-	3.0476	2.0393	-4.4242	-4.4484
	3.6578	-1.4732	0.0659		-2.1844	0.5254	-0.1185	0.5810
	0.7910	0.2729	1.9493	-	3.0476	2.0393	-4.4242	-4.4484
	1.3230	-2.9432	0.3769		-2.1844	0.5254	-0.1185	0.5810
	4.2287	0.1558	9.8505	-	3.0476	2.0393	-4.4242	-4.4484
	3.2051	-1.3073	0.9671		-2.1844	0.5254	-0.1185	0.5810
Selamat Malam	1.5252	1.8376	1.5144		1.3952	1.2975	1.1795	0.6141
	0.6666	-0.0862	0.2416		-0.1084	-0.6147	-0.6728	-0.3217
	0.5556	1.6928	1.1760		1.3952	1.2975	1.1795	0.6141
	0.2578	-0.3178	-0.2580		-0.1084	-0.6147	-0.6728	-0.3217
	-0.0534	1.6427	-1.8602	-	-0.2282	0.8108	-0.7328	-
	0.8108	0.5857	-0.1270		0.3643	0.3281	-0.5679	-0.7891
Selamat Pagi	1.4976	-0.2095	3.0365	-	1.4196	0.8408	1.1024	-0.1453
	0.8281	-0.9330	-0.2215		-0.8463	-0.6549	-0.8738	-0.1221
	3.1924	-2.2442	5.1346	-	3.2675	-0.7399	4.5034	-0.6672
	0.9151	-1.7495	-0.1678		-2.2294	0.1377	-2.1410	0.0366
	2.0948	1.6337	2.1533	-	2.3056	1.7859	1.2685	-1.2152
	1.2813	-2.3478	-0.2205		-3.2675	-1.4435	1.1225	0.0553
Hallo	4.0021	4.8572	-0.8894	-	4.0513	5.1681	-1.0440	-6.6598
	4.7925	-1.2838	0.2354		-2.4112	0.9011	-1.3025	-0.6497
	5.4519	4.7743	-3.7758	-	5.5189	5.0799	-4.4324	-14.1351
	10.1718	-1.2836	1.0392		-2.4108	3.9784	-0.2347	-0.0591
	1.8891	0	0	0	1.9123	0	0	0
	0	0			0	0	0	

4.2 Analisa Proses Jaringan Syaraf Tiruan *Backpropagation*

Beberapa parameter yang perlu diperhatikan dalam menyusun/merancang arsitektur Jaringan Syaraf Tiruan antara lain :

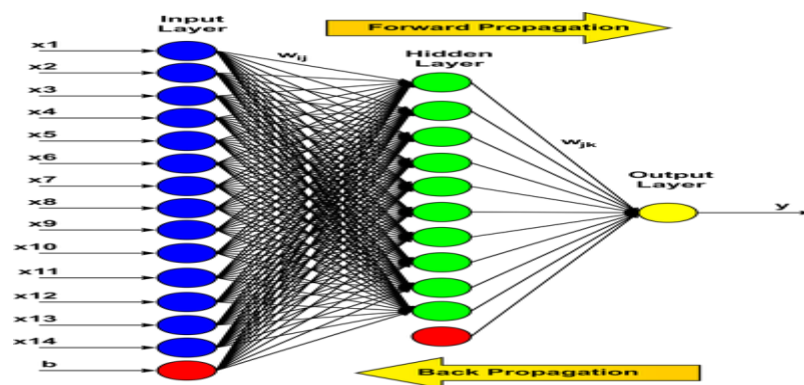
1. Jenis Jaringan Syaraf Tiruan yang digunakan = Backpropagation
2. Jumlah neuron hidden pada layer masukan
3. Jumlah Hidden layer pada layer tersembunyi
4. Jumlah neuron keluaran pada layer keluaran
5. Error limit (batas error) = Error goal
6. Siklus Perubahan Bobot = EPOCH (*Exponential Decay*)
7. Konstanta laju Pelatihan (*Learning Rate*) = α
8. Jumlah Koefisien MFCC
9. Jumlah Pelatihan

Proses Pelatihan pada metode *backpropagation* melibatkan 3 tahapan: pola training *feedforward*, penghitungan *error* dan penyesuaian bobot.

Data masukan dirambatkan maju, dihubungkan oleh bobot-bobot masukan yang sebelumnya telah diinisialisasi secara acak menuju neuron pada layer tersembunyi.

Pada layer tersembunyi, data masukan yang telah dihubungkan dengan bobot tersebut kemudian diproses menggunakan fungsi aktivasi. Selanjutnya data hasil olahan dari layer tersembunyi dihubungkan oleh bobot-bobot tersembunyi menuju neuron pada layer keluaran. Hasil yang diperoleh kemudian dibandingkan dengan data target

sehingga diperoleh kesalahan (error). Apabila tingkat kesalahan yang diperoleh lebih kecil daripada tingkat kesalahan yang sebelumnya telah ditetapkan (*target error*), maka proses perambatan akan berhenti. Namun apabila tingkat kesalahan masih lebih besar daripada tingkat kesalahan tetapan maka dilakukan proses perambatan balik dengan melakukan pembaharuan bobot. Gambar 4.8 berikut merupakan arsitektur Jaringan Syaraf Tiruan propagasi balik (backpropagation).



Gambar 4.8 Arsitektur Jaringan *Backpropagation*

4.2.1 Hasil Pelatihan Jaringan Syaraf Tiruan (*Backpropagation*)

Hasil pelatihan Jaringan Syaraf Tiruan *backpropagation* dengan inisialisasi perubahan bobot dari pengidentifikasian *voice* dengan menggunakan koefisien MFCC sebanyak 6, 8, 10, 20 dengan sampel 3 pria dan 3 wanita dapat diperlihatkan pada Tabel 4.4, Tabel 4.5 dan Lampiran 2. Pengaruh perubahan jumlah koefisien MFCC menunjukkan banyaknya vektor ciri yang akan digunakan. Pada umumnya pengaruh perubahan jumlah koefisien MFCC yang digunakan adalah semakin banyak koefisien

yang digunakan, akan memberikan hasil yang semakin baik karena informasi ciri yang dimiliki semakin banyak.

Tabel 4.4 Hasil Perubahan Bobot dengan 6 koefisien ciri ($C_n = 6$)

Gender	Suara	EPOCH	Nilai	EPOCH	Nilai
	<i>Voice</i>		MaxError		MaxError
Pria	Apa kabar	1 – 71	< 1	72 – 600	1
	Hallo	1 – 75	< 1	76 – 600	1
	Selamat Pagi	1 – 103	< 1	104 – 600	1
	Selamat Malam	1 – 60	< 1	61 – 600	1
Wanita	Apa kabar	1 – 59	< 1	60 – 600	1
	Hallo	1 – 57	< 1	58 – 600	1
	Selamat Pagi	1 – 350	< 1	351 - 600	1
	Selamat Malam	1 – 58	< 1	59- 600	1

Tabel 4.5 Hasil Perubahan Bobot dengan 8 koefisien ciri ($C_n = 8$)

Gender	Suara	EPOCH	Nilai	EPOCH	Nilai
	<i>Voice</i>		MaxError		MaxError
Pria	Apa kabar	1 – 71	< 1	72 – 600	1
	Hallo	1 – 65	< 1	66 – 600	1
	Selamat Pagi	1 -102	< 1	103 – 600	1
	Selamat Malam	1 – 60	< 1	61 – 600	1
Wanita	Apa kabar	1 – 67	< 1	68 – 600	1
	Hallo	1 – 55	< 1	56 – 600	1
	Selamat Pagi	1 – 66	< 1	67 – 600	1
	Selamat Malam	1 – 66	< 1	67 – 600	1

4.3 Analisa pengaruh nilai α terhadap Perubahan Bobot Jaringan Syaraf Tiruan

Backpropagation

Parameter yang perlu diperhatikan dalam menyusun/merancang arsitektur jaringan syaraf tiruan antara lain yang mempengaruhi perubahan bobot dalam pengidentifikasian dan pengenalan suara *voice recorded* terdapat pada *script* matlab Gambar 4.9 dan Lampiran 4.

```
num = 6;
n = 512; FFT
fn = 24;
l = length(x);
FrameStep=fs*Ts;    %Frame step in samples
a1=1;
b1=[1, -0.97];      % a and b are high pass filter
coefficients
e = 0.05; % error limit
n = 6; %INPUT
p = 2;
m = 1; % OUTPUT
alpha = 0.2;
disp('Loading the input vector x and target vectors')
x = [3.2366 -4.2290  4.0897  0.2167 -5.2322 -
0.3807; 1.4976 -0.2095  3.0365 -0.8281 -0.9330 -
0.2215];
trainRecords = size(x,1);
t = [0; 1];
```

Gambar 4.9 *Script* program Jaringan Syaraf Tiruan

Hasil pelatihan Jaringan syaraf tiruan *backpropagation* dengan inisialisasi perubahan bobot dari pengidentifikasian *voice* menggunakan koefisien MFCC sebanyak 6,8 dan perubahan konstanta laju pelatihan adalah $0 < \alpha < 1$ dengan sampel 2 pria dan 2 wanita dapat diperlihatkan pada Tabel 4.6 dan Tabel 4.7.

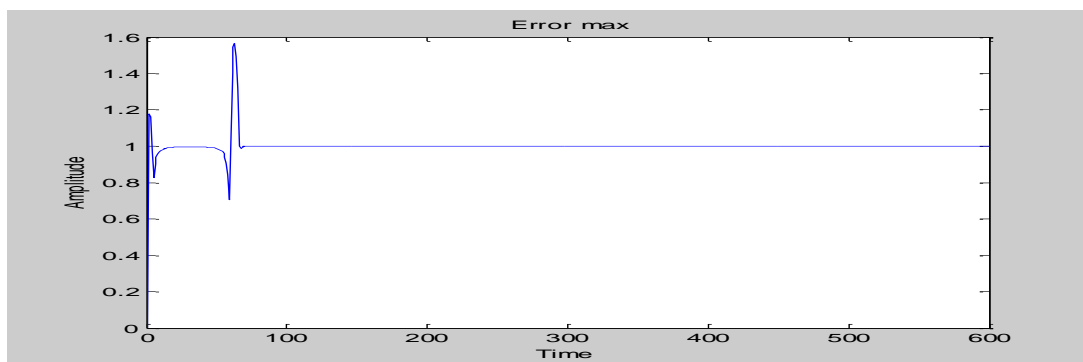
Tabel 4.6 Inisialisasi Perubahan Bobot dengan Jumlah Koefisien 6 ($C_n=6$)

Gender	Suara <i>Voice</i>	Input (C_n)	FFT	P	Error Min	α	EPOCH	Max Error
Pria	Selamat Pagi	6	512	2	0.05	0.2	1-94	<1
							95-600	=1
		6	512	2	0.05	0.4	1-66	<1
							67-600	=1
		6	512	2	0.05	0.7	1-37	<1
							38-600	=1
		6	512	2	0.05	0.9	1-23	< 1
							24-600	=1
Wanita	Selamat Pagi	6	512	2	0.05	0.2	1-63	< 1
							64-600	=1
		6	512	2	0.05	0.4	1-60	< 1
							61-600	=1
		6	512	2	0.05	0.7	1-37	< 1
							38 – 600	=1
		6	512	2	0.05	0.9	1-34	< 1
							35-600	=1

Tabel 4.7 Inisialisasi Perubahan Bobot dengan Jumlah Koefisien 8 ($C_n=8$)

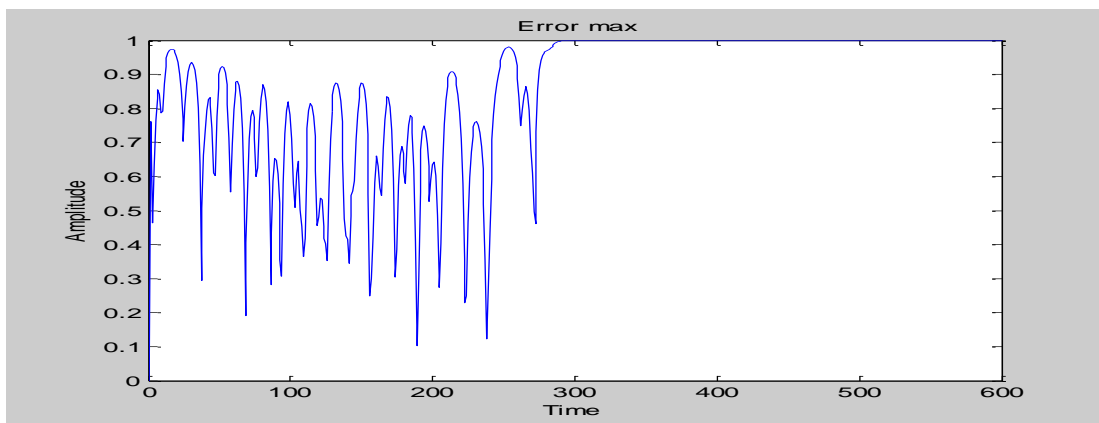
Gender	Suara <i>Voice</i>	Input (C_n)	FFT	P	Error Min	α	EPOCH	Max Error
Pria	Selamat Pagi	8	512	2	0.05	0.2	1-95	<1
							96-600	=1
		8	512	2	0.05	0.4	1-36	<1
							37-600	=1
		8	512	2	0.05	0.7	1-47	<1
							48-600	=1
		8	512	2	0.05	0.9	1-33	< 1
							34-600	=1
Wanita	Selamat Pagi	8	512	2	0.05	0.2	1-78	< 1
							79-600	=1
		8	512	2	0.05	0.4	1-60	< 1
							61-600	=1
		8	512	2	0.05	0.7	1-37	< 1
							38 – 600	=1
		8	512	2	0.05	0.9	1-34	< 1
							35-600	=1

Hasil pelatihan Jaringan syaraf tiruan *backpropagation* dengan inisialisasi perubahan bobot dari pengidentifikasian dan pengenalan suara *voice* pria dan wanita dengan menggunakan koefisien MFCC sebanyak 8, perubahan konstanta laju pelatihan antara $0 < \alpha < 1$, sampel 3 pria dan 3 wanita dapat diperlihatkan pada Tabel 4.8 dan Tabel 4.9. Perubahan bobot pada suara *voice* Pria mempengaruhi tingkat *error* dari pengenalan suara *voice*, dimana tingkat error yang digunakan antara $0 < Error < 1$ dengan nilai *error* minimal pengenalan *voice* 0.05 serta perubahan nilai konstanta laju pelatihan antara $0 < \alpha < 1$ dan nilai α yang digunakan pada simulasi 0.2. Dari beberapa teori yang digunakan dari bab sebelumnya dan simulasi yang digunakan diperoleh pengenalan suara *voice* pria memiliki nilai perubahan *Error* yang lebih besar atau sama dengan 1 dengan perubahan nilai bobot antara 96-600 dan nilai error kurang dari 1 dengan perubahan bobot antara 1-95 dengan proses iterasi $1 \leq EPOCH \leq 600$ dalam domain waktu. Proses perubahan error dan konstanta laju pelatihan ini mempengaruhi proses pengenalan suara, seperti yang diperlihatkan pada Gambar 4.10.



Gambar 4.10 Perubahan bobot suara *voice* Pria

Perubahan bobot pada suara *voice* wanita mempengaruhi tingkat *error* dari pengenalan suara *voice*, dimana tingkat error yang digunakan antara $0 < Error < 1$ dengan nilai *error* minimal pengenalan *voice* 0.05 serta perubahan nilai konstanta laju pelatihan antara $0 < \alpha < 1$ dan nilai α yang digunakan pada simulasi 0.2. Dari beberapa teori yang digunakan dari bab sebelumnya dan simulasi yang digunakan, diperoleh pengenalan suara *voice* wanita memiliki nilai perubahan *Error* yang lebih kecil dibandingkan perubahan error yang terjadi pada pengenalan suara *voice* pria dengan perubahan nilai bobot antara 376-600 atau sama dengan 1 dan nilai error kurang dari 1 dengan perubahan bobot antara 1-375 dengan proses iterasi $1 \leq EPOCH \leq 600$ dalam domain waktu. Proses perubahan error dan konstanta laju pelatihan ini mempengaruhi proses pengenalan suara, seperti yang diperlihatkan pada Gambar 4.11.



Gambar 4.11 Perubahan bobot suara *voice* Wanita

Tabel 4.8 Inisialisasi Perubahan Bobot dengan Jumlah Koefisien ($C_n=8$)

Gender	Suara Voice	Input (C_n)	FFT	P	Error Min	α	EPOCH	Max Error
Pria	Selamat Pagi	8	512	3	0.05	0.2	1-95	<1
							96-600	=1
		8	512	3	0.05	0.4	1-36	<1
							37-600	=1
		8	512	3	0.05	0.7	1-47	<1
							48-600	=1
		8	512	3	0.05	0.9	1-33	< 1
							34-600	=1
Wanita	Selamat Pagi	8	512	3	0.05	0.2	1-375	< 1
							376-600	=1
		8	512	3	0.05	0.4	1-65	< 1
							66-600	=1
		8	512	3	0.05	0.7	1-169	< 1
							170-600	=1
		8	512	3	0.05	0.9	1-146	< 1
							147-600	=1

Pada JST *Backpropagation*, terdapat beberapa parameter yang dapat menentukan kinerja JST dalam memproses input yang baru. Pada penelitian *voice identification* diuji parameter JST dengan merubah beberapa parameter ; jumlah pelatihan, konstanta laju pelatihan, nilai input layer, dan hidden layer. Terlihat bahwa beberapa parameter mempengaruhi perubahan bobot dari pengidentifikasian *voice* antara suara pria dan suara wanita. Siklus perubahan bobot dilakukan dengan pengubahan jumlah koefisien ciri (C_n) dengan sebagai masukan pada Jaringan Syaraf Tiruan. Berdasarkan pelatihan yang dilakukan menghasilkan hasil yang optimal dengan tingkat error yang melebihi batasan error (target error) dan nilai error lebih kecil dari satu ($0.05 < \text{error} < 1$) dengan jumlah pelatihan sebanyak tiga kali pada suara *voice* pria dan suara *voice* wanita, panjang *frame* 256, jumlah Mel Bank Filter 40, jumlah koefisien 8, FFT 512 dan nilai α adalah 0.2. Hasil proses iterasi yang dilakukan dengan menggunakan sampel suara dan disimpan dalam database dengan menggunakan software visual basic untuk menampilkan pengidentifikasian suara pria dan wanita yang terdapat pada Lampiran 5 dan Lampiran 6. Proses pengidentifikasian memiliki kemampuan pengidentifikasian suara *voice* dari gender wanita lebih baik dibandingkan pria. Error pada pendeteksian suara *voice* wanita lebih kecil dari proses inisialisasi jaringan syaraf tiruan *backpropagation*.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pada hasil penelitian, pendekatan teknik pengolahan sinyal dilakukan untuk mengenali suku kata bahasa lisan. Hal ini ditunjukkan pada observasi utama penelitian ini. Pendekatan ekstraksi ciri MFCC tidak cocok mengenali suara dengan jumlah target yang besar, namun ekstraksi ciri cocok untuk mengenali suara dengan jumlah data yang kecil. Adapun beberapa fakta lain yang muncul sebagai hasil penelitian pengenalan suara pengucapan (identifikasi *voice*) suku kata bahasa lisan Bahasa Indonesia adalah sebagai berikut.

1. Tingkat keakuratan pengenalan suara/ucapan dipengaruhi beberapa faktor eksternal dan internal. Faktor eksternal dipengaruhi kondisi lingkungan , letak mikrofon dan jenis mikrofon atau alat perekam yang digunakan, cara perekaman sinyal suara dan kondisi peralatan. Faktor Internal dipengaruhi oleh kondisi fisik dan psikis berupa intonasi suara, jenis kelamin dan usia responden.
2. Metode *Mel Frequency Cepstrum Coefficients* (MFCC) ini dapat dipakai sebagai metode yang baik dalam melakukan *Feature Extraction*.

Keluaran dari MFCC adalah koefisien ciri yang berisi nilai-nilai yang mewakili sinyal ucapan yang dapat digunakan sebagai ciri/feature dan pengenalan atau *identification* menggunakan jaringan syaraf tiruan.

3. Kemampuan pengenalan cenderung semakin baik ketika jumlah data pelatihan yang digunakan semakin besar.
4. Keberhasilan aplikasi dipengaruhi pada seberapa *noise* yang datang dan mempengaruhi tingkat akurasi dari pengenalan pola suara. Hal ini juga mempengaruhi Siklus perubahan bobot (EPOCH) yang kecil akan menghasilkan maximum Error yang semakin kecil sehingga pengenalan suara akan semakin bagus. Perubahan nilai target error ($0.05 \leq \text{target error} \leq 1$) yang semakin tinggi mempengaruhi siklus perubahan bobot (EPOCH) yang semakin kecil juga.
5. Semakin tinggi nilai konstanta laju pelatihan / *learning rate* (α) yang digunakan mempengaruhi tingkat Error yang semakin tinggi.
6. Eksplorasi parameter yang digunakan tergantung dari nilai yang digunakan, *framing*, jumlah koefisien ciri, siklus laju pelatihan dan jumlah data pelatihan, target error untuk menjelaskan hal-hal yang mempengaruhi kemampuan pengenalan suara *voice* terhadap suku kata bahasa lisan Bahasa Indonesia dari sumber suara yang digunakan sebagai masukan.

5.2 Saran

Dalam rangka memperbaiki kemampuan pengenalan terhadap suku kata bahasa lisan Bahasa Indonesia, penulis menyarankan untuk melakukan klasifikasi awal terlebih dahulu pada penelitian lebih lanjut. Klasifikasi awal dapat dimulai dengan mensurvei perbedaan bunyi suku kata. Hal ini didasarkan pada hasil pengujian yang menunjukkan kemampuan pengenalan yang lebih baik pada jumlah target yang kecil dan suku kata dengan bunyi yang berbeda.

Selain dengan menggunakan klasifikasi awal, cara lain yang dapat dilakukan adalah menguji kemampuan ciri lain sebagai wakil dari sinyal digital suara ucapan suku kata bahasa lisan Bahasa Indonesia.

1. Meningkatkan kemampuan penelitian untuk membedakan suara *noise*. Menambahkan modul-modul perintah seperti shortcut pada aplikasi Microsoft word.
2. Aplikasi *voice command* ini dapat bermanfaat bagi pengguna komputer yang mengalami cacat fisik maka diperlukan perluasan perintah pada aplikasi.
3. Menciptakan aplikasi sistem/perangkat hardware yang bisa mengidentifikasi *voice* dengan perbedaan suara pria dan wanita dengan penanggulangan pengaruh *noise* akibat distorsi alat baik eksternal dan internal.

DAFTAR PUSTAKA

- [1] A. H. Mansour, G. Z. A. Salh, and K. A. Mohammed, "Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms," *Int. J. Comput. Appl.*, vol. 116, no. 2, pp. 34–41, 2015.
- [2] K. Chakraborty Asmita Talele Savitha Upadhya, "Voice Recognition Using MFCC Algorithm," *Int. J. Innov. Res. Adv. Eng.*, vol. 1, no. 10, pp. 2349–2163, 2014.
- [3] M. E. Alhassan *et al.*, "Biometric voice recognition in security system," *Indian J. Sci. Technol.*, vol. 7, no. 2, pp. 104–112, 2013.
- [4] I. L. May, "Pengenalan vokal bahasa Indonesia dengan jaringan syaraf tiruan melalui transformasi wavelet diskret," pp. 1–5, 2002.
- [5] M. Devarajan, Fansheng Meng, P. Hix, and S. A. Zahorian, "HMM-neural network monophone models for computer-based articulation training for the hearing impaired," *2003 IEEE Int. Conf. Acoust. Speech, Signal Process. 2003. Proceedings. (ICASSP '03).*, vol. 2, p. II-369-72.
- [6] K. Kasi and S. a. Zahorian, "Yet Another Algorithm for Pitch Tracking," *2002 IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 1, p. I-361-I-364, 2002.
- [7] T. K. Das and K. M. O. Nahar, "A Voice Identification System using Hidden Markov Model," *Indian J. Sci. Technol.*, vol. 9, no. January, 2016.
- [8] A. Parab, "Speaker Recognition Using Mfcc and," vol. 3, no. 2, pp. 17–20, 2014.
- [9] M. Karnjanadecha and S. a. Zahorian, "Signal modeling for isolated word recognition," *1999 IEEE Int. Conf. Acoust. Speech, Signal Process. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 1, no. 6, pp. 647–654, 1999.
- [10] M. Nilsson, *Entropy and Speech*. 2006.
- [11] Z. B. Fiqhi, R. R. Isnanto, and M. Somantri, "Pengenalan Tanda Tangan Menggunakan Analisis Komponen Utama (Principal Component Analysis - PCA) Dan Metode Jaringan Saraf Tiruan Perambatan Balik," *Transient*, vol. 3, no. 2, pp. 1–8, 2014.
- [12] Abriyono and A. Harjoko, "Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST," *Indones. J. Comput. Cybern. Syst.*, vol. 6, no. 2, pp. 23–34, 2012.
- [13] S. Singh and E. . Rajan, "MFCC VQ based Speaker Recognition and Its Accuracy Affecting Factors," *Int. J. Comput. Appl.*, vol. 21, no. 6, pp. 1–6, 2011.

- [14] R. Faculty, "CURRICULUM 1 . CURRICULUM DESCRIPTION The list of compulsory and elective courses with the number of active classes required for their performance and ECTS credits," no. May, 2017.
- [15] Y. J. Khaw, T. Tan, and B. Ranaivo-malançon, "Automatic Phoneme Identification for Malay Dialects," vol. 9, no. 2, p. 2.
- [16] J. Sangeetha and S. Jothilakshmi, "Automatic continuous speech recogniser for Dravidian languages using the auto associative neural network," *Int. J. Comput. Vis. Robot.*, vol. 6, no. 1/2, p. 113, 2016.
- [17] J. H. L. Hansen, R. Huang, P. Mangalath, B. Zhou, M. Seadle, and J. John. R. Deller, "SPEECHFIND: spoken document retrieval for a national gallery of the spoken word," *Nord. Signal Process. Symp.*, vol. 13, no. 5, pp. 1–4, 2004.
- [18] J. Ramirez, J. M., and J. C., "Voice Activity Detection. Fundamentals and Speech Recognition System Robustness," *Robust Speech Recognit. Underst.*, no. June, pp. 1–22, 2007.
- [19] P. Borde, "Recognition of isolated words using Zernike and MFCC features for audio visual speech recognition," *Int. J. Speech Technol.*, vol. 18, pp. 167–175, 2015.
- [20] J. Ilmiah, I. Komputa, A. D. Andriana, J. Dipati, and U. Bandung, "METODE MEL FREQUENCY CEPSTRUM COEFFICIENTS Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)," *Ilm. Komput. dan Inform.*, vol. 2, no. 1, pp. 21–26, 2013.
- [21] C. Simamora, M. Pinem, and F. Fahmi, "Voice Identification Using Neural Network and Mel Frequency Cepstrum Coefficients."
- [22] M. F. BenZeghiba, "Joint speech and speaker recognition," 2005.
- [23] G. K. Venayagamoorthy, V. K. Moonasar Sandrasegaran, V. Moonasar, K. Sandrasegaran, K. Sandrasegaran, and massey ac nz, "Scholars' Mine Voice Recognition using Neural Networks VOICE RECOGNITION USING NEURAL NETWORKS," 1998.
- [24] A. D. Dongare, R. R. Kharde, and A. D. Kachare, "Introduction to Artificial Neural Network," *Int. J. Eng. Innov. Technol.*, vol. 2, no. 1, pp. 189–194, 2012.
- [25] W. Chan, N. Jaitly, Q. Le, and O. Vinyals, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4960–4964.

LAMPIRAN 1

Hasil vektor ciri suara *voiced pria Mel Frequency Cepstrum Coefficients*

dengan jumlah koefisien 6 dan 8

Suara Voice	Vektor Ciri (C _n =6)	Vektor Ciri (C _n =8)
Apakabar	-0.4797 -0.2424 -5.1478 -0.2081 -0.8504 0.0351	-0.4856 -0.2580 -6.0429 -0.2892 - 1.5972 0.1343 0.6174 -0.254
	2.8096 -0.4416 -8.8774 -1.4726 -2.1110 0.6250	2.8441 -0.4699 -10.4210 -2.0463 - 3.9650 2.3927 -0.7388 0.5470
	3.5699 -2.5954 -10.0991 -1.8565 -1.4437 0.6425	3.6138 -2.7616 -11.8552 -2.5799 - 2.7117 2.4599 0.0637 0.4953
Selamat Malam	4.1233 -2.4407 -8.1262 -2.1284 -3.0554 1.4346	4.1740 -2.5970 -9.5392 -2.9577 - 5.7386 5.4921 -2.1221 0.1443
	2.2063 -3.8183 -2.7967 0.9550 -0.3429 1.0123	2.2334 -4.0627 -3.2830 1.3270 - 0.6440 3.8757 -3.8071 -0.4591
	1.0431 -3.5865 -2.2493 0.9183 -0.8203 1.1422	1.0559 -3.8161 -2.6404 1.2761 - 1.5407 4.3728 -2.7898 -0.5396
Selamat Pagi	3.2366 -4.2290 4.0897 0.2167 -5.2322 -0.3807	3.2764 -4.4997 4.8008 0.3011 - 9.8272 -1.4573 -0.6246 0.0687
	3.6363 -5.7737 6.6668 -0.1531 -5.0774 -0.2149	3.6810 -6.1433 7.8260 -0.2127 - 9.5365 -0.8229 -1.7773 -0.1168
	1.7715 -6.9915 5.3845 2.1771 -4.3714 0.3047	1.7933 -7.4391 6.3208 3.0253 - 8.2103 1.1666 -2.3823 0.0432
Hallo	3.8303 6.0887 -3.4665 -5.9714 -2.3607 0.5788	1.7933 -7.4391 6.3208 3.0253 - 8.2103 1.1666 -2.3823 0.0432
	1.7650 3.7312 -5.0402 -6.1668 -1.2545 0.6083	1.7867 3.9700 -5.9165 -8.5696 - 2.3563 2.3290 0.4173 -0.1315
	1.4463 0 0	1.4641 0 0 0
	0 0 0	0 0 0 0

Hasil vektor ciri suara *voiced* wanita *Mel Frequency Cepstrum Coefficients*

dengan jumlah koefisien 6 dan 8

Suara Voice	Vektor Ciri (C _n =6)			Vektor Ciri (C _n =8)				
Apakabar	3.1367 -3.6578	2.7135 -1.4732	-4.8906 0.0659	3.0476 2.1844	2.0393 0.5254	-4.4242 -0.1185	-4.4484 0.5810	-
	0.7910 -1.3230	0.2729 -2.9432	-1.9493 0.3769	3.0476 2.1844	2.0393 0.5254	-4.4242 -0.1185	-4.4484 0.5810	-
	4.2287 -3.2051	0.1558 -1.3073	-9.8505 0.9671	3.0476 2.1844	2.0393 0.5254	-4.4242 -0.1185	-4.4484 0.5810	-
Selamat Malam	1.5252 0.6666	1.8376 -0.0862	1.5144 -0.2416	1.3952 0.1084	1.2975 -0.6147	1.1795 -0.6728	0.6141 -0.3217	-
	0.5556 0.2578	1.6928 -0.3178	1.1760 -0.2580	1.3952 0.1084	1.2975 -0.6147	1.1795 -0.6728	0.6141 -0.3217	-
	-0.0534 -0.8108	1.6427 0.5857	-1.8602 -0.1270	-0.2282 0.3281	0.8108 -0.5679	-0.7328 -0.7891	-0.3643 -0.0575	-
Selamat Pagi	1.4976 -0.8281	-0.2095 -0.9330	3.0365 -0.2215	1.4196 0.8463	0.8408 -0.6549	1.1024 -0.8738	-0.1453 -0.1221	-
	3.1924 -0.9151	-2.2442 -1.7495	5.1346 -0.1678	3.2675 2.2294	-0.7399 0.1377	4.5034 -2.1410	-0.6672 0.0366	-
	2.0948 -1.2813	1.6337 -2.3478	2.1533 -0.2205	2.3056 3.2675	1.7859 -1.4435	1.2685 1.1225	-1.2152 0.0553	-
Hallo	4.0021 -4.7925	4.8572 -1.2838	-0.8894 0.2354	4.0513 2.4112	5.1681 0.9011	-1.0440 -1.3025	-6.6598 -0.6497	-
	5.4519 10.1718	4.7743 -1.2836	-3.7758 1.0392	5.5189 2.4108	5.0799 3.9784	-4.4324 -0.2347	-14.1351 -0.0591	-
	1.8891 0	0 0	0	1.9123 0	0 0	0 0	0 0	

Hasil vektor ciri suara *voiced pria Mel Frequency Cepstrum Coefficients*

dengan jumlah koefisien 10 dan 20

Suara Voice	Vektor Ciri ($C_n=10$)	Vektor Ciri ($C_n=20$)
Apakabar	3.1932 2.9705 -6.1683 - 5.8509 -3.5357 0.3790 - 1.8367 2.3946 -2.1041 - 0.3896	3.2174 3.0846 -6.7733 -6.9921 - 4.7562 0.5986 -3.6078 6.3897 - 8.9924 -4.2856 5.5030 -2.3677 - 0.0360 -1.3647 3.2309 2.3474 2.5127 1.5213 -0.6610 -0.3456
	0.8052 0.2987 -2.4586 - 2.1162 -7.0638 2.1692 - 0.2721 1.2220 -0.0829 - 0.4400	0.8113 0.3102 -2.6997 -2.5290 - 9.5020 3.4261 -0.5344 3.2609 - 0.3543 -4.8402 5.5498 -3.8561 - 0.5230 3.4442 -0.4354 0.8628 1.0447 0.7335 -0.0629 -0.2296
	4.3049 0.1706 -12.4241 -5.1267 -3.1375 5.5660 -0.3366 2.2573 -2.2119 -0.5722	4.3375 0.1771 -13.6427 -6.1267 - 4.2205 8.7912 -0.6612 6.0233 - 9.4531 -6.2942 3.1944 -4.7020 2.9246 1.9780 0.4966 -0.2703 1.9493 1.3974 0.3314 -0.0295
Selamat Malam	1.5527 2.0116 1.9101 1.0662 -0.2069 -1.3906 - 2.0196 -1.9087 -1.2290 - 0.4000	1.5644 2.0889 2.0974 1.2742 - 0.2784 -2.1964 -3.9672 -5.0931 - 5.2523 -4.3999 -2.7783 -0.8330 0.9338 2.1203 2.5406 2.2601 1.5445 0.7448 0.1637 -0.0480
	0.5656 1.8531 1.4833 0.4124 -0.7628 -1.4849 - 1.5895 -1.2628 -0.7663 - 0.2666	0.5699 1.9243 1.6287 0.4929 - 1.0261 -2.3454 -3.1222 -3.3696 - 3.2749 -2.9324 -2.2544 -1.1384 0.2900 1.6141 2.3414 2.2355 1.4896 0.5948 0.0100 -0.1059
	-0.0544 1.7983 -2.3463 -1.2969 1.4056 -0.7312 -2.5225 -0.6968 0.2255	-0.0548 1.8673 -2.5764 -1.5499 1.8908 -1.1549 -4.9550 -1.8595 0.9636 -1.8612 -3.1931 -0.0856 1.6811 1.1599 1.8390 2.6947

	-0.1692	1.5533 0.0238 -0.3646 -0.2138
Selamat Pagi	3.1735 -3.5771 3.5908 0.5109 -10.7485 -1.0905 -2.6798 0.4596 0.5553 1.0384	1.4384 0.8983 1.3006 -0.1998 - 1.4547 -1.5551 -3.4216 -1.2831 - 2.8820 -0.6792 -0.7436 0.2219 2.1994 1.5555 3.1908 1.2124 1.0396 - 0.1675 -0.5112 -0.1804
	1.1735 -2.5771 2.5908 -1.5109 -8.7485 -2.0905 -1.6798 1.4596 -1.5553 -1.0384	3.3109 -0.7905 5.3132 -0.9178 - 3.8321 0.3269 -8.3838 0.3851 - 4.6188 -3.0823 -1.8487 -4.4185 - 0.1764 -0.2323 2.4548 2.7766 2.1455 1.4016 0.1222 0.0644
	2.1735 -2.5771 2.5908 -1.5109 -8.7485 -2.0905 -3.6798 -1.4596 -1.5553 0.0384	2.1487 1.8571 2.9823 -2.4493 - 7.5796 -2.0042 -6.3719 1.8243 - 1.1064 -1.9980 -1.6711 -3.5783 - 0.1072 1.5368 3.6101 3.6513 2.4218 0.4797 -0.5888 -0.3429
Hallo	4.0743 5.3173 -1.1218 - 7.6658 -3.0811 1.3547 - 2.5965 -2.5590 -1.6852 0.1919	4.1052 5.5215 -1.2318 -9.1610 - 4.1446 2.1397 -5.1003 -6.8283 - 7.2021 2.1104 1.2611 -4.6329 - 4.2755 -0.5489 3.9648 4.0208 1.6020 1.9609 0.3279 0.0850
	5.5502 5.2265 -4.7623 - 16.2703 -3.0806 5.9808 -0.4678 -0.2328 -2.3644 0.3945	5.5922 5.4272 -5.2294 -19.4438 - 4.1439 9.4463 -0.9189 -0.6212 - 10.1046 4.3396 -4.9674 -3.5105 - 4.2177 0.4787 2.8922 2.6162 2.6111 2.6184 0.8888 -0.0222
	1.9232 0 0 0 0 0 0 0 0	1.9377 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Hasil vektor ciri suara *voiced* wanita *Mel Frequency Cepstrum Coefficients*

dengan jumlah koefisien 10 dan 20

<i>Suara Voice</i>	Vektor Ciri ($C_n=10$)	Vektor Ciri ($C_n=20$)
Apakabar	-0.4884 -0.2654 - 6.4927 -0.3328 - 2.0410 0.2018 1.2309 -1.0003 1.1160 0.0989	-0.4921 -0.2756 -7.1295 -0.3978 - 2.7455 0.3188 2.4178 -2.6692 4.7696 -1.0876 -3.3281 3.6715 -1.6481 0.7441 0.8142 -1.0197 0.2557 0.1490 -0.1885 0.0618
	2.8603 -0.4835 - 11.1968 -2.3554 - 5.0665 3.5970 - 1.4728 2.1546 2.1994 -0.7127	2.8819 -0.5020 -12.2950 -2.8149 - 6.8153 5.6812 -2.8931 5.7493 9.3994 -7.8394 -6.6487 2.5181 0.3284 - 0.9979 -0.0711 -1.2629 2.4780 -0.0765 -0.2361 0.0638
	3.6343 -2.8413 - 12.7377 -2.9696 - 3.4650 3.6979 0.1269 1.9509 1.7270 -0.2163	3.6618 -2.9504 -13.9871 -3.5489 - 4.6610 5.8407 0.2493 5.2059 7.3804 -2.3790 -12.2446 6.6276 -3.8057 0.9638 -1.4683 -1.3550 1.4855 -1.1581 0.7710 0.0759
Selamat Malam	4.1977 -2.6719 - 10.2494 -3.4045 - 7.3329 8.2563 - 4.2304 0.5686 0.4015 0.6545	4.2294 -2.7745 -11.2546 -4.0685 - 9.8641 13.0405 -8.3098 1.5172 1.7158 7.1998 -5.7017 -3.3364 1.0969 -2.3366 3.0833 -3.1554 0.2974 -0.1088 -0.2865 -0.1311
	2.2461 -4.1800 - 3.5274 1.5275 - 0.8229 5.8263 - 7.5895 -1.8083 - 0.4838 0.7903	2.2631 -4.3405 -3.8734 1.8254 -1.1069 9.2024 -14.9081 -4.8253 -2.0677 8.6930 -4.1059 2.5030 0.3707 -5.0910 2.2556 -1.6723 -1.2351 0.4278 0.0367 -0.0033

	1.0619 -3.9263 - 2.8370 1.4688 - 1.9688 6.5736 - 5.5615 -2.1255 0.1101 0.8107	1.0699 -4.0770 -3.1152 1.7553 -2.6484 10.3827 -10.9244 -5.6717 0.4707 8.9173 -1.7035 -0.3778 -0.0427 - 3.2547 3.0022 -1.2420 -1.0852 -0.1658 0.0285 -0.1951
Selamat Pagi	3.1735 -3.5771 3.5908 0.5109 - 10.7485-1.0905 - 2.67980.4596 0.5553 1.0384	3.1975 -3.7145 3.9429 0.6106 -14.4587 1.7225 -5.2639 1.2264 2.3731 11.4219 3.3244 2.1286 -0.7292 -4.3527 0.9124 1.8789 -0.5717 1.4242 -0.3203 -0.2216
	3.7019 -6.3207 8.4086 -0.2449 - 12.1858 -1.2370 - 3.5430 -0.4601 - 0.5030 0.9309	3.7299 -6.5633 9.2333 -0.2927 - 16.3921 -1.9538 -6.9596 -1.2277 - 2.1497 10.2395 -5.7835 5.7349 - 0.3862 -2.0697 0.9124 1.8789 -0.5717 1.4242 -0.3203 -0.2216
	1.8035 -7.6539 6.7914 3.4823 - 10.4912 1.7538 - 4.7492 0.1700 - 1.7004 0.9727	1.8171 -7.9477 7.4574 4.1615 -14.112 62.7700 9.3289 0.4536 -7.2668 10.6997 - 5.7705 3.5743 1.8888 -2.015 0.0847 1.2693 -1.8612 0.5042 -0.2997-0.2007
Hallo	3.8994 6.6655 - 4.3722 -9.5516 - 5.6657 3.3314 - 0.1676 -1.6409 0.7146 0.3605	3.9289 6.9214 -4.8011 -11.4146 - 7.6214 5.2618 -0.3293 -4.3785 3.0539 3.9659 1.1850 -4.4204 -6.4646 0.5907 1.5864 -1.7266 -0.5446 -1.2544 - 0.1910 0.0011
	1.7968 4.0846 - 6.3570 -9.8640 - 3.0109 3.5012 0.8320 -0.5181 1.1411 0.7209	1.8104 4.2415 -6.9805 -11.7880 - 4.0502 5.5300 1.6343 -1.3824 4.8769 7.9296 0.2969 -6.5759 -5.4550 0.8999 3.0686 -1.0711 -0.5108 0.4114 0.4722 0.2129
	1.4724 0 0 0 0 0 0 0 0 0	1.4835 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

LAMPIRAN 2

Hasil Perubahan Bobot dengan 6 koefisien ciri ($C_n=6$)

Gender	Suara <i>Voice</i>	EPOCH	Nilai MaxError	EPOCH	Nilai MaxError
Pria	Apa kabar	1 – 71	< 1	72 – 600	1
	Hallo	1 – 75	< 1	76 – 600	1
	Selamat Pagi	1 – 103	< 1	104 - 600	1
	Selamat Malam	1 – 60	< 1	61 – 600	1
Wanita	Apa kabar	1 – 59	< 1	60 – 600	1
	Hallo	1 – 57	< 1	58 – 600	1
	Selamat Pagi	1 – 350	< 1	351 - 600	1
	Selamat Malam	1 – 58	< 1	59- 600	1

Hasil Perubahan Bobot dengan 8 koefisien ciri ($C_n=8$)

Gender	Suara <i>Voice</i>	EPOCH	Nilai MaxError	EPOCH	Nilai MaxError
Pria	Apa kabar	1 – 71	< 1	72 – 600	1
	Hallo	1 – 65	< 1	66 – 600	1
	Selamat Pagi	1 – 66	< 1	67 – 600	1
	Selamat Malam	1 – 60	< 1	61 – 600	1
Wanita	Apa kabar	1 – 67	< 1	68 – 600	1
	Hallo	1 – 55	< 1	56 – 600	1
	Selamat Pagi	1 – 102	< 1	103 - 600	1
	Selamat Malam	1 – 66	< 1	67 – 600	1

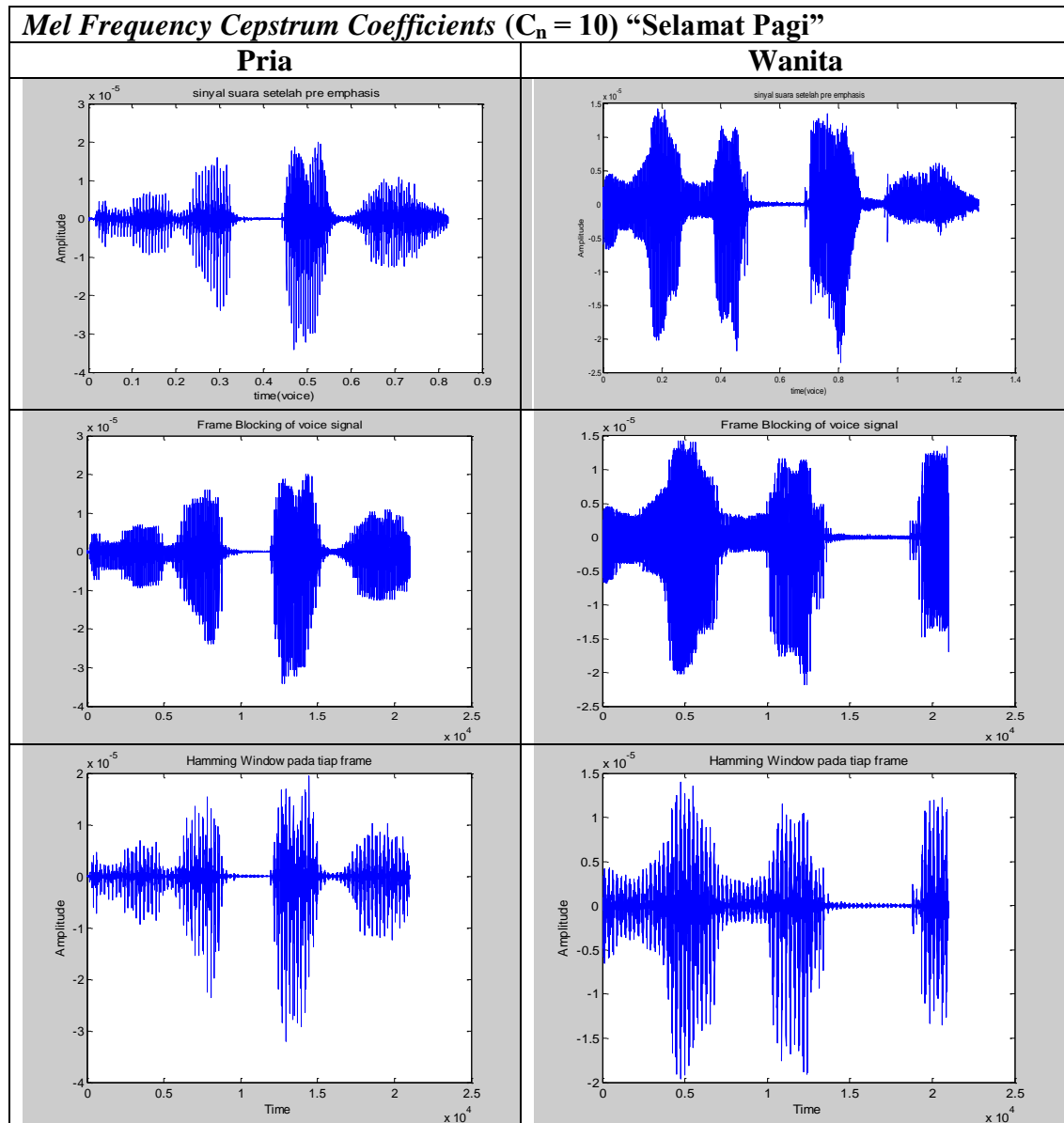
Hasil Perubahan Bobot dengan 10 koefisien ciri ($C_n=10$)

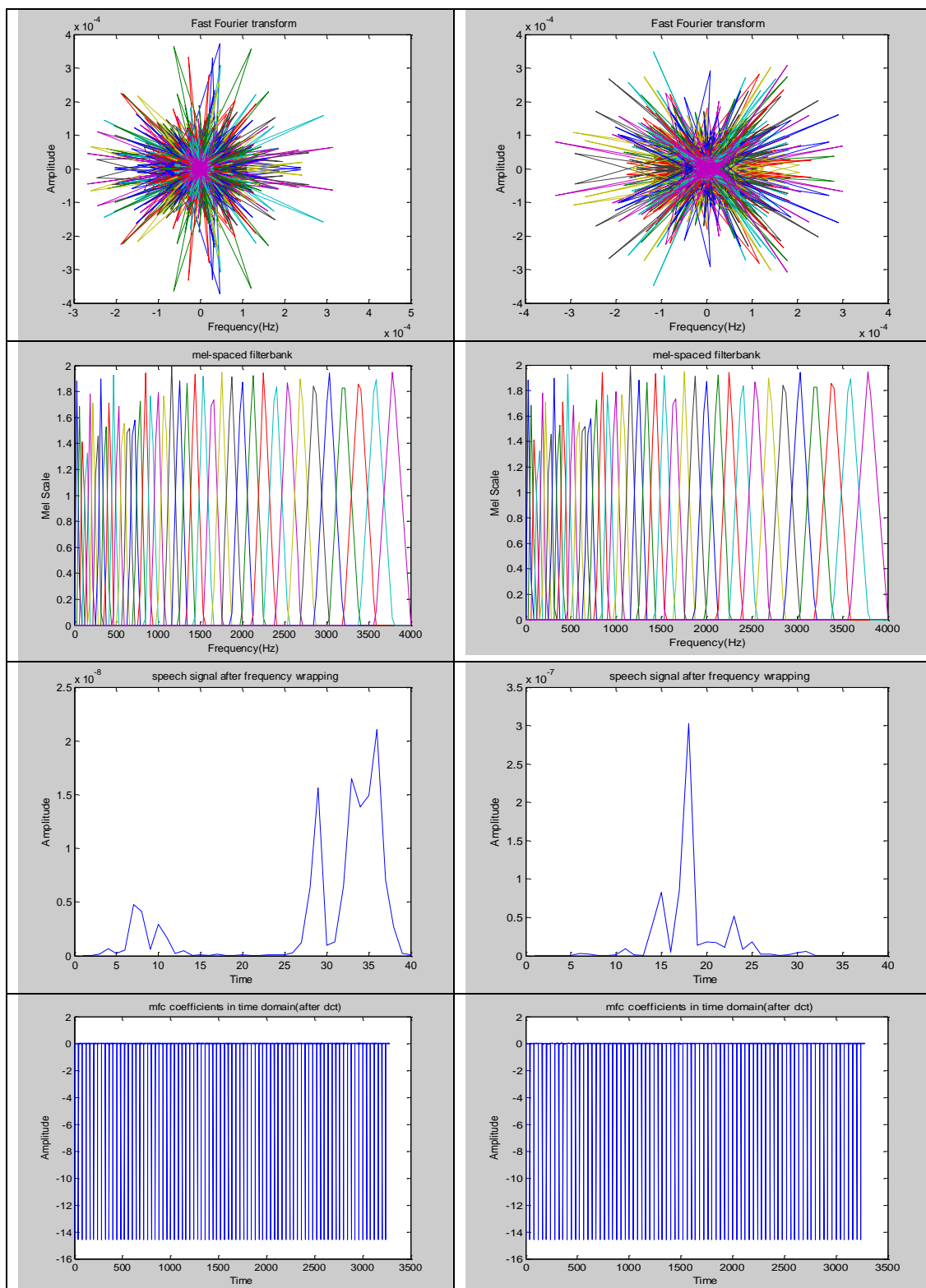
Gender	Suara <i>Voice</i>	EPOCH	Nilai MaxError	EPOCH	Nilai MaxError
Pria	Apa kabar	1-441	< 1	442 - 600	1
	Hallo	1-108	< 1	109 - 600	1
	Selamat Pagi	1-64	< 1	65 – 600	1
	Selamat Malam	1-94	< 1	95- 600	1
Wanita	Apa kabar	1-53	< 1	54 – 600	1
	Hallo	1-102	< 1	103 - 600	1
	Selamat Pagi	1-64	< 1	65 – 600	1
	Selamat Malam	1-54	< 1	55-600	1

Hasil Perubahan Bobot dengan 20 koefisien ciri ($C_n=20$)

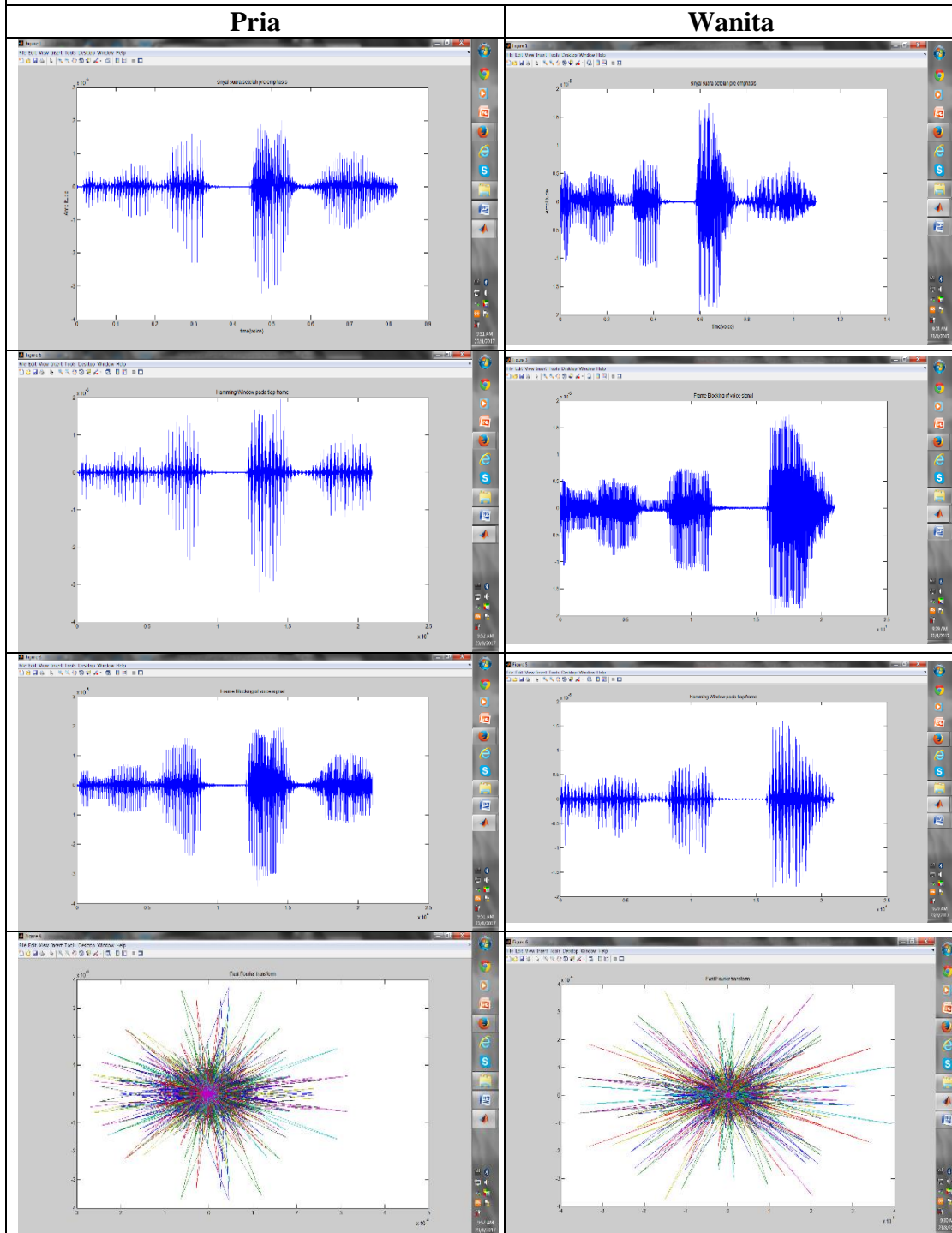
Gender	Suara <i>Voice</i>	EPOCH	Nilai MaxError	EPOCH	Nilai MaxError
Pria	Apa kabar	1-110	< 1	111-600	1
	Hallo	1-48	< 1	49-600	1
	Selamat Pagi	1-39	< 1	40-600	1
	Selamat Malam	1-168	< 1	169-600	1
Wanita	Apa kabar	1-49	< 1	50-600	1
	Hallo	1-66	< 1	67-600	1
	Selamat Pagi	1-59	< 1	60-600	1
	Selamat Malam	1-60	< 1	61-600	1

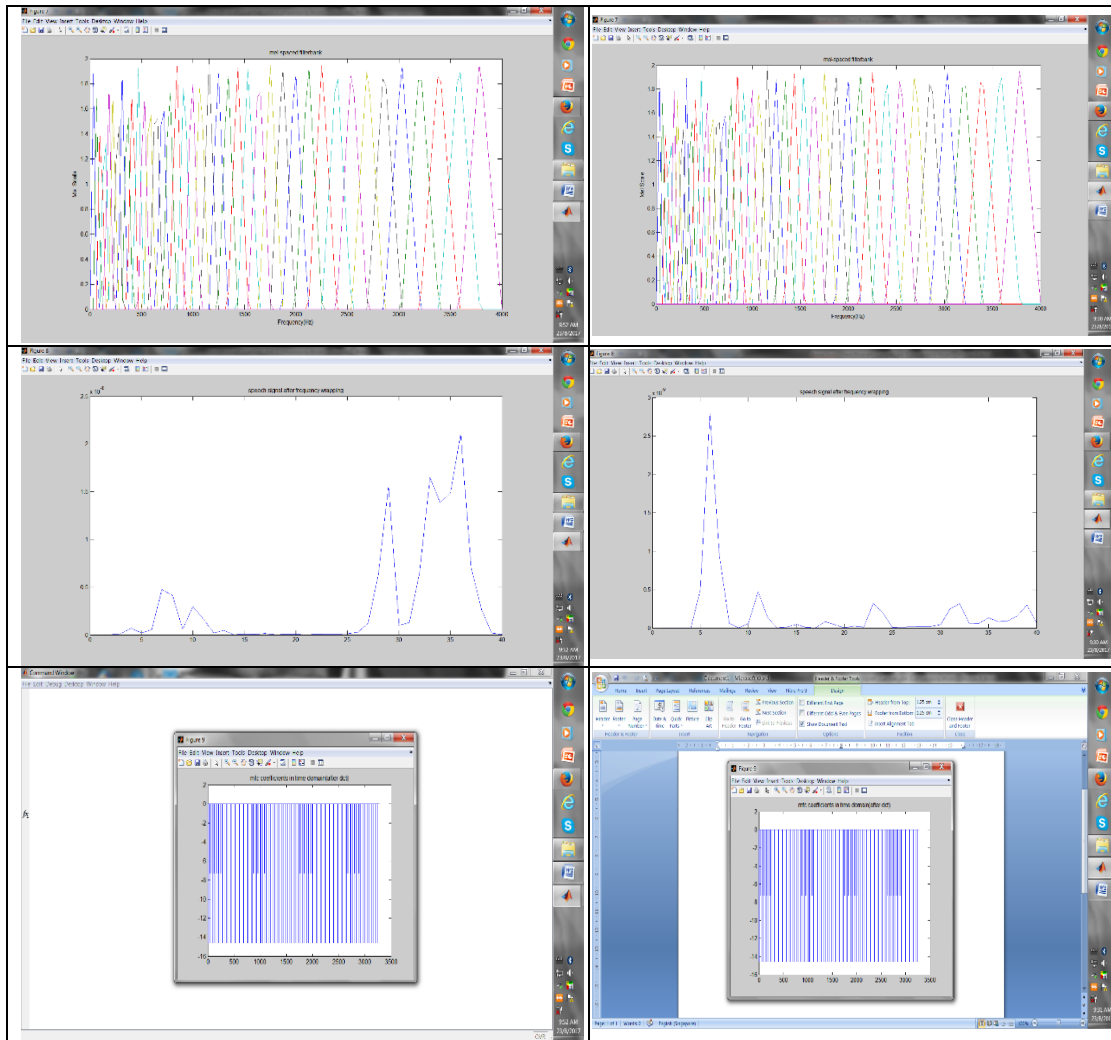
LAMPIRAN 3

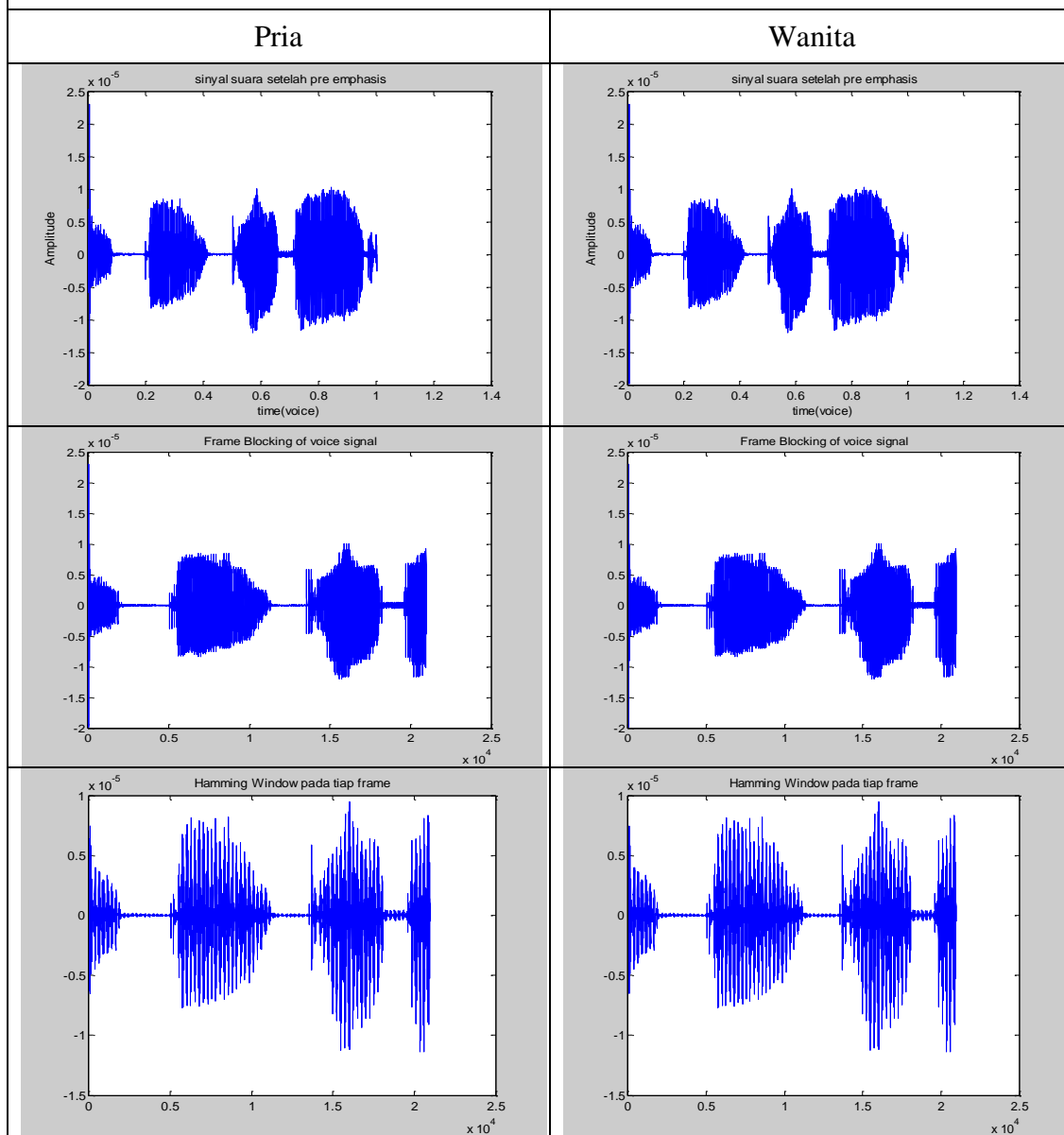


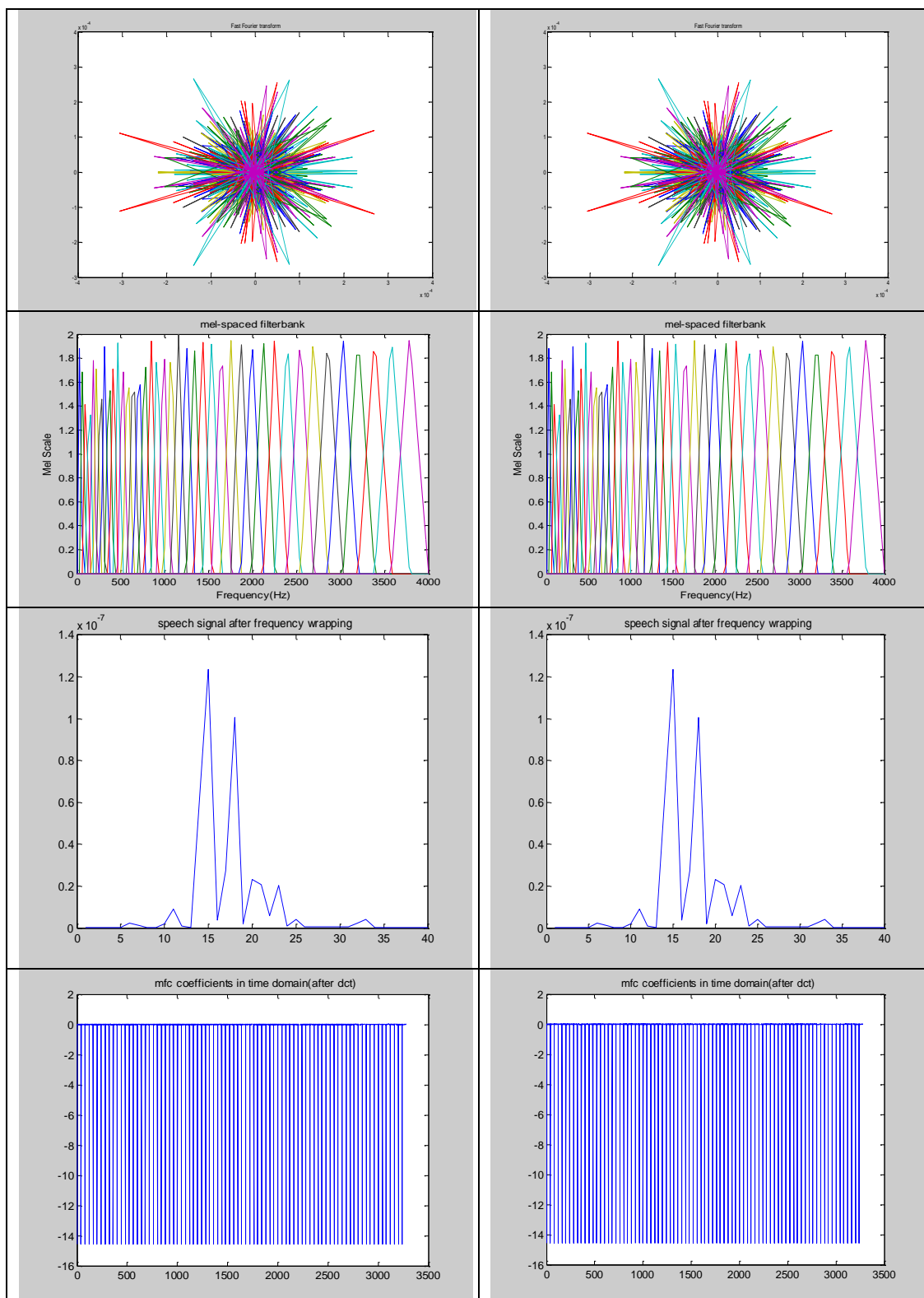


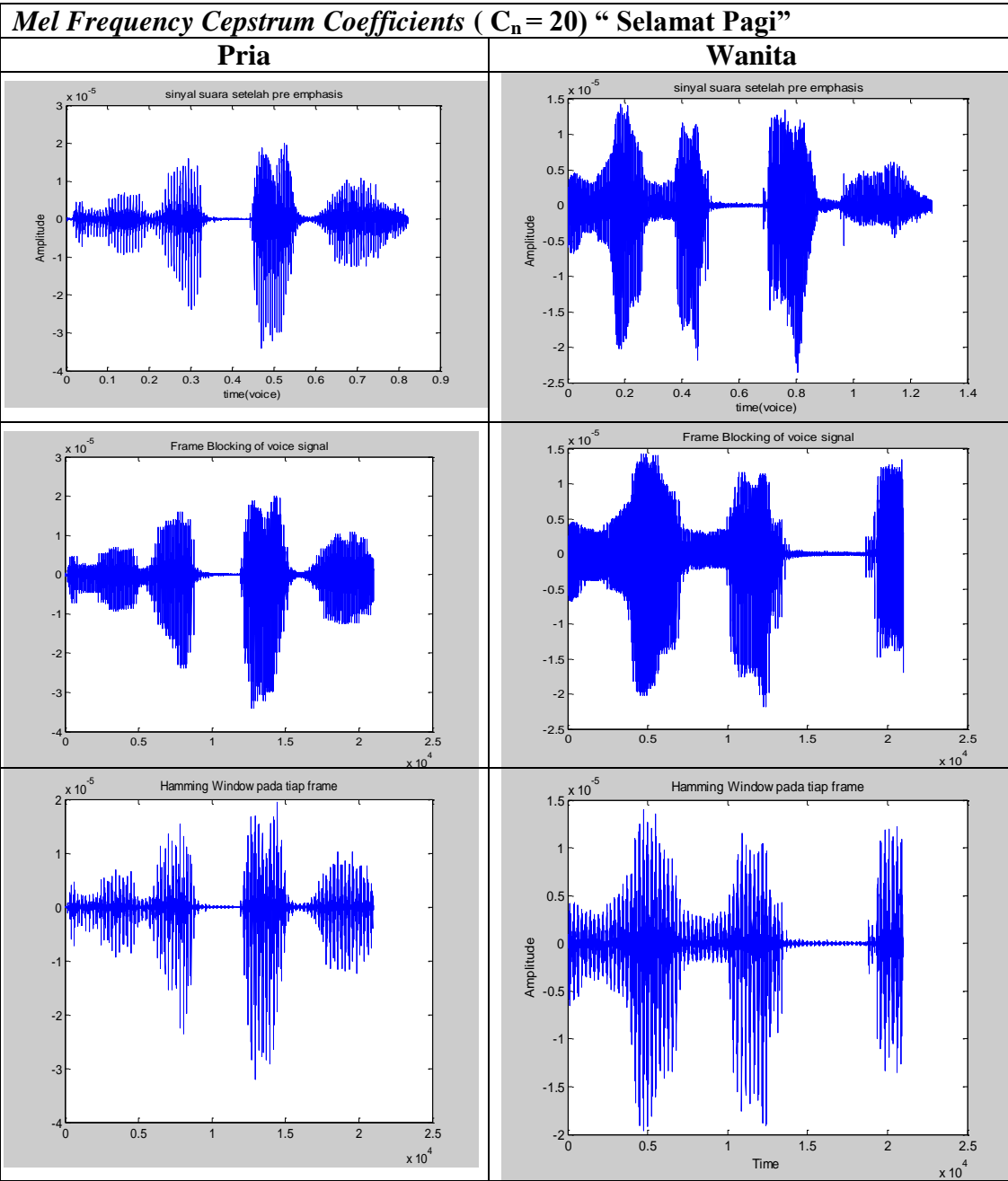
Mel Frequency Cepstrum Coefficients ($C_n = 8$)” Selamat Pagi”

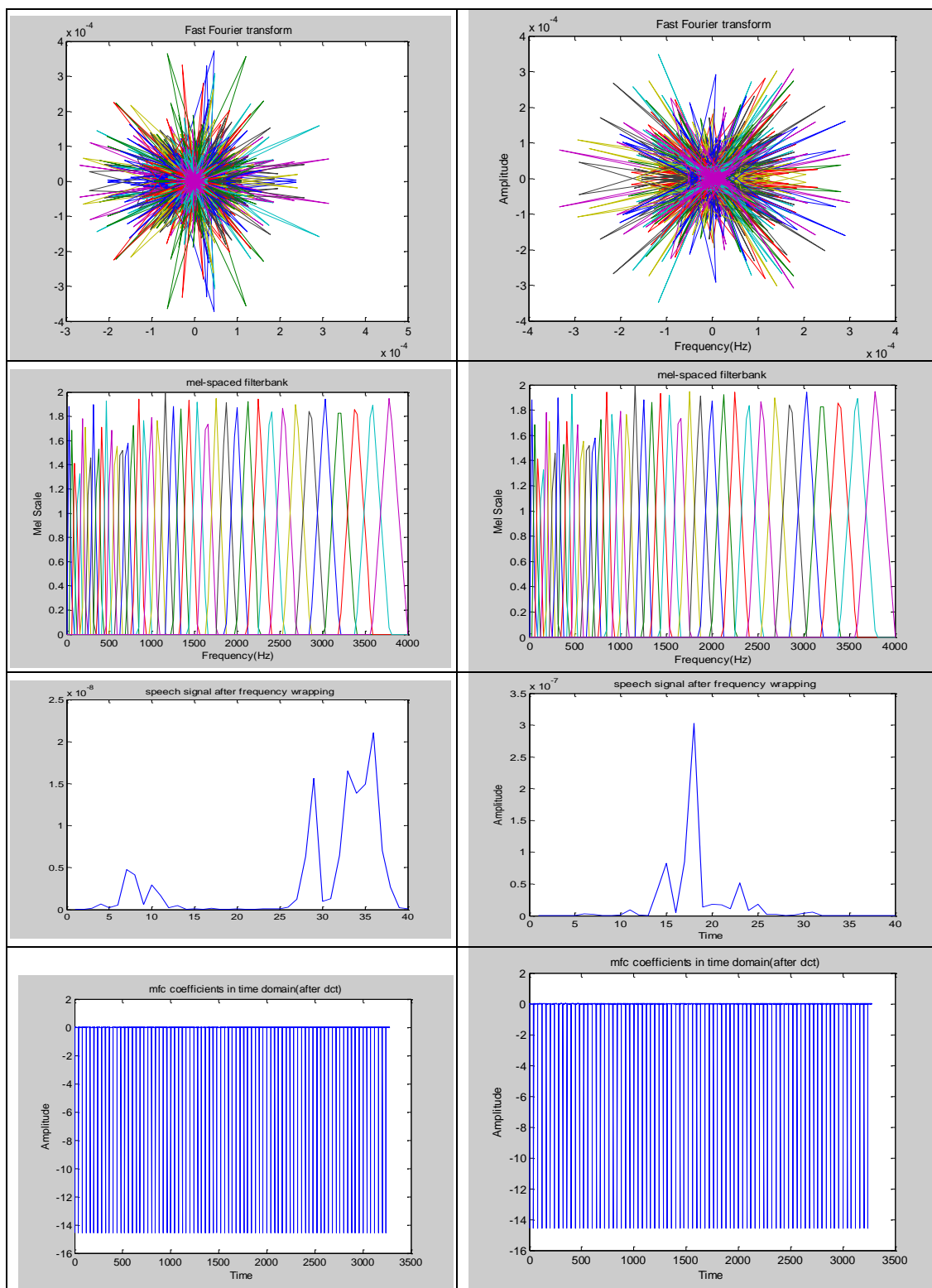




Mel Frequency Cepstrum Coefficients ($C_n = 8$)” Apa kabar”







LAMPIRAN 4

ALGORITMA *MEL FREQUENCY CEPSTRUM COEFFICIENTS*

```

fs=11000;
Nseconds=2;
%fprintf('Say Anything after Enter');
%x=wavrecord(Nseconds*fs,fs,'int16');
%wavwrite(x,fs,16,'audio');
x=wavread('SPagiW1.wav');
t=(0:length(x)-1)/fs;
figure(1);
plot(t,x);
wavplay(x,fs);
title('Original Wave');
xlabel('Time(x)');
ylabel('Amplitude');
num =8;
n =512;
fn =24;
l = length(x);
Ts=0.01;      %Frame step in seconds
FrameStep=fs*Ts; %Frame step in samples
a1=1;
b1=[1, -0.97]; % a and b are high pass filter coeffiends
noFrames=floor(l/FrameStep); %Maximum no of frames in speech sample
lifter=1:num; %Lifter vactor index
lifter=1+floor((num)/2)*(sin(lifter*pi/num)); %raised sine lifter version
if mean (abs(x))> 0.01
    x=x/max(x); % Normalises to compensate for mic vol differences
end

audio=double(x);
audio_normalized=audio/32768;
a=0.95;voice=filter([1,-a],1,audio_normalized);
time=(1:length(voice))/fs;
wavplay(voice,fs);
plot(time,voice);
title(sprintf('sinyal suara setelah pre emphasis',a));
xlabel('time(voice)');
ylabel('Amplitude');

```

ALGORITMA JARINGAN SYARAF TIRUAN *BACKPROPAGATION*

```

Function [v,v0,w,w0] = trainBPElementWise(v,v0,w,w0)
%% ==Step_0==Initialization=====
clc
%%% Enter the Architecture detail
disp('Enter the Architecture detail');
e = 0.05; % error limit
n = 8; %INPUT
p = 3;
m = 1; % OUTPUT
alpha = 0.2;
% -----
%%% Initialization of the Input / Target
%cd('E:\thesisExtra\nnet')
disp('Loading the input vector x and target vectors')
x = [1.4196  0.8408  1.1024 -0.1453 -0.8463 -0.6549 -0.8738 -0.1221;
3.2675 -0.7399  4.5034 -0.6672 -2.2294  0.1377 -2.1410  0.0366; 2.3056
1.7859 1.2685 -1.2152 -3.2675 -1.4435  1.1225  0.0553];
trainRecords = size(x,1);
t = [0; 1; 0];
% -----
%%% Initialization of the weights
if (nargin < 2)
    disp('weights v and w are getting initialised randomly');
end
%%%First hidden Layer
zin = zeros(trainRecords,p);
z = zeros(trainRecords,p);
dinj = zeros(trainRecords,p);
%%%Second hidden Layer
yin = zeros(trainRecords,m);

%% ==Step_1==While_stopping_condition_is_false==do_step_2-9==
while er==0
    errorMax(iteration) = max(max(error));
    disp(sprintf('Epoch : %4g, max err : %d',iteration, errorMax(iteration)));
    %%% ==Step_2==For_Each_Training_pair=====do_Steps_3-8==
    for Tp=1:trainRecords
        ==Step_4=====
        %%% First Layer
        for j=1:p

```

LAMPIRAN 5

ALGORITMA PENGUJIAN IDENTIFIKASI *VOICE*

```

Private Sub Timer1_Timer()
Form2.Show
Unload Me
End Sub

Private Sub Command1_Click()
End
End Sub

Private Sub Command2_Click()
Form4.Show
Else
username.Text = ""
End If
End Sub

Private Sub Command1_Click()
End
End Sub

Dim x(1 To 100000), Wanita, Pria As String
Dim r, g, b As Integer

Private Sub Command1_Click()
    Open "D:\Data_suara1.DAT " For Append As #1

```

LAMPIRAN 6

