

**APLIKASI *SPEECH TO TEXT* BERBAHASA INDONESIA
MENGUNAKAN *MEL-FREQUENCY CEPSTRAL COEFFICIENT*
DAN *HIDDEN MARKOV MODEL***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Jurusan Ilmu Komputer/ Informatika**

**Disusun Oleh:
EKO WIDIYANTO
24010310120049**

**JURUSAN ILMU KOMPUTER / INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2015

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

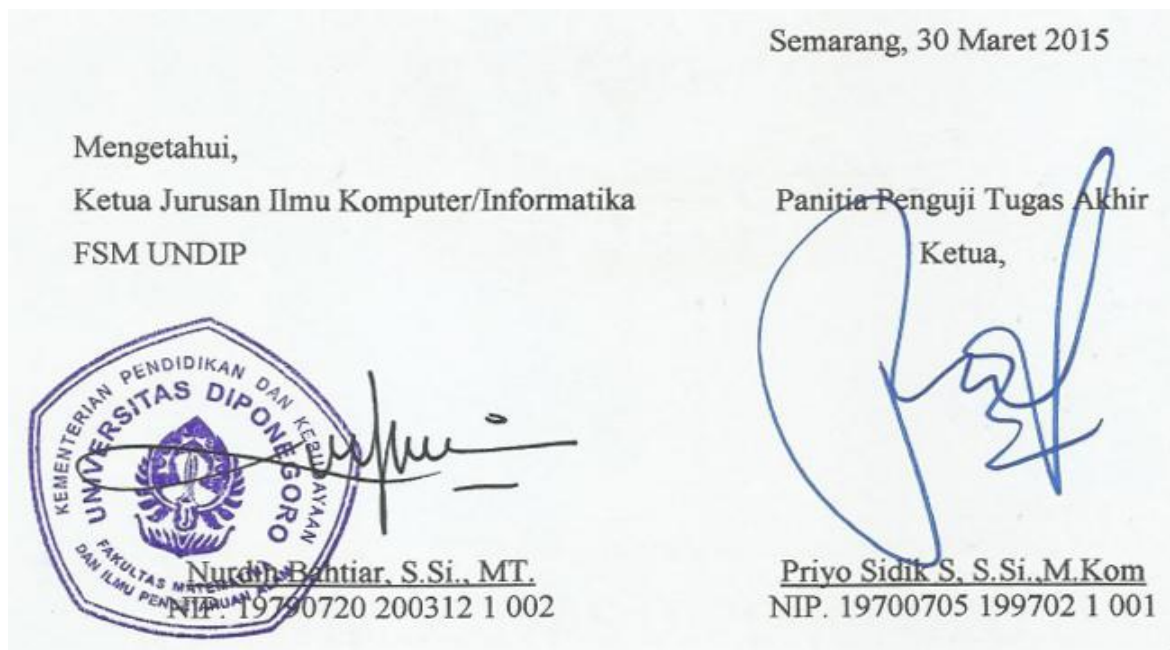
Dengan ini saya menyatakan bahwa dalam tugas akhir/skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka



HALAMAN PENGESAHAN

Judul : Aplikasi *Speech to Text* Berbahasa Indonesia Menggunakan *Mel-Frequency Cepstral Coefficient* Dan *Hidden Markov Model*
Nama : Eko Widiyanto
NIM : 24010310120049


Telah diujikan pada sidang tugas akhir pada tanggal 17 Maret 2015 dan dinyatakan lulus pada tanggal 26 Maret 2015



HALAMAN PENGESAHAN

Judul : Aplikasi *Speech to Text* Berbahasa Indonesia Menggunakan *Mel-Frequency Cepstral Coefficient* Dan *Hidden Markov Model*
Nama : Eko Widiyanto
NIM : 24010310120049

Telah diujikan pada sidang tugas akhir pada tanggal 17 Maret 2015

Semarang, 30 Maret 2015
Pembimbing Utama

Sukmawati Nur-Endah, S.Si, M.Kom
NIP. 19790720 200312 1 002

ABSTRAK

Pengenalan ucapan adalah suatu teknik yang memungkinkan sistem komputer untuk menerima input berupa kata yang diucapkan oleh seseorang tanpa memperdulikan identitas hingga dimengerti oleh komputer. Kata-kata yang ditangkap dan dikenali oleh komputer bisa jadi hasil akhir, untuk sebuah aplikasi seperti *command* dan *control*, penginputan data, dan persiapan dokumen. Aplikasi *speech to text* berbahasa Indonesia dapat menjadi solusi untuk mengenali kata dari sebuah ucapan. Aplikasi ini dikembangkan menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) yang digunakan untuk ekstraksi ciri dari sinyal ucapan dan *Hidden Markov Model* untuk pembentukan model dan menentukan nilai kemiripan kata yang tidak diucapkan dengan kata-kata yang ada dalam sebuah model. Data pengujian menggunakan 10 *speaker* (5 laki-laki dan 5 perempuan) yang masing-masing mengucapkan 15 kata dan untuk setiap katanya diucapkan sebanyak 4 kali. Pengujian dilakukan dengan mengubah-ubah nilai koefisien pada MFCC dan state pada HMM. Berdasarkan hasil pengujian hasil akurasi terbaik pada koefisien MFCC=20 dan state HMM=14 sebesar 90.83%.

Kata Kunci: Pengenalan Ucapan, *Mel-Frequency Cepstral Coefficient*, *Hidden Markov Model*

ABSTRACT

Speech recognition is a technique that allows a computer system to receive input in the form of words that had been spoken by a person disregarding the person identity. The words were captured and recognized by the computer can be the final result of an application such as command and control, data entry and document preparation. Speech to text applications in Indonesian language can be a solution to recognize the words of an utterance. This application was developed using Mel-Frequency Cepstral Coefficient Method (MFCC) for extraction feature from speech signals and Hidden Markov Model (HMM) to determine the similarity from unspoken words with the existance words. Testing data was built from 10 speaker (5 male and 5 female). Each of speakers will speak in 15 words and each words will be repeated 4th times. Testing have been done by manipulating the coefficient on MFCC value and state for HMM. Based on the test result, the best accuracy (90.83%) was acquired from MFCC coefficients=20 and state HMM=14.

Keywords Speech Recognition, Mel-Frequency Cepstral Coefficient, Hidden Markov Model

KATA PENGANTAR

Puji syukur penulis haturkan kehadirat Allah SWT atas rahmat, anugerah, dan kesempatan yang diberikan-Nya penulis dapat menyelesaikan laporan tugas akhir.

Tugas akhir yang berjudul “Aplikasi *Speech To Text* Berbahasa Indonesia Menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) Dan *Hidden Markov Model* (HMM)” ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Jurusan Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan tugas akhir ini penulis banyak mendapat bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan terima kasih kepada:

1. Dr. Muhammad Nur, DEA, selaku Dekan FSM UNDIP
2. Nurdin Bahtiar, S.Si, M.T selaku Ketua Jurusan Ilmu Komputer / Informatika dan dosen pembimbing II
3. Indra Waspada, S.T , M.TI, selaku Koordinator Tugas Akhir Jurusan Ilmu Komputer / Informatika
4. Sukmawati Nur Endah, S.Si, M.Kom, selaku dosen pembimbing
5. Semua pihak yang telah membantu hingga selesainya tugas akhir ini, yang tidak dapat penulis sebutkan satu persatu. Semoga Allah membalas segala kebaikan yang telah diberikan kepada penulis

Penulis menyadari bahwa dalam laporan ini masih banyak terdapat kekurangan baik dari penyampaian materi maupun isi dari materi itu sendiri. Hal ini dikarenakan keterbatasan kemampuan dan pengetahuan dari penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan.

Semoga laporan tugas akhir ini dapat bermanfaat bagi penulis dan juga pembaca pada umumnya.

Semarang, Maret 2015

Penulis

DAFTAR ISI

ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiv
DAFTAR SIMBOL	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan dan Manfaat	3
1.4. Ruang Lingkup	3
1.5. Sistematika Penulisan	4
BAB II DASAR TEORI.....	5
2.1. Sinyal Ucapan	5
2.1.1. Sistem Pembentukan Ucapan	5
2.1.2. Representasi Sinyal Ucapan	6
2.1.3. Karakteristik Sinyal Ucapan.....	8
2.2. Pengenalan Ucapan.....	11
2.2.1. Pengertian Pengenalan Ucapan	11
2.2.2. Skema Utama Pengenalan Ucapan	12
2.2.3. Parameter Pengenalan Ucapan	13
2.3. Ekstraksi Ciri <i>Mel-Frequency Cepstral Coefficient</i> (MFCC).....	14
2.3.1. <i>DC-Removal</i>	15
2.3.2. <i>Filter Pre-Emphasize Filtering</i>	15
2.3.3. <i>Frame Blocking</i>	16
2.3.4. <i>Proses Windowing</i>	17
2.3.5. <i>Fast Fourier Transform</i> (FFT)	18
2.3.6. <i>Mel-Frequency Wrapping</i>	19
2.3.7. <i>Discrete Cosine Transform</i> (DCT).....	20

2.3.8.	<i>Cepstral Liftering</i>	20
2.4.	<i>Hidden Markov Model (HMM)</i>	21
2.4.1.	Tipe-Tipe <i>Hidden Markov Model</i>	22
2.4.2.	Arsitektur <i>Hidden Markov Model</i>	22
2.4.3.	Implementasi HMM pada Pengenalan Ucapan	26
2.5.	Metode Pengembangan Perangkat Lunak <i>Unified Process</i>	29
2.5.1.	Struktur <i>Unified Process</i>	30
2.5.2.	Fase <i>Inception</i>	31
2.5.3.	Fase <i>Elaboration</i>	31
2.5.4.	Fase <i>Construction</i>	32
2.5.5.	Fase <i>Transition</i>	32
2.6.	<i>Unified Modeling Language (UML)</i>	32
2.6.1.	<i>Things</i>	32
2.6.2.	<i>Relationship</i>	33
2.6.3.	Diagram	36
BAB III FASE <i>INCEPTION</i> DAN FASE <i>ELABORATION</i>		40
3.1.	Iteration Plan	40
3.2.	Fase <i>Inception</i>	40
3.2.1.	Deskripsi Aplikasi	41
3.2.2.	<i>Business Rules</i>	42
3.2.3.	Kebutuhan Non-Fungsional	42
3.2.4.	Model <i>Use Case</i>	42
3.3.	Fase <i>Elaboration</i>	45
3.3.1.	<i>Elaboration</i> Iterasi Pertama	45
3.3.2.	<i>Elaboration</i> Iterasi Kedua	51
3.3.3.	Menyusun <i>Prototipe</i> Antarmuka	90
3.3.4.	Daftar Resiko	93
3.3.5.	Menyusun Rencana Pengujian	94
BAB IV FASE <i>CONSTRUCTION</i> DAN FASE <i>TRANSITION</i>		96
4.1.	Fase <i>Construction</i>	96
4.1.1.	Implementasi Sistem	96
4.1.2.	Implementasi Objek	96
4.1.3.	Implementasi Antarmuka	101

4.2. Fase Transition.....	104
4.2.1. Lingkungan Pengujian.....	105
4.2.2. Pengujian Fungsi	105
4.2.3. Pengujian Parameter	108
4.2.4. Analisa Pengujian	110
BAB V PENUTUP	112
5.1. Kesimpulan	112
5.2. Saran	112
DAFTAR PUSTAKA.....	113
LAMPIRAN	115

DAFTAR GAMBAR

Gambar 2.1	Foto sinar X penampang alat-alat ucap manusia	5
Gambar 2.2	Model sistem produksi ucapan manusia.	6
Gambar 2.3	Contoh sinyal ucapan “ <i>It’s Time</i> ”.	7
Gambar 2.4	Bentuk sinyal ucapan vokal Bahasa Indonesia pada suara pria	9
Gambar 2.5	Bentuk sinyal ucapan vokal Bahasa Indonesia pada suara wanita	9
Gambar 2.6	Blok diagram untuk MFCC	15
Gambar 2.7	Contoh dari <i>Pre-Emphasize</i> pada sebuah <i>frame</i>	16
Gambar 2.8	Diagram <i>block</i> proses penentuan <i>frame</i>	16
Gambar 2.9	<i>Spectral Analysis</i> jangka pendek	17
Gambar 2.10	Sinyal Ucapan Domain Waktu menjadi Domain Frekuensi	19
Gambar 2.11	Perbandingan Spektrum dengan dan tanpa <i>Cepstral Liftering</i>	21
Gambar 2.12	Contoh <i>Markov Chain</i>	22
Gambar 2.13	Arsitektur HMM	23
Gambar 2.14	Evolusi temporal dari <i>Hidden Markov Model</i>	24
Gambar 2.15	Contoh Pelatihan	28
Gambar 2.16	Pengenalan menggunakan HMM	28
Gambar 2.17	Diagram Blok Pengenalan Ucapan	29
Gambar 2.18	Alur kerja <i>Unified Process</i>	30
Gambar 2.19	Siklus hidup <i>Unified Process</i>	30
Gambar 2.20	Hubungan fase dengan <i>Workflow</i> dalam <i>Unified Process</i>	31
Gambar 2.21	<i>Dependency</i> antara <i>Class</i> ‘ <i>Filmclip</i> ’ dan ‘ <i>Channel</i> ’	34
Gambar 2.22	Contoh penggunaan <i>Name</i> Asosiasi antara <i>Class</i> ‘ <i>Person</i> ’ dan ‘ <i>Company</i> ’	34
Gambar 2.23	Contoh penggunaan <i>Role</i> dari Asosiasi antara <i>Class</i> ‘ <i>Person</i> ’ dan ‘ <i>Company</i> ’	34
Gambar 2.24	Contoh penggunaan <i>Multiplicity</i> dari Asosiasi antara <i>Class</i> ‘ <i>Person</i> ’ dan ‘ <i>Company</i> ’	35
Gambar 2.25	Contoh penggunaan <i>Aggregation</i> antara <i>Class</i> ‘ <i>Company</i> ’ dan ‘ <i>Department</i> ’	35
Gambar 2.26	<i>Generalization</i> : <i>Class</i> ‘ <i>Rectangle</i> ’, ‘ <i>Circle</i> ’, ‘ <i>Polygon</i> ’ Spesialisasi dari <i>Class</i> ‘ <i>Shape</i> ’	35

Gambar 2.27	Contoh <i>Class Diagram</i> Pemesanan Barang	36
Gambar 2.28	Simbol <i>Use Case</i>	37
Gambar 2.29	Simbol <i>Actor</i>	37
Gambar 2.30	Contoh <i>Sequence Diagram</i> untuk proses Pemesanan Barang	38
Gambar 2.31	Contoh <i>Activity Diagram</i> untuk proses Pemesanan Barang	39
Gambar 3.1	Alur proses Pengenalan dan Pelatihan Aplikasi <i>Speech to Text</i>	41
Gambar 3.2	<i>Use Case Diagram</i> Sistem	43
Gambar 3.3	<i>Domain Model</i> Aplikasi <i>Speech to Text</i> Berbahasa Indonesia	46
Gambar 3.4	<i>Class diagram</i> Aplikasi <i>Speech to Text</i> Berbahasa Indonesia	47
Gambar 3.5	<i>Sequence Diagram</i> Input Ucapan	48
Gambar 3.6	<i>Sequence Diagram</i> Pelatihan	49
Gambar 3.7	<i>Sequence Diagram</i> Pengenalan Suara	49
Gambar 3.8	<i>Activity Diagram</i> Aplikasi <i>Speech to Text</i> Berbahasa Indonesia	50
Gambar 3.9	<i>Flowchart</i> Proses Input Ucapan Untuk Data Pelatihan	52
Gambar 3.10	<i>Flowchart</i> Proses Input Ucapan Untuk Pengenalan Kata	52
Gambar 3.11	<i>Flowchart</i> Proses Pelatihan	53
Gambar 3.12	<i>Flowchart</i> MFCC	54
Gambar 3.13	<i>Flowchart</i> DC-Removal	55
Gambar 3.14	<i>Flowchart</i> Pre-Emphasize	56
Gambar 3.15	<i>Flowchart</i> Frame Blocking	57
Gambar 3.16	<i>Flowchart</i> Windowing	58
Gambar 3.17	<i>Flowchart</i> FFT	59
Gambar 3.18	<i>Flowchart</i> Filterbank	60
Gambar 3.19	<i>Flowchart</i> DCT	61
Gambar 3.20	<i>Flowchart</i> Cepstral Liftering	62
Gambar 3.21	<i>Flowchart</i> Pelatihan HMM	63
Gambar 3.22	<i>Flowchart</i> proses Pelatihan Parameter HMM	65
Gambar 3.23	<i>Flowchart</i> Inisialisasi Parameter B	66
Gambar 3.24	<i>Flowchart</i> proses Forward	67
Gambar 3.25	<i>Flowchart</i> proses Backward	68
Gambar 3.26	<i>Flowchart</i> proses Normalize	68
Gambar 3.27	<i>Flowchart</i> proses Symmetrize	69
Gambar 3.28	<i>Flowchart</i> proses Stochastic	69

Gambar 3.29	<i>Flowchart</i> Proses Pengenalan Ucapan.....	70
Gambar 3.30	Antarmuka Halaman Beranda	90
Gambar 3.31	Antarmuka Halaman Pelatihan	91
Gambar 3.32	Antarmuka Halaman Pengujian	91
Gambar 3.33	Antarmuka Detail Proses Pengenalan 1	92
Gambar 3.34	Antarmuka Detail Proses Pengenalan 2	92
Gambar 3.35	Antarmuka Detail Proses Pengenalan 3	92
Gambar 4.1	Tampilan Halaman Beranda	102
Gambar 4.2	Tampilan Halaman Pelatihan	102
Gambar 4.3	Tampilan Halaman Pengujian	103
Gambar 4.4	Tampilan Detail Proses Pengenalan 1	103
Gambar 4.5	Tampilan Detail Proses Pengenalan 2	104
Gambar 4.6	Tampilan Detail Proses Pengenalan 3	104
Gambar 4.7	Grafik perbandingan tingkat akurasi masing-masing Koefisien MFCC terhadap State yang digunakan.	110
Gambar 4.8	(A) Sinyal Ucapan Laki-laki: “Adik”, (B) Sinyal Ucapan Perempuan “Adik”	111

DAFTAR TABEL

Tabel 2.1	Fonem-fonem Bahasa Indonesia	8
Tabel 2.2	Tabel Parameter Pengenalan Suara	13
Tabel 2.3	Jenis <i>Relationship</i> pada <i>Use Case Diagram</i>	37
Tabel 2.4	Komponen <i>Activity Diagram</i>	38
Tabel 3.1	Daftar Aktor Sistem.....	43
Tabel 3.2	Daftar <i>Use Case</i> Aplikasi	43
Tabel 3.3	Detail <i>Use Case Input</i> Ucapan.....	44
Tabel 3.4	Detail <i>Use Case</i> Pelatihan	44
Tabel 3.5	Detail <i>Use Case</i> Pengenalan Ucapan.....	45
Tabel 3.6	Struktur Folder data_pelatihan	51
Tabel 3.7	Struktur database_pelatihan.....	51
Tabel 3.8	Hasil ekstraksi ciri data latih untuk sinyal ucapan “ayah”	76
Tabel 3.9	Hasil ekstraksi ciri data latih untuk sinyal ucapan “ibu”	77
Tabel 3.10	Hasil inialisasi awal Parameter Π	77
Tabel 3.11	Hasil inialisasi awal Parameter A	77
Tabel 3.12	Hasil inialisasi awal Parameter μ	77
Tabel 3.13	Hasil inialisasi nilai Parameter Σ untuk setiap State	78
Tabel 3.14	Nilai Parameter B	78
Tabel 3.15	Nilai α	79
Tabel 3.16	Nilai β	80
Tabel 3.17	Nilai γ	81
Tabel 3.18	Nilai $\gamma_{\text{observasi}}$ pada State 1	82
Tabel 3.19	Nilai $\gamma_{\text{observasi}}$ pada State 2	82
Tabel 3.20	Nilai $\gamma_{\text{observasi}}$ pada State 3	82
Tabel 3.21	Nilai μ baru	83
Tabel 3.22	Nilai Parameter Σ baru pada State 1	85
Tabel 3.23	Nilai Parameter Σ baru pada State 2	85
Tabel 3.24	Nilai Parameter Σ baru pada State 3	85
Tabel 3.25	Nilai Parameter A setelah Pelatihan	87
Tabel 3.26	Nilai Parameter Π setelah Pelatihan	87
Tabel 3.27	Nilai Parameter μ setelah Pelatihan	87

Tabel 3.28	Nilai Parameter Σ pada state 1 setelah Pelatihan	88
Tabel 3.29	Nilai Parameter Σ pada state 2 setelah Pelatihan	88
Tabel 3.30	Nilai Parameter Σ pada state 3 setelah Pelatihan	88
Tabel 3.31	Contoh sinyal ucapan yang diuji.....	88
Tabel 3.32	Nilai Parameter B berdasarkan Parameter μ dan Σ	88
Tabel 3.33	Nilai α sinyal ucapan terhadap Model	90
Tabel 3.34	Nilai <i>likelihood</i> sinyal ucapan terhadap Model	90
Tabel 3.35	Rencana Pengujian Fungsi Aplikasi <i>Speech to Text</i> Bahasa Indonesia.....	95
Tabel 3.36	Rencana Pengujian Parameter	95
Tabel 4.1	Implementasi <i>Class</i> Aplikasi <i>Speech to Text</i>	97
Tabel 4.2	Implementasi Atribut <i>Class</i> Pelatihan	97
Tabel 4.3	Implementasi Operasi <i>Class</i> Pelatihan	97
Tabel 4.4	Implementasi Atribut <i>Class</i> Pengenalan.....	98
Tabel 4.5	Implementasi Operasi <i>Class</i> Pengenalan.....	98
Tabel 4.6	Implementasi Atribut <i>Class</i> Masukan	98
Tabel 4.7	Implementasi Operasi <i>Class</i> Masukan.....	98
Tabel 4.8	Implementasi Atribut <i>Class</i> EkstraksiCiri	99
Tabel 4.9	Implementasi Operasi <i>Class</i> EkstraksiCiri	99
Tabel 4.10	Implementasi Atribut <i>Class</i> PengenalanPola	100
Tabel 4.11	Implementasi Operasi <i>Class</i> PengenalanPola.....	100
Tabel 4.12	Hasil dan Evaluasi Pengujian Fungsi Aplikasi <i>Speech to Text</i> Berbahasa Indonesia	106
Tabel 4.13	Hasil Pengujian Parameter pada Koefisien MFCC = 8	108
Tabel 4.14	Hasil Pengujian Parameter pada Koefisien MFCC = 12	109
Tabel 4.15	Hasil Pengujian Parameter pada Koefisien MFCC = 20	109

DAFTAR SIMBOL

DC_u	= Hasil proses <i>DC removal</i> pada sinyal ke u
x_u	= Sinyal ucapan ke u
\bar{x}	= Nilai rata-rata sinyal ucapan
U	= Jumlah sinyal ucapan
u	= 1, 2, 3... U
PR_u	= Sinyal hasil proses <i>pre-emphasize filter</i> pada sinyal ke u
kf	= Konstanta filter <i>pre-emphasis</i> , biasanya bernilai $0.9 \leq kf \leq 1.0$
W_h	= Nilai hasil proses <i>windowing</i> pada data ke h pada setiap frame
PR_h	= Sinyal hasil proses <i>pre-emphasize filter</i> pada sinyal ke h pada setiap frame
HM_h	= Fungsi <i>hamming window</i> pada data ke h pada setiap frame
H	= Jumlah data setiap frame
h	= 1, 2, 3... H
π	= 3.14
F_h	= Nilai dalam domain frekuensi pada data ke h
k	= <i>Variable frekuensi discrete</i> , dimana ($k = H/2, k \in H$)
C_h	= Hasil dari $\sum_{k=0}^{H-1} \left(W(k) * \cos \left(\frac{2\pi(h-1)k}{H} \right) \right)$ pada persamaan FFT
S_h	= Hasil dari $\sum_{k=0}^{H-1} \left(W(k) * \sin \left(\frac{2\pi(h-1)k}{H} \right) \right)$ pada persamaan FFT
Mf_g	= Keluaran dari proses <i>filterbank</i> pada data ke g
Hf_g	= Nilai filter segitiga ke g
fil	= Jumlah filter
g	= 1, 2, 3..., fil
F_m	= Nilai dalam domain frekuensi pada data ke m
$melf$	= Hasil dari fungsi <i>mel scale</i>
fs	= <i>Frequency Sampling</i>
HD_l	= Hasil dari proses DCT pada Koefisien l
$Koef$	= Jumlah koefisien yang diharapkan
l	= 1, 2, 3..., $Koef$
$Ceps_q$	= Hasil dari Fungsi <i>Cepstral liftering</i> pada data ke q
HD_q	= Hasil dari proses DCT pada Koefisien q

Q	= Jumlah <i>Cepstral Koefisien</i>
q	= 1, 2, 3..., Q
N	= Jumlah State
M	= Jumlah Simbol Pengamatan
Π	= <i>Initial State</i>
A	= Probabilitas Transisi
B	= Probabilitas Observasi
α	= Variable Algoritma <i>Forward</i>
β	= Variable Algoritma <i>Backward</i>
$\xi_t(i, j)$	= Peluang Transisi yang diharapkan dari <i>State</i> ke- i waktu ke- t menuju <i>State</i> ke- j Waktu ke- t
$\gamma_t(i)$	= Peluang Transisi yang diharapkan dari <i>State</i> ke- i pada waktu ke- t
T	= Total Waktu/ <i>Frame</i>
i	= Indeks i
j	= Indeks j
O	= Barisan Observasi
λ	= Model HMM
P	= Peluang
t	= Indeks Waktu/ <i>Frame</i>

DAFTAR LAMPIRAN

Lampiran 1. Implementasi Class	115
--------------------------------------	-----

BAB I

PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika dalam pembuatan tugas akhir mengenai pengembangan Aplikasi *Speech to Text* Berbahasa Indonesia Menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) Dan *Hidden Markov Model* (HMM).

1.1. Latar Belakang

Komunikasi bahasa antar manusia dengan manusia merupakan salah satu cara yang paling efektif untuk menyampaikan maksud dan tujuan seseorang dalam menyampaikan informasi untuk memudahkan seseorang dalam menyelesaikan pekerjaan. Informasi dapat disampaikan dalam bentuk teks, citra, suara, dan lain-lain. Pemakaian teknologi komputer untuk menjalin komunikasi manusia dengan komputer sudah menjadi suatu kebutuhan, jika komputer mengerti ucapan yang dimaksudkan manusia bisa menjadi suatu kemudahan dalam pengoperasian komputer, seperti *voice command*, akses kontrol sistem berbasis suara, dan identifikasi suara untuk keamanan sistem.

Perkembangan *speech recognition (speech to text)* berjalan cukup pesat pada saat ini dilihat dari banyaknya jurnal yang membahas mengenai *speech to text*. Suara manusia mempunyai karakteristik yang sangat kompleks, satu kata yang diucapkan oleh orang yang berbeda bisa menghasilkan karakteristik suara yang berbeda, namun suatu sistem diharuskan dapat mengenali sebagai suatu kata yang sama. Selain itu faktor yang mempengaruhi suara adalah kesehatan, psikologi, umur, dan jenis kelamin seseorang.

Speech to text memungkinkan suatu perangkat untuk mengenali dan memahami kata-kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Kata-kata yang diucapkan diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi sekumpulan angka yang kemudian disesuaikan dengan kode-kode tertentu untuk mengidentifikasi kata-kata tersebut. Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan.

Di Indonesia penelitian mengenai *speech to text* sudah mulai banyak dilakukan dilihat dari bermunculannya jurnal-jurnal mengenai *speech to text* bahasa Indonesia. Dari penelitian sebelumnya ada yang menggunakan metode ekstraksi ciri *Mel Frequency Cepstral Coefficients* (MFCC) dan metode pengenalan pola *Hidden Markov Model* (HMM) (Fawziah, et al., 2013), namun masih terbatas untuk *speaker* laki-laki saja dan hanya beberapa kata saja yang menjadi data pengujian. Dalam penelitian kali ini dicoba untuk menggunakan metode yang sama dengan menggunakan *speaker* laki-laki dan perempuan dengan data pengujian yang lebih banyak.

Mel Frequency Cepstrum Coefficients (MFCC) *feature extraction* mengkonversikan sinyal suara ke dalam beberapa vektor data berguna bagi proses pengenalan suara. Terdapat 7 tahapan dalam MFCC yaitu *Pre Emphasize, Frame Blocking, Windowing, Fast Fourier Transform, Mel Frequency Wrapping, Discrete Cosine Transform, dan Cepstral Liftering*. Metode ini memiliki beberapa kelebihan diantaranya adalah mampu menangkap informasi penting dalam sinyal suara, menghasilkan ciri sinyal ucapan seminimal mungkin tanpa menghilangkan informasi-informasi yang ada, dan mereplikasikan organ pendengaran manusia dalam melakukan persepsi terhadap sinyal suara (Andriana, 2011).

Hidden Markov Model (HMM) merupakan suatu metode pendekatan yang dapat mengelompokkan sifat-sifat *spectral* dari tiap bagian suara dengan beberapa pola. HMM memiliki 5 proses dasar dalam melakukan pengenalan suara, yaitu: *Feature Analysis, Unit Matching System, Lexical Decoding, Systactic Analysis, and Semantic Analysis*. Proses-proses itulah yang menyebabkan HMM mempunyai tingkat akurasi yang lebih tinggi dibanding metode lain, terbukti dengan banyaknya penelitian mengenai *speech recognition* yang menggunakan metode ini untuk pendekatan dalam mengenali pola suara (Rabiner & Juang, 1993).

Oleh karena itu dalam penelitian ini dilakukan pemodelan aplikasi *speech to text* berbahasa Indonesia menggunakan metode ekstraksi ciri *Mel-Frequency Cepstral Coefficient* (MFCC) dengan menggunakan *Hidden Markov Model* (HMM) untuk mengenali pola ucapannya.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, dapat dirumuskan permasalahan yang dihadapi yaitu bagaimana membuat suatu aplikasi *speech to text* yang dapat mengenali ucapan kata berbahasa Indonesia dengan menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) dan *Hidden Markov Model* (HMM).

1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai dalam penelitian tugas akhir ini adalah membuat aplikasi *speech to text* berbahasa Indonesia menggunakan metode *Mel-Frequency Cepstral Coefficient* (MFCC) dan *Hidden Markov Model* (HMM).

Manfaat dari penelitian tugas akhir ini adalah

1. Sebagai dasar dalam penelitian pengembangan aplikasi *speech to text*
2. Membantu pengguna yang membutuhkan konversi suara ke teks seperti orang yang tidak bisa menulis.

1.4. Ruang Lingkup

Adapun ruang lingkup dari aplikasi *Speech to Text* berbahasa Indonesia menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) dan *Hidden Markov Model* (HMM) adalah:

1. Suara yang di-*input*-kan dengan rekaman langsung atau file rekaman berekstensi *.wav dan *output* berupa teks yang ditampilkan pada sistem
2. Objek perekaman digunakan ucapan dengan dialek yang umum (jawa)
3. Objek perekaman berusia antara 19 tahun - 23 tahun
4. Objek perekaman untuk data pelatihan menggunakan 10 orang *speaker* (5 laki-laki dan 5 perempuan)
5. Kata yang digunakan dalam penelitian ini terdiri dari 15 kata yaitu “adik”, “ayah”, “botol”, “cerdas”, “dunia”, “ikan”, “jual”, “keluarga”, “lenyap”, “mimpi”, “minum”, “om”, “pasar”, “pergi”, dan “toko”
6. Perekaman inputan suara menggunakan *headset* Logitech Stereo H150 dengan jarak ± 3 cm dari sekitaran mulut seseorang
7. Inputan suara dilakukan pada *microphone* dengan kondisi normal
8. Sistem ini diimplementasikan berbasis desktop, dengan bahasa pemrograman MatLab 2012b

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi dalam beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika dalam pembuatan tugas akhir mengenai pengembangan Aplikasi *Speech to Text* Berbahasa Indonesia Menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) Dan *Hidden Markov Model* (HMM).

BAB II DASAR TEORI

Bab ini menyajikan dasar teori yang berhubungan dengan topik tugas akhir. Dasar teori yang digunakan dalam penyusunan tugas akhir ini meliputi Sinyal Ucapan, Pengenalan Ucapan, *Mel-Frequency Cepstral Coefficient* (MFCC), *Hidden Markov Model* (HMM), metode *Unified Process*, dan *Unified Modeling Language* (UML).

BAB III FASE *INCEPTION* DAN FASE *ELABORATION*

Bab ini menyajikan tahapan proses pembangunan perangkat lunak menggunakan model pengembangan *Unified Process*. Pada Bab ini disajikan dua fase awal yaitu *Inception* dan *Elaboration*.

BAB IV FASE *CONSTRUCTION* DAN FASE *TRANSITION*

Bab ini membahas mengenai tahapan akhir dari pembangunan perangkat lunak untuk model pengembangan *Unified Process*. Pada Bab ini disajikan fase *Construction* yaitu fase untuk melakukan implementasi dan fase *Transition* untuk melakukan pengujian sistem.

BAB V PENUTUP

Penutup berisi kesimpulan dari pengerjaan penelitian Tugas Akhir ini dan saran-saran penulis untuk pengembangan lebih lanjut dari penelitian serupa.