



DISERTASI - TE143597

**TALKING HEAD SYSTEM IN INDONESIAN
LANGUAGE WITH AFFECTIVE FACIAL
EXPRESSIONS SYNTHESIS**

**ENDANG SETYATI
NRP. 2211 301 005**

PROMOTOR:
Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D
Dr. Surya Sumpeno, S.T., M.Sc.

**PROGRAM DOKTOR
JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI INDUSTRI
INSTITUT TEKNOLOGI SEPULUH NOVEMBER
SURABAYA
2017**

خیر الناس لمن فتح لهم

Khairunnaṣ anfa'uhum linnas

Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia (orang) lain

Kupersembahkan

Kepada Bapak dan Ibu almarhum yang Ananda cintai dan rindukan.

Alhamdulillah gelar Doktor sudah Ananda raih berkat dukungan dan doa tulus dari Bapak dan Ibu yang sangat mengayangi Ananda.

Semoga pahala Bapak dan Ibu terus mengalir dari ilmu yang diberikan kepada Ananda.

Sekiranya keluarga Ananda menjadi keluarga yang sakinah mawadah warohmah,

Mohonkanlah kepada Allah SWT untuk bisa mengikuti suri tauladan yang Engkau ajarkan pada Ananda.

Aamiin Yaa Rabbal Allaamiin...

LEMBAR PENGESAHAN DISERTASI

Talking Head System in Indonesian Language with Affective Facial Expressions Synthesis

Disertasi ini disusun untuk memenuhi salah satu syarat memperoleh gelar
Doktor (Dr.)
di Institut Teknologi Sepuluh Noverember Surabaya

Oleh:
Endang Setyati
NRP: 2211 301 005

Tanggal Ujian: 3 Januari 2017
Periode Wisuda: Maret 2017

Disetujui oleh:

Prof. Ir. Mauridhi Hery P., M.Eng., Ph.D.
NIP: 19580916 198601 1 001

(Promotor)

Dr. Surya Sumpeno, ST., M.Sc.
NIP: 19690613 199702 1 003

(Co-Promotor)

Mochamad Hariadi, S.T., M.Sc., Ph.D.
NIP: 19691209 199703 1 002

(Penguji I)

I Ketut Edy Purnama, S.T., M.T., Ph.D.
NIP: 19690730 199512 1 001

(Penguji II)

Achmad Syarief, M.S.D., Ph.D.
NIP: 19690729 199702 1 001

(Penguji III)



Direktur Program Pascasarjana,

Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.
NIP. 19601202 19870110 01

[halaman ini sengaja dikosongkan]

PERNYATAAN KEASLIAN DISERTASI

Yang bertanda tangan di bawah ini:

Nama : Endang Setyati

Program Studi : Doktor Teknik Elektro

NRP : 2211 301 005

Dengan ini menyatakan bahwa isi sebagian maupun keseluruhan disertasi dengan judul:

TALKING HEAD SYSTEM IN INDONESIAN LANGUAGE WITH AFFECTIVE FACIAL EXPRESSIONS SYNTHESIS

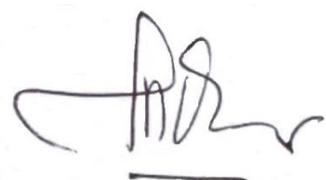
adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diizinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka.

Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, 10 Januari 2017

Yang Membuat Pernyataan,



Endang Setyati
NRP. 2211 301 005

[halaman ini sengaja dikosongkan]

TALKING HEAD SYSTEM IN INDONESIAN LANGUAGE WITH AFFECTIVE FACIAL EXPRESSION SYNTHESIS

Nama : Endang Setyati
NRP : 2211301005
Promotor : Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D.
Co-Promotor : Dr. Surya Sumpeno, S.T., M.Sc.

ABSTRAK

Dalam membuat persepsi dan tampilan dari sebuah antarmuka manusia-komputer alami (*Natural Human-Computer Interface*) dibutuhkan integrasi dari audio yang realistik dan informasi visual. Penelitian ini mengusulkan *Talking Head System* dalam bahasa Indonesia (*Talking Head System in Indonesian Language*) dengan sintesis dari ekspresi wajah berafeksi, berdasarkan 12 model viseme Indonesia. Model Viseme Indonesia ini didapatkan dari hasil pemetaan 49 fonem (termasuk *silent*) ke viseme Indonesia.

Sistem ini mengkonversikan input teks dari kalimat bahasa Indonesia yang berafeksi ke ucapan (*speech*) dengan sinkronisasi animasi dari bentuk mulut, gerakan bibir dan ekspresi emosional. Talking Head juga didasarkan pada sebuah model kepala manusia 3D generik. Untuk mengekspresikan visualisasi 3D wajah berafeksi dan pembangkitan urutan viseme (*viseme chain generation*) yang alami digunakan *Inesh² Corpora*, *Trigram Hidden Markov Model* dan Algoritma *Viterbi*.

Sedangkan untuk mendapatkan 3D model ekspresi wajah berafeksi didasarkan pada pengklasifikasian parameter afeksi/emosi dengan menggunakan *facial motion capture*. Akhir dari sistem ini, diharapkan dapat menghasilkan visualisasi 3D animasi dengan ucapan kalimat bahasa Indonesia berafeksi.

Kata kunci: Talking Head System in Indonesian Language, Viseme Chain Generation, Inesh² Corpora, Trigram Hidden Markov Model, Algoritma Viterbi.

TALKING HEAD SYSTEM IN INDONESIAN LANGUAGE WITH AFFECTIVE FACIAL EXPRESSION SYNTHESIS

Name : Endang Setyati
Student ID : 2211301005
Promotor : Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D.
Co-Promotor : Dr. Surya Sumpeno, S.T., M.Sc.

ABSTRACT

Natural Human-Computer Interface requires integration of realistic audio and visual information for perception and display. In this research, Talking Head System in Indonesian Language is proposed to affective facial expressions synthesis, based on 12 models viseme Indonesia. Model Viseme Indonesia is obtained from the mapping of 49 phonemes (including silent) to viseme Indonesia. The system converts text from affective Indonesian sentences to speech with synchronized animation of mouth shape, lip movements and emotion expressions. The Talking Head is based on generic 3D human head model. To express emotion and viseme chain generation using Inesh² Corpora, Trigram Hidden Markov Model and Viterbi Algorithm.

To get a 3D model of facial expressions based on classification parameters berafeksi affections / emotions using facial motion capture. The final system will produce animated 3D visualization of speech of sentences Indonesian with affection.

Keywords: Talking Head System in Indonesian Language, Viseme Chain Generation, Inesh² Corpora, Trigram Hidden Markov Model, the Viterbi Algorithm.

KATA PENGANTAR

Alhamdulillahirabbil'alaamiin kami panjatkan puji syukur kehadirat Allah SWT, atas rahmat, nikmat dan hidayah-Nya serta salawat dan salam kami kirimkan kepada Rasulullah SAW, sehingga kami dapat menyelesaikan disertasi program doktor ini. Sepenuhnya kami menyadari bahwa selama penulisan disertasi ini banyak pihak yang telah memberikan bantuannya, sehingga pada kesempatan ini kami ingin menyampaikan terima kasih yang tulus kepada:

1. Bapak Prof. Ir. Mauridhi Hery Purnomo, M.Eng, Ph.D. selaku Dosen wali dan Promotor yang telah memberikan bimbingan, saran, dorongan semangat dengan tulus dan penuh kesabaran, juga sering mengingatkan untuk segera menyelesaikan program doktor.
2. Bapak Dr. Surya Sumpeno, S.T., M.Sc. selaku Co-Promotor yang telah memberikan bimbingan, saran, koreksi tulisan, serta dorongan semangat dengan tulus dan penuh kesabaran.
3. Bapak Mochamad Hariadi, S.T., M.Sc., Ph.D., Bapak I Ketut Edy Purnama, S.T., M.T., Ph.D. dan Bapak Achmad Syarieff, M.S.D., Ph.D., sebagai tim penguji yang telah memberikan masukan, kritik dan saran dalam rangka penyempurnaan buku disertasi.
4. Prof. Kunio Kondo, Prof. Masanori Kakimoto, dan Prof. Koji Mikami, yang telah memberikan ide dan masukan untuk pembuatan jurnal internasional saat mengikuti Program Sandwich Luar Negeri dari Ditjen Dikti, di Tokyo University of Technology, Tokyo, Japan selama 4 bulan.
5. Bapak Dr. Ir. Yoyon K. Suprapto, M.Sc. dosen wali penulis pada tahun pertama kuliah, yang telah memberi semangat dan kesempatan diskusi bersama rekan mahasiswa S3 angkatan 2011, bu Diah, pak Aris, pak Cahyo, pak Solekhan, pak Susy, pak Firman, pak Basuki, pak Lie, pak Nyoman.
6. Bapak Prof. Dr. Ir. Imam Robandi, M.T., Ibu Prof. Ir. Handayani Tjandrasa, M.Sc., Ph.D., Bapak Prof. Ir. Joko Lianto Buliali, M.Sc., Ph.D., dan Prof. Dr.

Ir. Kuswara Setiawan, M.T. yang telah memberi rekomendasi penulis untuk studi lanjut Program Doktor di jurusan Teknik Elektro ITS.

7. Prof. Dr. Ir. Adi Supriyanto, M.T., Dr. Ir. Djoko Purwanto, M.Eng., Dr. Rony Seto Wibowo, S.T., M.T., pejabat program PPs Teknik Elektro ITS, yang membantu kelancaran proses studi lanjut; mbak Ranny, mas Hartono, pak Sapari, karyawan PPs Teknik Elektro; mas Anang, mas Iksan, pak Mamat, dan pak Pur karyawan program PPs ITS yang telah membantu kelancaran proses administrasi selama studi di ITS.
8. Beasiswa BPPDN DIKTI (2011-2015), Beasiswa Program Sandwich Luar Negeri Ditjen DIKTI (2013), Program Hibah Disertasi Doktor (2014) dan Program Hibah Bersaing dari DP2M DIKTI (2011-2016) yang telah memberi bantuan biaya studi selama kuliah di program Doktor Teknik Elektro ITS.
9. Bapak Ir. Arya Tandy Hermawan, M.T., selaku Ketua Sekolah Tinggi Teknik Surabaya (STTS) dan jajaran manajemennya, pak Gunawan, pak Herman, dan pak Ferdinandus, serta Ketua Yayasan PTTN Bapak Indra Setiawan, MM., MBA yang telah memberi kesempatan studi program Doktor, serta pemberian bonus akhir tahun dan bonus jalan-jalan ke Luar Negeri setiap tahun.
10. Rekan-rekan dosen STTS yang juga teman seperjuangan dalam menempuh program doktor, bu Francisca, pak Hari, pak Gunawan, pak Lukman, pak Hartarto, pak Yosi, pak Joan, pak Edwin, bu Esther, pak David yang telah memberikan semangat penulis untuk segera lulus jadi Doktor.
11. Pak Lukman, pak Julius, Bonie, Amel, Pujo, dan Ofel, yang support dalam *blend shape animation*; Milun Christianto Wibowo dan Joan Santoso, yang support dalam implementasi *Hidden Markov Model*; pak Yosi dan Herman Thuan yang membantu desain poster dan pamphlet.
12. Hosyi'ah Rosdiana dan Artiarini KN, mahasiswa S1 dan S2 Teknik Elektro ITS yang membantu perekaman *facial motion capture*, beberapa mahasiswa S1 Informatika, S1 DKV, dan S1 Industri STTS yang telah bersedia menjadi *model viseme Indonesia*.

13. Yang tercinta bapak dan ibu almarhum, Bapak K. S. Soemarno dan Ibu Hj. Oemijati, yang selama ini mendoakan penulis semasa hidupnya untuk meraih cita-cita setinggi-tingginya, namun penulis belum sempat mewujudkannya sebelum beliau dipanggil Allah SWT. Alhamdulillah gelar Doktor ini dapat diraih berkat dukungan dan doa tulus dari almarhum kedua orang tua yang sangat menyayangi penulis.
14. Suamiku tercinta H. Didik Wahyudi, yang dengan tulus telah memberikan dukungan, doa, pengorbanan, dan pengertian dengan kesabaran dan ketabahan yang sangat luar biasa, khususnya ketika ditinggal mengikuti *joint research* di luar negeri, sehingga penulis dapat menyelesaikan program doktor ini.
15. Terkhusus kepada ketiga puteri kami Anggraita Arindra Febrianti (Febby), Nadia Laksmita Sari (Mita), dan Amadhea Trizza Audina (Dhea), yang penulis kasih dan sayangi. Berkat doa tulus, dukungan semangat, dan pengertian yang luar biasa sehingga gelar Doktor ini dapat diraih. Terutama pengertian dan kesabarannya selama penulis tinggal untuk mengikuti program sandwich di Tokyo, Jepang selama 4 bulan.
16. Kakak-kakak penulis Hj. Harlistyati, Hj. Hardyahsanti, Hari Wahyudi, dan adik Suhardiyono (almarhum), yang ikut membantu doa dan semangat penulis dalam menyelesaikan pendidikan Sarjana, Magister hingga menjadi Doktor, serta Keluarga besar K.S. Soemarno dan Kaerun Tjiptodihardjo, yang bersedia hadir dalam sidang terbuka promosi doktor.
17. Pihak-pihak lain yang tidak dapat kami sebutkan satu-persatu dan telah memberikan bantuan dan doanya selama ini.

Akhirnya penulis pun sadar bahwa sebagai manusia biasa, kami memiliki keterbatasan dan kekurangan sehingga karya ini mungkin masih belum sempurna. Untuk itu saran dan kritik yang membangun akan diterima dengan tangan terbuka.

Surabaya, 10 Januari 2017

Endang Setyati

[halaman ini sengaja dikosongkan]

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PERSEMBAHAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN DISERTASI.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xix
DAFTAR TABEL.....	xxiii
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Permasalahan.....	3
1.3 Tujuan dan Manfaat Penelitian.....	3
1.4 Tahapan Penelitian.....	4
1.5 Kontribusi Penelitian.....	6
1.6 Orisinalitas Penelitian.....	6
1.7 Roadmap Penelitian dan Penelitian Sebelumnya.....	7
1.7.1 Roadmap Penelitian.....	7
1.7.2 Penelitian Sebelumnya.....	8
1.8 Sistematika Penulisan Disertasi.....	8
BAB II. PEMETAAN FONEM KE VISEME BAHASA INDONESIA BERDASARKAN <i>BLEND SHAPE ANIMATION</i>	19
2.1 Karakteristik Bahasa Indonesia.....	19
2.2 Deskripsi Fonem dalam Bahasa Indonesia.....	20

2.2.1 Fonologi, Fonetik, dan Fonemik.....	20
2.2.2 Pengertian Dasar Fonem.....	21
2.2.3 Pengucapan Artikulasi Fonem.....	22
2.3 Pemilihan Fonem Indonesia yang Digunakan.....	23
2.3.1 Fonem Vokal dan Konsonan.....	23
2.3.2 Himpunan Fonem Indonesia.....	23
2.4 Perbandingan antara Fonem Indonesia dan Fonem Bahasa Lainnya.....	26
2.5 Viseme.....	27
2.5.1 Definisi Viseme.....	27
2.5.2 Pemetaan Fonem ke Viseme.....	28
2.5.3 Pemetaan Fonem ke Viseme Bahasa Asing Lainnya...	29
2.6 Konstruksi Model Viseme Indonesia.....	30
2.6.1 Data Input.....	31
2.6.1.1 Definisi dari Aturan Prosodi Indonesia.....	31
2.6.1.2 Klasifikasi Suku Kata Bahasa Indonesia.....	32
2.6.1.3 Konsep dari Pembacaan Suku Kata Bahasa Indonesia.....	33
2.6.2 Perekaman dari Pembacaan Suku Kata Bahasa Indonesia.....	34
2.6.2.1 Perekaman Video.....	34
2.6.2.2 Pemilihan Parameter Geometrik dari <i>Blend Shape Animation</i>	35
2.7 Konstruksi Model 3D.....	37
2.7.1 <i>Blend Shape Model</i> Wajah.....	37
2.7.2 Transformasi Verteks.....	38
2.7.2.1 Matriks Transformasi.....	39
2.7.2.2 Koordinat.....	40
2.7.3 Pembentukan Animasi Mulut dengan Menggunakan <i>Blend Shape</i> Berdasarkan Perekaman Video.....	41
2.8 Hasil Eksperimen.....	44

2.8.1 Pemetaan Fonem-Viseme Indonesia.....	44
2.8.2 Validasi Berdasarkan Survei.....	46
2.8.2.1 Instrumen Survei.....	46
2.8.2.2 Pengumpulan Data.....	48
2.8.3 Uji Validitas dan Reliabilitas.....	50
2.8.4 Analisis Data dan Hasil Penelitian.....	50
2.8.4.1 Validitas.....	50
2.8.4.2 Reliabilitas.....	51
2.8.4.3 Rating Indeks, Mean Rating dan Mean Square Eror.....	51
2.9 Model Viseme Indonesia.....	54
 BAB III. AKUISISI DATA UNTUK EKSPRESI WAJAH 3D ANIMASI	57
3.1 Emosi dan Ekspresi Wajah.....	57
3.2 Deskripsi Tekstual dari Ekspresi Wajah.....	59
3.3 Afeksi dan Emosi.....	60
3.4 <i>Database Ekspresi Wajah</i>	60
3.5 <i>Indonesian Female Facial Expression-3D Database</i>	61
3.6 <i>Facial Motion Capture</i>	62
3.6.1 Penyusunan Kamera.....	64
3.6.2 Kalibrasi Kamera.....	65
3.6.3 Peletakan Penanda.....	68
3.6.4 Kalibrasi <i>Face Template</i>	69
3.7 Perekaman Ekspresi Wajah pada <i>Facial Motion Capture</i>	71
3.8 Pemilihan Afeksi.....	72
3.9 Pembuatan Ekspresi Wajah 3D Animasi.....	77
3.9.1 Animasi dengan <i>Motion Capture</i>	77
3.9.2 Pengeditan <i>Keyframe</i>	79
3.9.3 Rendering Realistik dari <i>Blend Shape</i>	80

BAB IV. HIDDEN MARKOV MODELS BERDASARKAN MODEL VISEME INDONESIA UNTUK UCAPAN ALAMI DENGAN AFEKSI.....	83
4.1 Arsitektur Sistem Keseluruhan.....	83
4.1.1 Desain Sistem Segmentasi Input Teks.....	84
4.1.2 Desain Sistem Implementasi <i>Hidden Markov Model</i> dan Agoritma <i>Viterbi</i>.....	85
4.2 Segmentasi Input Teks.....	87
4.2.1 Pemotongan Kalimat Menjadi Kata.....	87
4.2.2 Pemotongan Kata Menjadi Fonem.....	89
4.2.3 Pengubahan Fonem Menjadi Viseme.....	92
4.3 <i>Hidden Markov Model</i>.....	94
4.3.1 Terminologi HMM.....	95
4.3.1.1 Pola Deterministik dan Non-Deterministik.....	95
4.3.1.2 Pola Tersembunyi.....	97
4.3.2 Definisi HMM.....	99
4.3.3 Tiga Masalah HMM.....	100
4.3.4 Algoritma <i>Viterbi</i> dan HMM.....	101
4.3.5 Implementasi HMM dan Algoritma Viterbi.....	103
4.3.5.1 Pemisahan Data Training.....	103
4.3.5.1.1 Agoritma Pemisahan Data Training..	103
4.3.5.1.2 Penerapan Pemisahan Data Training.	104
4.3.5.2 Proses <i>Decoding</i> dengan Menggunakan Algoritma Viterbi.....	112
4.3.5.2.1 Algoritma Proses <i>Decoding</i> dengan Menggunakan Algoritma <i>Viterbi</i>.....	112
4.3.5.2.2 Penerapan Proses <i>Decoding</i> dengan Menggunakan <i>Trigram</i> HMM dan Algoritma <i>Viterbi</i>.....	115
4.4 <i>Inesh² Corpora</i>.....	125

BAB V. VISUALISASI WAJAH ANIMASI PADA TALKING HEAD SYSTEM DALAM BAHASA INDONESIA.....	129
5.1 Konstruksi dari Proses Visualisasi Urutan Viseme.....	129
5.2 Algoritma Visualisasi Urutan Viseme.....	131
5.3 Penerapan Visualisasi Urutan Viseme.....	132
5.4 Pengujian Pengucapan dari Hasil Eksperimen.....	134
5.4.1 Pengujian Pengucapan Urutan Viseme secara Alami...	135
5.4.2 Pengujian dengan Mean Square Error.....	137
5.5 Pengujian Representasi Visual.....	141
5.6 Perbandingan Jumlah Frame.....	144
5.7 Tampilan Urutan Viseme Alami Berafeksi.....	146
 BAB VI. PENUTUP.....	149
5.1 Kesimpulan Hasil Penelitian.....	149
5.2 Usulan Penelitian Selanjutnya.....	153
 DAFTAR PUSTAKA	155
BIODATA PENULIS.....	161
LAMPIRAN 1 DATA MODEL JIFFE-3D DATABASE.....	167

[halaman ini sengaja dikosongkan]

DAFTAR GAMBAR

Gambar	Halaman
Bab I	
Gambar 1.1 Fishbone Diagram dari Tahapan Penelitian Pembuatan THS-IL.....	4
Gambar 1.2 Bagan dari Roadmap Penelitian.....	6
Gambar 1.3 Bagan dari Roadmap Penelitian.....	7
Gambar 1.4 Sistematika Penulisan Disertasi.....	16
Bab II	
Gambar 2.1 Konstruksi dari Model Viseme Indonesia.....	30
Gambar 2.2 Posisi Duduk Model dari Arah Depan dan Samping Kamera	35
Gambar 2.3 Posisi Pembacaan Suku Kata Bahasa Indonesia.....	35
Gambar 2.4 Sistem Koordinat Cartesius (a) 2D dan (b) 3D	41
Gambar 2.5 Bentuk Standar Titik Kontrol yang dapat Digunakan Animator.....	43
Gambar 2.6 Rendering dari Blend Shape 3D Model Viseme Indonesia dengan Menggunakan MakeHuman: (a) Tekstur dari MakeHuman, (b) 3D dari Model Wajah, (c) Hasil <i>Rendering</i> dari (b), (d) Daerah Bibir Merah dari (c) dengan 9 CP dalam Daerah Mulut.....	44
Gambar 2.7 Grafik Rating Respon dari Responden.....	54
Bab III	
Gambar 3.1 Contoh Ekspresi Wajah dari JAFFE.....	57
Gambar 3.2 Konstruksi Pembentukan Model Ekspresi Wajah Berafeksi dengan IFFE-3D Database.....	62
Gambar 3.3 Contoh Pengaturan Posisi Multi Kamera pada Facial Motion Capture.....	64
Gambar 3.4 Ilustrasi Triangulasi Kamera Stereo.....	65
Gambar 3.5 Capture Volume yang Dihasilkan pada Proses Kalibrasi.....	66
Gambar 3.6 (a) Proses <i>Wanding</i> dan (b) <i>Optiwand Kit</i>	67

Gambar 3.7 <i>Screen shot</i> Hasil Akhir Proses <i>Wanding Enam Kamera</i>	67
Gambar 3.8 (a) <i>Penentuan ground plane</i> dan (b) <i>Calibration Square</i>	68
Gambar 3.9 Contoh (a) Ilustrasi Peletakan <i>Marker</i> pada Wajah dan (b) <i>Marker</i> yang Dipasang pada Wajah untuk <i>IFFE-3D Database</i>	69
Gambar 3.10 Contoh data 3D <i>marker based</i> pada <i>IFFE-3D Database</i> (a) Data tampak depan; (b) Data tampak samping.....	70
Gambar 3.11 Proses Penangkapan Ekspresi Wajah.....	71
Gambar 3.12 Contoh Data Survei.....	72
Gambar 3.13 Akuisisi Data Animasi dengan <i>Motion Capture</i> Menggunakan Blend Shape Model untuk Afeksi: (a) Netral, (b) Terkejut, (c) Malu, (d) Sedih.....	80
Gambar 3.13 Akuisisi Data Animasi dengan <i>Motion Capture</i> Menggunakan Blend Shape Model untuk Afeksi: (e) Takut, (f) Marah, (g) Senang, (h) Jijik.....	81
Bab IV	
Gambar 4.1 Desain Sistem Segmentasi Input Teks.....	85
Gambar 4.2 Desain Sistem Implementasi <i>Hidden Markov Model</i>	86
Gambar 4.3 Konstruksi Jalannya Penelitian Tahap Ketiga.....	87
Gambar 4.4 Urutan Warna yang Muncul pada Lampu Lalu Lintas.....	95
Gambar 4.5 <i>State Diagram</i> Perubahan Cuaca.....	96
Gambar 4.6 <i>HMM</i> Kasus Pertapa.....	97
Gambar 4.7 Confussion Matrix Hubungan Antara Keadaan Rumput Laut dan Cuaca.....	98
Gambar 4.8 Diagram Input-Output Algoritma <i>Viterbi</i>	101
Gambar 4.9 Frekuensi Panjang Frame dari Audacity pada Sebuah Contoh Kata “Keluar” untuk Masing-masing Parameter Afeksi (a) Jijik, (b) Malu, (c) Marah, (d) Netral, (e) Sedih, (f) Senang, (g) Takut, (h) Terkejut.....	108

Gambar 4.10 Informasi Urutan Viseme dan <i>Frame</i> dari Contoh data Training.....	110
Bab V	
Gambar 5.1 Proses Visualisasi Urutan Viseme.....	129
Gambar 5.2 Desain Sistem Visualisasi 3D Urutan Viseme Indonesia....	130
Gambar 5.3 Contoh Gambar Viseme A Hasil Rendering dari 8 Visualisasi Afeksi dengan Ekspresi Wajah Mulai dari Baris Pertama Kiri ke Kanan: Netral, Marah, Terkejut, Malu, Sedih, Jijik, Senang, Takut.....	132
Gambar 5.4 Contoh Visualisasi Afeksi Netral dari 12 Macam Jenis Bentuk Bibir Menurut Model Viseme Indonesia	133
Gambar 5.5 Contoh Gambar Transisi Viseme untuk Viseme O	134
Gambar 5.6 Contoh Sebuah Pertanyaan dalam Survei Online.....	142
Gambar 5.7 Grafik Panjang Frame Berdasarkan panjang Viseme.....	145

[halaman ini sengaja dikosongkan]

DAFTAR TABEL

Tabel	Halaman
Bab I	
Tabel 1.1 Matriks Penelitian Sebelumnya.....	9
Tabel 1.2 Topik Tahapan Penelitian Disertasi dan Publikasi.....	17
Bab II	
Tabel 2.1 Alfabet dalam Bahasa Indonesia.....	19
Tabel 2.2 Himpunan Fonem Bahasa Indonesia.....	25
Tabel 2.3 Pola Suku Kata Bahasa Indonesia.....	33
Tabel 2.4 Titik-titik Kontrol (<i>Control Points, CP</i>).....	37
Tabel 2.5 Himpunan Fonem Indonesia Berdasarkan Jenis Fonem Konsonan dan Vokal.....	45
Tabel 2.6 Contoh Isian Kuesioner dari Respon Survei.....	47
Tabel 2.7 Profil Responden.....	48
Tabel 2.8 Skor Respon dari Responden.....	49
Tabel 2.9 Prosentase Skor Respon dari Responden.....	49
Tabel 2.10 Rating Indeks Viseme yang Diberikan oleh Responden.....	52
Tabel 2.11 Mean Squared Error Respon dari Responden.....	53
Tabel 2.12 Hasil Model Animasi 3D dari Pemetaan Fonem ke Viseme Indonesia.....	55
Bab III	
Tabel 3.1 Klasifikasi Emosi Dasar Manusia Menurut Beberapa Psikolog.....	58
Tabel 3.2 Ekspresi Wajah dari Enam Emosi Dasar.....	59
Tabel 3.3 Perbandingan <i>Database</i> Ekspresi Wajah 2D dan 3D.....	61
Tabel 3.4 Profil Model IFFE-3D <i>Database</i>	62
Tabel 3.5 Bagian-bagian Wajah yang Dipasang Penanda secara Simetris.	69
Tabel 3.6 Contoh <i>Facebone</i> pada Ekspresi Wajah Model.....	70
Tabel 3.7 (a) Prosentase Hasil Survei Pose Terkejut.....	73
Tabel 3.7 (b) Prosentase Hasil Survei Pose Takut.....	73

Tabel 3.7 (c) Prosentase Hasil Survei Pose Sedih.....	74
Tabel 3.7 (d) Prosentase Hasil Survei Pose Senang.....	74
Tabel 3.7 (e) Prosentase Hasil Survei Pose Jijik.....	74
Tabel 3.7 (f) Prosentase Hasil Survei Pose Marah.....	75
Tabel 3.7 (g) Prosentase Hasil Survei Pose Netral.....	75
Tabel 3.8 Prosentase Terbesar Hasil Survei pada Setiap Pose.....	76
Bab IV	
Tabel 4.1 Tabel Pengklasifikasian Fonem Berkonsonan Ganda.....	89
Tabel 4.2 Pengklasifikasian Viseme.....	92
Tabel 4.3 Nama File Data Training untuk Parameter Afeksi.....	104
Tabel 4.4 Enam Contoh Kalimat Training yang Berafeksi Netral.....	105
Tabel 4.5 Representasi Frame dari Animasi Gerakan Bibir.....	106
Tabel 4.6 Contoh Panjang Frame rate dari Parameter Afeksi.....	109
Tabel 4.7 Jumlah <i>Trigram State</i> dan Jumlah <i>Bigram State</i> dari Urutan <i>Frame</i>	111
Tabel 4.8 Jumlah <i>State</i> → <i>Viseme</i> dan Jumlah <i>State</i>	111
Tabel 4.9 Rekapitulasi Hasil Perhitungan Viterbi.....	125
Bab V	
Tabel 5.1 Mean Square Error dari DT1 hingga DT6.....	138
Tabel 5.2 Hasil Perhitungan Uji Validasi Urutan Viseme Alami.....	139
Tabel 5.3 Profil dari Responden pada Survei Online.....	143
Tabel 5.4 Hasil Survei Online dari Representasi Visual.....	144
Tabel 5.5 Hasil Perbandingan Jumlah Frame.....	145
Tabel 5.6 Ilustrasi Barisan Gambar dari Output SASAG-IBU dengan Parameter Afeksi Marah.....	147

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, telah terjadi pertumbuhan minat dalam bidang interaksi antara manusia dan komputer (*human-computer interaction*) yang sangat pesat, terutama di industri kreatif (permainan, hiburan). Pengenalan emosi dan pembangkitannya adalah tema penelitian baru di Interaksi Manusia Komputer (IMK) yang didasarkan pada komputasi afektif yang diharapkan dapat meningkatkan kualitas IMK secara signifikan.

Pengembangan dari antarmuka manusia-komputer (*human-computer interface*) modern, seperti layanan berbasis web informasi, *e-commerce*, dan *e-learning*, di masa mendatang secara luas akan menggunakan animasi wajah yang dikombinasikan dengan sistem dialog (E. Cosatto, 2003) dan (Kang Liu, 2009).

Penelitian di bidang komputasi afektif (*affective computing*) saat ini berkonsentrasi pada analisis emosi manusia dengan mengukur sinyal fisiologis dan dengan mengekstraksi hubungan pola-pola afektif yang sesuai. Hal ini dilakukan misalnya dengan cara menganalisis ekspresi wajah (*facial expression*). Pengenalan dari kondisi emosional manusia ini, tidak hanya dapat meningkatkan kinerja komputer untuk mendeteksi keadaan afektif manusia, tetapi juga mampu menginterpretasikan dan merespon balik dari afektif penggunanya.

Untuk membangun relasi antara pengguna dan karakter virtual dilengkapi dengan pengenalan emosi. Pengenalan emosi ini dapat ditempuh dalam beberapa cara, antara lain pengenalan ekspresi wajah dari pengguna saat berinteraksi dengannya dan mengkomunikasikan melalui kata-kata (teks atau lisan).

Ekspresi wajah emosional seseorang dapat digunakan untuk membuat karakter virtual afektif yang bertindak sebagai sebuah antarmuka yang cerdas untuk melakukan interaksi secara alami dan ramah antara manusia dan mesin. Dalam membangun salah satu karakter virtual afektif, dapat digunakan sebuah sistem yang membuat sebuah visualisasi wajah animasi model kepala berbicara (*talking head system*) dengan menggunakan bahasa tertentu.

Talking head system merupakan sebuah sistem *text-to-audiovisual-speech* yang pertama kali diperkenalkan oleh Scott A. King pada tahun 2001. Sistem ini mengambil teks sebagai input dan secara otomatis membangkitkan audio dan visual sekaligus dari input teks (*text input*) tersebut.

Pada sebuah penelitian (J. Beskow, 2003), telah dikembangkan sebuah sistem aplikasi *chatting* yang menggunakan fungsi matematika polinomial. Sistem ini terdiri dari dua modul utama, yang pertama membangkitkan teks respon dan yang kedua membangkitkan emosi berdasarkan konteks dari sebuah percakapan. Teks respon dan emosi dikirim pada modul pembangkit ucapan (*speech generator*). Dalam modul ini, file audio ekspresif dan himpunan fonem dibangkitkan berdasarkan pada teks respon. Himpunan fonem dikirim pada modul pembangkit animasi (*animation generator*). Modul pembangkit animasi ini menerima daftar fonem dan membuat file animasi untuk *Talking Head System*.

Sedangkan dalam penelitian yang lain (Kang Liu and Joern Ostermann 2009), telah melakukan optimisasi *talking head system* berbasis gambar (*image*), yang mencakup dua bagian, yaitu: analisis dan sintesis. Bagian analisis audiovisual membuat sebuah model wajah dari subyek manusia yang direkam, berkaitan dengan informasi database besar dari gambar mulut. Bagian sintesis menghasilkan animasi wajah yang terlihat alami dari transkrip fonetik teks. Kondisi paling kritis dari sintesis ini adalah pemilihan unit (*unit selection*) yang memilih dan merangkai gambar-gambar bentuk mulut yang tepat dari database, sehingga sesuai dengan kata-kata yang diucapkan oleh *talking head system*. Pemilihan ini didasarkan pada sinkronisasi bibir dan kesamaan gambar bentuk mulut secara berturut-turut.

Untuk mengetahui bagaimana sinkronisasi bibir digerakkan secara aktual selama berekspresi, maka dibuatlah sebuah sistem untuk mengembangkan ekspresi wajah yang realistik. Setelah membuat ekspresi, berikutnya menambahkan ekspresi ke animasi ucapan sintesis, dengan menggunakan mulut untuk artikulasi sebuah ucapan. Ucapan yang ekspresif biasanya berdasarkan pada beberapa aturan dalam membuat ekspresif sintesis. Ekspresif sintesis ini digunakan untuk memodifikasi ucapan netral yang dibangkitkan oleh sistem *text-to-speech* (TTS) yang mengubah nada (*pitch*), intensitas, tempo dan durasi dari sebuah input teks kalimat.

Dalam berkomunikasi dengan menggunakan input teks, diperlukan sebuah pemetaan fonem (*phoneme*) ke viseme (*visual phoneme*). Fonem adalah unsur bahasa terkecil yang dapat membedakan arti (Gleason, 1961). Fonem adalah unit terkecil suara yang merupakan dasar untuk membangun sebuah ucapan manusia (*human speech*). Viseme diturunkan dari sekelompok fonem yang memiliki penampilan visual sama dan merupakan unit ekivalen dalam domain visual yang memodelkan sistem pengenalan ucapan secara audio-visual. Untuk membuat sistem yang memetakan fonem ke viseme, digunakan database file animasi.

Sampai dengan saat ini, belum ada peneliti lain yang tertarik menggunakan input teks dari kalimat Bahasa Indonesia berafeksi, untuk dipakai dalam pembuatan *talking head system*. Hal inilah yang mendorong peneliti saat ini untuk menggunakan model viseme Indonesia, yang dimanfaatkan dalam mengembangkan sistem yang menghasilkan visualisasi 3D wajah animasi ucapan dalam bahasa Indonesia (*Talking Head System in Indonesian Language*) dengan sintesis ekspresi wajah berafeksi (*affective facial expression synthesis*).

1.2 Permasalahan

Berdasarkan latar belakang yang telah dijelaskan di atas, maka permasalahan yang dihadapi dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengembangkan *talking head system* dengan menggunakan input teks dari sebuah kalimat Bahasa Indonesia?
2. Bagaimana cara mengembangkan sistem antarmuka cerdas berbentuk afektif yang menghasilkan visualisasi 3D wajah animasi ucapan dalam Bahasa Indonesia?
3. Bagaimana cara menambahkan ekspresi ke animasi wajah sintesis sesuai dengan input teks kalimat Bahasa Indonesia yang berafeksi?

1.3 Tujuan dan Manfaat Penelitian

Tujuan yang akan dicapai dari penelitian ini adalah :

1. Mengembangkan sistem afektif berwujud karakter virtual kepala berbicara (*talking head system*) dengan input teks kalimat Bahasa Indonesia berafeksi, berbasis teknik cerdas, pemodelan emosi dan visualisasi.

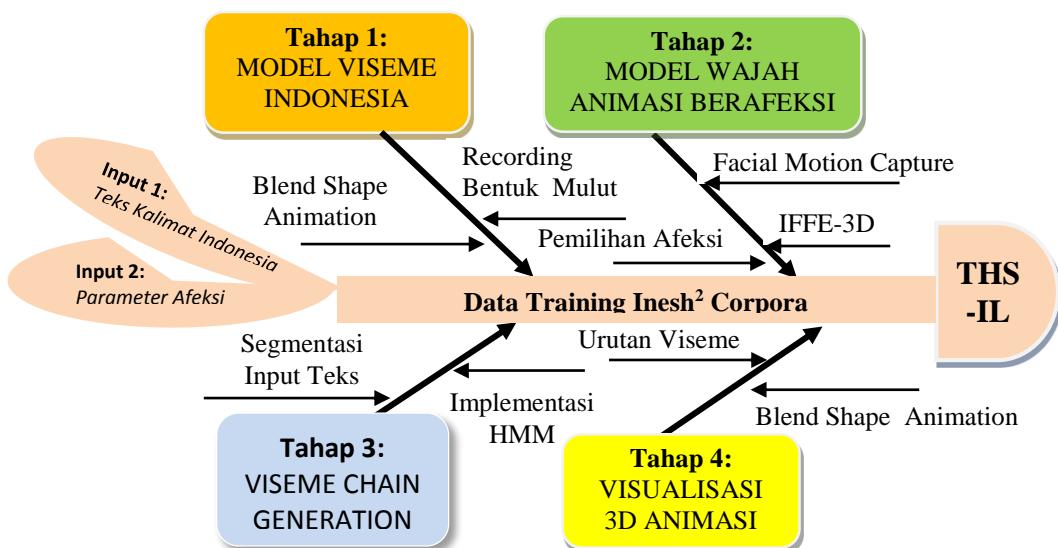
2. Mewujudkan karakter virtual dalam bentuk ekspresi wajah beremosi, antara lain dengan emosi marah, senang, sedih, takut, jijik, terkejut, netral dan malu.
3. Menghasilkan visualisasi 3D model wajah animasi dengan pengucapan urutan viseme Indonesia yang alami dan berafeksi.

Sedangkan manfaat yang diharapkan dari penelitian ini adalah:

1. Digunakannya perangkat lunak (*software*) berbasis sumber terbuka (*open source*) dan karya seni berlisensi *common creative* yang memungkinkan pengembangan lebih lanjut bagi peneliti-peneliti yang berminat.
2. Dihadirkannya teknologi antarmuka cerdas berafeksi yang tidak sulit dalam pelatihan dan penggunaannya.

1.4 Tahapan Penelitian

Tahapan dalam penelitian ini terbagi dalam 4 tahap, lihat Gambar 1.1 yang merupakan fishbone diagram dari tahapan penelitian.



Gambar 1.1 Fishbone Diagram dari Tahapan Penelitian Pembuatan Talking Head System in Indonesian Language (THS-IL)

Keempat tahapan penelitian pada Gambar 1.1 adalah sebagai berikut:

- Tahap pertama penelitian adalah pembuatan model viseme Indonesia (*Indonesian Viseme Model*), dengan menggunakan *blend shape animation*.

Pemodelan viseme Indonesia ini untuk mendapatkan bentuk mulut sesuai dengan pemilihan viseme Indonesia yang telah ditentukan, berdasarkan perekaman bentuk mulut dari beberapa model. Selain itu, model viseme Indonesia ini diharapkan dapat digunakan sebagai referensi dalam sintesis pengucapan dalam bahasa Indonesia (*Indonesian Speech Synthesis*) oleh peneliti berikutnya.

- Tahap kedua penelitian adalah pembentukan model ekspresi wajah berafeksi dengan menggunakan *facial motion capture*, untuk mendapatkan model *Indonesian Female Facial Expression (IFFE)-3D Database*. Dari tahap ini diharapkan terbentuk wajah animasi sesuai pemilihan afeksi.
- Tahap ketiga penelitian adalah pembangkitan urutan viseme (*Viseme Chain Generation*). Pada tahap ketiga ini, nantinya terbagi dalam 2 proses, yaitu: Segmentasi Input Teks dan Implementasi HMM, agar menghasilkan *talking head system* yang alami.
 - Tahap segmentasi input teks, memiliki 3 step, yaitu: melakukan pemotongan dari input teks kalimat Bahasa Indonesia menjadi kata, kemudian dari kata dipotong menjadi fonem, dan terakhir dari fonem diubah menjadi frame fonem atau viseme.
 - Implementasi HMM, memiliki 2 step, yaitu: membentuk model Trigram HMM yang sesuai dengan parameter afeksi, dengan melakukan pemisahan data training **Inesh² Corpora** (*Indonesian Emotional Speech-Indonesian Endang Surya Hery Corpora*) ke dalam trigram dan bigram HMM, dan kemudian melakukan proses decoding dengan menggunakan Algoritma Viterbi.
- Tahap keempat penelitian adalah tahap pembuatan visualisasi 3D animasi wajah berafeksi.

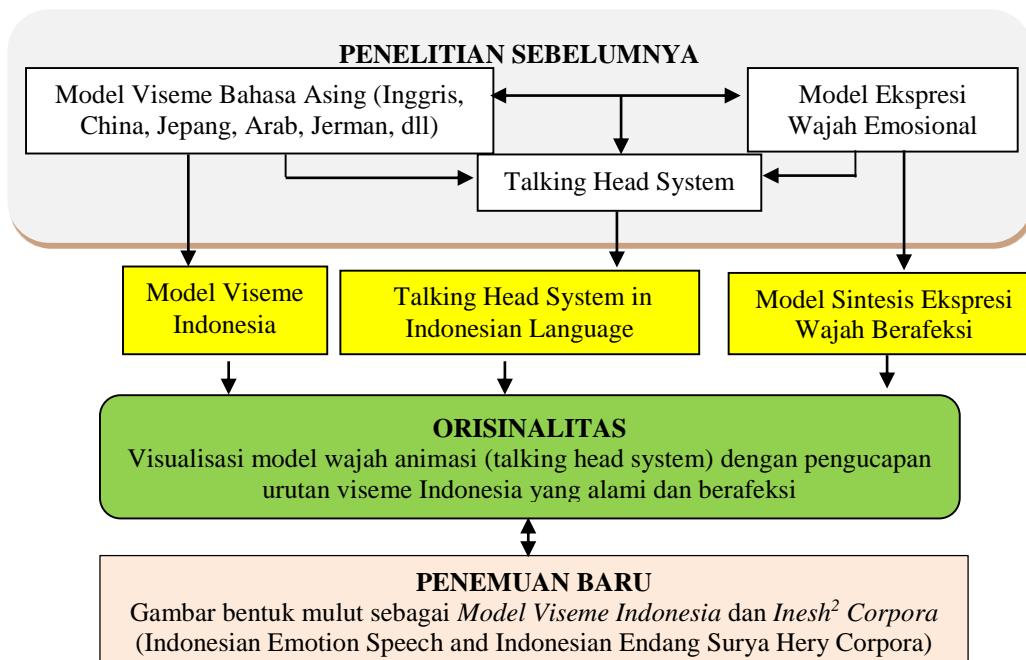
Hasil akhir dari seluruh tahapan penelitian ini adalah terbentuknya sebuah visualisasi 3D wajah animasi ucapan dalam bahasa Indonesia yang berafeksi (*Talking Head System in Indonesian Language with Affective Facial Expressions Synthesis*).

1.5 Kontribusi Penelitian

Kontribusi dari penelitian ini adalah dibangunnya satu sistem terintegrasi *Talking Head System* dalam Bahasa Indonesia yang membentuk sebuah visualisasi model wajah 3D animasi dengan pengucapan kalimat Bahasa Indonesia yang alami dan berafeksi.

1.6 Orisinalitas Penelitian

Berdasarkan uraian di atas, belum pernah ada penggunaan model viseme Indonesia dan urutan viseme alami, digabung menjadi visualisasi animasi ucapan berafeksi dari 3D model wajah animasi (*talking head system*) yang bisa menerima parameter ekspresi dengan input teks kalimat berbahasa Indonesia. Gambar 1.2 menunjukkan posisi orisinalitas penelitian dan penemuan baru dari penelitian ini.



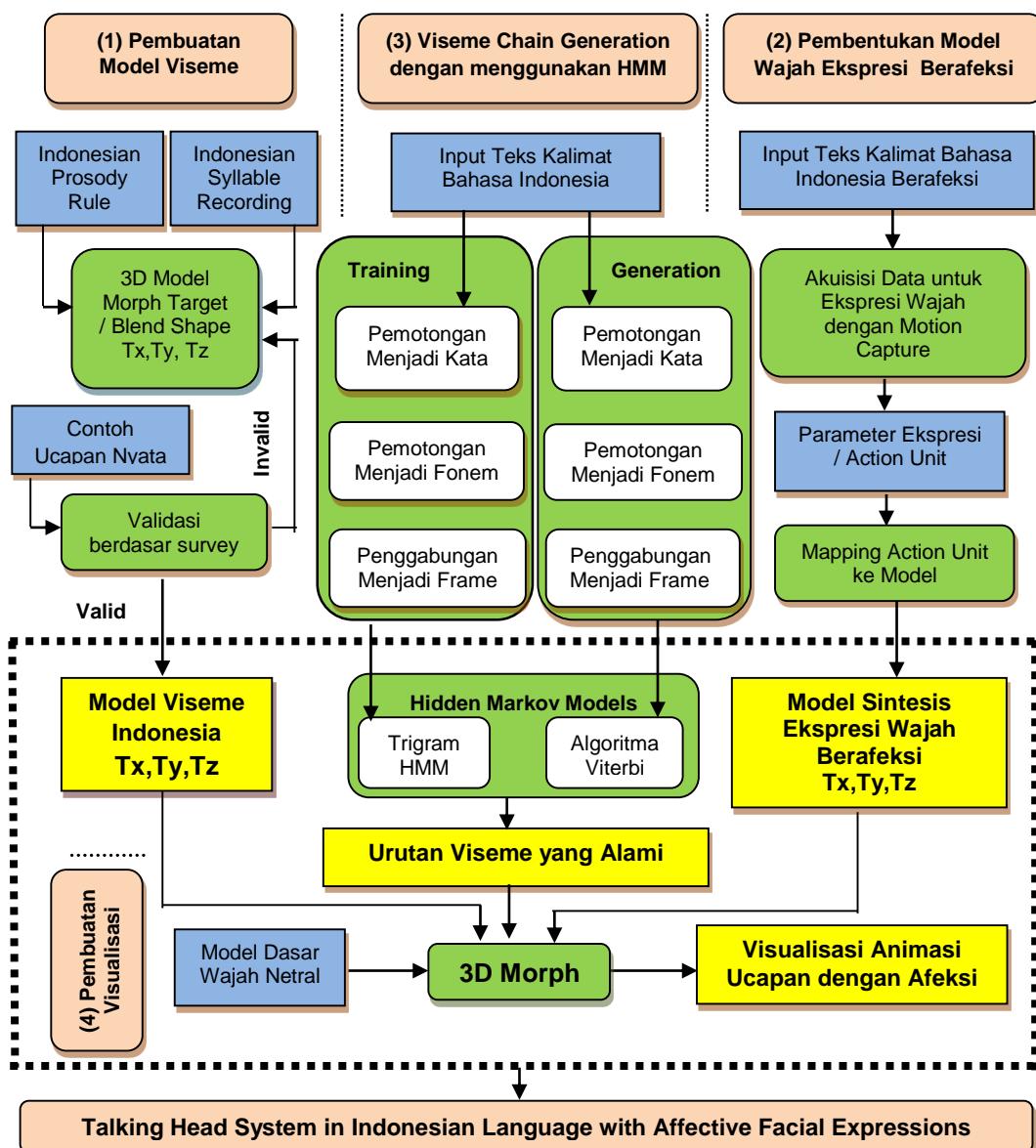
Gambar 1.2 Bagan Orisinalitas Penelitian

1.7 Roadmap Penelitian dan Penelitian Sebelumnya

Bagan pada Gambar 1.3 dan Tabel 1.1 dari roadmap penelitian, merupakan salah satu parameter yang menunjukkan bahwa penelitian ini masih orisinal dan bisa terus berkembang seiring dengan pesatnya perkembangan komputer, meskipun topik sejenis telah dibahas oleh banyak peneliti selama lebih dari lima belas tahun.

1.7.1 Roadmap Penelitian

Pada roadmap penelitian ini, tahapan yang dilakukan dalam pembuatan *Talking Head System in Indonesian Language* (THS-IL) terdiri dari empat tahap, yaitu: (1) Pembuatan model viseme; (2) Pembentukan model wajah berafeksi; (3) Pembentukan urutan viseme yang alami dengan menggunakan HMM; dan (4) Pembuatan visualisasi model wajah 3D.



Keterangan warna: Tahap penelitian (merah muda), Data input (biru), Proses (hijau), Sub proses (putih), Data output setiap tahap (kuning)

Gambar 1.3 Bagan dari Roadmap Penelitian

1.7.2 Penelitian Sebelumnya

Pengenalan dan pembangkitan emosi adalah tema penelitian baru di dalam IMK yang didasarkan pada komputasi afektif yang diharapkan dapat meningkatkan kualitas sistem dan komunikasi IMK secara signifikan. Banyak aplikasi untuk telekonferensi, animasi komputer yang membutuhkan reproduksi ekspresi wajah afektif dan gerak alami dari karakter virtual yang cerdas dan hidup, seperti sebuah “robot” yang berfungsi sebagai agen IMK.

Demikian juga dengan kebutuhan akan sebuah sistem membaca bibir otomatis yang saat ini semakin meningkat. Pada kenyataannya, ekstraksi dan analisis yang dapat diandalkan pada gerakan wajah merupakan bagian yang penting dalam sistem multimedia.

Penelitian-penelitian sebelumnya yang telah dilakukan oleh para peneliti untuk menunjang penelitian ini dapat dilihat pada Tabel 1.1 yang berisi matriks pelacakan publikasi dari penelitian sebelumnya.

1.8 Sistematika Penulisan Disertasi

Berdasarkan roadmap penelitian dan matriks pelacakan publikasi sebelumnya, maka sistematika penulisan disertasi disajikan dalam Gambar 1.4. Penelitian ini diawali dengan menggali studi literatur dan pengkajian penelitian sebelumnya tentang pengetahuan corpus, pemetaan fonem ke viseme dari berbagai bahasa asing, pembentukan model ekspresi wajah animasi berafeksi, pembangkitan urutan viseme dan cara membuat visualisasi 3D animasi dari model ekspresi wajah berafeksi/emosi hingga terwujudnya Talking Head System dalam Bahasa Indonesia dengan ucapan urutan viseme yang alami.

Sedangkan pada Tabel 1.2 berisi tentang topik dan hasil / target yang dicapai pada tahapan penelitian yang telah dipublikasikan di Seminar dan Jurnal Internasional.

Tabel 1.1 Matriks Pelacakan Publikasi dari Penelitian Sebelumnya

Ta hun	Pengarang dan Diterbitkan	Judul Penelitian	Pembentukan Model Viseme	Viseme Chain Generation using HMM	Pembentukan Model Ekspresi Wajah Berafeksi
2016	Endang Setyati, Joan Santoso, Surya Sumpeno, Mauridhi Hery P., 2016, Jurnal Kursor	Hidden Markov Model Based Indonesian Viseme Model For Natural Speech With Affection	Pembentukan Model Viseme Indonesia dari penelitian sebelumnya	Pembentukan Urutan Viseme yang Alami dengan menggunakan HMM	-
2015	Endang Setyati, Surya Sumpeno, Mauridhi Hery P., Mikami Koji, Masanori Kakimoto, Kunio Kondo, 2015, IAENG International Journal of Computer Science.	Phoneme-Viseme Mapping for Indonesian Language Based on Blend Shape Animation	Pemetaan fonem ke Viseme untuk Bahasa Indonesia, yang menghasilkan model viseme Indonesia dengan berdasarkan Blend Shape Animation	-	-
2015	David Alfter, Caroline Sporleder, Sven Naumann, 2015, Dissertation of Computerlinguistic & Digital Humanities, 1-149.	Language Segmentation	-	Proses pemotongan kalimat untuk data training	-
2015	Daniel Jurafsky & James H. Martin, 2015, Speech and Language Processing.	Part of Speech Tagging	-	Proses penugasan penanda “part of speech” untuk seiap kata dalam sebuah input teks	-
2012	<i>Tie Yun, Ling Guan</i> , 2012 Elsevier International Journal of Pattern Recognition 46 (2013), pp. 529-538.	Human Emotion State Recognition Using Real 3D Visual Features from Gabor Library	-	-	Membuat sistem pengenalan emosi wajah dengan menggunakan real 3D visual features yang diekstrak secara otomatis pada 3D database BU_3DFE (Database 3D Facial Expression dari Binghamton University).

Ta hun	Pengarang dan Diterbitkan	Judul Penelitian	Pembentukan Model Viseme	Viseme Chain Generation using HMM	Pembentukan Model Ekspresi Wajah Berafeksi
2012	<i>Rong Chuanzhen, Yue Zhenjun, Wang Yuan, the 2012 IEEE 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012), pp. 321-324</i>	Research on Chinese Viseme Based on Fuzzy Clustering and Grey Relation Analysis	Mengklasifikasi konsonan dan vokal ke dalam 13 viseme dasar untuk bahasa China	Mengekstrak fitur bibir dari konsonan dan vokal dengan menggunakan analisis relasi abu-abu untuk mengkonstruksi matriks relasi kemiripan fuzzy.	Merealisasikan sistem TTVS (<i>Text-To-Visual Speech</i>) untuk memverifikasi kinerja dari Chinese viseme.
2012	<i>Luca Cappelletta and Naomi Harte, the 2012 International Conference on Pattern Recognition Applications and Methods (ICPRAM 2012), February 7, 2012.</i>	Phoneme-to Viseme Mapping for Visual Speech Recognition	Pengenalan secara visual untuk memeriksa pemilihan peta viseme.	Membandingkan pemetaan dari 5 pengelompokan viseme ke dalam pengenalan ucapan kontinyu. Semua pemetaan didasarkan pada pendekatan fonem ke viseme dengan menggunakan metode linguistik.	-
2011	<i>Surya Sumpeno, IAENG International Journal of Computer Science, 38:2, IJCS_38_2_04.</i>	Facial Emotional Expressions of Life-like Character Based on Text Classifier and Fuzzy Logic	-	-	Membuat animasi life-like character yang mengekspresikan transisi emosi yang dapat dibangkitkan secara otomatis dengan sistem yang terdiri dari text classifier dan Fuzzy Inference System. Sistem ini untuk membangun life-like character yang mampu mengekspresikan emosi dari sebuah text input.
2011	<i>Anh Tuan Do, the 2011 IEEE Workshop on Digital Media and Digital Content Management, IEEE Computer Society, pp. 107-110</i>	Creating Emotional Speech for Conversational Agents	Tidak membentuk model viseme sendiri, akan tetapi menggunakan database J-ALICE berdasarkan AIML, dan menggunakan modul MARY-TTS untuk	Menyajikan sebuah pendekatan sistem yang real time dan otomatis yang mana mampu membuat ucapan ekspresif dengan menggunakan himpunan model matematika.	Memvisualisasikan emosi ucapan sintesis dalam video dan audio dengan mengumpulkan data pergerakan otot wajah melalui tracking system. Model emosi untuk mengendalikan bentuk wajah dan parameter prosodic

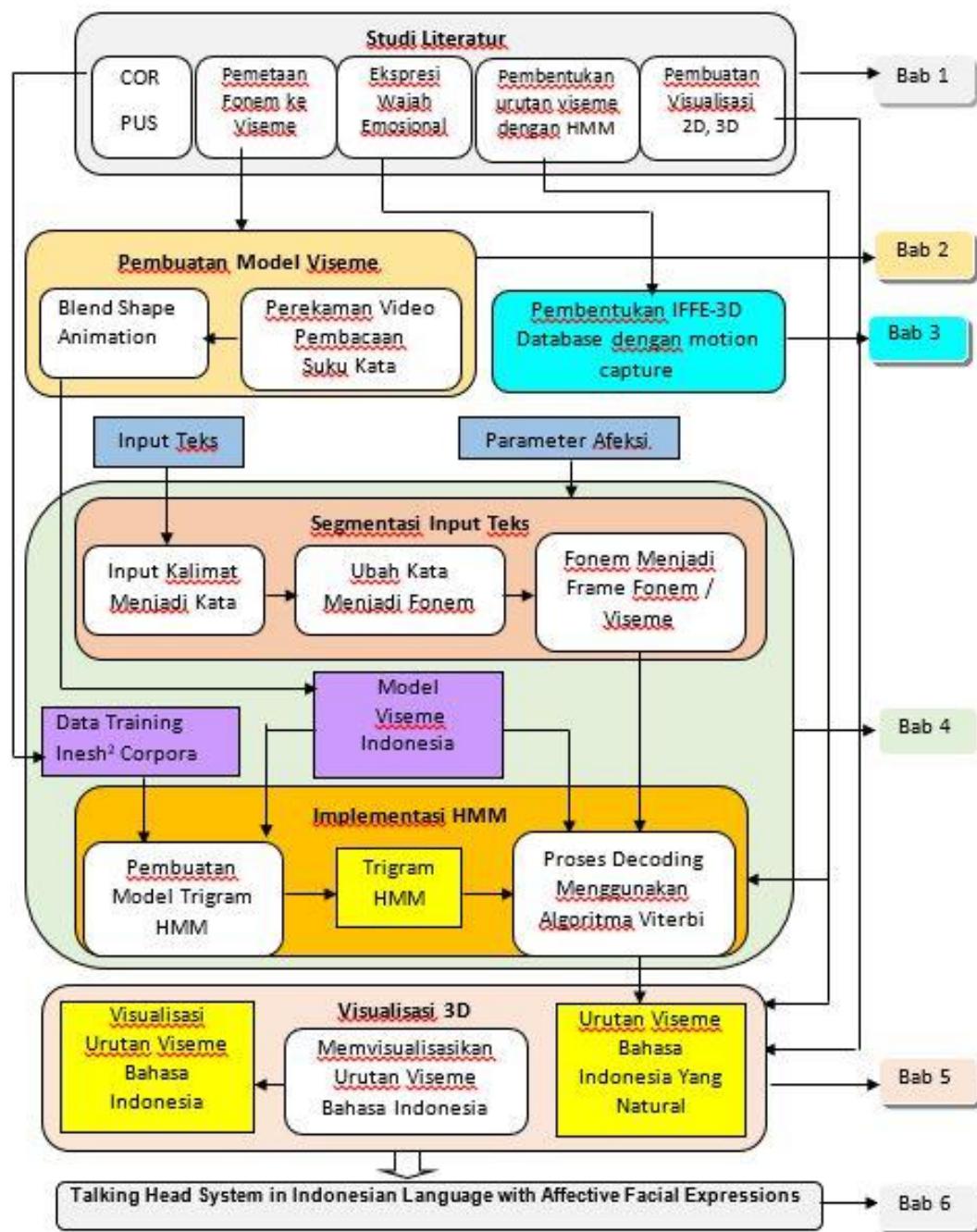
Ta hun	Pengarang dan Diterbitkan	Judul Penelitian	Pembentukan Model Viseme	Viseme Chain Generation using HMM	Pembentukan Model Ekspresi Wajah Berafeksi
			men-generate audio ekspresif.		untuk mengontrol pembangkitan sintesis.
2010	<i>Etienne B. Roesch, Lucas Tamarit, Lionel Reveret, Original Paper, Journal Nonverbal Behav, Springer Science+Business Media, LLC 2010, Published online: 19 November 2010. DOI 10.1007/s10919-010-0095-9</i>	FACSGen: A Tool to Synthesize Emotional Facial Expressions Through Systematic Manipulation of Facial Action Units	-	-	Mengembangkan FACSGen: alat temuan baru yang mengkreasikan stimuli wajah 3D sintetik: statik dan dinamik. FACSGen menyediakan penelitian dengan total kontrol facial action unit dan mengkorespondensikan informasi dalam wajah 3D sintetik.
2009	Kang Liu and Joern Ostermann, EURASIP Journal on Audio, Speech, and Music Processing Volume 2009, Article ID 174192, pp. 1-13	Optimization of an Image-Based Talking Head System		Membuat database bentuk mulut untuk digunakan sebagai	Memvisualisasikan Talking Head System berdasarkan gambar dengan menggunakan Bahasa China
2008	<i>Jacey-Lynn Minoi, S. Hassan Amin, Carlos Eduardo Thomaz, the 2008 2nd IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2008), September 29-October 1, 2008, pp. 1-6</i>	Synthesizing Realistic Expressions in 3D Face Data Sets	-	-	Membuat visualisasi ekspresi realistik sintesis dalam dataset wajah 3D visual features yang diekstrak secara otomatis, dari 3D database FERET.
2008	<i>Hao Tang, Yuxiao Hu, Yun Fu, the 2008 IEEE International Conference on Multimedia and Expo (ICME 2008), June 23-26, 2008, Hannover, Germany, pp. 1205-1208</i>	Real-Time Conversion from A Single 2D Face Image to A 3D Text-Driven Emotive Audio-Visual Avatar	-	Konversi real time dengan membentuk model wajah 3D yang dibangun berdasarkan gambar wajah 2D dengan menggunakan framework bentuk wajah 3D secara otomatis. Membentuk parameter animasi wajah, model wajah	Merekonstruksi 3D text-driven emotive audio avatar dari 2D gambar wajah frontal. Memvisualisasi animasi wajah yang dikombinasi dan disinkronisasi dengan membangkitkan ucapan emosional sintetik ke dalam teks pada sintesis ucapan netral.

Ta hun	Pengarang dan Diterbitkan	Judul Penelitian	Pembentukan Model Viseme	Viseme Chain Generation using HMM	Pembentukan Model Ekspresi Wajah Berafeksi
				dianimasi oleh channel viseme dan ekspresi menggunakan MPEG-4.	
2008	R. Cowie, E. Douglas-Cowie, K. Karpouzis, G. Caridakis, M. Wallace and S. Kollias, pdfs.semanticscholar.org, Department of Computer Science, University of Indianapolis, Athens Campus, 2008, pp. 1-39	Recognition of Emotional States in Natural Human-Computer Interaction	-	Membuat pengenalan dari state emosi dalam interaksi manusia komputer yang natural dengan menggunakan fitur ekstraksi FACS	Membuat aplikasi pengenalan state emosi secara otomatis
2008	<i>Meng Qingmei, Wu Weiguo, Zhong Yusheng, the 2008 8th IEEE-RAS International Conference on Humanoid Robots, December 1-3, 2008, Daejeon, Korea, pp. 603-608</i>	Research and Experiment of Lip Coordination with Speech for the Humanoid Head Robot-H&Frobot-III	Membangun 10 viseme bahasa Mandarin	Membuat modul talking bentuk mulut robot dengan menggunakan sistem ucapan visual	Membuat sistem ekspresi wajah robot dengan koordinasi bentuk mulut dan ucapan.
2007	<i>Dong-Mei Yu and Jian-An Fang, the 2007 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR 2007), Beijing, China, November 2-4, 2007, pp. 825-830</i>	Research on A Methodology to Model Speech Emotion	Mengembangkan metodologi pada model emosional yang terdiri dari ucapan sebagai fungsi linier dari fitur waktu dan ucapan.	Pada aplikasi yang dibuat tidak mendefinisikan kata-kata sederhana, tetapi menghitung dimensinya dan nilai diekspresikan dalam beberapa interval	-
2007	<i>Elif Bozkurt, Cigdem Eroglu Erdem, Engin Erzin, the 2007 IEEE International Conference</i>	Comparison of Phoneme and Viseme Based Acoustic Units for Speech Driven Realistic Lip Animation	Memetakan 16 klasifikasi fonem ke viseme	Mengevaluasi kinerja phone dan viseme berdasarkan unit akustik dengan dan tanpa informasi konteks untuk men-generate sinkronisasi bibir realistik dengan HMM.	-
2007	<i>Salah Werda, Walid Mahdi and Abdelmajid Ben Hamadou, International Journal of</i>	Lip Localization and Viseme Classification for	Membuat lokalisasi titik-titik fitur bibir secara otomatis dalam	-	Mengembangkan prototype Automatic Lip Feature Extraction (ALiFE).

Ta hun	Pengarang dan Diterbitkan	Judul Penelitian	Pembentukan Model Viseme	Viseme Chain Generation using HMM	Pembentukan Model Ekspresi Wajah Berafeksi
	Computing and Information Sciences, Vol. 5, No. 1, April 2007, On-Line, pp. 62-75	Visual Speech Recognition	wajah pembicara. Informasi visual yang telah diekstrak kemudian diklasifikasi dalam urutan viseme.		
2006	<i>Jianhua Tao, Yongguo Kang and Aijun Li</i> , IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, July 2006, pp. 1145-1154	Prosody Conversion from Neutral Speech to Emotional Speech	Memetakan fitur prosody dalam 3 derajat ekspresi berbeda	Mendeskripsikan eksperimen sebuah persepsi yang didesain untuk membuat klasifikasi suara emosional dengan derajat ekspresi yang berbeda.	-
2006	<i>Ming Xu and Ruimin Hu</i> , the 2006 IEEE First International Conference on Communications and Networking in China (ChinaCom 2006), October 25-27, 2006, pp. 1-5.	Mouth Shape Sequence Recognition Based on Speech Phoneme Recognition	Melakukan pengenalan terhadap barisan bentuk mulut berdasarkan pengenalan fonem ucapan yang efektif.	Membuat sistem framework dari pengenalan ucapan kontinyu dengan menggunakan training HMM berdasarkan tri-phone unit.	-
2006	<i>Nadia Mana and Fabio Pianesi</i> , the 8 th International Conference on Multimodal Interfaces (ICMI 2006), Banff, Alberta, Canada, November 2-4, 2006, pp. 380-387	HMM-based Synthesis of Emotional Facial Expressions during Speech in Synthetic Talking Heads	-	Menyajikan sistem berdasarkan HMM untuk sintesis ekspresi wajah emosional selama pengucapan.	Membuat Embodied Conversational Agent yang dapat mengkomunikasikan informasi dengan human like.
2005	Goranka Zorich, 2005 Master Thesis, Department of Telecommunications, Faculty of Electrical Engineering and Computing, pp. 1-87	Automatic Lip Synchronization by Speech Signal Analysis		Membuat sistem animasi wajah dengan menggunakan lip sync untuk kreasi dari kontes animasi	Menganalisa Signal Suara dengan melakukan sinkronisasi bibir secara otomatis
2005	Jianhua Tao, Tieniu Tan,	I. Affective Computing			Talking head system dalam Bahasa China dengan emosi yang

Ta hun	Pengarang dan Diterbitkan	Judul Penelitian	Pembentukan Model Viseme	Viseme Chain Generation using HMM	Pembentukan Model Ekspresi Wajah Berafeksi
	Proceedings of First International Conference, ACII 2005, Beijing, China, Oct 2005	and Intelligent Interaction			mentransformasikan talking head dari 3D ke 2D
2004	<i>Jianhua Tao, Le Xin, International Conference on Artificial Reality and Telexistence (ICAT) 2004.</i>	Mutual Features Controlling of Dynamic Visemes for Emotional Talking Head System	Membentuk 9 klasifikasi viseme China	Mendapatkan state emosi, dimana semua kata dibagi ke dalam content word (CW) dan emotion functional words (EFWs)	Membuat talking head system dinamik viseme dan ekspresi emosi yang diubah ke bentuk deretan action unit untuk proses detail dan finalisasi.
2004	<i>Jian-Qing Wang, Ka-Ho Wong, Pheng-Ann Heng, the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), May 17-21, 2004, Montreal, Quebec, Canada, pp. I-653-656.</i>	A Real Time Cantonese Text-To-Audiovisual Speech Synthesizer	Membentuk fonem dan 16 viseme bahasa Cantonese (Sylabel China)	-	Merancang dan mengembangkan synthesizer dari TTVS yang mengenerate ucapan sintetik alami.
2004	Jianhua Tao, Tieniu Tan, Proceedings of the 6th international conference on Multimodal Interfaces, (ICMI'14), pp. 273-280, USA, October 13-15, 2004	Emotional Chinese Talking Head System			Pembentukan Talking head system dalam Bahasa China dengan wajah emosi
2003	<i>Liang Dong, Say Wei Foo, and Yong Lian, the 2003 Fourth International Conference on Information, Communications and Signal Processing-Fourth IEEE Pacific-Rim Conference On Multimedia (ICICS-PCM 2003), Singapore, December 15-18, 2003, pp. 1394-1398.</i>	Modeling Continuous Visual Speech Using Boosted Viseme Models	Membuat model viseme yang menggunakan pendekatan dari 14 viseme bahasa Inggris	Membandingkan 2 metode Single HMM dan Boosted HMM	-

Ta hun	Pengarang dan Diterbitkan	Judul Penelitian	Pembentukan Model Viseme	Viseme Chain Generation using HMM	Pembentukan Model Ekspresi Wajah Berafeksi
2004	Jianhua Tao, Tieniu Tan, Proceedings of the 6th international conference on Multimodal Interfaces, (ICMI'14), pp. 273-280, USA, October 13-15, 2004	Emotional Chinese Talking Head System			Pembentukan Talking head system dalam Bahasa China dengan wajah emosi
2003	<i>The Duy Bui, Dirk Heylen</i> , IEEE the 16 th International Conference on Computer Animation and Social Agents (CASA 2003), IEEE Computer Society, pp. 1-8	Improvements on A Simple Muscle-based 3D Face for Realistic Facial Expressions	-	-	Mengembangkan kemampuan PC untuk merendering animasi wajah sehingga menghasilkan ekspresi wajah lebih realistik dengan memperbaiki kombinasi dari banyak action otot-otot.
2003	<i>Xie Lei, Jiang Dongmei, Ilse Ravyse</i> , the 4 th EURASIP Conference focused on Video/Image Processing and Multimedia Communication (EC-VIP-MC 2003), July 2-5, 2003, Zagreb, Croatia, pp. 649-654	Context Dependent Viseme Models for Voice Driven Animation	Membangun model viseme akustik yang bergantung pada vektor observasi akustik dari suara-suara ucapan yang berkorespondensi pada setiap 16 viseme.	Mengembangkan sistem berdasarkan HMM untuk vektor observasi akustik dari suara-suara ucapan yang berkorespondensi pada setiap 16 viseme.	-
2002	<i>Wang Zhiming, CAI Lianhong, AI Haizhou</i> , IEEE the 6 th International Conference on Signal Processing (ICSP 2002), August, 26-30, 2002, pp. 1015-1018.	A Dynamic Viseme Model for Personalizing A Talking Head	Membuat 19 model viseme dinamik bahasa China untuk sintesis ucapan visual	Membuat Facial Animation Parameters (FAPs) yang didefinisikan dalam MPEG-4 yang diestimasi sesuai deteksi titik fitur wajah.	Membuat talking head system yang tidak hanya pada head model tetapi juga pada talking manner-nya.
2001	<i>Fabio Lavagetto</i> , the IEEE International Conference on Image Processing (ICIP 2, 2001), Vol. 2, October 7-10, 2011, pp. 594-597	Emotional Representation and Animation of 3D facial Models: The Interface Approach	-	-	Membuat pendekatan Interface yang mengintegrasikan semua teknik analisis/ sintesis melalui percakapan dan ekspresi wajah



Gambar 1.4 Sistematika Penulisan Disertasi

Tabel 1.2 Topik Tahapan Penelitian Disertasi dan Publikasi

Tahap	Topik	Hasil/Target	Keterangan
-------	-------	--------------	------------

I	Pemetaan Fonem ke Viseme Indonesia berdasarkan Blend Shape Animation	Indonesian Viseme Model	IAENG International Journal of Computer Science, Vol. 42, No. 3, September 2015
II	Pengenalan Ekspresi Wajah Emosional: a. Berdasarkan Active Shape Model (ASM) dan Radial Basis Function Network (RBFN) b. Sintesis Wajah dengan Radial Basis Function Network	Partial Expression Parameter	Proceeding Seminar International 2012 IEEE CIMA, Tianjin, China, 2-4 July 2012, pp. 233-244
		Facial Animation Parameter	Proceeding Seminar Internasional ICSIIT 2012, Denpasar, 24-25 May 2012
III	Pemodelan urutan viseme alami dengan menggunakan HMM	Viseme Chain Generation	Jurnal Kursor, Scientific Journal on Information Technology, Vol. 8, No. 3, July 2016, pp. 215-222
IV	Talking Head System dalam Bahasa Indonesia dengan sintesis ekspresi wajah beremosi	Visualisasi 3D animasi wajah berekspresi	-
V	Penulisan Disertasi	Laporan Akhir (Buku Disertasi)	-

Sistematika pembahasan pada pembuatan laporan akhir dari buku Disertasi ini tersusun dalam enam bab, meliputi :

Bab I Pendahuluan

Pada bab satu ini berisi latar belakang, permasalahan, tujuan dan manfaat penelitian, kontribusi dan orisinalitas penelitian, dan roadmap serta tahapan penelitian.

Bab II Pemetaan Fonem ke Viseme Bahasa Indonesia Berdasarkan Blend Shape Animation

Pada bab dua ini berisi penjelasan pemilihan fonem dan viseme Indonesia, konstruksi model viseme Indonesia, konstruksi model 3D berdasarkan blend shape animation, hingga hasil eksperimen dari pemetaan fonem ke viseme Bahasa Indonesia, yang menghasilkan bentuk mulut model viseme Indonesia dan telah divalidasi berdasarkan survei.

Bab III Akuisisi Data untuk Ekspresi Wajah 3D Animasi

Pada bab tiga ini berisi tentang definisi emosi dan afeksi dari sebuah ekspresi wajah, pembuatan database *Indonesian Female Facial Expressions* (IFFE-3D Database), perekaman ekspresi wajah pada *facial motion capture*, pemilihan parameter afeksi yang sesuai.

Bab IV Hidden Markov Model Berdasarkan Model Viseme Indonesia untuk Ucapan Alami dengan Afeksi

Pada bab empat ini, berisi tentang arsitektur sistem secara keseluruhan dalam pembuatan talking head system, segmentasi input teks berdasarkan trigram HMM, serta penjelasan implementasi HMM yang terdiri dari Pemisahan Data Training berdasarkan *Inesh² Corpora* dan penerapan proses decoding dengan menggunakan Algoritma Viterbi dan HMM.

Bab V Visualisasi 3D Animasi Wajah Berafeksi

Pada bab lima ini, berisi tentang pembuatan visualisasi 3D Animasi wajah berafeksi dan pengujian pengucapan urutan viseme alami dari hasil eksperimen yang telah dilakukan.

Bab VI Penutup

Pada bab enam ini, merupakan bab penutup dari laporan akhir buku Disertasi yang terdiri dari kesimpulan hasil penelitian dan usulan penelitian selanjutnya.

BAB II

PEMETAAN FONEM KE VISEME BAHASA INDONESIA BERDASARKAN BLEND SHAPE ANIMATION

2.1 Karakteristik Bahasa Indonesia

Bahasa Indonesia adalah bahasa persatuan yang digunakan di Negara Kesatuan Republik Indonesia (James Neil Sneddon, 2004). Indonesia adalah bahasa resmi yang digunakan oleh hampir lebih dari 250 juta orang di 34 provinsi Republik Indonesia. Indonesia adalah negara terpadat keempat di dunia karena populasi yang besar, sehingga Bahasa Indonesia menjadi salah satu bahasa yang paling banyak digunakan di dunia (James Neil Sneddon, 2004). Sebagian besar orang Indonesia, selain berbicara menggunakan bahasa nasional, fasih juga dalam bahasa daerah sebagai bahasa ibu mereka, termasuk Bahasa Jawa, Madura, Sunda dan Bali, yang umum digunakan sehari-hari di rumah dan dalam masyarakat setempat.

Pendidikan formal dan hampir semua media nasional menggunakan bahasa Indonesia. Bahasa Indonesia standar dan formal digunakan dalam buku-buku, koran, televisi, dan siaran berita radio. Bahasa Indonesia standar terus dikembangkan. Meskipun di catatan terdahulu dalam prasasti Melayu bahwa basis suku kata ditulis dalam abjad Arab dan basis Bahasa Indonesia modern ditulis dalam abjad Romawi (George Quin, 1972).

Tabel 2.1 Alfabet dalam Bahasa Indonesia

No	Huruf	Ejaan	No	Huruf	Ejaan	No	Huruf	Ejaan
1	A a	ah	10	J j	jé	19	S s	es
2	B b	bé	11	K k	ka	20	T t	té
3	C c	ché	12	L l	el	21	U u	oo
4	D d	dé	13	M m	em	22	V v	fé
5	E e	é	14	N n	en	23	W w	wé
6	F f	ef	15	O o	oh	24	X x	iks
7	G g	gé	16	P p	pé	25	Y y	yé
8	H h	ha	17	Q q	ki	26	Z z	set
9	I i	ee	18	R r	air			

Bahasa Indonesia menggunakan 26 huruf seperti dalam alfabet bahasa Inggris / Belanda, seperti yang terlihat pada Tabel 2.1. Huruf “q”, “v”, “x”, dan “z”

digunakan oleh Bahasa Indonesia di dalam kata-kata serapan dari Eropa dan India, akan tetapi keempat huruf tersebut hampir tidak pernah digunakan di akhir kata bahasa Indonesia.

Bahasa Indonesia awalnya ditulis menggunakan sistem yang dikenal sebagai ejaan *van Ophuijsen*. Pada tahun 1947, ejaan itu berubah menjadi Ejaan Soewandi (nama Menteri Pendidikan saat itu Soewandi). Ejaan ini mengubah huruf yang sebelumnya dieja "oe" menjadi "u". Semua perubahan lain adalah bagian dari Ejaan Yang Disempurnakan (EYD), yang merupakan reformasi ejaan resmi dimandatkan pada tahun 1972, "tj" menjadi "c", "dj" menjadi "j", "j" menjadi "y", "ny" menjadi "ny", "sj" menjadi "sy", dan "ch" menjadi "kh". Beberapa ejaan lama (yang berasal dari ortografi Belanda) yang masih bertahan adalah nama orang, misalnya nama mantan presiden Indonesia masih ditulis Soekarno, Soeharto.

2.2 Deskripsi Fonem dalam Bahasa Indonesia

Sebelum membahas tentang pembentukan fonem dalam bahasa Indonesia secara lebih rinci, akan dijelaskan pengertian fonologi, fonetik, dan fonemik.

2.2.1 Fonologi, Fonetik, dan Fonemik

Dari beberapa sumber, pengertian fonologi (*phonology*) antara lain: Fonologi ialah bagian dari tata bahasa yang mempelajari bunyi-bunyi bahasa (Keraf, 1984). Fonologi ialah bidang dalam linguistik yang menyelidiki bunyi-bunyi bahasa menurut fungsinya (Kridalaksana, 1995). Fonologi ialah bidang linguistik yang mempelajari, menganalisis, dan membicarakan runutan bunyi-bunyi bahasa, yang secara etimologi terbentuk dari kata *fon* yaitu bunyi dan *logi* yaitu ilmu (Chaer, 1994). Dalam Kamus Besar Bahasa Indonesia (KBBI, 1997) dinyatakan bahwa fonologi adalah bidang dalam linguistik yang menyelidiki bunyi-bunyi bahasa menurut fungsinya. Sehingga secara garis besar, fonologi adalah merupakan sistem bunyi dalam bahasa Indonesia atau dapat juga dikatakan bahwa fonologi adalah ilmu tentang bunyi bahasa. Fonologi dalam tataran ilmu bahasa dibagi dua bagian, yaitu: fonetik dan fonemik.

Dari beberapa sumber, pengertian fonetik (*phonetic*) antara lain: Fonetik adalah cabang studi fonologi yang mempelajari bunyi bahasa tanpa memperhatikan apakah bunyi-bunyi tersebut mempunyai fungsi sebagai pembeda makna atau tidak

(Chaer, 1994). Fonetik adalah ilmu yang menyelidiki dan menganalisa bunyi-bunyi ujaran yang dipakai dalam tutur, serta mempelajari bagaimana menghasilkan bunyi-bunyi tersebut dengan alat ucapan manusia (Keraf, 1984). Fonetik adalah ilmu yang menyelidiki penghasilan, penyampaian, dan penerimaan bunyi bahasa; ilmu interdisipliner linguistik dengan fisika, anatomi, dan psikologi (Kridalaksana, 1995). Jadi, secara garis besar fonetik didefinisikan sebagai ilmu bahasa yang membahas bunyi-bunyi bahasa yang dihasilkan alat ucapan manusia, serta bagaimana bunyi itu dihasilkan.

Dari beberapa sumber, pengertian fonemik (*phonemic*) antara lain: Fonemik adalah cabang studi fonologi yang mempelajari bunyi bahasa dengan memperhatikan fungsi bunyi tersebut sebagai pembeda makna (Chaer, 1994). Fonemik adalah ilmu yang mempelajari bunyi ujaran dalam fungsinya sebagai pembeda arti (Keraf, 1984). Fonemik adalah penyelidikan mengenai sistem fonem suatu bahasa (Kridalaksana, 1995). Jadi, secara garis besar fonemik adalah ilmu bahasa yang membahas bunyi-bunyi bahasa yang berfungsi sebagai pembeda makna. Jika dalam fonetik mempelajari segala macam bunyi yang dapat dihasilkan oleh alat-alat ucapan serta bagaimana tiap-tiap bunyi itu dilaksanakan, maka dalam fonemik mempelajari dan menyelidiki kemungkinan-kemungkinan, bunyi-ujaran yang manakah yang dapat mempunyai fungsi untuk membedakan arti?

Bunyi /i/ yang terdapat pada kata-kata “intan”, “angin”, dan “batik” adalah tidak sama. Begitu juga bunyi /p/ pada kata dalam bahasa Inggris “pace”, “space”, dan “map”, juga tidak sama. Ketidaksamaan bunyi /i/ dan bunyi /p/ pada deretan kata-kata di atas itulah salah satu contoh obyek atau sasaran studi fonetik. Dalam kajiannya, fonetik akan berusaha mendeskripsikan perbedaan bunyi-bunyi itu serta menjelaskan sebab-sebabnya. Sebaliknya, perbedaan bunyi /p/ dan /b/ yang terdapat, misalnya, pada kata “paru” dan “baru” adalah menjadi contoh sasaran studi fonemik, sebab perbedaan bunyi /p/ dan /b/ itu menyebabkan berbedanya makna kata “paru” dan “baru” itu.

2.2.2 Pengertian Dasar Fonem

Sebelum ditemukan sejumlah fonem dalam bahasa Indonesia, terlebih dahulu dirumuskan pengertian tentang fonem. Fonem adalah unsur bahasa yang

terkecil dan dapat membedakan arti atau makna (Gleason, 1961). Berdasarkan definisi diatas maka setiap bunyi bahasa, apabila terbukti dapat membedakan arti dapat disebut fonem.

Setiap bunyi bahasa memiliki peluang yang sama untuk menjadi fonem. Namun, tidak semua bunyi bahasa pasti akan menjadi fonem. Bunyi itu harus diuji dengan beberapa pengujian penemuan fonem. Nama fonem, ciri-ciri fonem, dan watak fonem berasal dari bunyi bahasa. Adakalanya jumlah fonem sama dengan jumlah bunyi bahasa, tetapi sangat jarang terjadi. Pada umumnya fonem suatu bahasa lebih sedikit daripada jumlah bunyi suatu bahasa. Bunyi-bunyi yang secara fonetis mirip dan terdapat di dalam distribusi yang komplementer, harus dimasukkan ke dalam kelas-kelas bunyi (fonem) yang sama.

Untuk mengenal dan menentukan bunyi-bunyi bahasa yang bersifat fungsional (fonem), biasanya ditentukan melalui kontras pasangan minimal. Pasangan minimal ini adalah pasangan bentuk-bentuk bahasa yang terkecil dan bermakna pada sebuah bahasa atau kata tunggal yang secara ideal sama, kecuali satu bunyi berbeda. Contohnya : *dara* dan *tara* fonemnya /d/ dan /t/ atau *kalah* dan *galah* fonemnya /k/ dan /g/.

2.2.3 Pengucapan Artikulasi Fonem

Fonem adalah satuan bunyi bahasa terkecil yang bersifat fungsional, artinya satuan yang memiliki fungsi untuk membedakan makna. Fonem tidak dapat berdiri sendiri karena belum mengandung arti. Fonemisasi adalah usaha untuk menemukan bunyi-bunyi yang berfungsi dalam rangka pembedaan makna.

Dalam bahasa Indonesia bunyi /f/, /v/ dan /p/ pada dasarnya bukanlah tiga fonem yang berbeda. Kata *provinsi* apabila dilafalkan sebagai “propinsi”, “profinsi” atau “provinsi” tetap sama saja. Fonem tidak memiliki makna, tapi peranannya dalam bahasa sangat penting karena fonem dapat membedakan makna. Misalnya saja fonem /l/ dengan /r/. Jika kedua fonem tersebut berdiri sendiri, pastilah tidak akan menangkap makna. Akan tetapi lain halnya jika kedua fonem tersebut digabungkan dengan fonem lainnya seperti /m/, /a/, dan /h/, maka fonem /l/ dan /r/ bisa membentuk makna “marah” dan “malah”. Bagi orang Jepang kata marah dan malah mungkin mereka anggap sama karena dalam bahasa mereka tidak ada fonem

/l/. Contoh lain: “mari”, “lari”, “dari”, “tari”, “sari”, jika satu unsur diganti dengan unsur lain maka akan membawa akibat yang besar yakni perubahan arti.

2.3 Pemilihan Fonem Indonesia yang Digunakan

2.3.1 Fonem Vokal dan Konsonan

Sebuah Asosiasi Fonetik Internasional (*International Phonetic Association*) menghimpun huruf-huruf dari Abjad-abjad Fonetik Internasional (*International Phonetic Alphabet, IPA*) ke dalam dua kategori, yaitu: vokal dan konsonan (A. Anugraha, 2008). Dalam Bahasa Indonesia, abjad latin atau huruf terbagi menjadi 5 vokal dan 21 konsonan, sehingga total terdapat 26 jenis abjad, lihat Tabel 2.1.

Sebuah vokal adalah suara yang dihasilkan oleh aliran terbatas dari udara di pita suara. Fungsi vokal sangat penting dalam sebuah kata, karena hampir setiap kata memiliki setidaknya satu buah. Tanpa vokal dalam sebuah kata, maka menjadi tidak bermakna. Sedangkan konsonan adalah suara yang dihasilkan oleh aliran terbatas dari udara di pita suara. Konsonan disusun secara tunggal atau berpasangan dari bunyi tak bersuara dan bunyi bersuara.

Sebuah fonem, bisa terdiri dari satu atau dua abjad. Jadi antara jumlah fonem dan jumlah abjad dalam Bahasa Indonesia tidak sama. Hal ini disebabkan dalam bahasa Indonesia juga dipengaruhi oleh kata-kata serapan (*absorp*) dari bahasa daerah, seperti: Jawa, Sunda, dan Bali maupun kata-kata serapan dari bahasa asing, seperti bahasa Arab dan bahasa Inggris. Contoh fonem-fonem serapan dari Bahasa Arab, antara lain: /ch/ di “achmad”, /dh/ di “dhuafa”, /dl/ “ramadlan”, /dz/ di “muadzin”, /sh/ di “sholat”, /gh/ di “maghrib” dan fonem serapan dari Bahasa Inggris, antara lain: /th/ di “therapy”, /ts/ di “tsunami”, /ps/ di psychology, /ph/ di “photography”.

2.3.2 Himpunan Fonem Indonesia

Setelah menganalisis beberapa literatur dan mengamati frekuensi penggunaan kalimat bahasa Indonesia secara umum, fonem terbagi menjadi fonem vokal dan konsonan, baik berbentuk monoftong (huruf tunggal) dan diftong (huruf rangkap/ganda).

Ada 5 abjad yang mewakili vokal Indonesia (/a/, /e/, /i/, /o/, /u/), tetapi ada 10 suara yang berbeda terkait dengan kelima vokal tersebut, antara lain: /a/, /ə/, /e/, /ɛ/, /i/, /ɪ/, /o/, /ɔ/, /u/, /ʊ/), termasuk alofon. Alofon adalah suara yang dihasilkan oleh alat ucapan manusia, yang merupakan varian dari fonem yang sama (A. Anugraha, 2008). Lima vokal utama dalam Bahasa Indonesia diwakili oleh 5 fonem: /a/, /e/, /i/, /o/, dan /u/. Lima lainnya adalah alofon: "ə" (/ə/), "ɛ" (/ɛ/), "ɪ" (/ɪ/), "ɔ" (/ɔ/), dan "ʊ" (/ʊ/). Jumlah fonem vokal tunggal (monoftong) seluruhnya ada 10 buah. Selain 10 vokal tunggal yang telah disebutkan di atas, Bahasa Indonesia juga memiliki 3 vokal ganda (diftong), yaitu: /ai/, /au/, /oi/. Sehingga jumlah fonem vokal baik tunggal maupun ganda terdapat 13 buah.

Tidak semua fonem vokal dapat digunakan di awal, di tengah dan di akhir sebuah kata. Seperti fonem vokal diftong /ai/ di “pan-tai”, /au/ di “pu-lau”, dan /oi/ di “a-soi”, hanya bisa dipakai di akhir kata. Vokal diftong bila diletakkan di awal sebuah kata, maka dibaca sebagai vokal monoftong, misal: “aida” (a-i-da, bukan ai-da), “aura” (a-u-ra, bukan au-ra), “soimah” (so-i-mah, bukan soi-mah). Fonem vokal /ə/, /ɛ/, /ɪ/, /ɔ/, /ʊ/, juga tidak pernah digunakan di akhir kata dalam bahasa Indonesia, akan tetapi sering dipakai sebagai pengganti ucapan, seperti pantai (dibaca pants), dan pulau (dibaca pulɔ), sedangkan diftong oi sangat jarang dipakai dalam bahasa Indonesia.

Fonem konsonan terdiri dari 21 konsonan tunggal (/b/, /c/, /d/, /f/, /g/, /h/, /j/, /k/, /l/, /m/, /n/, /p/, /q/, /r/, /s/, /t/, /v/, /w/, /x/, /y/, /z/) dan 15 konsonan ganda (/ch/, /dh/, /dl/, /dz/, /gh/, /kh/, /ks/, /ph/, /ps/, /sh/, /sy/, /th/, /ts/, /ng/, /ny/). Fonem konsonan yang berfungsi sebagai akhiran yang tidak (jarang) terpakai dalam bahasa Indonesia terdiri dari 18 fonem konsonan, antara lain: /c/, /ch/, /dh/, /dl/, /dz/, /j/, /ph/, /ps/, /q/, /sh/, /sy/, /th/, /ts/, /v/, /w/, /x/, /ny/, /y/, /z/ dan 4 fonem vokal /ə/, /ɛ/, /ɪ/, dan /ʊ/. Kesepuluh akhiran fonem konsonan yang tidak (jarang) terpakai ini, seringkali cara membacanya diganti menjadi fonem lainnya seperti berikut ini:

1. Akhiran /c/, /j/ dan /th/ sering kali dibaca menjadi /t/.
2. Akhiran /ch/ sering kali dibaca menjadi /kh/.
3. Akhiran /dh/, /dl/ dan /dz/ sering kali dibaca menjadi /d/.
4. Akhiran /q/ sering kali dibaca menjadi /k/.

5. Akhiran /sh/, /sy/, /ts/ dan /z/ sering kali dibaca menjadi /s/.
6. Akhiran /ph/, /v/ sering kali dibaca menjadi /f/.
7. Akhiran /ʊ/, /w/ sering kali dibaca menjadi /u/.
8. Akhiran /x/ sering kali dibaca menjadi /ks/.
9. Akhiran /ɪ/, /y/ sering kali dibaca menjadi /i/
10. Akhiran /ə/, /ɛ/ dan /ny/, tidak pernah ada.

Tabel 2.2 Himpunan Fonem Bahasa Indonesia

No	Huruf		Fonem	Contoh Penggunaan Dalam Kata		
	V	K		Di Awal	Di Tengah	Di Akhir
1	a		/a/	Air	buaya	rahasia
2	ai		/ay/	-	-	santai
3	au		/aw/	-	-	pulau
4		b	/b/	bahasa	sabtu	adab
5		c	/c/	cantik	kaca	-
6		ch	/ch/	chairil	achmad	-
7		d	/d/	Dua	ada	abad
8		dh	/dh/	dhuafa	bagdhad	mudlaradh*
9		dl	/dl/	dluhur	ramadlan	faridl*
10		dz	/dz/	dzikir	muadzin	mahfudz
11	ə		/ə/	Emas	internasional	kode
12	e		/e/	Eja	faedah	sate
13	ɛ		/ɛ/	elektronik	internet	-
14		f	/f/	Fakir	tafsir	maaf
15		g	/g/	Guna	tiga	gubug
16		gh	/gh/	Ghaib	maghrib	mubaligh
17		h	/h/	halaman	saham	celah
18	i		/i/	Ikan	kaisar	dai
19	ɪ		/ɪ/	-	kering	-
20		j	/j/	Jalan	manja	bajaj*
21		k	/k/	Kami	paksa	katak
22		kh	/kh/	khaliq	akhir	tarikh
23		ks	/ks/	ksatria	ekstra	teks
24		l	/l/	Lain	kalau	halal
25		m	/m/	makan	kamu	salam
26		n	/n/	Nama	pintar	kawan
27		ng	/ŋ/	Ngilu	angkasa	siang
28		ny	/ň/	nyonya	senyap	-
29	o		/o/	Obat	saolin	bakso
30	ɔ		/ɔ/	Orang	borang	-
31	oi		/oy/	-	-	amboi
32		p	/p/	paman	sopan	gelap

No	Huruf		Fonem	Contoh Penggunaan Dalam Kata		
	V	K		Di Awal	Di Tengah	Di Akhir
33		Ph	/ph/	photocopy*	graphology*	graph*
34		Ps	/ps/	psikologi*	-	tips
35		Q	/q/	qur'an	Wuquf	baliq*
36		r	/r/	ramah	Hari	pudar
37		s	/s/	suka	Kasih	kasus
38		sh	/sh/	sholat	Nashab	-
39		sy	/š/	syukur	Termasyur	arsy*
40		t	/t/	tuhan	Hutan	kabut
41		th	/th/	therapy*	-	-
42		ts	/ts/	tsunami*	batsuni	hadits*
43	u		/u/	udang	Pusat	tahu
44	ʊ		/ʊ/	-	bertarung	-
45		v	/v/	vitamin	Lava	-
46		w	/w/	wanita	Hawa	-
47		x	/x/	xerox	-	xerox*
48		y	/y/	yang	payah	-
49		z	/z/	zaman	lezat	lafaz*
Total	13	36	49			

Keterangan : *) sangat jarang atau hampir tidak pernah digunakan di akhir kata

Pada Tabel 2.2 tersebut di atas, terlihat bahwa tidak semua fonem dapat digunakan di awal, di tengah dan di akhir sebuah kata. Seperti fonem vokal diftong /ai/ (pan-tai), /au/ (pu-lau), dan /oi/ (a-soi), hanya bisa dipakai di akhir kata. Vokal diftong bila diletakkan di awal sebuah kata, maka dibaca sebagai vokal monoftong, misal: aida (a-i-da, bukan ai-da), aura (a-u-ra, bukan au-ra), soimah (so-i-mah, bukan soi-mah). Fonem vokal /ɛ/, /ɪ/, /ʊ/, juga tidak pernah digunakan di akhir kata dalam bahasa Indonesia, akan tetapi sering dipakai sebagai pengganti ucapan, seperti pantai (dibaca pante), dan pulau (dibaca pulo), sedangkan diftong /oi/ sangat jarang dipakai dalam bahasa Indonesia, hanya ada beberapa kata saja, seperti: amboi, asoi, konvoi. Dengan demikian, jumlah fonem Indonesia yang digunakan dalam penelitian ini sejumlah 49 fonem, seperti yang disajikan pada Tabel 2.2.

2.4 Perbandingan antara Fonem Indonesia dan Fonem Bahasa Lainnya

Himpunan fonem Indonesia ditentukan berdasarkan Tata Bahasa Indonesia yang dijelaskan dalam Tata Bahasa Baku Bahasa Indonesia (Alwi, 2003). Enam vokal utama dalam bahasa Indonesia (Sakriani Sakti, 2008), direpresentasikan

sebagai enam fonem: /a/, /e/, /ə/, /i/, /o/, dan /u/. Keenam vokal utama tersebut, mirip dengan huruf hidup dalam bahasa Inggris, yaitu, /a/ (seperti “a” dalam “*father*”), /e/ (seperti “e” dalam “*bed*”), /ə/ (suara *schwa*, seperti “e” di “*learn*”), /i/ (seperti “ee” di “*meet*”), /o/ (seperti “o” dalam “*stop*”), /u/ (seperti “oo” di “*soon*”), dan tiga diftong, /ai/, /au/, dan /oi/.

Ada vokal dalam lafal Bahasa Indonesia yang mirip dengan huruf hidup dalam pengucapan bahasa lainnya. Misalnya, vokal /a/ dalam bahasa Jerman (seperti “a” dalam “*mann*”), vokal dalam bahasa Jepang, yaitu, /a/ (seperti “a” dalam “*aka*”), /e/ (seperti “e” dalam “*eki*”), /i/ (seperti “i” di “*ika*”), /o/ (seperti “o” dalam “*oto*”), /u/ (seperti “u” di “*ushi*”), vokal dalam bahasa Arab, yaitu, /a/ (seperti “a” dalam “*aba*”), /i/ (seperti “i” di “*ummi*”), /u/ (seperti “u” di “*fulan*”). Tidak ada /a:/, /i:/, dan /u:/ dalam bahasa Indonesia.

Menariknya, Bahasa Jepang asli tidak memiliki suara “r”, “f”, “th”, “v”. Dalam Bahasa Jepang tidak bisa membedakan antara “r” dan “l”. Bahasa Jepang tidak menggerak bibir secara berlebihan ketika berbicara, sebab saat berbicara banyak dilakukan oleh gerakan lidah. Kata bahasa Inggris “*weekend*” menjadi enam suku kata ketika diucapkan dalam bahasa Jepang – “*u-ii-ku-e-n-do*” (*oo-ee-koo-en-doh*). Ketika diucapkan dalam bahasa Indonesia menjadi kata dua suku kata – “*wik-en*” (week-end) (Alvin, 2002).

2.5 Viseme

2.5.1 Definisi Viseme

Viseme (*Visual-Phoneme*) diwakili oleh bentuk mulut. Viseme adalah unit ekivalen dalam domain visual yang memodelkan sistem pengenalan pengucapan secara audio-visual (C. Benoit, dkk., 1990). Viseme mempunyai banyak interpretasi dalam beberapa literatur dan ada yang tidak setuju pada cara mendefinisikan viseme-viseme tersebut. Dua cara mendefinisikan viseme secara praktis dan logis adalah sebagai berikut:

- Viseme dapat diasumsikan sebagai gerakan artikulasi, seperti menutup bibir bersama-sama, menggerakkan dagu, dan lain-lain.
- Viseme diturunkan dari sekelompok fonem yang memiliki penampilan visual sama.

Cara kedua adalah yang paling banyak digunakan (Potamianos, 2003; Saenko, 2004; Neti, 2000; Bozkurt, 2007). Dengan menggunakan pendekatan kedua, viseme dan fonem dikorelasikan melalui pemetaan fonem ke viseme.

Beberapa fonem, misal /b/, /m/, /p/, dapat berkorespondensi pada sebuah viseme tunggal, karena fonem ini tidak bisa secara visual dibedakan pengucapannya antara “ba”, “ma” dan “pa”, ketiganya dianggap memiliki realisasi berdasarkan bibir yang sama.

Pada tahun 1960, Woodward dan Barber mengusulkan gerakan artikulasi fonem dalam hirarki visual yang dihasilkan oleh seorang pembicara. Pada tahun 1968, Fischer memperkenalkan viseme untuk mewakili suatu gerakan bibir terkecil dan artikulasi yang sesuai dengan suara (T. Chen, 1998; Potamianos, dkk., 2003). Sampai saat ini, belum ada klasifikasi tunggal atau himpunan viseme standar yang ditetapkan untuk fonem pada bahasa tertentu (Potamianos, dkk., 2003).

2.5.2 Pemetaan Fonem ke Viseme

Setiap bunyi dari bahasa, jika terbukti dapat membedakan arti, dapat dianggap sebagai sebuah fonem. Fonem adalah unit terkecil dari bunyi yang menjadi dasar untuk membangun sebuah ucapan manusia. Pemetaan fonem ke viseme adalah penting untuk visual pengenalan ucapan dan sintesis dari kepala berbicara (*talking head*).

Fonem dan viseme memiliki korelasi yang tinggi (Luca, 2012), dan viseme dapat diturunkan dengan menggunakan pemetaan fonem ke viseme. Pemetaan ini harus menjadi sebuah pemetaan dari banyak ke satu (*many-to-one*), karena banyak fonem yang tidak dapat dibedakan dengan menggunakan isyarat visual.

Ada dua cara untuk membangun pemetaan fonem ke viseme:

1. *Linguistik*, kelas-kelas viseme didefinisikan melalui pengetahuan linguistik dan melalui intuisi dari kemungkinan munculnya fonem yang sama secara visual.

Pendefinisian kelas viseme melalui intuisi adalah kemampuan untuk mengarahkan pengetahuan atau kognisi suatu bentuk mulut dari satu fonem dengan fonem lainnya, apakah memiliki tingkat kemiripan atau bahkan bisa serupa satu sama lain ketika diamati secara visual, tanpa menggunakan pikiran

- rasional. Sehingga dari tampilan visual tersebut, banyak fonem bisa dipetakan menjadi satu viseme.
2. *Data Driven*, kelas-kelas viseme yang dibentuk dengan melakukan pengelompokan fonem berdasarkan fitur yang diekstrak dari aturan-aturan penalaran.

Metode data driven mempunyai beberapa keuntungan, antara lain: yang pertama, sejak sistem pengenalan viseme banyak menggunakan model data pelatihan statistik, maka secara otomatis kelas-kelas viseme yang terbentuk akan belajar dari data-data secara alamiah. Yang kedua, kelas-kelas viseme dapat menghitung variasi dan selisih kontekstual antar pembicara, tetapi berlaku hanya jika database yang tersedia besar (Saenko, 2004).

Dalam penelitian ini, mengusulkan pemetaan fonem ke viseme Indonesia dengan menggunakan pendekatan linguistik, kemudian memvalidasi hasil pemetaan melalui survei. Sampai saat ini belum ada penelitian yang membahas pemetaan fonem ke viseme Indonesia melalui pengetahuan linguistik, padahal pemetaan dalam bahasa lain seperti Arab (Pascal Damien, dkk., 2009), Jerman (Bianca, dkk., 2005), English (Salah Werda, dkk, 2007), telah diselidiki oleh banyak peneliti. Oleh karena itu, sebagai langkah awal dari pengembangan sistem ini, telah dicoba untuk membuat pemetaan fonem ke viseme Indonesia yang membantu untuk menciptakan gerakan bibir yang tepat dari kata-kata yang diucapkan. Penelitian ini membangun bentuk mulut model wajah animasi 3D dengan menggunakan *blend shape animation* berdasarkan rekaman video.

2.5.3 Pemetaan Fonem ke Viseme Bahasa Asing Lainnya

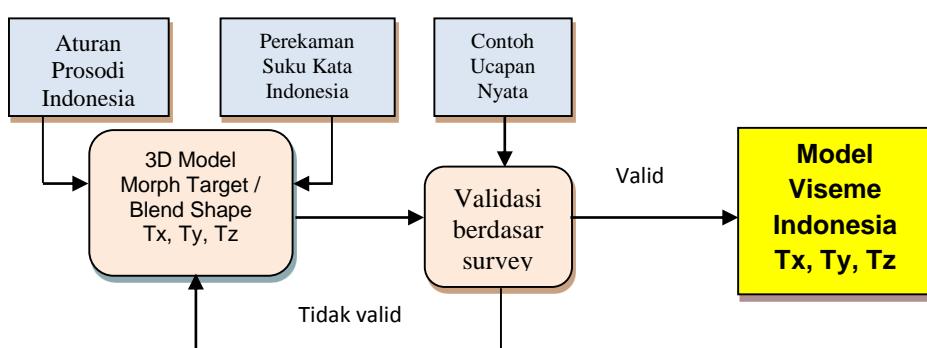
Semua jenis pemetaan fonem ke viseme, seperti yang dinyatakan oleh Luca Cappelletta, dkk., 2012; Potamianos, 2003; Saenko, 2004; Bianca, 2005; McGurk, 1976; dan Cohen, 1993, untuk viseme vokal dalam Bahasa Inggris sangat berbeda dari satu peneliti dengan peneliti lainnya. Jumlah viseme vokal bervariasi dari 4 sampai 7, tidak ada pola khusus dalam pemetaan. Beberapa kesamaan yang sangat jelas, terutama pemetaan antara Bianca, dkk., 2005 dan McGurk, dkk., 1976. Dalam dua pemetaan ini, terdapat 5 kelas viseme konsonan yang identik. Di semua pemetaan, kelas viseme konsonan menunjukkan pemisahan kelas yang sama.

Pada Bahasa Jepang terdiri dari beberapa ucapan yang sangat dasar yang dibangun untuk membuat lebih banyak ucapan. Bahasa Jepang memiliki 5 viseme dasar (Alvin Sebastian, 2002), hal ini sangat memungkinkan untuk membangun setiap bentuk mulut yang diperlukan dalam ucapan Bahasa Jepang. Sebab, dalam ucapan Bahasa Jepang, lidah lebih sering digunakan daripada bahasa asing lainnya dan membutuhkan sedikit gerakan bibir untuk membuat fonem dasar bahasa ini.

Bianca, dkk., 2005, Luca Cappelletta, dkk., 2012, Potamianos, 2003, dan K. Nielsen, 2004, menunjukkan bahwa sejumlah viseme dalam bahasa asing lainnya memiliki sejumlah fonem yang berbeda untuk setiap viseme, akan tetapi pola pemetaan fonem ke viseme masih tetap dari banyak ke satu. Contohnya adalah bahasa Arab yang terdiri dari 10 viseme dari 31 fonem, tidak termasuk diam, sedangkan bahasa Jerman terdiri dari 15 viseme dari 42 fonem.

2.6 Konstruksi Model Viseme Indonesia

Metode yang diusulkan dalam penelitian ini bertujuan untuk menghasilkan visualisasi animasi 3D yang dapat diterapkan pada bentuk mulut model viseme Indonesia. Konstruksi model viseme Indonesia merupakan yang pertama dalam penelitian pembuatan Sistem Bahasa Indonesia Kepala Berbicara (*Indonesian Language Talking Head System*) dan diharapkan menjadi referensi untuk penelitian di masa mendatang.



Gambar 2.1 Konstruksi dari Model Viseme Indonesia

Gambar 2.1 adalah konstruksi model viseme Indonesia. Proses konstruksi untuk membangun animasi bentuk mulut menggunakan blend shape adalah sebagai berikut:

1. Konstruksi dari Model 3D menggunakan dua data input: aturan prosodi Indonesia dan perekaman suku kata bahasa Indonesia yang dibaca oleh sejumlah pembicara atau pengisi suara.
2. Setelah pemahaman literatur dan perekaman pembicara telah dilakukan, Model 3D dikonstruksi dengan menggunakan program MakeHuman, yang dibaca oleh program Blender. Proses ini diterapkan pada bentuk mulut dari viseme Indonesia, termasuk viseme diam, dan diterjemahkan dalam parameter blend shape Tx, Ty, Tz sesuai model viseme yang diharapkan.
3. Proses selanjutnya adalah validasi berdasarkan survei menggunakan data input dari contoh nyata sejumlah pembicara. Jika model viseme Indonesia yang dibangun tidak valid, proses ini diulang. Jika sudah valid, maka dapat diperoleh output berupa model viseme Indonesia.

2.6.1 Data Input

Data input yang digunakan dalam konstruksi 3D Model terdiri dari dua data, yaitu: aturan prosodi Indonesia dan perekaman pembacaan suku kata Bahasa Indonesia. Pada sub-bab di bawah ini, akan dijelaskan kedua jenis data input tersebut.

2.6.1.1 Definisi dari Aturan Prosodi Indonesia

Secara tradisional, fonetik adalah studi tentang bunyi ucapan, yaitu bunyi yang dihasilkan oleh organ vokal manusia, dalam konteks bahasa lisan. Namun, pada saat ini telah diterima secara luas bahwa fonetik ditekankan tidak hanya sekedar untuk mempelajari sifat-sifat dari vokal dan konsonan yang membentuk kalimat yang akan diucapkan.

Prosodi terdiri dari semua sifat pengucapan yang tidak dapat dipahami secara langsung dari urutan linier segmen-semen ucapan. Fungsi linguistik dari prosodi adalah: untuk menandai domain dalam waktu (misalnya paragraf, kalimat, frase), untuk mengklasifikasikan informasi yang disajikan dalam domain (misal sebagai pernyataan, pertanyaan), dan untuk menyoroti aksentuasi dari konstituen tertentu dalam domain ini (Vincent, 1994).

Domain terkecil yang dapat ditandai adalah berupa suku kata. Dalam tulisan ini, suku kata yang berakhir dengan vokal disebut suku kata terbuka, dan ketika berakhir dengan konsonan, disebut suku kata tertutup. Suku kata terbuka diucapkan lebih lama dari suku kata tertutup. Prosodi harfiah berarti iringan. Hal ini menunjukkan bahwa struktur segmental mendefinisikan isi verbal dari kata-kata, sementara prosodi memberikan melodi dan irama. Prosodi sering dibagi menjadi dua kategori besar dari fenomena: struktur temporal dan melodi. Struktur temporal bahasa adalah seperangkat keteraturan yang menentukan durasi dari bunyi ucapan dan berhenti di ucapan-ucapan lisan. Struktur melodi didefinisikan sebagai seperangkat aturan yang mencirikan variasi *pitch* dari ucapan lisan dalam bahasa tertentu, tidak ada dua bahasa memiliki sifat melodi yang sama (Jialin Zhong, dkk., 1994).

2.6.1.2 Klasifikasi Suku Kata Bahasa Indonesia

Menurut Kamus Bahasa Indonesia, suku kata adalah struktur dari satu atau urutan fonem sebagai bagian dari sebuah kata. Setiap suku kata ditandai dengan vokal (termasuk diftong). Memisahkan kata menjadi suku kata menggunakan pedoman berikut:

1. Jika di tengah kata, ada dua vokal yang berurutan, pemisahan ditempatkan sebelum vokal kedua, seperti “ba-ik”;
2. Jika ada konsonan di antara dua vokal, pemisahan ditempatkan sebelum konsonan, seperti “i-bu”;
3. Untuk semua konsonan ganda (*compounds*) yang terdiri dari dua huruf, yaitu: “ng”, “ny”, “sy”, “kh” yang mewakili konsonan, pemisahan dilakukan sebelum atau setelah serangkaian huruf, seperti “ang-ka”, “me-nya-nyi”, “sya-rat”, “a-khir”;
4. Jika ada dua konsonan berturut-turut, pemisahan dilakukan antara konsonan, seperti “sam-ping”;
5. Jika ada tiga atau lebih konsonan, pemisahan itu ditempatkan antara konsonan pertama (termasuk “-ng”) dan konsonan kedua, seperti: “ben-trok”, “bang-krut”, “sin-kron”; dan

6. Jika kata itu memiliki awalan, sisipan, atau akhiran, kata dasar dipisahkan mengikuti aturan di atas, seperti: “mem-prak-tek-kan”, “ge-me-tar-an”.

Untuk klasifikasi suku kata bahasa Indonesia dengan vokal (V) dan konsonan (C), terdapat sebelas pola suku kata Bahasa Indonesia yang dapat dilihat pada Tabel 2.3.

Tabel 2.3 Pola Suku Kata Bahasa Indonesia

No	Pola Suku Kata	Contoh Pemenggalan Kata	
		Di depan	Di belakang
1	V	<i>i-bu</i>	<i>de-a</i>
2	VC	<i>an-da</i>	<i>ma-in</i>
3	CV	<i>sa-ya</i>	<i>fe-bi</i>
4	CVC	<i>pin-tu</i>	<i>ru-sak</i>
5	CCV	<i>dra-ma</i>	<i>in-dus-tri</i>
6	CCVC	<i>trak-tor</i>	<i>e-lek-trik</i>
7	VCC	<i>eks-por</i>	<i>ons</i>
8	CVCC	<i>laks-mi-ta</i>	<i>kon-teks</i>
9	CCVCC	<i>trans-fer</i>	<i>kom-pleks</i>
10	CCCV	<i>stra-tegi</i>	<i>de-mon-stra-si</i>
11	CCCVC	<i>struk-tur</i>	<i>kon-struk-si</i>

2.6.1.3 Konsep dari Pembacaan Suku Kata Bahasa Indonesia

Rekaman pembacaan suku kata Bahasa Indonesia dilakukan oleh 10 pembaca yang terdiri dari 5 mahasiswa laki-laki dan 5 mahasiswa perempuan berkewarganegaraan Indonesia. 10 mahasiswa tersebut mengucapkan suku kata-suku kata Bahasa Indonesia yang dikelompokkan dalam pola V, CV, CCV, VC, dan VCC. Pola suku kata ini dikelompokkan menjadi tiga tipe, yang bertujuan untuk menganalisis pengaruh bentuk mulut, ketika vokal ditempatkan di depan dan konsonan ditempatkan di belakang, dan sebaliknya. Konsonan tunggal dan ganda dianggap memiliki tipe yang sama, oleh karena itu dalam penelitian ini hanya menunjukkan fonem yang diambil dari himpunan fonem Indonesia. Ketiga tipe yang digunakan dalam perekaman antara lain:

1. Tipe 1: diam V diam
2. Tipe 2: diam VC dan VCC diam
3. Tipe 3: diam CV dan CCV diam

Pada tipe pertama, vokal ditempatkan sebelum dan sesudah diam (*silent*). Untuk tipe kedua, vokal ditempatkan sebelum konsonan. Sedangkan untuk tipe

ketiga, vokal ditempatkan setelah konsonan. Semua suku kata dibaca oleh 10 mahasiswa, dengan ketentuan sebagai berikut:

1. Tipe 1: membaca fonem yang terdiri dari 13 vokal (termasuk diftong)
2. Tipe 2: membaca fonem yang terdiri dari 10 vokal (tidak termasuk diftong) yang disambung dengan 36 konsonan (termasuk konsonan ganda), sehingga banyaknya suku kata yang dibaca oleh setiap mahasiswa pada tipe kedua ini adalah 360 suku kata (10 vokal x 36 konsonan). Pada tipe kedua ini, hanya 10 jenis vokal yang digunakan karena diftong Indonesia tidak pernah muncul di depan suku kata, tetapi selalu di akhir kata.
3. Tipe 3: membaca fonem yang terdiri dari semua suku kata yang mengandung 36 konsonan (termasuk konsonan ganda) yang disambung dengan 13 vokal, sehingga banyaknya suku kata yang dibaca oleh setiap mahasiswa pada tipe kedua ini adalah 468 suku kata (36 konsonan x 13 vokal).

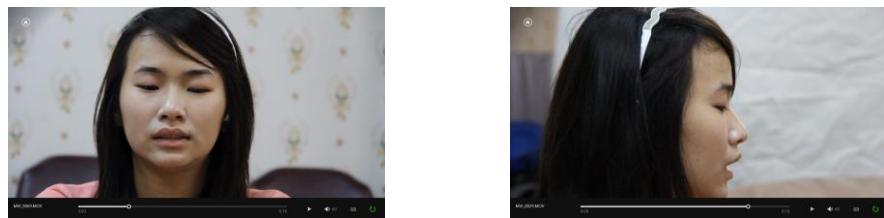
Jadi banyaknya rekaman pembacaan suku kata dari setiap mahasiswa dengan tiga tipe ini adalah sebanyak 841 suku kata, yang diperoleh dari $13 + 360 + 468$ suku kata yang dibaca. Urutan pembacaan suku kata dimulai dan ditutup dengan diam.

2.6.2 Perekaman dari Pembacaan Suku Kata Bahasa Indonesia

2.6.2.1 Perekaman Video

Rekaman video berlangsung di laboratorium komputer STTS, dengan menggunakan dua buah kamera digital Canon Type EOS 600D model SLR yang memiliki ukuran layar 3.0 inci dengan resolusi tinggi 18.0 Megapiksel, dalam lingkungan yang terkendali. 10 pembicara duduk tegak berjarak sekitar 1 meter di depan sebuah kamera dan menghadap kamera video yang dipasang pada tripod. Sedangkan kamera video satu lagi dihadapkan di samping model, lihat Gambar 2.2. Posisi vertikal kamera disesuaikan dengan tingkat wajah pembicara. Perekaman ini dilakukan pada pagi hari, jam 08.00-11.00 WIB untuk memberikan pencahayaan optimal, kurang lebih 300-700 lux dan meminimalkan bayangan dari mulut, agar representasi bentuk mulut yang ditangkap kamera tidak terganggu. Tujuannya untuk membantu menentukan bentuk mulut yang mewakili sebuah viseme dari

model wajah animasi yang diharapkan menyerupai model wajah nyata manusia, sehingga diharapkan hasil pemetaan fonem ke viseme menjadi lebih akurat.



Gambar 2.2 Posisi Duduk Model dari Arah Depan dan Samping Kamera

Penampilan pada saat rekaman video dari pembacaan suku kata Bahasa Indonesia pada 10 mahasiswa ditunjukkan pada Gambar 2.3. Pembacaan suku kata diambil dalam 13 adegan untuk tipe pertama dan tipe ketiga, sedangkan tipe kedua diambil dalam 10 adegan. Setiap perekaman akan berhenti jika pembicara telah membaca semua konsonan (tunggal dan ganda) dengan vokal yang sama. Dengan demikian, masing-masing pembicara membacakan 841 suku kata yang berbeda dari semua tipe dalam 36 adegan untuk setiap tipe. Fase artikulasi dari fonem ini dibangun dari satu mulut yang terbuka dan satu mulut yang tertutup. Tujuannya adalah untuk mendapatkan bentuk mulut yang sempurna.



Gambar 2.3 Posisi Pembacaan Suku Kata Bahasa Indonesia

2.6.2.2 Pemilihan Parameter Geometrik dari *Blend Shape Animation*

Blend Shape Animation adalah metode pilihan untuk frame penentu (*keyframe*) dari animasi wajah, yaitu sebuah himpunan dari *blend shape* (ekspresi wajah penentu) yang digunakan untuk mendefinisikan ruang dari ekspresi wajah

secara linier (Pushkar Joshi, 2003). Namun, *blend shape* perlu disegmentasi menjadi daerah yang lebih kecil. Dalam penelitian ini mengusulkan segmentasi yang termotivasi secara fisik yang mempelajari kontrol dan parameter secara langsung dari himpunan *blend shape*. Di dalam penelitian ini juga menyediakan sebuah algoritma *rendering* untuk meningkatkan aliran realistik visual dari model *blend shape* (Elif Bozkurt, dkk., 2007).

Wajah manusia selalu memegang peran tertentu bagi komunitas komputer grafik. Animasi wajah membutuhkan sebuah pembentukan kembali model dari wajah untuk mengekspresikan berbagai konfigurasi wajah terkait dengan ucapan atau emosi. Sebuah model *blend shape* secara langsung menganggap setiap ekspresi wajah sebagai sebuah kombinasi linier dari beberapa ekspresi wajah yang dipilih. Dengan bobot kombinasi linier yang bervariasi, seluruh ekspresi wajah dapat dibuat dengan sedikit perhitungan komputasi.

Untuk menunjukkan ekspresi yang sangat rinci, animator digital harus membuat basis data *blend shape* yang besar. Dengan memanipulasi sebuah area yang lebih kecil, pengguna dijamin bahwa modifikasi akan berdampak hanya pada bagian tertentu dari wajah, misalnya mulut. Segmentasi seharusnya mencerminkan keistimewaan dari wajah yang dimodelkan, menyediakan detil pengeditan dan tingkat yang berbeda. Sebuah contoh purwa rupa (*prototype*) merupakan segmentasi dari sebuah wajah ke dalam daerah atas dan daerah yang lebih rendah, yaitu: daerah atas digunakan untuk mengekspresikan emosi, sementara daerah yang lebih rendah mengekspresikan ucapan. Model *blend shape* telah membantu komunitas visi komputer dalam menganalisis gambar wajah dan video (Pushkar Joshi, 2003).

Bibir adalah bagian dari sebuah mulut. Bibir bersifat sangat elastis, karena dapat bergerak menuju arah yang berbeda dan membangun bentuk kerucut pada saat mengucapkan sebuah fonem. Bibir sangat ekspresif dalam pengucapan artikulasi, sehingga pemodelan bibir dan gerakan bibir adalah syarat utama untuk sistem animasi wajah berkualitas tinggi (K. Nielsen, 2004; C. Benoit, dkk., 1990). T. Chen, dkk., 1998, memisahkan bibir dari model wajah dan model khusus bibir untuk meningkatkan animasi wajah yang realistik.

Tabel 2.4 Titik-titik Kontrol (*Control Points, CP*)

Model Wajah 3D dengan 19 CP	CP	CP di Daerah Mulut	CP	CP di Daerah Wajah
CP-1	P_mulut_kanan	CP-10	P_alis_kanan	
CP-2	P_mulut_kiri	CP-11	P_alis_kiri	
CP-3	P_mulut_tengah	CP-12	P_antar_dua_alis	
CP-4	P_bibir_atas_kanan	CP-13	P_pelupuk_mata_atas_kanan	
CP-5	P_bibir_atas_kiri	CP-14	P_pelupuk_mata_atas_kiri	
CP-6	P_bibir_atas_tengah	CP-15	P_pelupuk_mata_bawah_kanan	
CP-7	P_bibir_bawah_kanan	CP-16	P_pelupuk_mata_bawah_kiri	
CP-8	P_bibir_bawah_kiri	CP-17	P_pipi_kanan	
CP-9	P_bibir_bawah_tengah	CP-18	P_pipi_kiri	
		CP-19	P_hidung	

Model bibir dikendalikan secara geometris, secara eksternal dan secara internal, pada 9 titik kontrol (*control points, CP*), seperti yang terlihat pada Tabel 2.4. Geometri eksternal meliputi semua daerah bibir merah. Geometri internal meliputi selaput lendir di dalam bibir. Geometri ini sangat penting untuk menjaga realistik ketika mulut dibuka. Titik-titik kontrol dari model wajah animasi 3D disesuaikan dengan program MakeHuman, yang dibagi menjadi 10 CP di sekitar wajah dan 9 CP di sekitar bibir. Posisi dan nama-nama istilah yang digunakan pada titik-titik kontrol model wajah 3D, disajikan pada Tabel 2.4.

2.7 Konstruksi Model 3D

2.7.1 Blend Shape Model Wajah

Sebuah model wajah *blend shape* didefinisikan sebagai sebuah kombinasi linier konveks dari n vektor basis dan setiap vektornya merupakan salah satu *blend shape* (Pushkar Joshi, 2003). Kombinasi linier konveks adalah kombinasi linier yang titik-titiknya merupakan semua koefisien non negative dengan jumlah 1. Setiap *blend shape* merupakan model wajah yang mencakup geometri dan tekstur. Semua *mesh blend shape* pada model yang diberikan membagi topologi yang sama. *Mesh* adalah ukuran dari jumlah lubang suatu jaring atau kasa pada luasan 1 inchi persegi jaring/kasa, yang bisa dilalui oleh suatu material, misal *mesh* 4 memiliki arti terdapat 4 lubang pada bidang jaring/kasa seluas 1 inchi, demikian seterusnya. Tekstur pada titik tertentu dari model *blend shape* mirip dengan kombinasi linier

dari tekstur *blend shape* dengan bobot *blend* sama seperti yang digunakan pada geometri.

Koordinat sebuah verteks V dari model blend shape, dapat ditulis sebagai persamaan (2.1):

$$V = \sum_{i=1}^n \alpha_i V_i \quad (2.1)$$

di mana skalar α_i adalah bobot proses *blend* ke- i , V_i adalah lokasi dari verteks dalam *blend shape* i , dan n jumlah *blend shape*. Bobot-bobot ini harus memenuhi kendala konveks dan disamakan dengan 1 untuk invarian rotasi dan translasi, seperti yang terlihat pada Persamaan (2.2):

$$\alpha_i \geq 0, \text{ untuk semua } i \text{ dan } \sum_{i=1}^n \alpha_i = 1 \quad (2.2)$$

Transformasi ini diterapkan pada sekelompok titik yang membangun sebuah objek. Jika objek berubah, lokasi titik-titik juga akan berubah dari lokasi awal ke lokasi baru. Transformasi verteks 3D pada dasarnya adalah sama dengan transformasi verteks pada objek 2D, yang terdiri dari sumbu x (panjang) dan sumbu y (lebar). Dalam transformasi 3D, ditambahkan sumbu z, yang memberikan kesan mendalam di mata manusia (T_x, T_y, T_z). Kesan yang mendalam ini dapat dianggap sebagai jarak dari mata terhadap sudut pandang objek.

2.7.2 Transformasi Verteks

Transformasi verteks diartikan sebagai suatu metode yang dapat digunakan untuk memanipulasi lokasi dari sebuah titik. Apabila transformasi dikenakan terhadap sekumpulan titik yang membentuk sebuah benda (obyek), maka obyek tersebut akan mengalami perubahan. Perubahan dalam hal ini adalah perubahan dari lokasi awal suatu obyek menuju lokasi yang baru dari obyek tersebut. Transformasi verteks 3D pada dasarnya tidak jauh berbeda dengan transformasi verteks pada obyek 2D, yang terdiri dari sumbu x (panjang) dan sumbu y (lebar). Pada transformasi 3D menggunakan 1 buah sumbu tambahan, yaitu sumbu z, yang berfungsi untuk menampilkan kesan kedalaman (*depth*) yang akan dilihat oleh mata

manusia. Kedalaman ini dapat diartikan sebagai jarak pandang antar mata (*viewer*) dengan obyek yang dilihat.

2.7.2.1 Matriks Transformasi

Transformasi verteks 3D direpresentasikan sebagai sebuah matriks berukuran 4x4. Matriks ini membentuk operasi dasar dari pergeseran (*translation*), rotasi (*rotation*), dan penskalaan (*scaling*). Transformasi 3D merupakan operasi yang menyebabkan perubahan objek tiga dimensi. Perubahan ini berlaku dalam arah sumbu x, y, z menjadi titik (x', y', z'). Pada transformasi 3D, diperlukan kemampuan untuk merotasi sebuah titik (x, y, z) di dalam penambahan penskalaan dan pergeserannya, dengan cara membuat transformasi verteks matriks berukuran 4x4 dan matriks baris berukuran 1x4, yang disajikan dalam persamaan 2.3 berikut:

$$\begin{bmatrix} x' & y' & z' & 1 \end{bmatrix} = \begin{bmatrix} x & y & z & 1 \end{bmatrix} \cdot \begin{bmatrix} m_{11} & m_{12} & m_{13} & 0 \\ m_{21} & m_{22} & m_{23} & 0 \\ m_{31} & m_{32} & m_{33} & 0 \\ b_1 & b_2 & b_3 & 1 \end{bmatrix} \quad (2.3)$$

Perubahan pada matriks transformasi 3D dapat berupa:

- Translasi, T(x, y, z), berfungsi untuk menggeser obyek dari suatu posisi awal menuju posisi baru sesuai dengan keinginan pengguna, seperti persamaan 2.4 berikut:

$$T(x, y, z) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x & y & z & 1 \end{bmatrix} \quad (2.4)$$

- Skala, S(s_x, s_y, s_z), berfungsi untuk memperbesar atau memperkecil obyek sesuai dengan ukuran yang diinginkan oleh pengguna, seperti persamaan 2.5 berikut:

$$S(s_x, s_y, s_z) = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.5)$$

- Rotasi, R(θ), berfungsi untuk memutar obyek menggunakan sumbu koordinat sebagai pusat perputaran. Ada 3 macam. yaitu :

- a. Rotasi sumbu x terhadap θ , menggunakan matriks seperti persamaan 2.6 berikut:

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.6)$$

- b. Rotasi sumbu y terhadap θ , menggunakan matriks seperti persamaan 2.7 berikut:

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.7)$$

- c. Rotasi sumbu z terhadap θ , menggunakan matriks seperti persamaan 2.8 berikut:

$$R_z(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.8)$$

2.7.2.2 Koordinat

Koordinat dapat diartikan sebagai tata keseimbangan yang membantu di dalam menentukan suatu kondisi dengan nilai dan batas dalam konteks geometri. Koordinat diciptakan dengan tujuan untuk membantu manusia agar dapat dengan lebih mudah mengenal, menguasai serta mengendalikannya. Pada saat menggambar sebuah koordinat, mungkin akan timbul sebuah pertanyaan, dimana letak koordinat ini berada, sebab koordinat satu dengan yang lain hampir sama. Untuk itu perlu diketahui dimana letak koordinat yang sedang digambar. Hal ini untuk menghindari kerancuan yang mungkin terjadi.

Pada komputer grafik, ada 3 macam sistem koordinat yang harus diperhatikan, yaitu: Koordinat Nyata (*World Coordinate*), Koordinat Sistem (*System Coordinate*) dan Koordinat Layar (*Screen Coordinate*). Berikut ini adalah penjelasan untuk masing-masing koordinat:

1. Koordinat Nyata (*World Coordinate*)

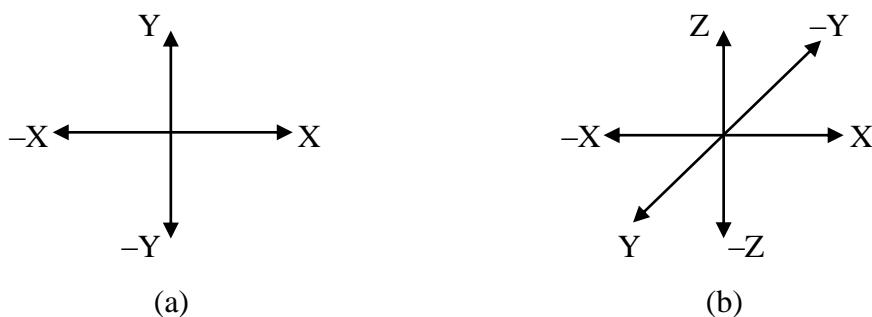
Yang dimaksud dengan koordinat nyata adalah koordinat yang menyatakan lokasi benda atau obyek itu di “dunia”. Dunia yang dimaksud adalah dimana benda itu berada. Koordinat nyata mempunyai batas $-\infty$ (minus tak hingga) sampai $+\infty$ (plus tak hingga) untuk sumbu x dan sumbu y.

2. Koordinat Sistem (*System Coordinate*)

Yang dimaksud dengan koordinat sistem adalah koordinat yang biasanya digunakan untuk perhitungan data yang hasilnya perlu ditampilkan kembali pada layar maupun diatas kertas. Koordinat ini lebih banyak digunakan di bidang geometri. Koordinat sistem sendiri terbagi menjadi dua macam, yaitu:

a. Sistem Koordinat Cartessius

Sistem koordinat Cartessius menggunakan pasangan sumbu x dan sumbu y untuk menyatakan lokasi benda di bidang 2D. Sedangkan untuk benda di bidang 3D menggunakan pasangan sumbu x, y, z.



Gambar 2.4 Sistem Koordinat Cartesius (a) 2D dan (b) 3D

b. Sistem Koordinat Polar

Sistem koordinat polar menggunakan sudut terhadap garis horisontal (θ) dan jarak dari titik pusat (R) untuk menunjukkan lokasi sebuah benda.

3. Koordinat Layar

Koordinat Layar atau (*screen coordinate*) adalah koordinat yang dipakai untuk mengatur penampilan suatu obyek pada layar, baik itu layar komputer ataupun layar lainnya.

Sebuah benda yang berada di ruang 3D dapat disusun dengan menghubungkan sebuah titik dengan titik-titik lainnya. Kumpulan titik yang saling terhubung itu disebut *wire frame* (kerangka). Sedangkan benda yang disusun dari

kumpulan titik dan permukaan di sebut dengan *mesh*. Yang paling menentukan pada pembuatan benda yang berada pada ruang tiga dimensi adalah definisi sumbu koordinat yang digunakan untuk mengidentifikasikan benda tersebut.

2.7.3 Pembentukan Animasi Mulut Menggunakan Blend Shape Berdasarkan Perekaman Video

Langkah-langkah dalam membentuk animasi mulut menggunakan blend shape yang didasarkan pada perekaman video adalah sebagai berikut:

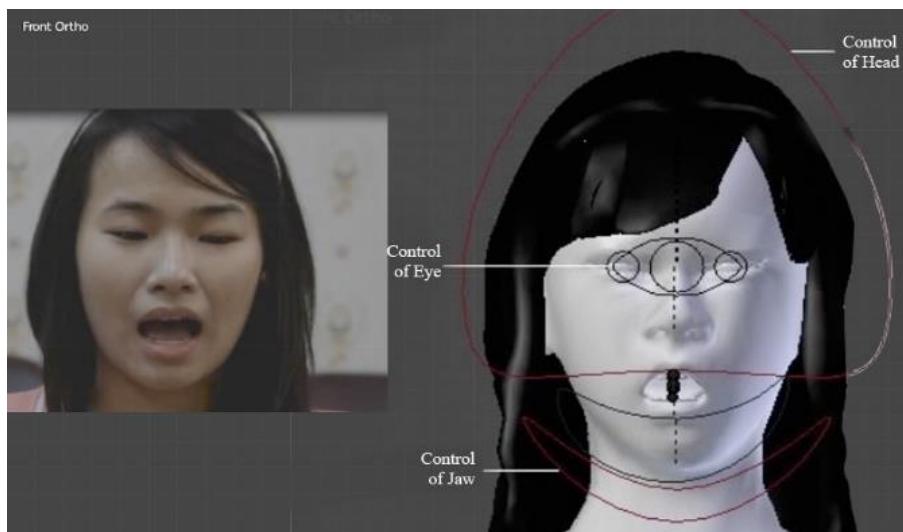
1. Fitur-fitur pada Gambar 2.2 yang berasal dari rekaman video digunakan sebagai acuan untuk bentuk mulut alami (*natural*) dari artikulasi fonem-fonem.
2. Model mesh 3D dari bentuk mulut netral (diam) akan digunakan sebagai awal dari blend shape.
3. Membuat bentuk mulut untuk setiap fonem dengan melakukan deformasi verteks-verteks, secara manual atau dengan menurunkan deformasi dari perangkat *motion capture* wajah. *Deformasi* adalah perubahan bentuk, posisi, dan dimensi dari suatu bentuk alamiah menjadi bentuk baru. Input dari proses ini adalah bentuk netral dari langkah 2. Deformasi manual seharusnya menggunakan rekaman video dari langkah 1 sebagai referensi.
4. Bentuk mulut dari langkah 3 dapat dikombinasikan dengan proporsi tertentu untuk membuat kombinasi bentuk.
5. Untuk menggabungkan bentuk pada langkah 4, titik-titik kontrol seperti pada Tabel 2.4 dapat dimanfaatkan untuk kenyamanan. Titik-titik kontrol dapat ditarik (*drag*) untuk menyesuaikan bentuk mulut.

Pada langkah 3 didapatkan variasi bentuk mulut yang akan digunakan sebagai *morph target* untuk setiap fonem. Dari bentuk-bentuk ini, bentuk lainnya dapat dibangun dengan menggabungkan verteks-verteks dengan proporsi tertentu (langkah 4 sampai 5).

Animasi *morph target* atau animasi per verteks merupakan sebuah metode dari 3D animasi komputer yang digunakan sebagai pengganti atau alternatif dari teknik animasi sekeletal. Animasi *morph target* tersimpan sebagai sebuah deretan

posisi verteks. Pada setiap keyframe, animasi verteks-verteks tersebut bergerak sesuai dengan posisi yang berbeda sesuai target perubahan pergerakannya. Animasi *morph* ini digunakan misalnya untuk membuat ekspresi wajah.

Animasi skeletal adalah sebuah teknik dalam animasi komputer, khususnya animasi yang melibatkan struktur pergerakan susunan tulang, yang dibagi menjadi 2 bagian, yaitu: representasi permukaan yang digunakan untuk menggambarkan karakter (biasa disebut skin/kulit/muka) dan susunan tulang digunakan hanya untuk meng-animasikan. Teknik ini digunakan untuk membuat atau merancang struktur pertulangan, dengan setiap unsur tulang mempunyai 3 dimensi transformasi yaitu posisi, skala dan orientasi dan sebuah opsi atau pilihan untuk tulang yang berhubungan dengannya. Aplikasi teknik ini sering digunakan untuk diaplikasikan pada obyek-obyek mekanik, misalnya untuk membuat animasi robot.



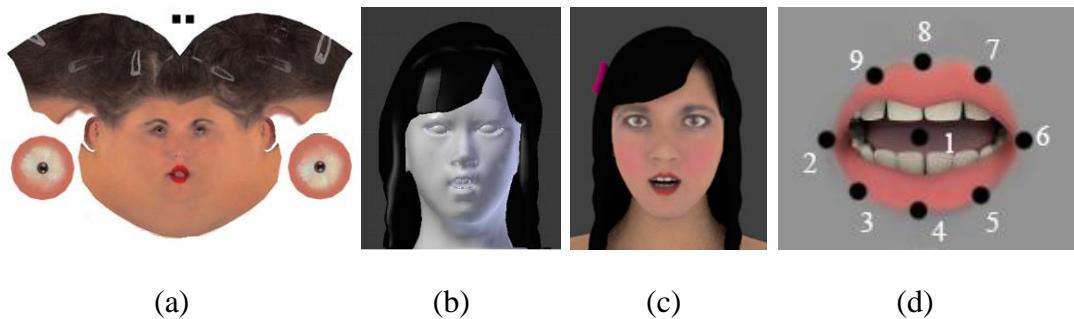
Gambar 2.5 Bentuk Standar Titik Kontrol yang dapat Digunakan Animator

Bentuk standar untuk titik-titik kontrol dapat digunakan sebagai antarmuka (*interface*) yang lebih ramah untuk animator, misalnya titik kontrol untuk rahang yang lebih rendah dapat divisualisasikan sebagai busur *jawlike* yang ditunjukkan seperti pada Gambar 2.5.

Secara matematis, hasil dari langkah 3 adalah himpunan dari translasi verteks dari bentuk netral ke setiap bentuk *morph target*. Himpunan translasi verteks mirip dengan hasil perangkat otomatis penangkap wajah. Untuk membuat bentuk mulut yang lebih kompleks, beberapa bentuk mulut pada langkah 3 dapat

dikombinasikan dengan proporsi tertentu. Kombinasi dari dua atau lebih bentuk mulut dapat dilakukan dengan mengambil rata-rata bobot tertentu untuk setiap translasi. Langkah ini dimungkinkan karena bentuk mulut adalah himpunan dari translasi verteks.

Membangun animasi mulut menggunakan *blend shape* berdasarkan fitur-fitur dari rekaman video pada Gambar 2.3, ditunjukkan pada Gambar 2.6 yang merepresentasikan hasil *render* dari *blend shape* 3D model viseme Indonesia dengan menggunakan MakeHuman. Gambar 2.5 (a) merupakan bagian dari tekstur MakeHuman, (b) adalah salah satu contoh 3D model wajah untuk animator, (c) adalah hasil *render* dari 3D model wajah dan (d) daerah bibir berwarna merah dari 3D model wajah dengan 9 CP di daerah mulut (seperti yang terlihat pada Tabel 2.4).



Gambar 2.6 *Rendering* dari *Blend Shape* 3D Model Viseme Indonesia dengan Menggunakan MakeHuman: (a) Tekstur dari MakeHuman, (b) 3D dari Model Wajah, (c) Hasil *Rendering* dari (b), (d) Daerah Bibir Merah dari (c) dengan 9 CP dalam Daerah Mulut

2.8 Hasil Eksperimen

2.8.1 Pemetaan Fonem-Viseme Indonesia

Menurut daftar himpunan fonem Bahasa Indonesia yang terdapat pada Tabel 2.2, diperlukan pemetaan unit-unit fonem ini ke sebuah himpunan simbol-simbol viseme yang merepresentasikan urutan visual sesuai dengan ucapan lisan dari contoh ucapan nyata. Selain itu, akan ditampilkan langkah-langkah yang terurut untuk membuat pemetaan yang tepat dari fonem dan segmentasi animasi 3D dari viseme berdasarkan pendekatan linguistik. Himpunan fonem Bahasa Indonesia yang didasarkan pada klasifikasi konsonan dan vokal ditunjukkan pada Tabel 2.5.

Tabel 2.5 Himpunan Fonem Berdasarkan Jenis Fonem Konsonan dan Vokal

Klasifikasi Jenis Fonem	Jumlah Fonem	Himpunan Fonem Bahasa Indonesia	Kelas Viseme
Konsonan	3	/b/, /m/, /p/	V1
	4	/f/, /v/, /w/, /ph/	V2
	8	/d/, /dh/, /dl/, /dz/, /l/, /n/, /t/, /th/	V3
	1	/r/	V4
	12	/c/, /j/, /ks/, /ps/, /s/, /sh/, /sy/, /ts/, /x/, /y/, /z/, /ny/	V5
	8	/ch/, /g/, /gh/, /h/, /k/, /kh/, /q/, /ng/	V6
	-	-	-
Total	49		

Berikut ini adalah pembahasan yang dimulai dengan jenis fonem-fonem konsonan. Kelas viseme pertama mencakup tiga fonem /b/, /m/, dan /p/, mirip dengan sistem sintesis bahasa Inggris. Penggabungan tiga fonem ini, karena memiliki tempat artikulasi yang sama dan semua bibir menutup di posisi akhir, meskipun secara visual bunyi dari /m/ berbeda dari dua fonem lainnya.

Di kelas viseme kedua pada Tabel 2.5, fonem /ph/, dimasukkan sebagai bagian dari sebuah kata serapan dalam Bahasa Indonesia, yang dibaca /f/. Fonem /ph/ ini digabung dengan dua fonem lainnya, yaitu: /f/ dan /v/. Sementara fonem /w/ juga dimasukkan dalam kelas ini, karena fonem /w/ memiliki bentuk mulut sedikit terbuka di tengah, mirip dengan tiga fonem /f/, /v/ dan /ph/.

Fonem /d/, /dh/, /dl/, /dz/, /t/, /th/, /n/ dan /l/ diletakkan bersama-sama di kelas viseme ketiga, meskipun fonem-fonem ini berbeda perilaku artikulasinya, tetapi karena perbedaan ini dihasilkan dalam mulut dan di depan lidah, maka tidak dibedakan secara visual. Sedangkan kelas viseme keempat hanya memiliki satu fonem /r/, karena /r/ menghasilkan bentuk mulut, dimana bagian depan lidah sebagai artikulator berada di langit-langit mulut dan mulut sedikit terbuka.

Kelas viseme kelima berisi dua belas fonem, yaitu: /c/, /j/, /ks/, /ps/, /s/, /sh/, /sy/, /ts/, /x/, /y/, /z/, /ny/ yang memiliki tempat dan cara artikulasi berbeda.

Menariknya, keduabelas fonem ini dipetakan oleh satu viseme. Dalam bahasa Indonesia, fonem /c/, /j/, /z/, /x/, /y/, /ny/ tidak pernah muncul di akhir kata. Selanjutnya, fonem /ts/, /ps/, /ks/, /sh/, /sy/ juga sangat jarang digunakan pada akhir kata dalam bahasa Indonesia, karena fonem-fonem ini adalah bagian dari kata serapan. Namun, jika harus membuat bentuk mulut dari fonem-fonem ini, maka dapat digantikan oleh fonem-fonem lain yang memiliki bentuk mulut yang mirip dengan fonem /s/, sebab artikulator didominasi oleh bagian depan dan tengah lidah.

Kelas viseme keenam sebagai jenis terakhir fonem konsonan adalah /ch/, /g/, /gh/, /h/, /k/, /kh/, /q/, /ng/ yang memiliki tempat artikulasi berbeda. Namun, penempatan dari delapan fonem berbeda ke dalam satu kelas viseme ini, karena bentuk mulut dari fonem ini tidak dapat dibedakan secara visual.

Berikut ini adalah pembahasan tentang jenis fonem-fonem vokal. Vokal adalah bunyi yang dihasilkan oleh aliran air terbatas di pita suara. Vokal memiliki peran penting dalam pembentukan sebuah kata, karena sebuah kata tanpa disertai vokal tidak memiliki makna apapun dan hampir setiap kata memiliki setidaknya satu vokal. Dalam penelitian ini, penggabungan fonem berdasarkan pada tinggi badan lidah dan posisi depan-belakang. Kualitas vokal umumnya ditentukan oleh tiga hal, yaitu: (1) bentuk bibir terbuka bulat, (2) posisi lidah atas-bawah, dan (3) gerakan lidah maju-mundur.

Lima huruf vokal dalam Bahasa Indonesia, yang terdiri dari “a”, “e”, “i”, “o”, “u”, digunakan sebagai pengelompokan kelas fonem vokal dari tiga belas fonem vokal termasuk diftong, yang disesuaikan dengan kesamaan pengucapan. Jadi pengelompokan kelas fonem vokal seperti yang terlihat pada Tabel 2.5.

2.8.2 Validasi Berdasarkan Survei

2.8.2.1 Instrumen Survei

Dalam penelitian ini, telah dibuat kuesioner dengan menggunakan metode skala sebagai alat statistik utama dalam pengujian model. Penskalaan adalah sebuah proses pengukuran terhadap atribut kuantitatif atau sifat. Untuk memastikan skala validitas, pengukuran dikembangkan berdasarkan studi yang relevan.

Contoh sebagian form isian kuesioner dari respon survei disajikan dalam Tabel 2.6. Pada Tabel 2.6 memuat sejumlah pertanyaan dalam kuesioner dan setiap

pertanyaan mewakili satu kelas viseme. Dalam kuesioner tersebut, terdapat 12 pertanyaan yang mewakili visualisasi viseme ke bentuk mulut model animasi 3D.

Penggunaan visualisasi 3D dari bentuk mulut yang mengandung 50 himpunan fonem harus diisi berdasarkan skala *Likert*. Kelima puluh himpunan fonem ini mewakili 49 fonem dan 1 silent (diam), yang mana responden diminta untuk mencocokkan 50 himpunan fonem dengan bentuk visualisasinya. Setiap himpunan fonem diukur oleh 4 poin skala *Likert* dengan poin akhir: (1) sangat tidak setuju, (2) tidak setuju, (3) setuju dan (4) sangat setuju pada penilaian dari Pemetaan Fonem ke Viseme Indonesia.

Tabel 2.6 Contoh Isian Kuesioner dari Respon Survei

Kelas Vise me	Bentuk Mulut dari Animasi 3D	Fonem ke-n	Fonem (cara membaca)	Visualisasi 3D dari Bentuk Mulut			
				1 Sangat Tidak Setuju	2 Tidak Setuju	3 Setuju	4 Sangat Setuju
V1		1	p (əp)				✓
		2	b (əb)			✓	
		3	m (əm)			✓	
V2		4	f (éf)				✓
		5	v (év = éf)		✓		
		6	w (éw)			✓	
		7	ph (éf)		✓		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
V9		41	ə (əh)				✓
		42	e (eh)				✓
		43	ɛ (éh)			✓	
		44	ai (pantai=panté)		✓		
V10		45	o (oh)				✓
		46	O (Oh)			✓	
		47	au (pulO)				✓
V11		48	u (uh)				✓
		49	U (Uh=uh)			✓	
V12		50	Silent (netral)				✓

Perbedaan penting harus dibuat antara skala *Likert* dan *Likert* himpunan fonem. Skala *Likert* adalah jumlah dari tanggapan pada beberapa *Likert* himpunan fonem, sementara *Likert* himpunan fonem merupakan pernyataan dari responden yang diminta untuk mengevaluasi sesuai dengan kriteria tertentu. Tingkat setuju atau tidak setuju diukur seperti pada umumnya. Skala *Likert* yang baik dianggap

simetris, karena terdapat jumlah pilihan positif dan negatif yang sama untuk sebuah pernyataan. Dalam kuesioner ini, tidak disediakan pilihan tengah "agak setuju atau agak tidak setuju" ketika responden tidak yakin dengan pilihannya.

2.8.2.2 Pengumpulan Data

Cara pengumpulan data yang dilakukan dalam penelitian ini adalah dengan mengirimkan file kuesioner melalui email ke mahasiswa program magister dan program doktor di dua universitas terkenal di Surabaya, dan juga dikirim ke beberapa lulusan dari kedua universitas tersebut. Dalam file yang dikirim, telah tersedia informasi awal tentang penelitian ini, sehingga responden diminta untuk mencocokkan dalam mengisi kuesioner dan menjawab pertanyaan dengan benar.

120 file kuesioner yang dikirimkan ke responden melalui email, ternyata terdapat 100 file yang dikembalikan. Semua file yang dikembalikan oleh responden, telah diisi lengkap secara konsisten, sehingga tidak satupun yang dibuang. Jadi efektifitas rata-rata tingkat pengembalian file kuesioner oleh responden adalah 83,33%. Semua data yang telah dikumpulkan adalah valid. Demografi responden yang telah berpartisipasi dalam pengumpulan data ini disajikan pada Tabel 2.7.

Tabel 2.7 Profil Responden

Jenis Kelamin	Pria: 77 (77%) dan Wanita: 23 (23%)
Usia (tahun)	Rata-rata: 30,61 dan Standar Deviasi: 2,22
Status Pendidikan	Mahasiswa: 70 (70%) dan Bukan Mahasiswa: 30 (30%)
Status Mahasiswa	S1: 27 (38,57%), S2: 32 (45,71%), S3: 11 (15,71%)
Status non Mahasiswa	S1: 9 (30%), S2: 21 (70%)
Pekerjaan	Dosen/Guru: 54 (54%), Bidang IT: 11 (11%), Lain: 35 (35%)

Hasil skor dari pengumpulan data kuesioner dapat dilihat pada Tabel 2.8. Skor data di tabel ini diberikan oleh responden. Kolom (1) sampai (4) pada Tabel 2.8 menunjukkan jumlah respon dari responden dengan menggunakan skala *Likert* mulai dari sangat tidak setuju sampai sangat setuju. Pada kolom total dari setiap kelas viseme, menunjukkan perkalian dari jumlah fonem dengan jumlah responden. Pada baris total (5000), menunjukkan perkalian dari jumlah fonem (50, termasuk diam) dengan jumlah responden yang telah mengisi kuesioner (100 responden).

Tabel 2.8 Skor Respon dari Responden

Kelas Viseme	Jumlah Fonem	Jumlah Respon				Total
		(1)	(2)	(3)	(4)	
V1	3	27	46	56	171	300
V2	4	32	87	105	176	400
V3	8	62	127	240	371	800
V4	1	2	17	20	61	100
V5	12	64	243	374	519	1200
V6	8	41	135	225	399	800
V7	1	1	5	13	81	100
V8	3	32	71	68	129	300
V9	4	34	75	114	177	400
V10	3	6	20	66	208	300
V11	2	5	1	25	169	200
V12	1	1	9	19	71	100
Total	50	307	836	1325	2532	5000

(1) Sangat tidak setuju, (2) Tidak setuju, (3) Setuju, (4) Sangat setuju

Sementara itu, Tabel 2.9 adalah ringkasan dari Tabel 2.8 dalam bentuk prosentase respon dari 100 responden dengan menggunakan skala Likert untuk 50 fonem (termasuk diam) di dalam 12 pertanyaan (kelas viseme).

Tabel 2.9 Prosentase Skor Respon dari Responden

Kelas Viseme	Jumlah Fonem	% Jumlah Respon				% Total
		(1)	(2)	(3)	(4)	
V1	3	9	15.33	18.67	57	100
V2	4	8	21.75	26.25	44	100
V3	8	7.75	15.88	30.97	44.65	100
V4	1	2	17	20	61	100
V5	12	5.33	20.25	31.17	43.25	100
V6	8	5.13	16.88	28.13	49.88	100
V7	1	1	5	13	81	100
V8	3	10.67	23.67	22.67	43	100
V9	4	8.50	18.75	28.5	44.25	100
V10	3	2	6.67	22	69.33	100
V11	2	2.50	0.50	12.5	84.50	100
V12	1	1	9	19	71	100
% Total	50	63,13	171,17	272,84	692,86	1200
% Rata-rata		5,26	14,26	22,74	57,74	100

(1) Sangat tidak setuju, (2) tidak setuju, (3) setuju, (4) sangat setuju

Prosentase hasil skor respon dari responden pada Tabel 2.9, terlihat bahwa jumlah responden yang menyatakan pilihan positif (setuju dan sangat setuju) tentang visualisasi 3D dari bentuk mulut yang mewakili masing-masing viseme adalah 80,48%. Hal ini berarti bahwa pemetaan fonem ke viseme Indonesia yang menggunakan animasi blend shape telah mewakili bentuk visualisasi 3D dari 50 himpunan fonem Indonesia menjadi 12 viseme.

2.8.3 Uji Validitas dan Reliabilitas

Sifat-sifat instrumen yang digunakan dalam penelitian ini dinilai dengan menggunakan konsistensi internal, yaitu: validitas dan reliabilitas. Nilai-nilai koefisien reliabilitas berkisar dari 0 sampai 1,0. Sebuah koefisien 0 berarti tidak reliabilitas (handal) dan ketika semua tes memiliki beberapa kesalahan, maka koefisien reliabilitas tidak akan pernah mencapai 1,0. Pada umumnya, reliabilitas dari tes standar adalah di atas 0,8, yang mana dikatakan menjadi sangat handal. Jika di bawah 0,5, maka tidak dianggap sebagai tes yang reliabel (dihandalkan).

Validitas mengacu pada akurasi dari sebuah penilaian, apakah data yang diukur sudah valid atau tidak? Bahkan jika tes yang dilakukan dapat dihandalkan, akan tetapi mungkin saja tidak memberikan ukuran yang valid. Validitas disimpulkan dari penilaian yang sangat penting. Instrumen Validitas dan Reliabilitas diukur dengan menggunakan persamaan 2.9, dengan r adalah korelasi antara variabel x dan y , x_k dan y_k merupakan variabel diantara fonem-fonem, $k = 1, 2, \dots, N$, dan N adalah jumlah fonem (termasuk diam).

$$r = \frac{N \sum_{k=1}^N x_k y_k - \left(\sum_{k=1}^N x_k \right) \left(\sum_{k=1}^N y_k \right)}{\sqrt{N \sum_{k=1}^N x_k^2 - \left(\sum_{k=1}^N x_k \right)^2} \sqrt{N \sum_{k=1}^N y_k^2 - \left(\sum_{k=1}^N y_k \right)^2}} \quad (2.9)$$

2.8.4 Analisis Data dan Hasil Penelitian

2.8.4.1 Validitas

Validitas mengacu pada sejauh mana derajat penelitian mencerminkan konsep tertentu, sehingga upaya peneliti untuk mengukur secara akurat.

Berdasarkan data dari responden, hasil penelitian menunjukkan bahwa semua data yang dimasukkan adalah valid. Hasil yang berasal dari perhitungan 50 pertanyaan ini, masing-masing terdiri dari skor 1, 2, 3 atau 4.

Untuk 100 responden, jumlah data yang dimasukkan adalah $50*100$ atau 5000. Dengan menghitung korelasi r dan membandingkannya dengan nilai dari momen produk yang menggunakan tingkat signifikansi 5%, yaitu 0,576, maka hasilnya diperoleh 96,19%. Hal ini menunjukkan bahwa klasifikasi viseme yang dibuat memiliki validitas tinggi dan kuat.

2.8.4.2 Reliabilitas

Reliabilitas atau kehandalan adalah sejauh mana percobaan, tes, atau suatu prosedur pengukuran menghasilkan hasil yang sama pada uji coba yang berulang-ulang. Kehandalan mengacu pada sejauh mana penilaian yang telah dilakukan itu konsisten ataukah tidak.

Di dalam penelitian ini, hasil pengukuran tidak membandingkan *Alpha Cronbach* untuk mengukur kehandalan, karena data dari kuesioner terdiri dari skor 1 sampai 4 saja, atau mulai dari sangat tidak setuju sampai sangat setuju. Hasil uji reliabilitas adalah 96,85%. Hasil ini menunjukkan bahwa data yang digunakan sangat handal.

2.8.4.3 Rating Indeks, Mean Rating dan Mean Square Eror

Rating Indeks, umumnya dikenal sebagai RI, adalah besaran yang digunakan untuk menentukan peringkat item tertentu, tetapi tidak memiliki pembedaran teoritis dari sudut pandang statistik. Tabel 2.10 menunjukkan data rating indeks dari viseme berdasarkan jumlah fonem. Namun, RI dapat digunakan untuk mengukur indeks keberhasilan dalam penelitian yang melibatkan pendapat responden, sedangkan mean rating indeks (MRI) adalah rata-rata semua RI yang diukur.

Tabel 2.10 menunjukkan bahwa rata-rata dari MRI adalah 3,33. Hasil dari nilai rata-rata ini menunjukkan bahwa kisaran MRI adalah antara 3,00 dan 4,00. Berdasarkan skala Likert, hal ini menunjukkan bahwa hasilnya adalah baik karena nilai rata-rata MRI adalah berada diantara pendapat setuju dan sangat setuju

(pendapat positif) yang artinya memiliki tingkat keberhasilan dalam pembuatan model animasi 3D dari hasil pemetaan fonem ke viseme Indonesia adalah 83,25%. Dengan demikian, representasi bentuk mulut viseme dengan visualisasi 3D dapat dikatakan mirip atau sesuai.

Tabel 2.10 Rating Indeks Viseme yang Diberikan oleh Responden

Kelas Viseme	Jumlah Fonem	RI dari Viseme untuk Setiap Fonem	Mean Rating Index
V1	3	3.17; 3.07; 3.47	3,24
V2	4	3.15; 3.21; 2.96; 2.93	3,06
V3	8	3.27; 3.13; 3.02; 3.09; 2.96; 3.17; 3.24; 3.32	3,12
V4	1	3.40	3,40
V5	12	3.25; 3.26; 3.13; 3.33; 3.24; 3.33; 2.96; 2.97; 3.04; 2.95; 3.07; 2.95	3,12
V6	8	3.35; 3.30; 3.17; 3.26; 3.24; 3.20; 3.14; 3.16	3,23
V7	1	3.74	3,74
V8	3	3.15; 3.08; 2.71	2,98
V9	4	3.22; 3.14; 2.99; 2.99	3,09
V10	3	3.65; 3.62; 3.49	3,59
V11	2	3.83; 3.75	3,79
V12	1	3.60	3,60
Rata-rata			3,33

Mean Squared Error (MSE) digunakan untuk mengukur kinerja dari RI antara bentuk mulut dan pemetaan fonem ke viseme. MSE ditentukan berdasarkan selisih antara nilai observasi dan nilai harapan dari MRI. Semakin rendah nilai MSE, semakin baik kinerjanya.

RI, MRI dan MSE dapat diukur dengan menggunakan masing-masing persamaan 2.10, persamaan 2.11, dan persamaan 2.12. ω_j adalah bobot skala *Likert* ($\omega_1 = 1$, $\omega_2 = 2$, $\omega_3 = 3$, $\omega_4 = 4$, yang menunjukkan rating, mulai dari sangat tidak setuju sampai dengan sangat setuju); S_j adalah jumlah responden di setiap bobot; $j = 1, 2, 3, 4$; $k = 1, 2, \dots, 50$; $z = 1, 2, \dots, 12$; M adalah jumlah dari semua responden; $P(z)$ adalah jumlah fonem di kelas viseme ke-z, termasuk diam; dan $Q(k)$ adalah jumlah himpunan fonem termasuk diam di kelas viseme.

$$RI(z) = \frac{1}{M} \sum_{j=1}^4 \omega_j S_j \quad \text{dengan } M = \sum_{j=1}^4 S_j \quad (2.10)$$

$$MRI(z, k) = \frac{1}{P(z)} \left[\sum_{z=1}^{12} RI(z) \right] \quad (2.11)$$

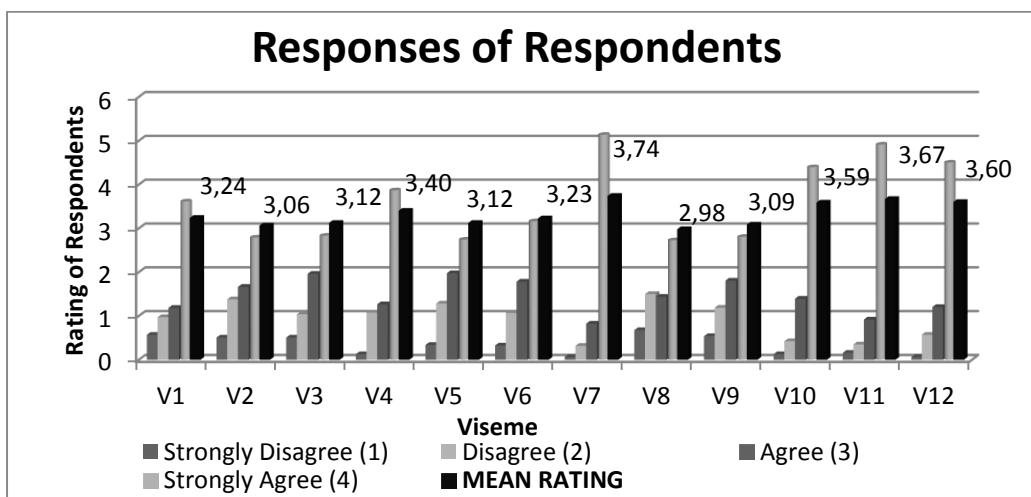
$$MSE = \frac{1}{P(z)Q(k)} \sum_{z=1}^{12} \sum_{k=1}^{50} [MRI_{obs}(z, k) - MRI_{exp}(z, k)]^2 \quad (2.12)$$

Tabel 2.11 menunjukkan MSE untuk setiap kelas viseme. Tabel ini menghitung rating indeks viseme. Hal ini sangat penting, karena bisa menjawab pertanyaan-pertanyaan berikut: (1) Berapakah mean rating indeks dari setiap viseme? Dan (2) Apakah visualisasi dari animasi 3D viseme mewakili pemetaan fonem ke viseme dengan tepat?

Tabel 2.11 Mean Squared Error Respon dari Responden

Kelas Viseme	MRI	Mean Rating Expected	Selisih	Nilai Mutlak	Square Error	MSE
V1	3,24	4	-0,76	0,76	0,5776	0,0116
V2	3,06	4	-0,94	0,94	0,8836	0,0177
V3	3,12	4	-0,88	0,88	0,7744	0,0155
V4	3,4	4	-0,60	0,60	0,3600	0,0072
V5	3,12	4	-0,88	0,88	0,7744	0,0155
V6	3,23	4	-0,77	0,77	0,5929	0,0119
V7	3,74	4	-0,26	0,26	0,0676	0,0014
V8	2,98	4	-1,02	1,02	1,0404	0,0208
V9	3,09	4	-0,91	0,91	0,8281	0,0166
V10	3,59	4	-0,41	0,41	0,1681	0,0034
V11	3,79	4	-0,21	0,21	0,0441	0,0009
V12	3,6	4	-0,40	0,40	0,1600	0,0032
Rata-rata	3,33	4	-0,67	0,67	0,4489	0,01045

Hasil nilai rata-rata MSE diperoleh 0,01045, yang menunjukkan bahwa respon dari responden memenuhi syarat sesuai dengan kondisi nilai kondisi, yang hanya sekitar 1% atau kurang dari 5% ($MSE \leq 0,05$). Dengan demikian, hal ini menunjukkan bahwa kesalahan dalam pengukuran adalah kecil atau rendah.



Gambar 2.7 Grafik Rating Respon dari Responden

Gambar 2.7 diperoleh dari Tabel 2.9 dan Tabel 2.10. Hal ini menunjukkan bahwa grafik prosentase respon dari responden, memiliki hubungan antara viseme dan skor yang diberikan oleh responden. Sebagian besar responden memberikan skor yang baik antara setuju dan sangat setuju untuk setiap viseme (antara 3,00 sampai 4,00). Gambar 2.7 juga menunjukkan rating rata-rata dari setiap viseme. Peringkat rata-rata tinggi menunjukkan bahwa pemetaan fonem ke viseme untuk bahasa Indonesia telah terwakili.

2.9 Model Viseme Indonesia

Berdasarkan sembilan parameter geometri dari bentuk perpaduan animasi bentuk mulut sekitar bibir dari model animasi 3D, dapat membangun pemetaan fonem ke viseme yang berbeda untuk bahasa Indonesia, yang dapat dilihat pada Tabel 2.12.

Hal ini dimungkinkan untuk dapat mengenali vokal dan menentukan konsonan. Untuk penelitian di masa mendatang, berencana menggunakan hasil penelitian ini untuk mengeksplorasi parameter geometri lain untuk mengembangkan kelas viseme Indonesia yang lebih halus atau lebih baik. Hasil pemetaan fonem ke viseme untuk bahasa Indonesia didasarkan pada pendekatan linguistik dan divalidasi melalui survei, sehingga diperoleh dua belas viseme

Indonesia yang telah dihasilkan termasuk diam. Dua belas viseme ini dianggap cukup lengkap untuk mewakili semua empat puluh sembilan fonem Indonesia. Kesalahan pengukuran relatif kecil, rata-rata dari mean rating indeks untuk semua viseme adalah baik, dan nilai korelasi yang diperoleh sangat tinggi, mendekati nilai 1, sehingga hasil korelasi antara animasi 3D dari bentuk mulut yang dibuat terhadap pengelompokan kelas-kelas viseme menunjukkan hasil yang baik.

Tabel 2.12 Hasil Model Animasi 3D dari Pemetaan Fonem ke Viseme Indonesia

Ke las Vis eme	Him punan Fo nem	M R I	Bentuk Mulut dari Model Manusia	Bentuk Mulut Animasi 3D	Ke las Vise me	Him punan Fo nem	M R I	Bentuk Mulut dari Model Manusia	Bentuk Mulut Animasi 3D
V1	b, m, p	3,24			V7	a	3,74		
V2	f, v, w, ph	3,06			V8	i, I, oi	2,98		
V3	d, dh, dl, dz, l, n, t, th	3,12			V9	ə, e, ε, ai	3,09		
V4	R	3,4			V10	o, O, au	3,59		
V5	c, j, s, z, ts, ps, ks, sh, sy, x, y, ny	3,12			V11	u, U	3,79		
V6	ch, g, gh, h, k, kh, q, ng	3,23			V12	Silent	3,6		

[halaman ini sengaja dikosongkan]

BAB III

AKUISISI DATA UNTUK EKSPRESI WAJAH 3D ANIMASI

3.1 Emosi dan Ekspresi Wajah

Ekspresi wajah (*facial expression*) dapat mengekspresikan emosi (*emotion*) dan suasana hati. Emosi dan suasana hati juga dapat menyampaikan informasi tentang kepribadian atau karakteristik seseorang. Informasi ini dapat menunjukkan sisi dalam yang tersembunyi dari orang tersebut, yang tidak dapat diakses dari saluran verbal. Masalah pemodelan wajah manusia dan menghasilkan gerakan wajah oleh komputer adalah menjadi tantangan dalam komunitas riset komputer grafis.

Pada kasus ekspresi wajah, ras atau suku yang berbeda memiliki budaya atau kebiasaan tersendiri di dalam mengekspresikan dan menafsirkan ekspresi wajah ke dalam bentuk emosi. Berdasarkan penelitian yang dilakukan oleh tim Universitas Glasgow (Judith Burns, 2013), Orang Timur (negara-negara Asia) cenderung menafsirkan emosi berdasarkan ekspresi mata seseorang, sedangkan Orang Barat (negara-negara Eropa dan Amerika) menafsirkan emosi berdasarkan ekspresi keseluruhan wajah, termasuk mata dan mulut. Contoh ekspresi wajah dari wanita Jepang (*Japanese Female Facial Expression, JAFFE*), dapat dilihat pada Gambar 3.1.



Gambar 3.1 Contoh Ekspresi Wajah dari JAFFE

Emosi merupakan reaksi terhadap seseorang atau kejadian. Emosi dapat ditunjukkan ketika merasa senang mengenai sesuatu, marah kepada seseorang, ataupun takut terhadap sesuatu. Oleh karena itu, emosi memainkan peran penting dalam komunikasi dan interaksi antara manusia dengan manusia. Hal ini memungkinkan orang untuk mengekspresikan diri di luar domain verbal dan memahami satu sama lain dengan bantuan dari informasi kontekstual.

Para pakar Psikologi telah mencoba untuk menjelaskan macam-macam emosi dasar manusia dalam beberapa dekade terakhir. Akan tetapi, sampai saat ini belum ada kata sepakat tentang berapa jumlah eksak dari afeksi atau emosi dasar manusia. Tabel 3.1 menjelaskan klasifikasi emosi dasar yang telah dibahas oleh banyak pakar Psikologi.

Tabel 3.1 Klasifikasi Emosi Dasar Manusia Menurut Beberapa Psikolog

Psikolog	Klasifikasi Emosi Dasar Manusia
Plutchick	Marah, Antisipasi, Jujur, Jijik, Senang, Takut, Sedih, Terkejut
Ekman, Friesen, Ellsworth	Marah, Jijik, Senang, Takut, Sedih, Terkejut
Frijda	Ingin, Bahagia, Tertarik, Terkejut, Heran, Sedih
Izard	Marah, Terhina, Jijik, Sedih, Takut, Rasa bersalah, Tertarik, Sukacita, Rasa malu, Terkejut
James	Takut, Sedih, Sayang, Marah
Mowrer	Nyeri, Senang
Banse & Scherer	Marah, Takut, Sedih, Senang, Cemas, Jijik, Malu
Cowie, et al.	Marah, Takut, Sedih, Senang, Cemas
Oatley and Johnson-Laird	Marah, Jijik, Cemas, Senang, Sedih

Pada tahun 1978, Ekman dan Friesen, merumuskan enam emosi dasar yang masing-masing memiliki sebuah ciri khas dengan ekspresi wajah yang unik. Keenam emosi dasar tersebut adalah senang, sedih, takut, jijik, terkejut, dan marah. Beberapa ilmuwan menentukan emosi pada suatu penelitian adalah sebagai berikut: (1) Emosi bukan fenomena tetapi konstruksi yang sistematis, dihasilkan oleh proses kognitif, perasaan subjektif, gairah fisiologis, kecenderungan motivasi, dan reaksi perilaku (Jonghwa Kim, 2008); (2) Sebuah emosi biasanya dialami sebagai ciri khas dari kondisi mental, kadang-kadang disertai atau diikuti oleh perubahan tubuh, dan ekspresi tindakan (Oatley dan Jenkins, 1996). Istilah ekspresi menyiratkan adanya sesuatu yang diungkapkan. Terlepas dari pendekatan, ekspresi wajah tertentu dikaitkan dengan emosi manusia tertentu. Penelitian menunjukkan bahwa orang mengkategorikan emosi wajah dalam sebuah kesamaan lintas budaya, dimana ekspresi wajah yang sama cenderung terjadi dalam merespon munculnya peristiwa dari emosi tertentu.

3.2 Deskripsi Tekstual dari Ekspresi Wajah

Emosi merupakan reaksi terhadap seseorang atau kejadian. Emosi dapat ditunjukkan ketika merasa senang mengenai sesuatu, marah kepada seseorang, ataupun takut terhadap sesuatu. Oleh karena itu, emosi memainkan peran penting dalam komunikasi dan interaksi antara manusia dengan manusia. Hal ini memungkinkan orang untuk mengekspresikan diri di luar domain verbal dan memahami satu sama lain dengan bantuan dari informasi kontekstual.

Pada literatur tentang emosi wajah manusia, diketahui terdapat sebuah hubungan antara ekspresi wajah dan pernyataan emosional (*emotional state*). Dukungan tentang pandangan emosi dasar, diasumsikan terdapat himpunan kecil emosi dasar yang dapat diekspresikan secara khas dari satu ekspresi wajah ke ekspresi wajah lainnya. Sebagai contoh, ketika orang sedang marah maka akan mengkerut dan ketika mereka senang maka akan tersenyum. Deskripsi tentang ekspresi wajah dari delapan emosi dasar dapat dilihat pada Tabel 3.2.

Tabel 3.2 Ekspresi Wajah dari Delapan Emosi Dasar

No	Emosi Dasar	Deskripsi Tekstual dari Ekspresi Wajah
1	Senang	Kedua alis mata dirilekskan. Mulut dibuka dan ujung mulut ditarik menuju ke belakang telinga.
2	Sedih	Kedua alis mata sebelah dalam diengkokkan ke atas. Kedua mata sedikit ditutup. Mulut dirilekskan.
3	Takut	Kedua alis mata dinaikkan dan ditarik bersama-sama. Kedua alis mata sebelah dalam dibengkokkan ke atas. Kedua mata tegang dan pandangan tajam.
4	Marah	Kedua alis mata sebelah dalam ditarik ke bawah bersama-sama. Kedua mata dibuka lebar. Kedua bibir ditekan berlawanan satu sama lain atau dibuka untuk menampilkan gigi.
5	Jijik	Kedua alis mata dan kedua kelopak mata dirilekskan. Bibir atas dinaikkan dan dikerutkan secara simetris.
6	Terkejut	Kedua alis mata dinaikkan. Kedua kelopak mata atas dibuka lebar, sedikit dirilekskan. Dagu dibuka.
7	Malu	Malu adalah ekspresi universal yang terlihat saat mata menghadap ke bawah dengan tampilan sedih atau khawatir.
8	Netral	Alis dan kelopak mata dirilekskan. Bibir atas dan bawah sedikit tertutup.

Pada penelitian ini, ekspresi wajah yang akan digunakan adalah enam emosi dasar oleh Paul Ekman, dkk ditambah dengan dua ekspresi lagi, yaitu: malu dan netral, seperti yang disajikan pada Tabel 3.2 (Tekalp dan Ostermann, 2000). Sehingga total ekspresi wajah yang digunakan ada delapan.

3.3 Afeksi dan Emosi

Afeksi biasanya diidentifikasi sebagai emosi, akan tetapi sebenarnya keduanya adalah fenomena yang sangat berbeda meskipun terkait erat. Perbedaan mendasar antara emosi dan afeksi adalah bahwa emosi merupakan sesuatu yang terjadi di dalam seseorang, sedangkan afeksi adalah sesuatu yang mengalir dan bergerak dari satu orang ke orang lain, memproduksi beberapa emosi.

Mendeteksi dimensi emosional (M. Schroeder, 2014) dalam sebuah ucapan merupakan tujuan penelitian yang besar sebagai sarana untuk meningkatkan interaksi manusia komputer. Dengan mempertimbangkan penggunaan dari kata afeksi dalam kehidupan sehari-hari, dapat disimpulkan bahwa afeksi adalah sesuatu yang bisa diberikan kepada orang lain. Sehingga bisa dikatakan bahwa seseorang akan "memberikan afeksi" atau bisa juga "menerima afeksi". Dengan cara ini, tampaknya bahwa afeksi mungkin sesuatu yang dapat memberikan dan menerima "dari" atau "ke" pada seseorang. Sebaliknya, emosi tidak diberikan atau diambil, emosi hanya dialami oleh diri sendiri tanpa persyaratan orang lain.

3.4 Database Ekspresi Wajah

Ekspresi wajah memiliki peranan yang sangat penting di dalam interaksi sosial antar manusia, karena ekspresi wajah mampu menunjukkan banyak informasi tentang keadaan emosi seseorang dibandingkan komunikasi verbal dan vokal. Apabila dikembangkan lebih lanjut, ekspresi wajah dapat digunakan untuk penelitian mengenai animasi, game, industri kreatif, dan penelitian lainnya.

Salah satu bentuk penelitian mengenai ekspresi wajah adalah berupa pembuatan *database* ekspresi wajah. *Database* ekspresi wajah merupakan kumpulan ekspresi wajah manusia yang berfungsi sebagai dataset untuk penelitian-penelitian yang berhubungan dengan emosi dan ekspresi wajah. *Database* ekspresi wajah menampilkan emosi netral dan emosi dasar manusia. Sebagian besar

database ekspresi wajah menggunakan pembagian emosi dasar manusia yang dilakukan oleh Ekman, Freisen, dan Tomkins. (Ekman, Freisen, dan Tomkins, 1971) membagi emosi dasar manusia menjadi enam kategori, yaitu: marah, sedih, senang, terkejut, jijik, dan takut. Keenam emosi dasar ini tidak mencakup seluruh emosi yang ada, akan tetapi keenam emosi tersebut cukup untuk digunakan sebagai pembeda antara emosi yang satu dengan emosi yang lainnya.

Database ekspresi wajah terbagi atas dua jenis dimensi, yaitu: 2D dan 3D. Data 2D hanya dapat menampilkan data dalam dimensi panjang dan tinggi, sedangkan pada data 3D dapat menampilkan data dalam dimensi panjang, lebar, dan tinggi. Kedua *database* tersebut memiliki kelebihan dan kekurangan masing-masing. Perbandingan *database* ekspresi wajah 2D dan 3D dapat ditunjukkan pada Tabel 3.3 (Arman Savran, dkk., 2012).

Tabel 3.3 Perbandingan *Database* Ekspresi Wajah 2D dan 3D

Dasar Perbandingan	<i>Database</i> Ekspresi Wajah	
	2D	3D
Ukuran <i>database</i>	Relatif besar	Relatif kecil
Penggunaan dalam penelitian	Sering digunakan, karena sudah cukup lama dikembangkan	Masih jarang digunakan, karena teknologi yang mendukung baru dikembangkan
Kriteria pada saat dilakukan pengambilan data	<ul style="list-style-type: none"> • Bergantung pada pencahayaan • Terpengaruh ketika ada gerakan atau perubahan posisi data 	<ul style="list-style-type: none"> • Tidak bergantung pada pencahayaan • Tidak terpengaruh dengan adanya perubahan posisi data
Contoh dataset	JAFFE	BU-3DFE

3.5 Indonesian Female Facial Expression-3D Database

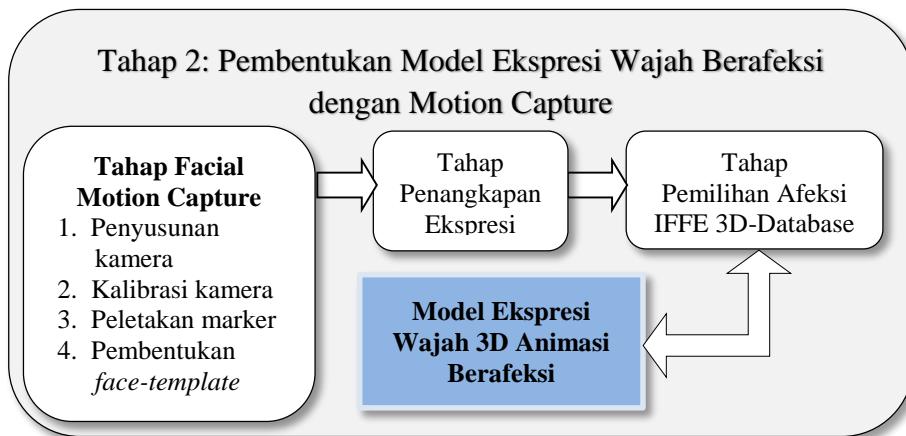
Pada penelitian ini, sistem dirancang untuk menghasilkan *database* ekspresi wajah perempuan Indonesia berbasis 3D. Database ini selanjutnya disebut sebagai *Indonesian Female Facial Expression-3D Database (IFFE-3D Database)*. *IFFE-3D Database* ini bermanfaat untuk membantu penelitian yang berhubungan dengan pengenalan emosi dasar manusia untuk ekspresi wajah 3D animasi (Ekman, Freisen, dan Tomkins, 1971).

Terdapat ekspresi netral dan tujuh emosi dasar yang ditampilkan oleh *IFFE-3D Database*, yaitu emosi marah (*angry*), sedih (*sad*), senang (*happy*), terkejut (*surprise*), jijik (*disgust*), malu (*shame*), dan takut (*fear*). *IFFE-3D Database* menggunakan model perempuan Indonesia sebanyak 10 orang. Keterangan lengkap mengenai profil dari model *IFFE-3D Database* dapat dilihat pada Tabel 3.4.

Tabel 3.4 Profil Model IFFE-3D Database

Nomor	Kriteria Model	Keterangan
1	Jenis Kelamin	Perempuan
2	Usia	18 – 25 tahun
3	Warga Negara	Indonesia
4	Suku Bangsa	Jawa dan etnis China
5	Pekerjaan	Mahasiswa S1 dan S2

Dalam konstruksi pembentukan model ekspresi wajah berafeksi dengan motion capture, seperti yang ditunjukkan dalam Gambar 3.2, ada 3 tahap yang dilakukan, antara lain: Tahap Facial Motion Capture, Tahap Penangkapan / Perekaman Ekspresi Wajah, dan terakhir Tahap Pemilihan Afeksi dari *IFFE-3D Database*.



Gambar 3.2 Konstruksi Pembentukan Model Ekspresi Wajah Berafeksi dengan IFFE-3D Database

3.6 Facial Motion Capture

Motion capture (Mocap) merupakan proses merekam suatu gerakan dan menerjemahkannya ke dalam persamaan matematika dengan cara melacak

sejumlah titik-titik penting pada ruangan sepanjang waktu, dan menggabungkannya dalam bentuk representasi gerakan 3D (Alberto Menache, 2000). Teknologi *Mocap* ini dapat menangkap gerakan di dunia nyata dan kemudian menerjemahkannya ke dalam gerakan digital. Subjek yang ditangkap bisa berupa benda apapun, selama benda tersebut memiliki gerakan. Hal penting yang dibutuhkan pada proses ini adalah pemilihan titik atau area yang menunjukkan gerakan subyek, yaitu bagian dari subyek yang menunjukkan perpindahan.

Motion capture terdiri atas dua macam, yaitu *markered* dan *markerless motion capture*. Sistem *markered motion capture* membutuhkan penanda (*marker*) untuk menangkap gerakan. Penanda tersebut dipasangkan di seluruh bagian yang ingin ditangkap gerakannya. Sedangkan sistem *markerless motion capture* tidak menggunakan penanda sebagai penangkap gerakan, melainkan menggunakan algoritma komputer yang mampu menandai area gerak secara otomatis.

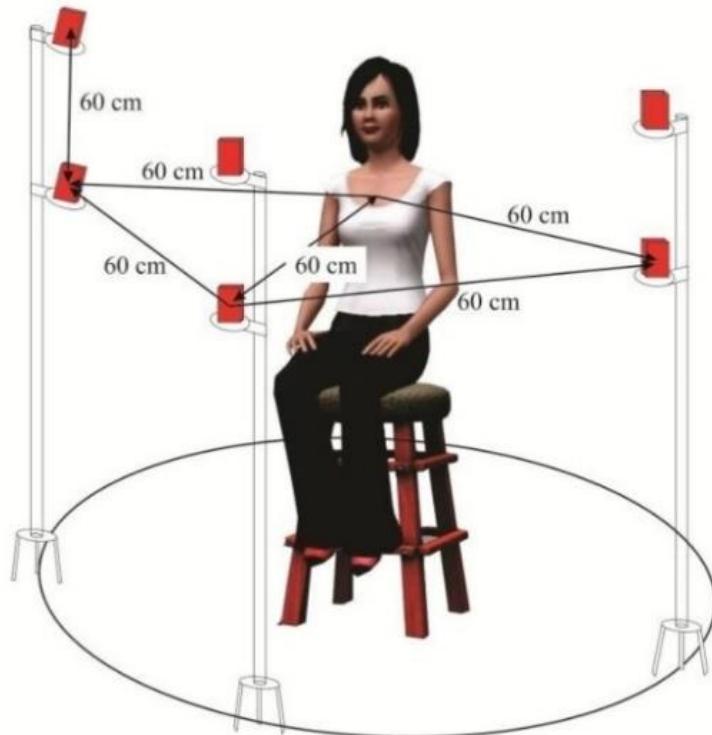
Sistem *markered mocap* lebih banyak digunakan untuk bidang-bidang yang membutuhkan keakuratan gerakan, seperti bidang perindustrian film dan penelitian. Sedangkan *markerless mocap* lebih banyak digunakan untuk bidang-bidang yang tidak membutuhkan keakuratan gerakan, melainkan membutuhkan kepraktisan dalam penggunaannya, seperti game. Hasil *capture* dari *markerd mocap* sangat akurat dan presisi, sedangkan hasil *capture* dari *markerless mocap* kurang akurat dan kurang presisi.

Salah satu sistem *mocap* yang diciptakan oleh NaturalPoint Inc. yang berhubungan dengan *facial motion capture* adalah OptiTrackExpressionTM, yang memiliki sistem penangkapan ekspresi wajah lengkap, yaitu mulai dari *software*, kamera, hingga perlengkapan tambahan, seperti alat kalibrasi, alat sinkronisasi, penanda, dan peralatan lainnya. OptiTrackExpressionTM menggunakan *software* bernama ARENA Expression, yaitu program khusus penangkapan ekspresi wajah yang berbasis pada penanda.

Proses penangkapan ekspresi wajah yang dilakukan adalah menggunakan sistem multi kamera. Kamera-kamera tersebut diletakkan pada posisi tertentu sehingga dapat menangkap seluruh gerakan subyek dari sudut pandang yang berbeda. Selain itu, kamera-kamera tersebut harus tersinkronisasi satu sama lain

sebelum menangkap gerakan, agar video-video yang didapat memiliki waktu secara simultan sehingga dapat memunculkan gerakan digital secara utuh.

3.6.1 Penyusunan Kamera

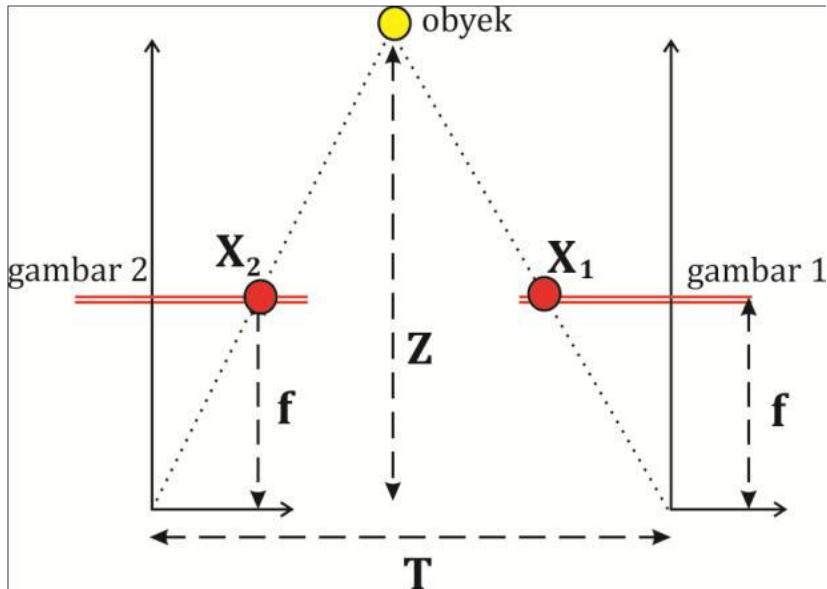


Gambar 3.3 Contoh Pengaturan Posisi Multi Kamera pada Facial Motion Capture

Jumlah kamera minimal yang digunakan pada penangkapan ekspresi adalah enam buah kamera. Kamera-kamera tersebut disusun menyerupai busur lingkaran dengan sudut kurang lebih 120 derajat. Tiga kamera disusun di atas kepala dan tiga kamera lainnya disusun setinggi dada. Jarak antar kamera dengan kamera dan kamera dengan model sebesar 60 cm. Masing-masing kamera tersebut difokuskan ke arah hidung model. Gambar 3.3 menunjukkan ilustrasi penyusunan kamera pada tahap pengambilan data.

Pada sistem *mocap* yang menggunakan teknik multi kamera, setiap kamera harus memiliki kesamaan persepsi dalam menentukan posisi suatu titik pada koordinat dunia nyata. Penentuan posisi titik ini didapatkan melalui metode triangulasi kamera stereo (Muhammad Annis, 2012). Triangulasi kamera stereo biasanya terjadi pada dua kamera yang saling tersinkronisasi atau kamera stereo.

Namun, metode ini dapat diterapkan pada proses penangkapan gambar oleh *motion capture* yang menggunakan lebih dari dua kamera. Triangulasi kamera merupakan penentuan perbedaan jarak obyek melalui realitas geometri obyek. Adapun ilustrasi proses triangulasi kamera stereo dapat ditunjukkan pada Gambar 3.4.



Gambar 3.4 Ilustrasi Triangulasi Kamera Stereo

Berdasarkan Gambar 3.4 di atas, obyek ditangkap dan direkam menggunakan dua buah kamera, sehingga menghasilkan dua buah gambar, yaitu gambar 1 dan gambar 2. Posisi obyek pada koordinat dunia nyata dapat diketahui menggunakan rumus pada Persamaan 3.1.

$$\frac{T - x_2 - x_1}{Z - f} = \frac{T}{Z} \rightarrow Z = \frac{fT}{x_2 - x_1} \quad (3.1)$$

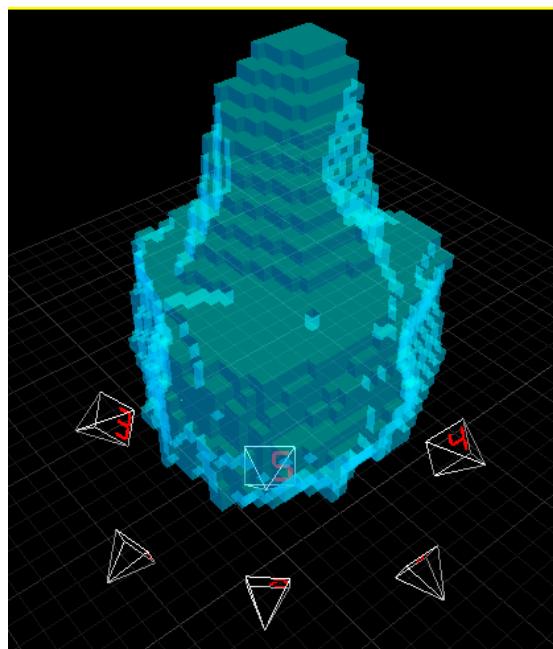
dengan:

Z : kedalaman atau *depth*, T : jarak antar dua kamera, f : panjang fokus kamera.
 x_1 : posisi obyek pada gambar 1, x_2 : posisi obyek pada gambar 2.

3.6.2 Kalibrasi Kamera

Optiwand dan *Calibration Square* merupakan perlengkapan tambahan pada sistem OptiTrack™ yang berfungsi sebagai alat kalibrasi kamera. Hasil akhir dari proses kalibrasi kamera adalah berupa *capture volume* yang bisa ditangkap oleh keseluruhan kamera. Pada proses kalibrasi, *OptiWand* berfungsi untuk

mengkorespondensikan antar satu kamera dengan kamera lainnya agar setiap kamera memiliki pemahaman yang sama di dalam mendefinisikan suatu titik. Sedangkan *Calibration Square* berfungsi untuk menyesuaikan sumbu koordinat pada program agar sesuai dengan sumbu koordinat dunia.



Gambar 3.5 Capture volume yang dihasilkan pada proses kalibrasi

Kalibrasi kamera terdiri atas empat tahap, yaitu penentuan arah kamera, pengaturan kamera, proses *wanding*, dan penentuan *ground plane*. Hasil akhir dari keseluruhan tahap kalibrasi menghasilkan *capture volume* yang mampu ditangkap oleh kamera. Gambar 3.5 menunjukkan *capture volume* yang dihasilkan pada proses kalibrasi. Penjelasan lebih lanjut mengenai tahapan dari proses kalibrasi kamera adalah sebagai berikut.

1. Penentuan Arah Kamera

Tahap pertama pada kalibrasi kamera adalah memastikan bahwa setiap kamera telah mengarah pada model, karena pada saat tahap penyusunan kamera, posisi dan arah kamera belum tentu sesuai dengan yang diharapkan.

2. Pengaturan Kamera

Pengaturan kamera dilakukan agar kamera dapat menangkap penanda dengan sempurna.

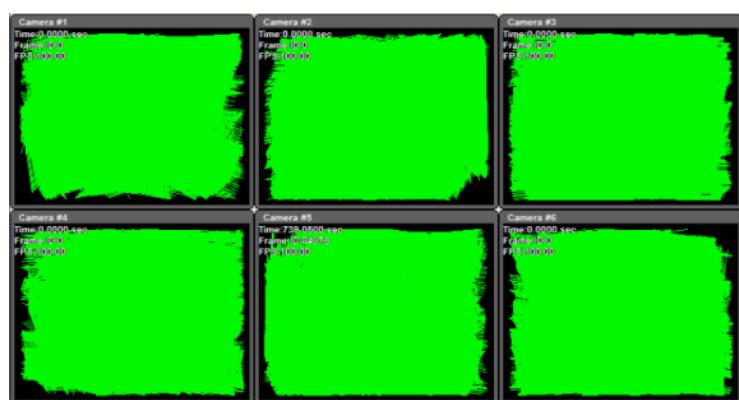
3. Proses *Wanding*

Proses *wanding* berguna untuk mengkorespondensikan antar satu kamera dengan kamera lainnya. Hal ini perlu dilakukan karena pada awalnya setiap kamera memiliki pemahaman yang berbeda di dalam mendefinisikan suatu titik, sehingga perlu dilakukan korespondensi antar keenam kamera tersebut agar dihasilkan satu pemahaman di dalam mendefinisikan titik tersebut.



Gambar 3.6. (a) Proses *wanding* dan (b) *Optiwand kit*

Proses *wanding* dilakukan dengan cara mengarahkan *optiwand kit* ke tengah-tengah kamera, kemudian mengayunkannya ke seluruh daerah *capture* dengan kecepatan normal. Gambar 3.6 adalah proses dan peralatan yang digunakan pada proses *wanding*. Proses ini selesai ketika seluruh kamera menampilkan *Screen shot* seperti pada Gambar 3.7.



Gambar 3.7 *Screen shot* Hasil Akhir Proses *Wanding* Enam Kamera

4. Penentuan *Ground Plane*

Penentuan *ground plane* ini digunakan untuk menyesuaikan sumbu koordinat pada program agar sesuai dengan sumbu koordinat dunia. Penentuan *ground plane* dilakukan dengan menyearahkan sumbu-x dan sumbu-y pada *calibration square* agar sesuai dengan sumbu-x dan sumbu-y pada koordinat dunia. Gambar 3.8 menunjukkan peralatan *calibration square* dan proses penentuan *ground plane*.



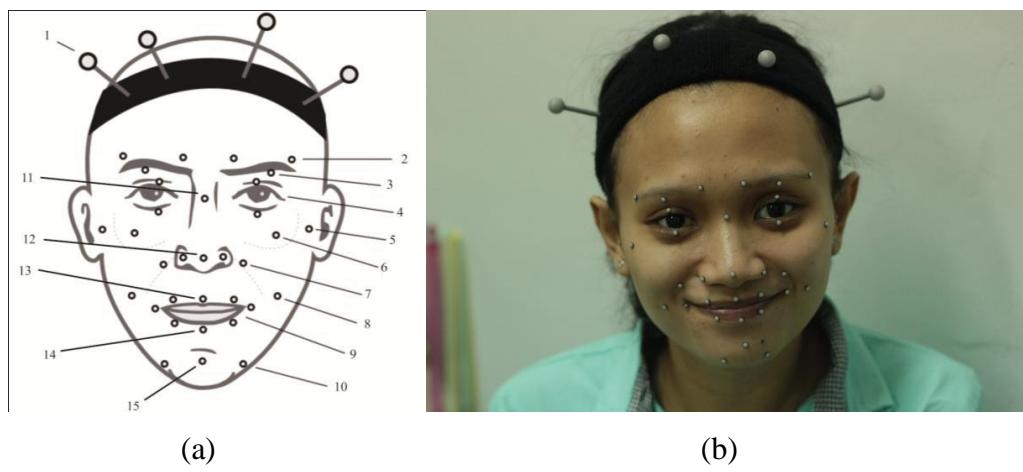
Gambar 3.8 (a) *Penentuan ground plane* dan (b) *Calibration Square*

3.6.3 Peletakan Penanda

Peletakan penanda merupakan suatu alat penanda yang digunakan untuk menandai bidang gerak dari obyek yang telah disediakan oleh OptiTrack™. OptiTrack™ menyediakan penanda yang terbuat dari bahan dan rancangan yang khusus. Pada *facial motion capture*, dibutuhkan *reflective* penanda *headband set* yang dipasang di kepala dan *reflective* penanda *set* yang berukuran 4 mm. Penggunaan *reflective* penanda *set* adalah dengan cara merekatkan penanda tersebut pada wajah, seperti yang ditunjukkan pada Gambar 3.9 berdasarkan Tabel 3.5.

Tabel 3.5 Bagian-bagian Wajah yang Dipasang Penanda secara Simetris

Nomor	Bagian-bagian Wajah	Jumlah Penanda		
		Kiri	Tengah	Kanan
1	Head	-	4	-
2	Eyebrow	2	-	2
3	Orbital Upper	1	-	1
4	Eyelids	2	-	2
5	Ear	1	-	1
6	Orbital Lower	1	-	1
7	Nostril base & Bulge	2	-	2
8	Puffer	1	-	1
9	Lip & Mouth	3	-	3
10	Jaw end	1	-	1
11	<i>Nose bridge</i>	-	1	-
12	<i>Nose tip</i>	-	1	-
13	<i>Upper lip</i>	-	1	-
14	<i>Lower lip</i>	-	1	-
15	<i>Chin</i>	-	1	-
Sub Total Penanda		14	9	14
Total Penanda				37



Gambar 3.9 Contoh (a) Ilustrasi Peletakan Marker pada Wajah dan (b) Marker yang Dipasang pada Wajah untuk IFFE-3D Database

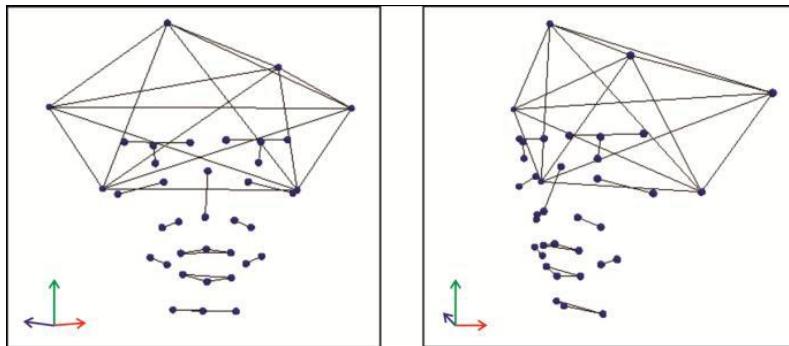
3.6.4 Kalibrasi Face Template

Face template digunakan untuk mendefinisikan penanda pada wajah model agar sesuai dengan penanda template pada program. Hasil kalibrasi *face template* disebut sebagai *facebone*. *Facebone* dapat terbentuk apabila penanda pada wajah model tertangkap secara lengkap, yaitu sebanyak 37 buah. Tahapan yang perlu dilakukan untuk kalibrasi *face template* adalah sebagai berikut:

- Perekaman untuk setiap kali ekspresi wajah model selama 10 detik.
- Pemilihan kesesuaian posisi penanda dari hasil poin (a), dengan syarat posisi yang dipilih harus memiliki jumlah penanda sebanyak 37 buah.
- Pembentukan *facebone* dilakukan dengan cara memblok empat penanda kepala, kemudian menekan tombol “create”. *Screen shot* pembentukan *facebone* dapat dilihat pada Tabel 3.6.

Tabel 3.6 Contoh *Facebone* pada Ekspresi Wajah Model

Ekspresi Wajah	Model Manusia	Facebone	Ekspresi Wajah	Model Manusia	Facebone
Netral (Neutral)			Takut (Fear)		
Terkejut (Surprise)			Marah (Angry)		
Malu (Shame)			Senang (Happy)		
Sedih (Sad)			Jijik (Disgust)		

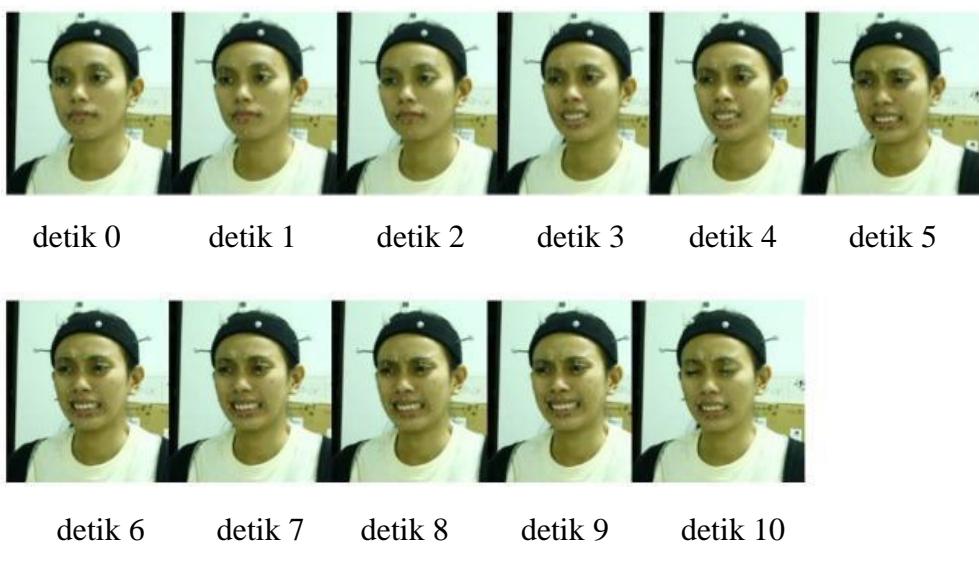


Gambar 3.10 Contoh data 3D *marker based* pada IFFE-3D Database: (a) Data tampak depan; (b) Data tampak samping

Data *IFFE-3D Database* berbentuk pose 3D dan berbasis pada penanda (*marker based*), ditunjukkan pada Gambar 3.10. Data pose 3D merupakan data ekspresi wajah yang menampilkan afeksi tertentu selama satu frame dalam bentuk data 3D. Data 3D ini berbasis pada penanda, yaitu seluruh bagian pada data mengacu pada penanda, seperti acuan berupa posisi ataupun bentuk struktur wajah. Gambar-gambar hasil *face template* dari 10 orang model pada data *facial motion capture* tersedia di Lampiran 1 (Hosyiah Rusdiana, 2015).

3.7 Perekaman Ekspresi Wajah pada *Facial Motion Capture*

Pada tahap perekaman ekspresi wajah, ekspresi wajah model direkam dengan menggunakan teknik *facial motion capture*. Proses perekaman ini dilakukan dengan melacak sejumlah titik penting yang menandai daerah pergerakan obyek. Pada *IFFE-3D Database*, fokus obyek adalah wajah model dan titik-titik yang menandai daerah pergerakan disebut penanda.



Gambar 3.11 Proses Penangkapan Ekspresi Wajah

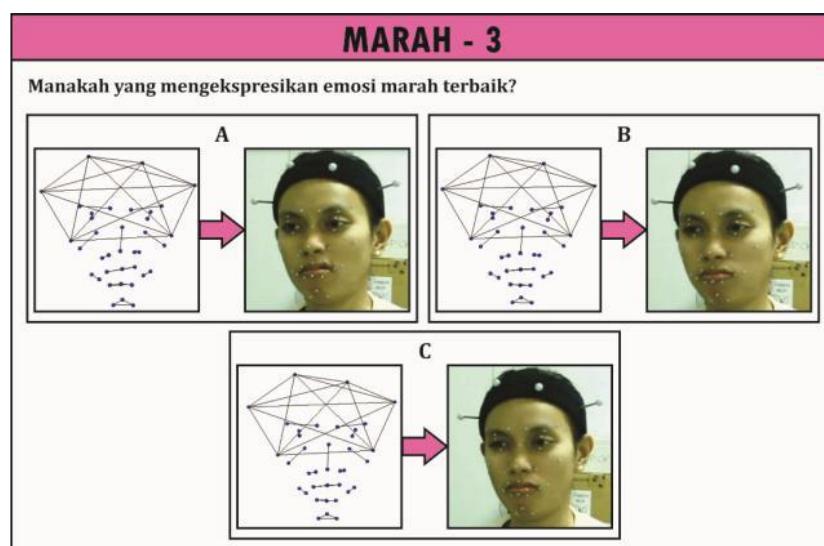
Tahap perekaman ekspresi wajah dilakukan dengan cara merekam ekspresi wajah model selama 10 detik. Pada 3 detik pertama model berekspresi netral, lalu pada 7 detik selanjutnya model berekspresi sesuai ekspresi yang ingin ditampilkan, yaitu ekspresi marah, sedih, senang, terkejut, jijik, takut, malu atau netral. Setiap ekspresi diambil sebanyak 4 kali. Bila ada kerusakan *facebone* atau ekspresi tidak

sesuai dengan yang diharapkan, maka perlu dilakukan proses penangkapan ekspresi wajah ulang. Gambar 3.11 menunjukkan proses penangkapan ekspresi wajah jijik.

Data yang dihasilkan oleh penangkapan ekspresi wajah berupa *file extension pt2*. File tersebut berupa data penanda 2D yang berasal dari masing-masing kamera. Data-data penanda disinkronisasi menjadi satu-kesatuan, sehingga mendapatkan data penanda seperti pada koordinat di dunia nyata. Data yang dihasilkan pada proses ini memiliki *frame* sebanyak 1000 *frame*. Data ini diolah dan dipilih, sehingga hanya ada 1 *frame* yang dijadikan data untuk *IFFE-3D Database*.

3.8 Pemilihan Afeksi

Tahap pemilihan afeksi dilakukan untuk mendapatkan ekspresi wajah yang menghasilkan afeksi terbaik. Hal ini perlu dilakukan karena seorang model tidak selalu menampilkan afeksi yang terbaik selama proses penangkapan ekspresi. Tahap pemilihan afeksi terbaik dilakukan dengan menggunakan teknik survei. Salah satu contoh data survei yang digunakan untuk pemilihan afeksi dapat ditunjukkan pada Gambar 3.12.



Gambar 3.12 Contoh Data Survei

Data survei yang digunakan adalah sebanyak 840 pose (Hosyi'ah Rusdiana, 2015), dengan rincian setiap data hasil perekaman ekspresi diwakili oleh tiga pose. Tiga pose tersebut merupakan pose yang dianggap mewakili afeksi yang

ditampilkan. Pemilihan pose tersebut mengacu pada parameter ekspresi wajah berdasarkan Tabel 3.2 (Tekalp dan Ostermann, 2000).

Proses survei ini dilakukan dengan cara setiap koresponden memilih salah satu pose dengan afeksi terbaik dari beberapa pilihan pose yang ditawarkan. Persentase terbesar dari hasil survei pada masing-masing afeksi yang terpilih pada posisi frame tertentu, dipakai sebagai titik-titik referensi untuk menentukan pose afeksi yang dijadikan *IFFE-3D Database*. Data hasil survei secara keseluruhan ditunjukkan pada Tabel 3.7.

Tabel 3.7 (a) Prosentase Hasil Survei Pose Terkejut

Rekaman ke-	TERKEJUT (SURPRISE)	Model Manusia IFFE-3D Database ke-									
		1	2	3	4	5	6	7	8	9	10
1	Frame ke-	742	401	320	297	540	508	550	330	380	460
	Pemilih (%)	64%	73%	55%	73%	55%	73%	45%	55%	64%	55%
2	Frame ke-	501	448	540	580	640	384	390	230	380	460
	Pemilih (%)	45%	64%	45%	73%	45%	82%	45%	45%	45%	55%
3	Frame ke-	307	495	460	470	550	681	860	490	380	550
	Pemilih (%)	55%	45%	36%	36%	55%	45%	45%	36%	55%	45%
4	Frame ke-	288	597	540	460	550	370	230	390	370	660
	Pemilih (%)	45%	91%	55%	45%	36%	45%	55%	45%	36%	64%

Tabel 3.7 (b) Prosentase Hasil Survei Pose Takut

Rekaman ke-	TAKUT (FEAR)	Model Manusia IFFE-3D Database ke-									
		1	2	3	4	5	6	7	8	9	10
1	Frame ke-	776	416	440	390	360	320	450	330	360	280
	Pemilih (%)	45%	45%	91%	36%	55%	36%	45%	45%	45%	64%
2	Frame ke-	204	285	410	470	650	682	540	410	460	430
	Pemilih (%)	55%	55%	36%	45%	36%	55%	55%	45%	36%	45%
3	Frame ke-	229	360	480	560	670	290	550	360	660	550
	Pemilih (%)	55%	55%	45%	45%	36%	73%	45%	45%	45%	45%
4	Frame ke-	555	822	390	370	450	932	450	450	360	640
	Pemilih (%)	45%	82%	64%	64%	36%	55%	55%	36%	45%	45%

Tabel 3.7 (c) Prosentase Hasil Survei Pose Sedih

Rekaman ke-	SEDIH (SAD)	Model Manusia IFFE-3D Database ke-									
		1	2	3	4	5	6	7	8	9	10
1	Frame ke-	766	442	440	960	724	149	510	280	350	560
	Pemilih (%)	73%	45%	55%	55%	36%	45%	55%	55%	45%	45%
2	Frame ke-	653	603	550	650	750	604	290	353	480	700
	Pemilih (%)	45%	36%	55%	36%	45%	45%	45%	55%	55%	45%
3	Frame ke-	651	1	540	880	450	498	950	400	320	750
	Pemilih (%)	45%	36%	36%	55%	55%	36%	36%	55%	64%	55%
4	Frame ke-	395	581	650	470	425	753	450	420	340	650
	Pemilih (%)	55%	82%	55%	64%	45%	45%	36%	36%	64%	64%

Tabel 3.7 (d) Prosentase Hasil Survei Pose Senang

Rekaman ke-	SENANG (HAPPY)	Model Manusia IFFE-3D Database ke-									
		1	2	3	4	5	6	7	8	9	10
1	Frame ke-	747	423	285	750	340	368	550	560	440	470
	Pemilih (%)	64%	45%	55%	55%	36%	73%	36%	45%	36%	45%
2	Frame ke-	86	721	350	590	350	859	240	470	400	440
	Pemilih (%)	73%	73%	45%	45%	55%	55%	36%	45%	73%	55%
3	Frame ke-	183	844	470	950	650	934	950	490	340	540
	Pemilih (%)	55%	45%	64%	64%	36%	73%	45%	45%	36%	55%
4	Frame ke-	295	860	360	499	560	306	460	470	480	860
	Pemilih (%)	45%	55%	45%	45%	55%	45%	36%	36%	45%	36%

Tabel 3.7 (e) Prosentase Hasil Survei Pose Jijik

Rekaman ke-	JIJIK (DISGUST)	Model Manusia IFFE-3D Database ke-									
		1	2	3	4	5	6	7	8	9	10
1	Frame ke-	798	898	480	330	899	540	440	300	230	660
	Pemilih (%)	45%	55%	45%	55%	45%	55%	55%	36%	45%	64%
2	Frame ke-	745	724	411	310	460	357	750	330	285	400
	Pemilih (%)	45%	36%	45%	55%	55%	64%	36%	45%	36%	45%
3	Frame ke-	205	398	450	349	740	420	575	580	299	570
	Pemilih (%)	55%	91%	45%	55%	45%	64%	45%	55%	55%	45%
4	Frame ke-	231	441	480	350	550	632	540	440	360	850
	Pemilih (%)	55%	82%	55%	82%	55%	55%	45%	64%	55%	45%

Tabel 3.7 (f) Prosentase Hasil Survei Pose Marah

Rekaman ke-	MARAH (ANGRY)	Model Manusia IFFE-3D Database ke-									
		1	2	3	4	5	6	7	8	9	10
1	Frame ke-	399	410	360	406	750	7	550	450	380	450
	Pemilih (%)	64%	55%	45%	82%	45%	45%	55%	45%	82%	82%
2	Frame ke-	991	186	530	270	895	293	260	440	370	490
	Pemilih (%)	55%	55%	45%	55%	45%	45%	45%	36%	55%	45%
3	Frame ke-	232	473	460	290	380	333	370	350	550	660
	Pemilih (%)	36%	45%	73%	55%	45%	64%	36%	55%	45%	45%
4	Frame ke-	860	862	590	340	498	647	460	350	349	650
	Pemilih (%)	55%	55%	64%	55%	55%	45%	45%	36%	55%	55%

Tabel 3.7 (g) Prosentase Hasil Survei Pose Netral

Rekaman ke-	NETRAL (NEUTRAL)	Model Manusia IFFE-3D Database ke-									
		1	2	3	4	5	6	7	8	9	10
1	Frame ke-	298	452	260	50	470	750	40	440	460	390
	Pemilih (%)	55%	55%	55%	45%	55%	45%	64%	45%	45%	45%
2	Frame ke-	400	149	160	360	540	460	60	440	450	450
	Pemilih (%)	64%	45%	45%	45%	55%	45%	55%	64%	36%	36%
3	Frame ke-	948	326	440	260	440	560	340	850	450	440
	Pemilih (%)	55%	45%	45%	36%	36%	55%	55%	45%	45%	45%
4	Frame ke-	800	322	150	170	340	450	60	40	440	540
	Pemilih (%)	73%	45%	36%	64%	55%	36%	55%	55%	55%	55%

Pose dengan prosentase pemilih tertinggi yang ditampilkan pada setiap afeksi, diberi blok kuning dengan angka tertulis **tebal (bold)**, diwakili oleh:

- Afeksi Terkejut, frame ke-597 dari model 2, pada pengambilan rekaman keempat, dengan prosentase pemilih 91%.
- Afeksi Takut, frame ke-440 dari model 1, pada pengambilan rekaman pertama, dengan prosentase pemilih 91%.
- Afeksi Sedih, frame ke-581 dari model 2, pada pengambilan rekaman keempat, dengan prosentase pemilih 91%.
- Afeksi Senang, frame ke-934 dari model 6, pada pengambilan rekaman ketiga, dengan prosentase pemilih 73%. Pada afeksi senang ini, sebenarnya ada 4 model yang memiliki prosentase sama 73%, yaitu model 1, model 2,

model 9 pada pengambilan rekaman kedua dan model 6 pada pengambilan rekaman pertama dan ketiga. Akan tetapi, karena model 6 memiliki prosentase 73% sebanyak 2 kali, maka yang terpilih adalah model 6 dengan pengambilan rekaman yang terakhir atau ketiga.

- e. Afeksi Jijik, frame ke-398 dari model 2, pada pengambilan rekaman ketiga, dengan prosentase pemilih 91%.
- f. Afeksi Marah, frame ke-406 dari model 4, pada pengambilan rekaman pertama, dengan prosentase pemilih 82%. Sama halnya dengan afeksi senang di atas, terdapat 3 model yang memiliki prosentase pemilih sama 82%, yaitu model 4, model 9 dan model 10 pada pengambilan rekaman pertama semua. Oleh karena model 1 memiliki rata-rata terbesar dalam pengambilan semua rekaman, maka yang terpilih adalah model 4 dengan pengambilan rekaman pertama.
- g. Afeksi Netral, frame ke-800 dari model 1, pada pengambilan rekaman keempat, dengan prosentase pemilih 73%.
- h. Afeksi Malu, frame ke-122 dari model 9, pada pengambilan rekaman kedua, tanpa dilakukan survei pada responden, hanya diamati sendiri oleh peneliti.

Hasil survei keseluruhan dari pemilihan afeksi *IFFE-3D Database* yang memiliki prosentase terbesar di Tabel 3.7, diringkas menjadi Tabel 3.8 dan dinyatakan terpilih sebagai titik-titik referensi pada pose afeksi *IFFE-3D Database*.

Tabel 3.8 Prosentase Terbesar Hasil Survei pada Setiap Pose

No mor	Pose Afeksi	Frame ke-	Model IFFE-3D Database ke-	Pengambilan Rekaman ke-	Prosentase Pemilih
1	Terkejut	597	2	4	91%
2	Takut	440	1	1	91%
3	Sedih	581	2	4	91%
4	Senang	934	6	3	73%
5	Jijik	398	2	3	91%
6	Marah	406	4	1	82%
7	Netral	800	1	4	73%
8	Malu	122	9	2	100%
			Prosentase Rata-rata Pemilih		86,5%

3.9 Pembuatan Ekspresi Wajah 3D Animasi

3.9.1 Animasi dengan *Motion Capture*

Gerakan di dalam data *motion capture* diekspresikan dengan menggunakan *blend shape model*. Gerakan itu diasumsikan sebagai gerakan (atau posisi per-frame) dari sebuah penanda (*marker*) gerak yang dapat dinyatakan sebagai sebuah kombinasi linear dari titik-titik yang bersesuaian dalam *blend shape*, yaitu:

$$M_j = \sum_{i=1}^n \alpha_i V_{ij} \quad (3.2)$$

dengan M_j adalah lokasi pada wajah yang gerakannya dicatat, $j = 1, 2, \dots, m$. V_{ij} adalah lokasi yang bersesuaian pada *blend shape* ke- i , $i = 1, 2, \dots, n$. m adalah banyaknya penanda gerak dan n adalah banyaknya *blend shape* (Choe, dkk., 2001).

Diberikan beberapa persamaan seperti pada persamaan 3.2, tentukan bobot blending dari α_i . Kemudian susun kembali persamaan 3.2 ini sebagai masalah minimisasi, diperlukan untuk meminimumkan jumlah dari selisih seperti yang tertulis pada persamaan 3.3 berikut:

$$\sum_{j=1}^m \left(M_j - \left(\sum_{i=1}^n \alpha_i V_{ij} \right) \right)^2 \quad (3.3)$$

Keseluruhan sistem ini merupakan sistem persamaan linier, dimana α_i adalah bobot-bobot dalam gabungan *blend shape* yang tidak diketahui.

Dengan menggunakan penyelesaian program kuadrat iteratif (Nocedal dan Wright, 1999), diperoleh nilai optimal dari bobot *blending* α_i dengan menggunakan metode kuadrat terkecil (*least square method*). Penyelesaian dari sistem ini ekivalen dengan memproyeksikan gerak secara ortogonal pada himpunan *blend shape*. Pada persamaan 3.3 secara umum, tidak memiliki solusi yang tepat, karena gerakan-gerakannya bisa lebih ekspresif dari pada himpunan *blend shape* yang mungkin.

Untuk menghasilkan *mesh* animasi yang mengikuti gerakan lebih tepat, dilengkapi proyeksi pada basis *blend shape* dengan menerjemahkan verteks-verteks di dalam *mesh* dengan sisa $= \left(M_j - \sum_{i=1}^n \alpha_i V_{ij} \right)$. Sisa ini, yang hanya dikenal untuk sebuah himpunan kecil dari titik-titik, diinterpolasikan ke seluruh *mesh* wajah dengan menggunakan *Radial Basis Function* (Nielson, 1993).

Koordinat terakhir, V_j , dari sebuah verteks pada wajah ini kemudian dikonstruksi dengan menggunakan:

$$V_j = P_j + RBF(P_j) \quad (3.4)$$

P_j adalah proyeksi pada himpunan *blend shape*:

$$P_j = \sum_{i=1}^n \alpha_i V_{ij} \quad (3.5)$$

Dan $RBF(P_j)$ adalah interpolasi sisa pada verteks P_j :

$$RBF(P_j) = \sum_{i=1}^m \exp(-\|M_i - P_j\|).C_i \quad (3.6)$$

Dalam persamaan 3.6, vektor C_i dihitung dengan menggunakan nilai-nilai yang diketahui dari sisa di M_i . Sejak sistem persamaan linier ini tidak diketahui, dengan menggunakan metode linier kuadrat terkecil menyediakan perkiraan yang tidak diketahui (Pighin, Hecker, dkk, 1998). Perhatikan bahwa hanya dengan menerapkan sisa akan memiliki pengaruh yang berbeda, dengan terlebih dahulu memproyeksikan pada himpunan *blend shape*, diperoleh sebuah geometri wajah yang mencerminkan *blend shape-blend shape* ini, kemudian terapkan sisa yang membawa geometri lebih dekat dengan gerakannya.

Pendekatan (Choe, dkk, 2001) untuk pemetaan gerak ke satu himpunan *blend shape* sangat mirip. Perbedaan utamanya adalah bagaimana sisa yang diambil pada perhitungan. Dalam pendekatan ini, *blend shape* dimodifikasi untuk menyesuaikan dengan gerakan. Di sisi lain, menggunakan *Radial Basis Function* untuk modifikasi geometri pada sebuah basis per-frame. Metode ini mungkin akan lebih efektif untuk memproses sejumlah besar gerakan, sedangkan *blend shape-blend shape* itu akan tampil lebih baik pada dataset yang kecil.

Selain menyelesaikan sistem di atas untuk seluruh model, diselesaikan juga untuk setiap daerah yang dibuat dengan menggunakan proses segmentasi otomatis, agar hasilnya lebih memuaskan dari kendala *spasial*. Hal ini juga memungkinkan untuk diekspresikan ke berbagai gerakan dengan hanya menggunakan sejumlah *blend shape* yang terbatas.

Untuk setiap *frame* dan untuk setiap daerah (*region*), konstruksikan masalah minimisasi di atas dan dapatkan bobot-bobot *blending*. Bobot-bobot yang sama

kemudian digunakan untuk mendapatkan, baik untuk semua verteks daerah (*region*), maupun posisi baru yang cocok dengan gerakannya. Jadi, untuk setiap *frame* gerakan, dapat diselesaikan sebuah masalah minimisasi untuk memperoleh bobot-bobot *blending* dan secara konsekuensi *mesh* wajah mengikuti data *motion capture*.

3.9.2 Pengeditan Keyframe

Dengan menggunakan *blend shape model*, *mesh* wajah dapat dikonstruksi secara interaktif yang dapat digunakan sebagai *keyframe-keyframe* dalam sebuah *keyframing* berdasarkan program animasi wajah. Membuat sebuah *keyframe* mirip dengan memproduksi sebuah frame di sebuah urutan *motion capture* yang memerlukan titik kontrol khusus (penanda), pemetaan masing-masing, dan kendala *spasial* (yaitu posisi dari penanda). Dalam sebuah antarmuka (*interface*), pengguna dapat menentukan semua kontrol di atas dengan *klik* dan *drag* menggunakan *mouse* pada model wajah. Seperti dalam proses yang digunakan dalam aplikasi *motion capture*, berdasarkan persamaan 3.2, mengkonstruksi sebuah masalah minimisasi menggunakan kendala tertentu dan akan mendapatkan bobot-bobot *blend shape* secara interaktif.

Blend shape model dapat disegmentasi ke dalam daerah-daerah (*regions*) menggunakan teknik segmentasi otomatis. Dalam rangka untuk mengedit *keyframe* yang mungkin pada berbagai tingkat detil, akan dibangun sebuah hierarki daerah. Hierarki ini dibuat terlebih dahulu dengan menjalankan algoritma segmentasi dengan nilai ambang batas (*threshold*) tinggi sehingga dapat menghasilkan daerah kecil dan daerah lokal. Daerah ini merupakan tingkat wilayah paling bawah. Kemudian daerah-daerah dapat digabungkan secara iteratif, sehingga daerah berdekatan digabungkan bersama-sama seperti menghasilkan tingkat daerah yang lebih tinggi.

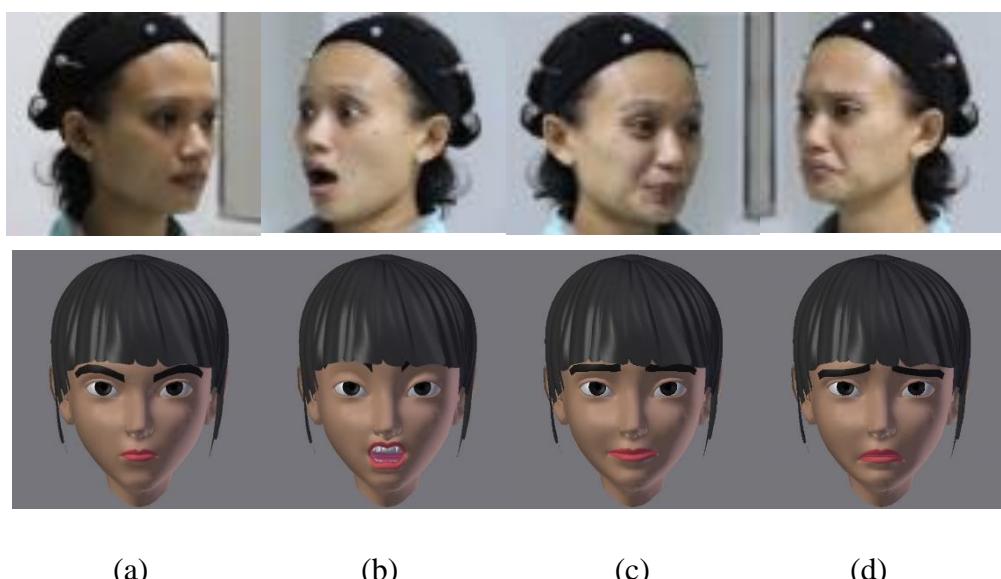
Beberapa lokasi di wajah tidak bergerak secara signifikan sepanjang himpunan *blend shape* (misal ujung hidung). Jika sebuah lokasi tersebut dipilih dan mencoba untuk merusaknya, sejauh ini dengan menggunakan gambaran antar muka, sebuah gerakan kecil *mouse* akan memicu perubahan dramatis dalam ekspresi wajah. Untuk mengurangi sensitivitas sistem, skala perpindahan *mouse* disesuaikan pada faktor yang berbanding terbalik dengan perpindahan maksimum dalam *blend shape model* di titik yang dipilih pada *mesh*. Pengeditan *keyframe*

berturut-turut dilakukan dengan meningkatkan tingkat detil untuk memperbaiki ekspresi wajah secara lokal.

3.9.3 *Rendering Realistik* dari Blend Shape

Proses *rendering* dasar dari *blend shape model* cukup sederhana dan dapat dilakukan dalam dua langkah: pertama, geometrinya dievaluasi, dan kemudian dirender sebanyak *blend shape* yang ada dalam model untuk memetakan tekstur *blend*. Langkah terakhir ini dilakukan dengan menentukan setiap *alpha channel* verteks yang sesuai bobotnya untuk *blend shape* yang diberikan. Untuk memperbaiki *rendering*, diputuskan tidak mencampur peta tekstur pada bagian wajah yang tekstur seharusnya tidak bervariasi sebagai fungsi dari ekspresi wajah, khususnya, di rambut, leher, dan daerah telinga.

Daerah ini ditekstur dengan menggunakan peta tekstur dari sembarang *blend shape* (biasanya satu yang sesuai dengan ekspresi netral). Kesalahan pada registrasi tekstur adalah masalah umum *rendering blend shape* untuk animasi wajah yang realistik. Jika tekstur tidak sesuai pada setiap titik geometri wajah, maka penggabungan secara linier akan menghasilkan *rendering* kabur. Akuisisi data untuk ekspresi wajah 3D animasi pada delapan afeksi, dapat dilihat pada Gambar 3.13 (a) hingga (h).



Gambar 3.13 Akuisisi Data Animasi dengan Motion Capture Menggunakan Blend Shape Model untuk Afeksi: (a) Netral, (b) Terkejut, (c) Malu, (d) Sedih



(e)

(f)

(g)

(h)

Gambar 3.13 Akuisisi Data Animasi dengan Motion Capture Menggunakan Blend Shape Model untuk Afeksi: (e) Takut, (f) Marah, (g) Senang, (h) Jijik

Gambar 3.13 di atas merupakan hasil dari data facial motion capture yang diekspresikan dengan menggunakan *blend shape animation model*. Setiap gerakan atau posisi per-frame dari sebuah *marker* pada masing-masing ekspresi wajah model dinyatakan sebagai sebuah kombinasi linier titik-titik yang bersesuaian dalam *blend shape* dari setiap ekspresi wajah.

[halaman ini sengaja dikosongkan]

BAB IV

HIDDEN MARKOV MODEL BERDASARKAN MODEL VISEME INDONESIA UNTUK UCAPAN ALAMI DENGAN AFEKSI

Pembahasan pada bab sebelumnya mengenai proses pembuatan model viseme Indonesia pada Bab 2 dan proses pembentukan model wajah berafeksi pada Bab 3, merupakan proses dari dua tahapan awal pada penelitian ini. Berdasarkan roadmap penelitian yang sudah dijelaskan pada sistematika penulisan disertasi Gambar 1.4, maka pada Bab 4 ini akan dijelaskan proses dari tahapan ketiga, yang membahas tentang pembentukan urutan viseme yang alami dengan menggunakan *Hidden Markov Model* (HMM). Selanjutnya, tahapan akhir tentang sistem representasi visual, dibahas pada Bab 5.

Pembentukan urutan viseme yang alami (natural), dimulai dengan membahas arsitektur sistem secara keseluruhan, agar diperoleh sebuah gambaran sistem secara jelas. Arsitektur sistem ini terbagi dalam dua bagian, yaitu: sistem segmentasi input teks dan sistem implementasi HMM.

4.1 Desain Sistem Tahap Ketiga

Pada penelitian ini telah dirancang sebuah sistem yang saling berhubungan satu sama lain. Sistem ini dibagi menjadi dua sistem besar, yaitu sistem segmentasi input teks dan sistem implementasi HMM. Seluruh sistem yang telah dibuat, berkaitan secara berurutan dan setiap hasil dari sistem ini digunakan untuk sistem selanjutnya hingga sistem ini selesai dikerjakan semua.

Sistem besar pertama yang akan dilakukan adalah segmentasi input teks. Sistem segmentasi input teks ini berfungsi untuk melakukan pemisahan kata yang nantinya akan digunakan pada proses selanjutnya. Sistem ini akan dilakukan pada awal setelah user menginputkan sebuah kalimat Bahasa Indonesia yang disertai dengan parameter afeksi. Sistem ini dibagi menjadi beberapa proses lagi yang bertujuan untuk mengubah kalimat yang diinputkan menjadi sebuah frame viseme.

Sistem besar selanjutnya yang akan dilakukan adalah implementasi HMM. Sistem ini berfungsi untuk mengimplementasikan algoritma HMM pada frame

viseme yang sudah dihasilkan dari proses sebelumnya. Pada proses ini juga akan dilakukan pembuatan model HMM dari data training yang sudah dikumpulkan sebelumnya bersamaan dengan model viseme Indonesia yang sudah ada. Sistem implementasi HMM merupakan sistem yang bertujuan untuk menghasilkan state model HMM yang terbaik, sehingga akan menghasilkan urutan viseme Bahasa Indonesia yang natural dari input user. Arsitektur dari seluruh sistem pada tahap ketiga ini ditunjukkan pada Gambar 4.1 yang menjelaskan urutan proses yang dilakukan pada sistem secara keseluruhan.

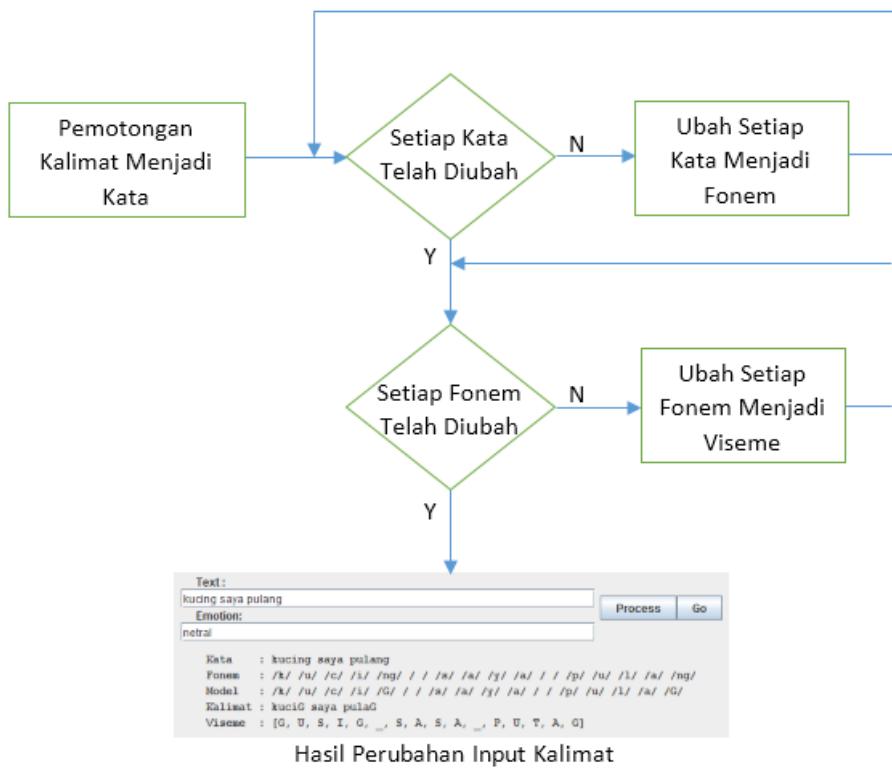
4.1.1 Desain Sistem Segmentasi Input Teks

Sistem segmentasi input teks merupakan sistem besar pertama yang dilakukan. Sistem ini dilakukan sekali saja, yaitu setelah user menginputkan kalimat Bahasa Indonesia dan parameter afeksinya. Sistem ini berfungsi untuk merubah kalimat input teks menjadi urutan viseme yang dipergunakan pada proses implementasi HMM. Hasil akhir dari sistem ini adalah hasil perubahan input kalimat menjadi urutan viseme Bahasa Indonesia. Sistem segmentasi input teks digambarkan dalam bentuk diagram yang ditunjukkan pada Gambar 4.1.

Segmentasi Input Teks dilakukan setelah proses pembacaan input kalimat dan afeksi. Proses pertama yang dilakukan adalah melakukan pemotongan kalimat menjadi kata. Kalimat yang sudah diinputkan dipotong menjadi potongan kata guna memudahkan proses selanjutnya, yaitu pemotongan kata menjadi urutan fonem.

Proses pemotongan kata menjadi fonem, dilakukan untuk mengklasifikasikan kata yang ada menjadi fonem, sehingga dapat dengan mudah diubah menjadi urutan viseme dari kalimat input yang dilakukan pada proses selanjutnya, yaitu proses pemotongan fonem menjadi viseme Indonesia.

Proses pemotongan fonem menjadi viseme Indonesia, dilakukan untuk merubah fonem dari proses sebelumnya menjadi viseme Indonesia, sehingga dapat digunakan untuk menuju sistem besar selanjutnya, yaitu implementasi HMM. Perubahan fonem menjadi viseme Indonesia ini disesuaikan dengan bentuk model viseme Indosesia, seperti yang tertuang di Tabel 2.12 pada bab sebelumnya.



Gambar 4.1 Desain Sistem Segmentasi Input Teks

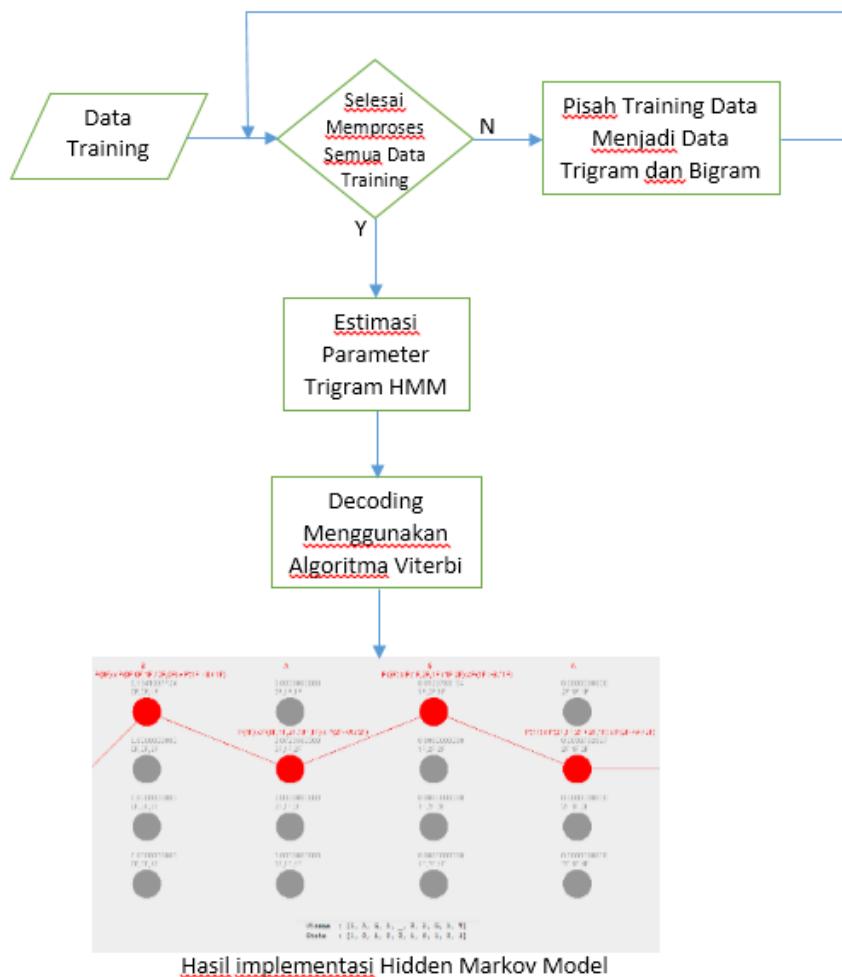
4.1.2 Desain Sistem Implementasi Hidden Markov Model dan Algoritma Viterbi

Sistem Implementasi HMM merupakan sistem yang dilakukan pada hasil output sistem segmentasi input teks, yaitu urutan viseme Indonesia dari input teks kalimat Bahasa Indonesia, serta parameter afeksi yang telah diinputkan sebelumnya. Implementasi HMM ini dilakukan pada output tersebut untuk menentukan panjang-pendeknya urutan viseme Indonesia dari kalimat input sesuai dengan parameter afeksi yang telah diinputkan pula, sehingga didapat urutan viseme Indonesia yang natural, baik dan benar. Sistem implementasi HMM dapat dilihat pada Gambar 4.2.

Sistem implementasi HMM, diawali dengan memproses data training yang telah dikumpulkan. Setiap data training yang ada, dipisah menjadi *trigram* dan *bigram* HMM untuk digunakan pada proses selanjutnya. Setelah semua data training dipisah menjadi data *trigram* dan *bigram* HMM, proses selanjutnya

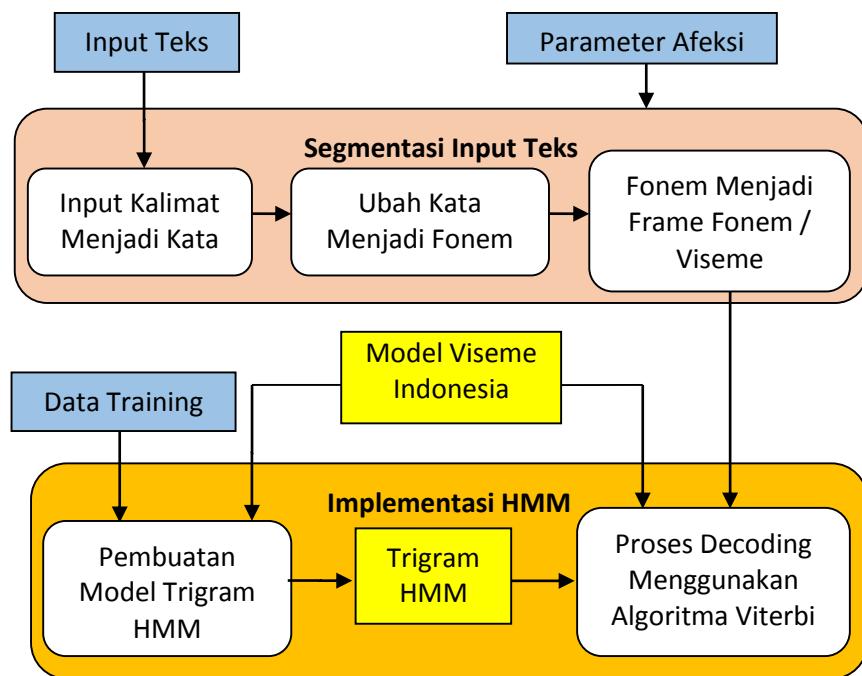
melakukan proses estimasi parameter *trigram* HMM dan proses decoding menggunakan algoritma Viterbi.

Proses perhitungan estimasi parameter dan *decoding algoritma Viterbi* akan dijelaskan lebih lanjut pada sub-bab berikutnya. Hasil dari sistem implementasi HMM ini adalah model state Viterbi serta urutan state panjang viseme Indonesia yang alami.



Gambar 4.2 Desain Sistem Implementasi Hidden Markov Model

Gabungan dari dua desain sistem secara keseluruhan pada tahap ketiga ini dikonstruksika seperti pada Gambar 4.3. Selanjutnya, pada sub-bab berikut ini akan dilakukan pembahasan secara lebih detail untuk setiap proses yang dikerjakan.



Gambar 4.3 Konstruksi Jalannya Penelitian Tahap Ketiga

4.2 Segmentasi Input Teks

Dalam segmentasi input teks ini menjelaskan mengenai proses apa saja yang dilakukan agar proses selanjutnya dapat dijalankan dengan baik dan hasil yang didapatkan sesuai harapan. Proses dalam segmentasi input teks ini dibagi menjadi 3 bagian, yaitu: (1) pemotongan kalimat menjadi kata, (2) pemotongan kata menjadi fonem, dan (3) pengubahan fonem menjadi viseme. Berikut ini, akan dijelaskan masing-masing bagian dari proses segmentasi input teks pada setiap sub-bab.

4.2.1 Pemotongan Kalimat Menjadi Kata

Sebuah kalimat adalah sebuah kumpulan kata yang ditata sedemikian rupa sehingga menjadi sebuah urutan kata yang baik dan benar. Sebuah kalimat dalam Bahasa Indonesia, bisa terdiri dari 1 kata, 2 kata atau bahkan bisa lebih dari 3 kata. Setiap kata yang ada dalam 1 kalimat dipisahkan oleh spasi (*space*), yang memisahkan antar kata. Kalimat Bahasa Indonesia yang menjadi input teks ini diproses sehingga dapat diubah menjadi kumpulan kata yang berguna untuk proses selanjutnya. Berikut adalah algoritma pengambilan kata pada sebuah kalimat.

Algoritma 4.1 Pemotongan Kalimat

- 1: Lakukan pembacaan pada kalimat yang diinputkan oleh user
- 2: Lakukan pengambilan kata secara berkala
- 3: Simpan kata yang telah diambil pada sebuah vektor
- 4: Ulangi proses 2 dan 3 hingga semua kata telah diambil

Pada algoritma 4.1, langkah pertama yang dilakukan adalah melakukan pembacaan kalimat Bahasa Indonesia yang telah diinputkan oleh user melalui parameter *inputText*. Setelah itu, setiap kata dari kalimat Bahasa Indonesia tersebut diambil dari kalimat yang diinputkan tadi.

Proses pengambilan kata dilakukan secara berkala dan setiap kata yang didapatkan disimpan pada sebuah vektor. Proses pengambilan kata dilakukan secara terus menerus hingga tidak ada lagi kata yang tersedia. Proses penyimpanan setiap kata dari sebuah kalimat input teks disimpan pada sebuah vektor yang diberi nama “*inputWords*”, dengan cara menggunakan fungsi “*split()*” pada Bahasa pemrograman *Java*.

Berikut ini adalah beberapa contoh hasil uji coba pemotongan kalimat Bahasa Indonesia menjadi kata yang dilakukan oleh user melalui *inputText* yang ditampung dalam vektor *inputWord*.

1. *InputText* : Saya suka makan nasi

Vector inputWords:

saya
suka
makan

2. *InputText* : Singa itu berlari kencang

Vector inputWords:

singa
berlari
kencang

3. *InputText* : Saya sakit kepala

Vector inputWords:

saya
sakit
kepala

4.2.2 Pemotongan Kata Menjadi Fonem

Untuk kebutuhan dalam pengubahan menjadi viseme di proses selanjutnya, maka setiap kata yang sudah didapatkan di proses sebelumnya, yaitu pemotongan kalimat menjadi kata, akan diklasifikasikan ke dalam fonem.

Pada pemotongan kata menjadi fonem ini, berdasarkan struktur tata bahasa dalam Bahasa Indonesia, setelah kata adalah suku kata, yang pemisahannya mengikuti aturan kata dasar dengan 11 pola suku kata seperti yang tertuang di Tabel 2.3. Setelah kata dipotong menjadi suku kata, barulah dipotong menjadi fonem.

Di dalam penelitian ini, untuk mempersingkat proses pemotongan kata menjadi fonem, peneliti telah menggunakan sebuah aturan tambahan dari peneliti sebelumnya yang perhatian terhadap *text to speech* Bahasa Indonesia (Arry Akhmad Arman, 2011). Aturan yang ditambahkan dalam pemotongan kata menjadi fonem adalah memperhatikan konsonan ganda, yang disebut sebagai kluster. Kluster ini yang kemudian dipakai untuk mengklasifikasi kata menjadi fonem.

Dalam penelitian ini, berdasarkan himpunan fonem Indonesia yang telah ditetapkan pada bab sebelumnya, terdapat 15 konsonan ganda, seperti yang tertulis di dalam Tabel 2.5. Dari 15 konsonan ganda ini yang selanjutnya disebut sebagai 15 kluster, diklasifikasikan ke dalam sebuah fonem dengan menggunakan simbol huruf besar/kapital. Pengklasifikasian konsonan ganda ke dalam kluster, dapat dilihat pada Tabel 4.1.

Tabel 4.1 Tabel Pengklasifikasian Fonem Berkonsonan Ganda

Nomor	Jumlah Fonem Berkonsonan Ganda	Jenis Kluster	Klasifikasi
1	1	ph	F
2	4	gh, ch, kh, ng	G
3	4	th, dh, dl, dz	T
4	6	ks, ps, sh, sy, ts, ny	S
Total	15		

Selain 15 kluster di atas, sebenarnya masih banyak sekali contoh kluster pada Bahasa Indonesia yang tidak digunakan pada penelitian ini, seperti “dr” (drama), “tr” (trofi), “sr” (sri), “kr” (kriya), “fr” (frustasi), “kl” (reklamasi), dan lain-lain. Hal ini disebabkan karena dalam penelitian pemetaan fonem ke viseme

Indonesia telah menggunakan 49 fonem, yang terdiri dari 13 fonem vokal (10 vokal tunggal dan 3 vokal ganda/*diftong*) dan 36 fonem konsonan (21 konsonan tunggal dan 15 konsonan ganda/*compunds*), termasuk *silent* (diam), seperti yang sudah dijelaskan pada Bab 2 sebelumnya.

Pengklasifikasian ini sangat berguna untuk menentukan jenis viseme yang digunakan pada tahap pemotongan kata menjadi fonem, karena pengklasifikasian fonem ini didasarkan pada bentuk bibir pengucapan pada saat visualisasi dari model viseme Indonesia. Berikut adalah algoritma pemotongan kata menjadi fonem dengan menggunakan *class extractor Java* dengan memanfaatkan pointer *LinkedList* pada *input string*.

Algoritma 4.2 Pemotongan Kata Menjadi Fonem

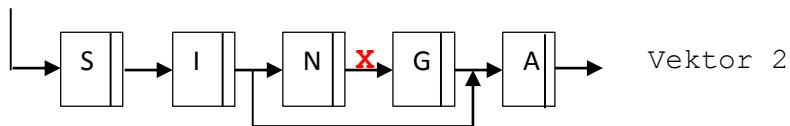
- 1: Lakukan pengambilan kata dari vektor kata
- 2: Lakukan pengambilan setiap huruf dari kata yang telah diambil
- 3: Lakukan pengecekan untuk huruf yang diambil apakah memenuhi kriteria pengklasifikasian kluster menjadi fonem yang bersangkutan
- 4: Simpan fonem yang diambil dalam sebuah *LinkedList*
- 5: Ulangi langkah 2 dan 3 sampai setiap huruf dalam kata selesai diklasifikasikan
- 6: Ulangi langkah 1 sampai 4 sampai setiap kata selesai diubah menjadi urutan fonem

Pada algoritma 4.2, langkah pertama yang dilakukan tentunya adalah pengambilan kata dari vektor kata yang dihasilkan dari proses sebelumnya. Setelah itu, kata yang terambil tersebut akan diambil hurufnya satu persatu dan akan dicek apakah kata tersebut mengandung kluster, dan pengklasifikasianya sesuai dengan Tabel 4.1. Proses ini diulang secara berkala hingga setiap huruf dari kata yang ada di dalam vektor kata selesai diproses. Setelah selesai, urutan fonem yang didapat dalam *LinkedList* fonem digunakan untuk proses selanjutnya.

Contoh hasil uji coba pemotongan kata menjadi fonem, diambilkan dari contoh kedua di atas yang telah menampung *inputWord* “Singa berlari kencang”. Berikut ini adalah ilustrasi hasilnya:

inputWord: “Singa itu berlari kencang”

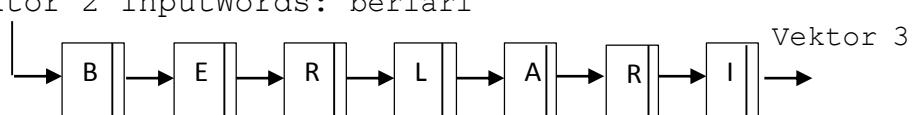
1. Vektor 1 InputWords: Singa



Ditemukan kluster dari konsonan ganda “ng” (**X** = hapus pointer N ke G)

```
LinkedList : /s//i//ng//a//_/
              ↓ Kluster 2, ng → G
Fonem       : /s//i//G//a//_/
```

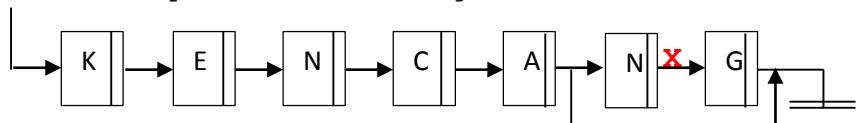
2. Vektor 2 InputWords: berlari



Tidak ditemukan kluster dari konsonan ganda

```
LinkedList : /b//e//r//l//a//r//i//_/
Fonem       : /b//e//r//l//a//r//i//_/
```

3. Vektor 3 InputWords: kencang



Ditemukan kluster dari konsonan ganda “ng” (**X** = hapus pointer N ke G)

```
LinkedList : /k//e//n//c//a//ng/
              ↓ Kluster 2, ng → G
Fonem       : /k//e//n//c//a//G/
```

Perpindahan dari satu vektor ke vektor berikutnya, dianggap sebagai spasi antar kata, yang disimbolkan sebagai silent (diam). Setelah memotong kata pertama menjadi beberapa fonem pada vektor 1, dilanjutkan hingga vektor terakhir, maka input teks kalimat: “Singa berlari kencang”, fonem seluruhnya adalah:

```
/s/ /i/ /G/ /a/ /_/_/ /b/ /e/ /r/ /l/ /a/ /r/ /i/ /_/_/ /k/ /e/ /n/
/c/ /a/ /G/
```

4.2.3 Pengubahan Fonem Menjadi Viseme

Untuk keperluan visualisasi 3D pada proses selanjutnya, maka setiap fonem yang sudah didapat dari pemotongan kata sebelumnya, harus diubah menjadi viseme dahulu sesuai dengan klasifikasinya. Berdasarkan Tabel 2.12 pada Bab 2 sebelumnya, telah dibuat 12 macam klasifikasi viseme Indonesia, dengan setiap fonem akan diklasifikasikan ke dalam 12 macam viseme seperti yang terlihat pada Tabel 4.2.

Tabel 4.2 Pengklasifikasian Viseme

Kelas Viseme	Himpunan Fonem	Klasifikasi Viseme
V1	/b/, /m/, /p/	P
V2	/f/, /v/, /w/, /ph/	F
V3	/d/, /dh/, /l/, /dz/, /l/, /n/, /t/, /th/	T
V4	/r/	R
V5	/c/, /j/, /ks/, /ps/, /s/, /sh/, /sy/, /ts/, /x/, /y/, /z/, /ny/	S
V6	/ch/, /g/, /gh/, /h/, /k/, /kh/, /q/, /ng/	G
V7	/a/	A
V8	/i/, /I/, /oi/	I
V9	/ə/, /e/, /ɛ/, /ai/	E
V10	/o/, /ɔ/, /au/	O
V11	/u/, /ʊ/	U
V12	silent “/”	—

Pengubahan fonem menjadi viseme berdasarkan klasifikasi dari kelas viseme yang diwakili oleh sebuah huruf kapital, didasarkan pada bentuk mulut yang sama untuk setiap fonem. Klasifikasi viseme ini digunakan untuk proses implementasi HMM dan proses Visualisasi 3D. Proses pengubahan fonem menjadi viseme dari kalimat Bahasa Indonesia yang telah diinputkan dapat dituliskan sebagai algoritma 4.3.

Algoritma 4.3 Pengubahan Fonem Menjadi Viseme

- 1: Lakukan pengambilan fonem dari LinkedList fonem
- 2: Lakukan pengecekan untuk fonem yang diambil apakah memenuhi kriteria pengklasifikasian viseme menjadi viseme yang bersangkutan
- 3: Simpan viseme yang diambil dalam sebuah LinkedList
- 4: Ulangi langkah 2 dan 3 sampai setiap huruf dalam kata selesai diklasifikasikan
- 5: Ulangi langkah 1 sampai 4 sampai setiap fonem selesai diubah menjadi viseme

Pada algoritma 4.3, langkah pertama yang dilakukan tentunya adalah pengambilan fonem dari *LinkedList* fonem yang dihasilkan dari proses sebelumnya. Setelah itu, fonem yang terambil tersebut akan diklasifikasikan sesuai dengan Tabel 4.2. Proses ini akan diulang secara berkala hingga setiap yang ada didalam *LinkedList* fonem selesai diproses. Setelah selesai, urutan viseme yang didapat dalam *LinkedList* viseme akan digunakan untuk proses selanjutnya.

Dengan cara yang sama seperti pemotongan kata menjadi fonem, maka pengubahan fonem menjadi viseme juga memanfaatkan pointer pada LingkedList. Sehingga dari *LinkedList* Fonem ini selanjutnya menjadi *LinkedList* Viseme, dengan memanfaatkan klasifikasi viseme seperti pada Tabel 4.2, yaitu mengubahnya fonem menjadi huruf besar.

Lanjutan dari contoh pemotongan kata menjadi fonem dan dilanjutkan menjadi viseme adalah sebagai berikut:

Input Teks: Singa berlari kencang
LinkedList Fonem:
/s//i//ng//a//_/_b//e//r//l//a//r//i//_/_k//e//n//c//a//ng/

Fonem:
/s//i//G//a//_/_b//e//r//l//a//r//i//_/_k//e//n//c//a//G/

Viseme:
S I G A _ P E R T A R I _ G E T S A G

Contoh beberapa input teks kalimat lain yang telah dilakukan dengan pengubahan fonem menjadi viseme adalah:

1. Input Teks: Kucing saya hilang
LinkedList Fonem:
/k//u//c//i//ng//_/_s//a//y//a//_/_h//i//l//a//ng/
2. Input Teks: Febi Menulis Surat
LinkedList Fonem:
/f//e//b//i//_/_m//e//n//u//l//i//s//_/_s//u//r//a//t/

Fonem:

/f//e//b//i//_//m//e//n//u//l//i//s//_//s//u//r//a//t/

Viseme:

F E P I _ P E T U T I S _ S U R A T

3. Input Teks: Saya suka makan nasi padang

LinkedList Fonem:

/s//a//y//a/ /_ /s//u//k//a/ /_ /m//a//k//a//n/ /_ /
/n//a//s//i/ /_ /p//a//d//a//ng/

Fonem:

/s//a//y//a/ /_ /s//u//k//a/ /_ /m//a//k//a//n/ /_ /
/n//a//s//i/ /_ /p//a//d//a//G/

Viseme:

S A S A _ S U G A _ P A G A T _ T A S I _ P A T A G

4.3 Hidden Markov Model

Tidak seperti *template* citra lainnya, viseme mempunyai identitas di dalam dua media yang berbeda. Media yang pertama adalah dalam domain audio. Viseme seringkali direlasikan pada fonem sebagai unit linguistik. Sedangkan media yang kedua adalah dalam domain citra. Viseme didefinisikan oleh citra dari artikulator manusia, seperti bentuk bibir, pergerakan dagu, dan lain-lain. Viseme adalah unit ekivalen dalam domain visual, tetapi ada yang kurang setuju pada apa yang dimaksud dengan viseme atau berapa banyak model viseme yang tepat untuk sistem pengenalan ucapan dalam audio-visual (LR Rabiner dan Juang, 1986).

Dalam penelitian ini, dipresentasikan sebuah pendekatan yang mengekstraksi viseme-viseme dari kedua domain di atas. Pada domain citra, bentuk mulut direpresentasikan oleh sembilan titik fitur pada kontur di dalam bibir, yang diekstraksi melalui *tracking* wajah dan analisis gambar mulut. Sedangkan dalam domain akustik, segmentasi viseme diperoleh secara otomatis dengan menyelaraskan string fonem pada sinyal audio melalui proses penyelarasan HMM dan algoritma Viterbi. Algoritma Viterbi digunakan untuk menebak urutan state yang paling mungkin dan dapat digunakan untuk menghitung probabilitas yang cocok dari observation state (Ryan dan Nudd, 1993).

Pada Sub-bab ini akan dijelaskan mengenai HMM yang digunakan sebagai *basis engine* pengenalan ucapan. Pembahasan akan meliputi teori tentang

kemunculan pola, definisi HMM, tiga permasalahan dasar dalam HMM, algoritma yang umum digunakan untuk menyelesaikan ketiga masalah tersebut, Implementasi HMM dan Algoritma Viterbi.

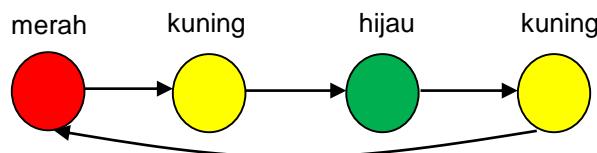
4.3.1 Terminologi HMM

4.3.1.1 Pola Deterministik dan Non-Deterministik

Sering seseorang tertarik untuk mengetahui tentang kemunculan pola yang terjadi dalam perubahan ruang dan waktu. Pola-pola ini muncul di berbagai bidang, misalnya: urutan kata dalam pembicaraan, perubahan cuaca dan sebagainya. Secara garis besar, sifat kemunculan pola dibagi dua jenis, yaitu: pola deterministik dan pola non-deterministik.

1. Pola Deterministik

Sebagai contoh adalah perubahan warna lampu lalu lintas. Urutan munculnya warna lampu seperti pada Gambar 4.4. Berdasar gambar ini seseorang dapat menentukan warna apa yang muncul berikutnya bila dia mengetahui warna yang telah muncul sebelumnya. Jadi keadaan saat ini hanya bergantung pada keadaan sebelumnya.



Gambar 4.4 Urutan Warna yang Muncul pada Lampu Lalu Lintas

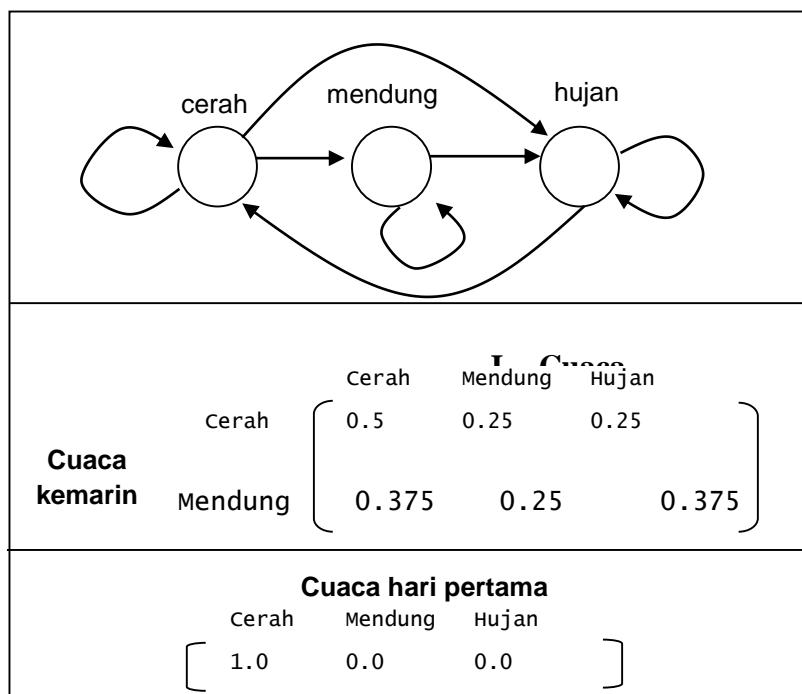
2. Pola Non-Deterministik

Pada suatu sistem yang lain, seseorang tidak dapat menebak secara pasti keadaan saat ini walaupun dia tahu keadaan sebelumnya. Sebagai contoh ialah pada pola perubahan cuaca. Ada tiga jenis cuaca yang hendak diamati, yaitu: cerah, mendung dan hujan. Seseorang tidak akan dapat menggambarkan state diagram seperti gambar 4.4. Salah satu cara untuk menggambarkan sistem ini adalah dengan menggunakan asumsi Markov dan menyederhanakan masalah secara besar-besaran. Asumsi yang akan digunakan adalah: cuaca saat ini tergantung pada cuaca beberapa hari sebelumnya. Dengan kata lain, bila

seseorang memiliki data cuaca beberapa hari yang lalu, maka dia dapat menebak cuaca hari ini.

Dalam asumsi Markov ada nilai yang disebut sebagai order. Yang dimaksud dengan order adalah jumlah keadaan sebelumnya yang mempengaruhi keadaan saat ini. Misalkan asumsi Markov order ke-2, berarti keadaan saat ini dipengaruhi oleh dua keadaan sebelumnya. Untuk pembahasan selanjutnya, yang selalu digunakan adalah asumsi Markov order pertama, yaitu: keadaan saat ini hanya bergantung pada satu keadaan sebelumnya.

Dengan bantuan asumsi Markov, contoh sistem cuaca di atas dapat digambarkan dalam state diagram seperti Gambar 4.5. State diagram ini dilengkapi dengan matriks probabilitas yang menyatakan probabilitas keadaan cuaca hari ini berdasarkan cuaca kemarin. Sebagai contoh, probabilitas cuaca hari ini hujan bila kemarin mendung adalah 0.375.



Gambar 4.5 State Diagram Perubahan Cuaca

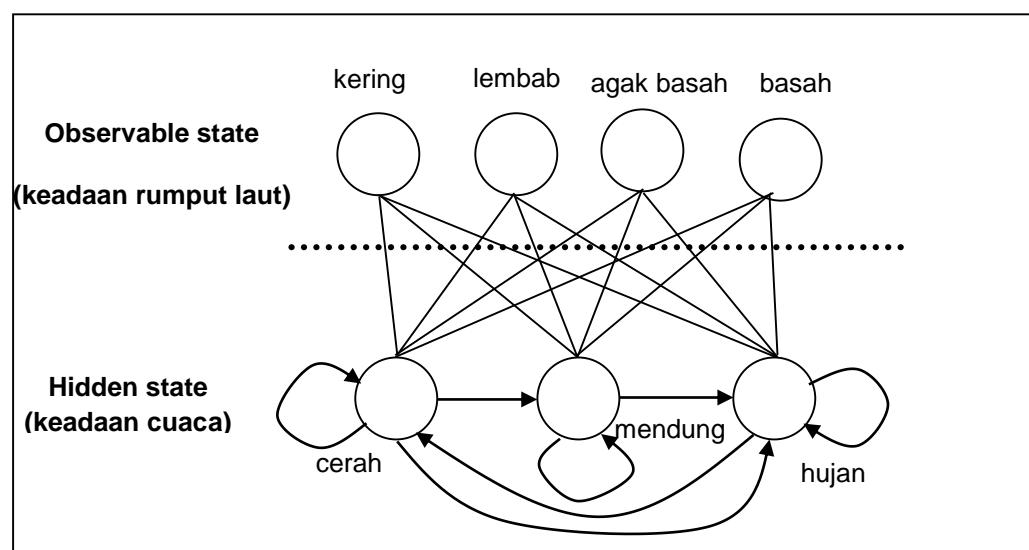
State diagram juga dilengkapi dengan sebuah vektor yang menyatakan keadaan awal (cuaca pada hari pertama). Berdasar vektor tersebut, dapat diketahui bahwa pada hari pertama cuacanya cerah. Sistem cuaca di atas

merupakan contoh dari proses Markov order pertama. Semua sistem yang sejenis dapat didefinisikan dengan cara yang sama. Proses Markov order pertama selalu dinyatakan dalam: State (pada contoh di atas: cerah, mendung, hujan), π vector (pada contoh di atas adalah vektor keadaan awal), State transition matrix.

4.3.1.2 Pola Tersembunyi

Kadangkala proses Markov tidak dapat digunakan untuk menggambarkan suatu sistem. Sebagai contoh kasus seorang pertapa yang tinggal di gua. Si pertapa memiliki sepotong rumput laut yang digunakan untuk mengetahui keadaan cuaca tanpa perlu keluar gua. Rumput laut memiliki empat keadaan: kering, lembab, agak basah dan basah. Keadaan rumput laut ini dipengaruhi oleh keadaan cuaca di luar. Berdasar pengalamannya maka pertapa bisa menebak cuaca di luar hanya dengan melihat keadaan rumput lautnya.

Agar dapat menyatakan sistem seperti ini maka proses Markov harus dikembangkan dengan menambahkan suatu *hidden state*. Pengembangan ini disebut sebagai Hidden Markov Model (HMM). Gambar dari HMM untuk sistem ini dapat dilihat pada Gambar 4.6.



Gambar 4.6 HMM Kasus Pertapa

Pada Gambar 4.6 dapat dilihat bahwa dalam HMM terdapat dua jenis state, observable state dan hidden state:

1. Observable State

Observable state menyatakan data observasi atau data pengamatan. Dalam contoh kasus ini yang menjadi data pengamatan adalah keadaan rumput laut.

2. Hidden State

Hidden state adalah keadaan yang harus ditebak (tidak dapat diamati secara langsung). Dalam contoh kasus ini, yang menjadi hidden state adalah cuaca di luar gua.

Contoh lebih nyata dari sistem yang bisa dimodelkan dengan HMM adalah pengenalan pengucapan dan pengenalan coretan tulisan tangan. Pada sistem pengenalan pengucapan, observable state adalah fitur dari setiap fonem, sedangkan hidden statenya adalah kata apa yang harus dikenali. Pada sistem pengenalan coretan tulisan tangan, observable state adalah kode arah (chain code) dari coretan, dan hidden state-nya adalah huruf apa yang harus dikenali.

		Kering	Lembab	Agak basah	Basah
Cuaca	Cerah	0.60	0.20	0.15	0.05
	Mendung	0.25	0.25	0.25	0.25
	Hujan	0.05	0.10	0.35	0.50

Gambar 4.7 Confusion Matrix Hubungan Antara Keadaan Rumput Laut dan Cuaca

Seperti halnya proses Markov, maka sistem ini juga dinyatakan dalam state, vektor π dan *transition matrix*. Selain itu diperlukan lagi sebuah matriks yang menyatakan hubungan antara observable state dan hidden state. Matriks ini disebut sebagai *confusion matrix*. Confusion matrix yang menggambarkan hubungan antara keadaan rumput laut dan cuaca di luar gua dapat dilihat pada Gambar 4.7. Sebagai contoh, dari matriks tersebut dapat diketahui bahwa probabilitas cuaca hujan bila rumput laut basah adalah 0.5.

Jadi, semua sistem yang bisa dimodelkan dalam suatu HMM selalu terdiri dari: *Observable state*, *Hidden state*, π vector, *State transition matrix*, *Confusion matrix*.

4.3.2 Definisi HMM

Barisan observasi yang dinyatakan sebagai: $\bar{o} = o_1, o_2, \dots, o_T$ dimana $o_i \in \{O_1, O_2, \dots, O_M\}$. Probabilitas yang menyatakan bahwa state S_i akan menghilangkan observasi O_j adalah $P(O_j|S_i)$, dimana $i = 1, 2, \dots, N$ dan $j = 1, 2, \dots, M$. N adalah jumlah state yang mungkin dan M adalah jumlah observasi yang mungkin.

HMM selalu dinyatakan dalam tiga bagian: *vector π* , *transition matrix A*, dan *observation matrix B*, sehingga parameter model HMM dapat ditulis sebagai persamaan 4.1:

$$\lambda = (A, B, \pi) \quad (4.1)$$

Definisi dari masing-masing bagian akan dijelaskan sebagai berikut:

1. *Vector π (initial state distribution vector)*, $\pi_i = P(S_i)$.

Vector π , menyatakan probabilitas awal data observasi pertama berada pada state nomer berapa. Ukuran vektornya adalah N , dimana N adalah jumlah state. Total jumlah nilai vektor ini pasti sebesar 1.0, yang data dinyatakan seperti pada persamaan 4.2:

$$\pi = \{\pi_1, \dots, \pi_N\} \text{ dan } \sum_{i=1}^N \pi_i = 1 \quad (4.2)$$

2. *Matrix A (transition state probability matrix)*, $A_{ij} = P(S_j|S_i)$.

Matriks A menyatakan probabilitas perpindahan state. Setiap elemen A_{ij} dari matriks ini menyatakan: berapa probabilitas bila observasi saat ini berada pada state-i dan observasi berikutnya berada pada state-j. Sesuai asumsi Markov order pertama, nilai probabilitas tidak bergantung pada waktu observasi. Nilai ini hanya bergantung pada satu observasi sebelumnya. Ukuran matriks A adalah $N \times N$, dimana N adalah jumlah state. Total nilai

suatu baris dalam matriks ini pasti sebesar 1.0, yang dinyatakan seperti pada persamaan 4.3:

$$A = \{a_{ij}\}, 1 \leq i, j \leq N \text{ dan } \sum_{j=1}^N a_{ij} = 1 \quad (4.3)$$

3. *Matrix B, (observation probability matrix), $B_{ij} = P(O_j/S_i)$.*

Matriks B menyatakan berapa probabilitas suatu simbol observasi berada dalam suatu state. Ukuran matriks B adalah $N \times M$, dimana N adalah jumlah state dan M adalah jumlah jenis simbol yang unik dalam data observasi. Setiap elemen matrix B_{ij} menyatakan berapa probabilitas simbol j berada dalam state- i , yang dinyatakan seperti pada persamaan 4.4:

$$B = \{b_i(v_j)\}, 1 \leq i \leq N, 1 \leq j \leq M \text{ dan } \sum_{k=1}^M b_i(v_k) = 1 \quad (4.4)$$

4.3.3 Tiga Masalah HMM

Ada tiga permasalahan dasar yang muncul pada HMM. Permasalahan itu adalah:

1. Permasalahan Evaluation

Masalah ini timbul bila terdapat beberapa HMM dan sebuah urutan data observasi (*observable state*). Pertanyaan yang muncul adalah: dari HMM-HMM itu, manakah yang paling “cocok” untuk data observasi yang ada? HMM yang paling cocok adalah HMM yang paling mungkin untuk menghasilkan urutan state yang serupa dengan data observasi. Sebagai contoh, bila terdapat sehimpunan HMM dan terdapat urutan data keadaan rumput laut (basah, basah, kering, lembab,...), pertanyaan yang harus dijawab adalah: HMM mana yang kiranya menghasilkan urutan seperti itu?

2. Permasalahan Decoding

Permasalahan decoding muncul bila terdapat sebuah HMM dan sebuah urutan data observasi (*observable state*). Pertanyaan yang harus dijawab adalah: urutan hidden state seperti apa yang paling cocok untuk pasangan HMM dan observable state yang ada? Sebagai contoh, terdapat sebuah HMM dan sebuah

urutan data keadaan rumput laut (basah, basah, kering, lembab,...).

Pertanyaannya adalah, bagaimanakah urutan pergantian cuaca di luar gua?

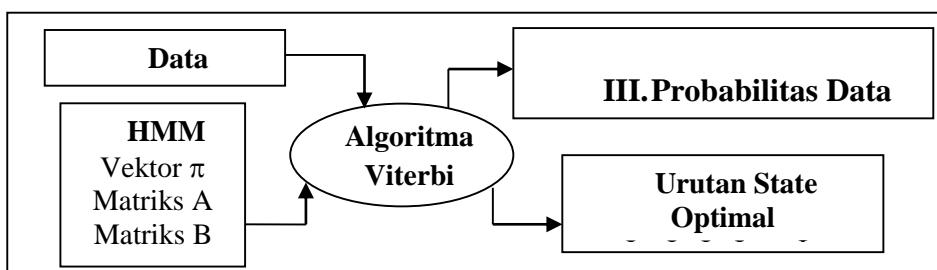
3. Permasalahan Training

Permasalahan training adalah permasalahan yang paling kompleks. Bila terdapat sehimpunan urutan observable state dan sehimpunan urutan hidden state, yang harus dipecahkan adalah: buatlah sebuah HMM yang paling cocok untuk kedua himpunan itu. Sebagai contoh, si pertapa memiliki banyak data yang berupa pasangan antara keadaan rumput laut dan keadaan cuaca di luar gua. Berdasar data yang dimilikinya ia harus membuat sebuah HMM yang bisa memodelkan hubungan antar data tersebut.

Untuk menyelesaikan ketiga permasalahan di atas, telah dikembangkan beberapa algoritma. Untuk permasalahan evaluasi, algoritma yang bisa dipakai: algoritma Forward atau algoritma Backward dan algoritma Viterbi. Untuk permasalahan decoding, hanya algoritma Viterbi yang bisa digunakan. Untuk masalah training bisa digunakan algoritma Baum-Welch atau Segmental k-Means. Dalam penelitian ini, masalah yang muncul adalah masalah recognition dan training, sehingga pembahasan akan lebih difokuskan pada algoritma Viterbi dan Segmental k-Means saja.

4.3.4 Algoritma Viterbi dan HMM

Algoritma Viterbi digunakan untuk menebak urutan state yang paling mungkin dari suatu data observasi dan HMM. Selain itu algoritma ini juga bisa menghitung berapa nilai probabilitas kecocokan antara data observasi dan HMM. Diagram input output untuk algoritma ini dapat dilihat pada Gambar 4.8.



Gambar 4.8 Diagram Input-Output Algoritma Viterbi

Kriteria optimalitas, terdiri dari pemilihan urutan state (atau path) yang memiliki probabilitas maksimum berkaitan dengan model yang diberikan. Urutan ini dapat ditentukan secara rekursif melalui algoritma Viterbi. Algoritma Viterbi digunakan untuk menghasilkan urutan paling mungkin dari urutan observasi state yang diberikan dan memiliki model seperti Persamaan (4.5):

$$\arg \max_S P(S | O, A, B, \pi) \quad (4.5)$$

Pada setiap time frame dalam persamaan (4.5), struktur data diperbarui dengan mencari jalan yang memiliki probabilitas tertinggi untuk setiap state. Ketika seluruh urutan observasi telah dihabiskan, lakukan backtrack untuk memulihkan urutan state. Sekarang beralih ke masalah penentuan urutan state optimal, gunakan algoritma Viterbi sesuai dengan langkah-langkah algoritma Viterbi berikut:

1. Inisialisasi

Hitung probabilitas O_1 pada state tertentu. Probabilitas ini dihitung dengan mengalikan vektor π dengan matriks B. Kolom B yang dipilih harus sesuai dengan simbol pada O_1 . Jadi O_1 pada state ke- i dapat ditunjukkan pada persamaan 4.6.

$$O_1(i) = \pi_i \cdot B_{i,S_1} \quad i = 1, 2, \dots, N_s \quad (4.6)$$

dengan π_i adalah probabilitas sebelum berada di state S pada saat $n = 1$.

2. Induksi atau Rekursi

Hitung probabilitas dari O_2, O_3, \dots, O_T di beberapa state tertentu. Probabilitas ini dihitung dari perkalian matriks A dan B dengan probabilitas sebelumnya. State mana yang harus dikunjungi, ditentukan dari hasil perkalian yang memberikan nilai terbesar. Oleh karena itu, jalan terbaik ditentukan oleh induksi, yang ditunjukkan pada persamaan 4.7.

$$O_n(j) = \max_{1 \leq i \leq N_s} (O_{n-1}(i) \cdot A_{i,j} \cdot B_{j,S_n}), \quad 1 \leq i \leq N_s, 1 \leq j \leq N_s, 2 \leq n \leq N \quad (4.7)$$

3. Terminasi

Urutan state optimal adalah urutan state yang dikunjungi pada langkah sebelumnya. Nilai probabilitas akhir juga diperlakukan sebagai output. Cari kemungkinan terbaik ketika di akhir urutan observasi tercapai, seperti persamaan 4.8:

$$P(S | A, B, \pi) = \max_{1 \leq i \leq N_s} O_N(i) \quad (4.8)$$

Untuk penjelasan tentang langkah-langkah algoritma Viterbi yang digunakan dalam metodologi penelitian ini, akan diuraikan pada Implementasi HMM dan Algoritma Viterbi di bawah ini.

4.3.5 Implementasi HMM dan Algoritma Viterbi

Pada sub-bab ini menjelaskan bagaimana proses implementasi Hidden Markov Model (HMM) dan algoritma Viterbi dilakukan. Proses implementasi HMM dan algoritma Viterbi ini dilakukan setelah proses segmentasi input teks. Implementasi HMM dan algoritma Viterbi dilakukan pada setiap viseme kalimat input untuk mendapatkan state panjang frame viseme yang terbaik. Implementasi HMM dan algoritma Viterbi terdiri dari 2 proses utama, yaitu pemisahan data training *Inesh² Corpora*, serta proses decoding menggunakan algoritma Viterbi.

4.3.5.1 Pemisahan Data Training *Inesh² Corpora*

4.3.5.1.1 Algoritma Pemisahan Data Training

Langkah pertama pada proses implementasi HMM dan algoritma Viterbi adalah pemisahan data training *Inesh² Corpora*. Penjelasan tentang *Inesh² Corpora* dibahas pada subbab tersendiri. Pemisahan data training ini didasarkan pada data yang sudah diperoleh, yaitu kalimat dan urutan statenya. Pemisahan data training ini bergantung pada parameter afeksi/emosi yang diinputkan, sebab data training untuk setiap parameter afeksi/emosi yang diinputkan berbeda-beda.

Format data training yang sudah disimpan dalam bentuk *textfile* akan terdiri dari 2 baris sekali, dimana baris pertama adalah kalimat data training, dan baris kedua adalah urutan state yang benar untuk kalimat pada baris pertama. Berikut adalah algoritma pemisahan data training yang dilakukan.

Algoritma 4.4 Pemisahan Data Training

- 1: Pilih file data training sesuai dengan parameter afeksi/emosi
- 2: Baca nama file data training
- 3: Ambil baris saat ini dari file yang sedang dibaca
- 4: Simpan kalimat data training pada sebuah ArrayList
- 5: Ambil baris berikutnya dari file yang sedang dibaca
- 6: Simpan jumlah Trigram State pada sebuah HashMap

```
7: Simpan jumlah Bigram State pada sebuah HashMap
8: Simpan jumlah State->Viseme pada sebuah HashMap
9: Simpan jumlah State pada sebuah HashMap
10: Lanjutkan ke baris selanjutnya pada file yang sedang dibaca
11: Ulangi proses 3 hingga 10 sampai keseluruhan file selesai
dibaca
```

Pada algoritma 4.4, langkah pertama yang dilakukan adalah memilih file data training yang sesuai dengan parameter afeksi/emosi yang sudah diinputkan oleh user. Selanjutnya file data training tersebut akan dibaca dan diambil barisnya satu persatu, dimana setiap baris ganjil merupakan kalimat training, dan baris genap merupakan urutan state yang benar untuk kalimat pada baris sebelumnya.

Kalimat training akan disimpan dalam sebuah *ArrayList*, kemudian, pada baris genap akan diambil jumlah Trigram State, jumlah Bigram State, jumlah State terhadap Viseme (*State → Viseme*) dan jumlah State yang ada dan disimpan pada sebuah *HashMap*. Kemudian akan dilanjutkan ke baris selanjutnya. Proses ini akan berulang sampai keseluruhan file data training sudah selesai dibaca, dimana setiap *HashMap* akan sudah diisi oleh data dari keseluruhan file.

4.3.5.1.2 Penerapan Pemisahan Data Training *Inesh² Corpora*

Berdasarkan algoritma 4.4, langkah-langkah dalam pemisahan data training yang pertama kali dilakukan adalah memilih file data training sesuai dengan parameter afeksi/emosi yang diinginkan, misalkan input parameter afeksi adalah “Netral”, beri nama file data trainingnya “*datanetral.txt*”. Untuk input parameter afeksi “Marah”, nama file-nya “*datamarah.txt*”, dan seterusnya. Nama file dari 8 parameter afeksi lainnya, seperti terlihat pada Tabel 4.3. Default nama file data training *Inesh² Corpora* adalah “*data.txt*”.

File “*data.txt*” menampung data urutan viseme yang memiliki parameter sesuai afeksinya. Setiap file data training memuat 2 baris informasi, baris pertama menunjukkan parameter input kalimat yang sudah menjadi urutan viseme dan baris kedua merupakan parameter afeksi/emosi yang disimpan dalam urutan state (*frame*) yang benar untuk kalimat pada baris pertama. Urutan viseme dan *frame* di dalam file data training untuk membantu mendapatkan urutan viseme yang alami.

Tabel 4.3 Nama File Data Training *Inesh² Corpora* untuk Parameter Afeksi

Nomor	Parameter Afeksi	Nama File Data Training
1	Jijik	<i>Datajijik.txt</i>
2	Malu	<i>Datamalu.txt</i>
3	Takut	<i>Datatakul.txt</i>
4	Sedih	<i>Datasedih.txt</i>
5	Netral	<i>Datanetral.txt</i>
6	Marah	<i>Datamarah.txt</i>
7	Senang	<i>Datasenang.txt</i>
8	Terkejut	<i>Dataterkejut.txt</i>

Tabel 4.4 Enam Contoh Kalimat Training yang Berafeksi Netral

Barisan Informasi Data Training	Parameter Input Kalimat Training (<i>Inesh² Corpora</i>) Berafeksi Netral
<i>Data Training 1</i>	**SASA_SAKIT_PERUT
Urutan Viseme	**SASA_SAGIT_PERUT
Frame	*0F,*0F,S1F,A2F,S1F,A2F,_2F,S1F,A2F,G1F,I2F,T1F,_2F,P1F,E2F,R1F,U4F,T1F
<i>Data Training 2</i>	**KUCING_SAYA_MATI
Urutan Viseme	**GUSIG_SASA_PATI
Frame	*0F,*0F,G1F,U2F,S1F,I2F,G1F,_2F,S1F,A2F,S1F,A2F,_2F,P1F,A2F,T1F,I4F
<i>Data Training 3</i>	**SAYA_SANGAT_MENYESAL_DENGAN_PERBUATAN_SAYA
Urutan Viseme	**SASA_SAGAT_PESESAT_TEGAT_PERPUATAT_SASA
Frame	*0F,*0F,S1F,A2F,S1F,A2F,_2F,S1F,A2F,G1F,A2F,T1F,_2F,P1F,E2F,S1F,E2F,S1F,A3F,T1F,_3F,T1F,E2F,G1F,A2F,T1F,_2F,P1F,E2F,R1F,P1F,U1F,A2F,T1F,A2F,T1F,_2F,S1F,A2F,S1F,A2F
<i>Data Training 4</i>	**KUCING_SAYA_MATI
Urutan Viseme	**GUSIG_SASA_PATI
Frame	*0F,*0F,G2F,U1F,S1F,I2F,G2F,_2F,S1F,A2F,S1F,A2F,_2F,P1F,A2F,T1F,I4F
<i>Data Training 5</i>	**ANJING_SAYA_TIGA
Urutan Viseme	**ATSIG_SASA_TIGA
Frame	*0F,*0F,A2F,T1F,S1F,I2F,G1F,_2F,S1F,A2F,S1F,A2F,_2F,T1F,I2F,G1F,A4F
<i>Data Training 6</i>	**SAYA_TIDAK_NAIK_KELAS
Urutan Viseme	**SASA_TITAG_TAIG_GETAS
Frame	*0F,*0F,S1F,A2F,S1F,A2F,_2F,T1F,I2F,T1F,A2F,G1F,_2F,T1F,A2F,I2F,G1F,_2F,G1F,E2F,T1F,A2F,S1F

Tabel 4.4 menunjukkan 6 contoh kalimat training yang memuat baris informasi dalam parameter afeksi Netral. Setiap data training memberikan urutan viseme dan urutan state dalam satuan *frame* dari sebuah contoh kalimat training.

Pada Tabel ini, terlihat bahwa data training pada parameter input kalimat training, terdapat “*” (tanda bintang) atau *star marker*. *Star marker* ini membantu sebagai penanda awal untuk dimulainya gerak bibir dari posisi diam (*silent*), yang seolah-olah membutuhkan 2 *frame* (* = 1 *frame*, ** = 2 *frame*). Dari observasi yang telah dilakukan, tanpa ada penambahan *starter marker* ini, membuat gerakan bibir pada saat pengucapan menjadi tidak sinkron (ada penundaan atau *delay*), sehingga pada setiap data training yang diinputkan, maka selalu diberi tambahan “**” pada awal kalimat training dan awal *state*.

Urutan state dengan satuan *frame*, merepresentasikan kebutuhan *frame* dalam setiap viseme yang muncul. Setiap *frame* yang muncul (0, 1, 2, 3 atau 4) mewakili *range* kebutuhan dari sejumlah *frame* berdasarkan panjang durasi pengucapan. Tabel 4.5 berikut merepresentasikan panjang durasi pengucapan dari pergerakan animasi gerakan bibir (*lip movement*) pada setiap *frame* untuk setiap viseme dengan satuan tertentu.

Tabel 4.5 Representasi Frame dari Animasi Gerakan Bibir

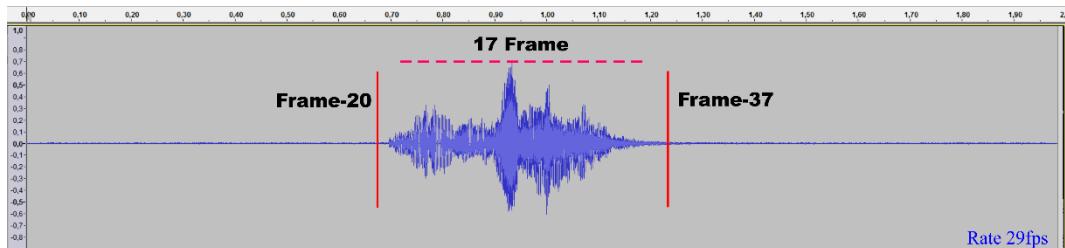
Frame	Range (per detik)
0	0
1	1 – 3
2	4 – 6
3	7 – 9
4	10 – 12

Frame 0 memiliki range 0 detik, yang merepresentasikan viseme diam atau (*silent*). Frame 1 memiliki range 1 sampai 3 detik, yang merepresentasikan pengucapan urutan viseme dalam sebuah parameter afeksi. Sedangkan Frame 2 merepresentasikan sebuah pengucapan urutan viseme yang memiliki range 4 sampai 6 detik. Frame 3 dan Frame 4 masing-masing merepresentasikan pengucapan urutan viseme yang memiliki range 7 sampai 9 detik dan 10 sampai 12 detik. Penentuan range ini dilandasi dari hasil sampel perhitungan frekuensi yang menggunakan *panjang frame* dan observasi dari perekaman beberapa suara model seperti yang ditunjukkan pada Gambar 4.9.

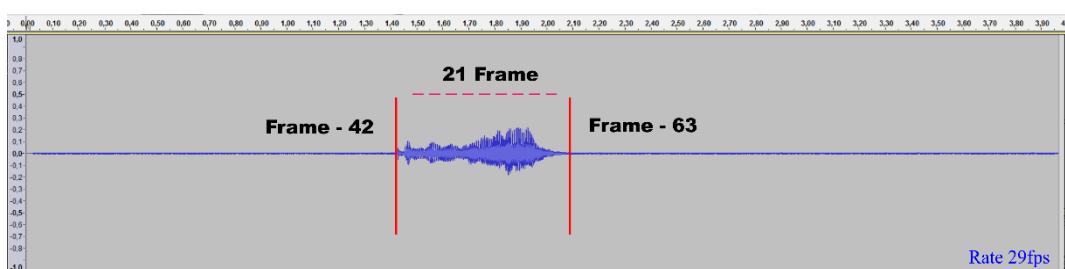
Pada Gambar 4.9 ini, merupakan salah satu contoh perhitungan frekuensi *panjang frame* pada sebuah contoh kata “keluar” yang diucapkan oleh seorang

model untuk masing-masing parameter afeksi. Kata “keluar” terdiri dari 6 fonem atau 6 viseme, sehingga dari panjang *frame* yang dihasilkan pada setiap parameter afeksi dibagi dengan 6. Keseluruhan parameter afeksi yang digunakan adalah 8 parameter afeksi.

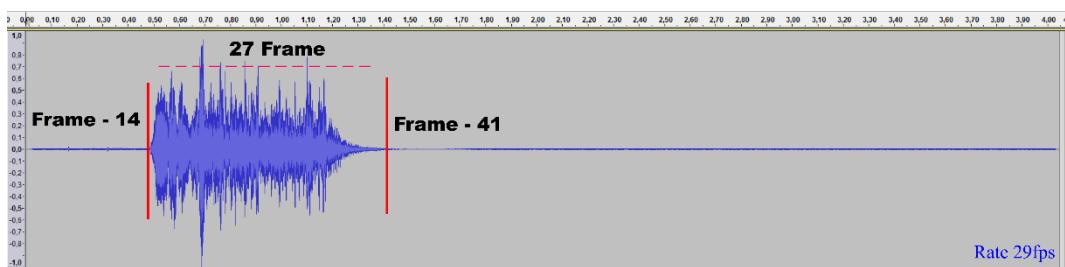
Frekuensi panjang *frame* berasal dari sampel pengucapan suara model dari parameter afeksi “jijik” dapat dilihat pada Gambar 4.9 (a). Sedangkan Gambar 4.9 (b) merupakan panjang frame untuk parameter afeksi “malu”, (c) untuk parameter afeksi “marah”, (d) untuk parameter afeksi “netral”, (e) untuk parameter afeksi “sedih”, (f) untuk parameter afeksi “senang”, (g) untuk parameter afeksi “takut”, (h) untuk parameter afeksi “terkejut”. Berikut ini adalah semua Gambar 4.9 dari (a) sampai dengan (h).



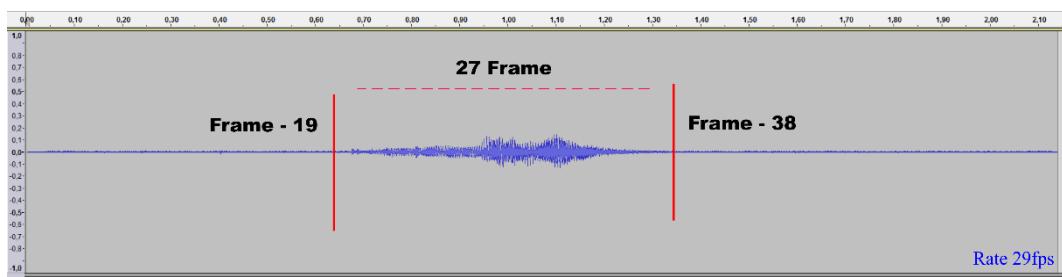
(a) parameter afeksi “jijik”



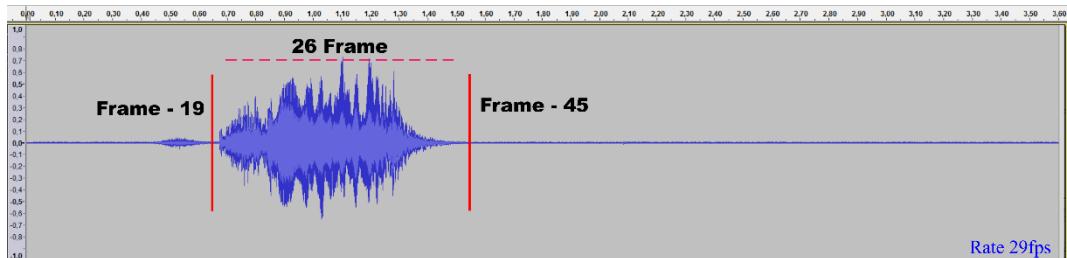
(b) parameter afeksi “malu”



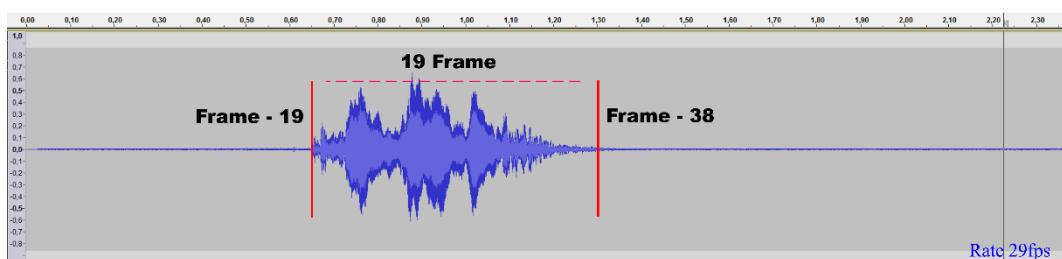
(c) parameter afeksi “marah”



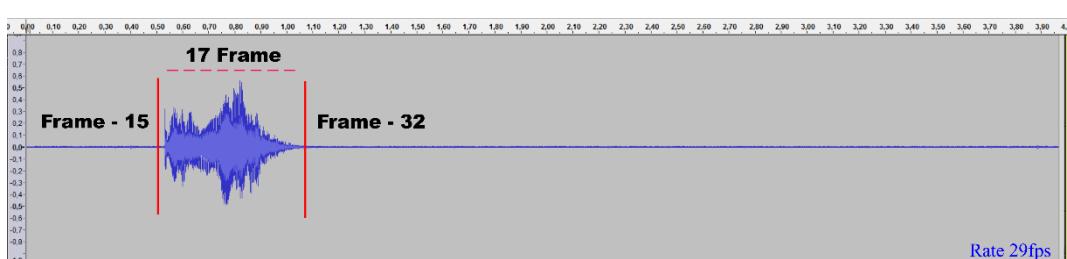
(d) parameter afeksi “netral”



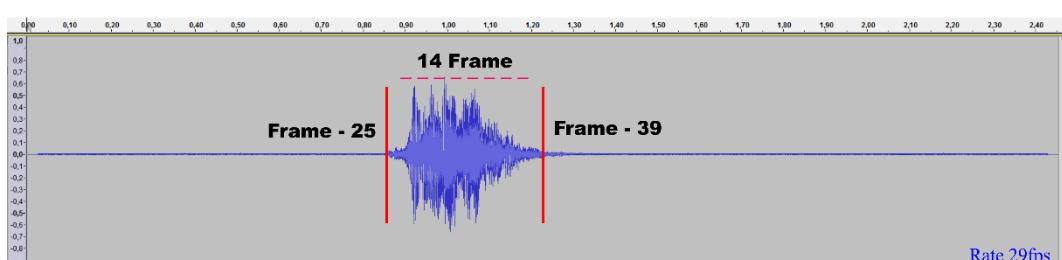
(e) parameter afeksi “sedih”



(f) parameter afeksi “senang”



(g) parameter afeksi “takut”



(h) parameter afeksi “terkejut”

Gambar 4.9 Frekuensi Panjang *Frame* dari Audacity pada Sebuah Contoh Kata “Keluar” untuk Masing-masing Parameter Afeksi

Dengan menggunakan hasil *capture* pada Gambar 4.9, ringkasan data panjang *frame rate* untuk setiap parameter afeksi dituliskan pada Tabel 4.6. Tabel ini menunjukkan panjang *frame rate* dari sebuah kata “*keluar*” (6 viseme) yang diucapkan oleh seorang model untuk masing-masing parameter afeksi.

Tabel 4.6 Contoh Panjang Frame Rate dari Parameter Afeksi

No	Parameter Afeksi	Frame Rate dari 6 viseme (satuan frame)	Rata-rata Frame Rate dari 1 viseme (satuan frame)
1	Jijik	17	2,83
2	Malu	21	3,50
3	Marah	27	4,50
4	Netral	27	4,50
5	Sedih	26	4,30
6	Senang	19	3,17
7	Takut	17	2,83
8	Terkejut	14	2,33

Dari hasil observasi ini, maka penentuan setiap urutan *state* yang diinputkan diwakili oleh 0, 1, 2, 3, dan 4 frame. Penentuan jenis frame ini didasarkan pada Gambar 4.9 dan Tabel 4.6, dengan rata-rata dari frame rate adalah 4,5 atau dibulatkan ke bawah menjadi 4. Sehingga urutan state dengan 4 frame merupakan frame maksimum. Sedangkan untuk panjang durasi pengucapan antara 10 - 12 detik merupakan durasi terpanjang. Angka ini diperoleh dari observasi rekaman beberapa model.

Langkah-langkah selanjutnya dalam pemisahan data training berdasarkan algoritma 4.4 adalah menyimpan urutan state dalam satuan *frame* pada sebuah variabel bertipe *HashMap*. Urutan *frame* ini kemudian ditampung dan disimpan dalam jumlah *Trigram State*, Jumlah *Bigram State*, Jumlah (*State* → *Viseme*) dan Jumlah *State*.

Contoh perhitungan Jumlah *Trigram State*, Jumlah *Bigram State*, Jumlah (*State* → *Viseme*) dan Jumlah *State* diambilkan dari data pada Tabel 4.4 yang berisi urutan viseme dan urutan frame dari file data training (*Inesh² Corpora*) yang disimpan sebagai “*datanetral.txt*” seperti Gambar 4.10.

Gambar 4.10 ini berisi 2 baris informasi dari 6 contoh kalimat training. Total baris keseluruhan sebanyak 12 baris. Baris-baris ganjil atau baris pertama pada

setiap kalimat training menunjukkan urutan viseme, sedangkan baris-baris genap atau baris kedua pada setiap kalimat training menunjukkan urutan state (*frame*).

```
**SASA_SAGIT_PERUT
*0F,*0F,S1F,A2F,S1F,A2F,_2F,S1F,A2F,G1F,I2F,T1F,_2F,P1F,E2F,R1F,U4F,T1F
**GUSIG_SASA_PATI
*0F,*0F,G1F,U2F,S1F,I2F,G1F,_2F,S1F,A2F,S1F,A2F,_2F,P1F,A2F,T1F,I4F
**SASA_SAGAT_PESESAT_TEGAT_PERPUATAT_SASA
*0F,*0F,S1F,A2F,S1F,A2F,_2F,S1F,A2F,G1F,A2F,T1F,_2F,P1F,E2F,S1F,E2F,S1F
,A3F,T1F,_3F,T1F,E2F,G1F,A2F,T1F,_2F,P1F,E2F,R1F,P1F,U1F,A2F,T1F,A2F,T1
F,_2F,S1F,A2F,S1F,A2F
**GUSIG_SASA_PATI
*0F,*0F,G2F,U1F,S1F,I2F,G2F,_2F,S1F,A2F,S1F,A2F,_2F,T1F,A2F,T1F,I4F
**ATSIG_SASA_TIGA
*0F,*0F,A2F,T1F,S1F,I2F,G1F,_2F,R1F,A2F,T1F,O2F,_2F,T1F,I2F,G1F,A4F
**SASA_TITAG_TAIG_GETAS
*0F,*0F,S1F,A2F,S1F,A2F,_2F,T1F,I2F,T1F,A2F,G1F,_2F,T1F,A2F,I2F,G1F,_2F
,G1F,E2F,T1F,A2F,S1F
```

Gambar 4.10 Informasi Urutan Viseme dan Frame dari *Inesh2 Corpora* pada Contoh Data Training

Jumlah *Trigram State*, diperoleh dengan cara membaca 3 *string* secara berurutan dan membentuk kombinasi *string*. Kombinasi *string* merupakan gabungan beberapa *text string* ke dalam 1 *text string* yang ditampung pada variabel *HashMap*. Hasil urutan masing-masing kombinasi *string* dihitung frekuensi jumlah urutannya. Penghitungan Jumlah *Bigram State* memiliki pola yang sama dengan Jumlah *Trigram State*, akan tetapi cara membacanya dengan 2 *string* berurutan. Kombinasi dari 2 *string* ini, kemudian dihitung Jumlah *Bigram State* yang dihasilkan. Rekapitulasi Jumlah *Trigram State* dan Jumlah *Bigram State* pada setiap kombinasi yang dihasilkan dari urutan frame dapat dilihat pada Tabel 4.7.

Setelah perhitungan Jumlah *Trigram State* dan Jumlah *Bigram State* selesai dilakukan, langkah selanjutnya adalah menghitung Jumlah *State → Viseme* dan Jumlah *State*. Jumlah *State → Viseme* diperoleh dengan cara menghitung frekuensi dari *Frame* ke *Viseme*. Sedangkan perhitungan Jumlah *State*, diperoleh dari menghitung banyaknya *state* pada setiap jenis frame 0, 1, 2, 3, dan 4 dari seluruh data dalam kalimat training. Hasil perhitungan Jumlah *State → Viseme* dan Jumlah *State* ditampung di variabel *HashMap*, seperti yang ditunjukkan pada Tabel 4.8.

Tabel 4.7 Jumlah *Trigram State* dan *Bigram State* dari Urutan *Frame*

Jumlah Trigram State			Jumlah Bigram State					
String ke		Kombinasi String	Jumlah	String ke		Kombinasi String	Jumlah	
1	2	3		1	2			
0	0	1	001	5	0	00	6	
0	0	2	002	1	0	01	5	
0	1	2	012	5	0	02	1	
0	2	1	021	1	1	11	3	
1	1	1	111	1	1	12	49	
1	1	2	112	2	1	13	2	
1	2	1	121	41	1	14	4	
1	2	2	122	7	2	21	49	
1	3	1	131	2	2	22	7	
1	4	1	141	1	3	31	2	
2	1	1	211	2	4	41	1	
2	1	2	212	41				
2	1	3	213	1				
2	1	4	214	4				
2	2	1	221	7				
3	1	2	312	1				
3	1	3	313	1				

Tabel 4.8 Jumlah *State* → *Viseme* dan Jumlah *State*

Jumlah State ke Viseme					Jumlah State	
Frame	→	Viseme	Kombinasi String	Jumlah	Frame	Jumlah Frame/State
0	→	*	0*	12	0	12
1	→	U	1U	1	1	60
1	→	T	1T	17	2	57
1	→	S	1S	22	3	2
1	→	R	1R	2	4	4
1	→	P	1P	7		
1	→	G	1G	11		
2	→	_	2_	15		
2	→	I	2I	6		
2	→	E	2E	7		
2	→	A	2A	28		
2	→	U	2U	1		
3	→	A	3A	1		
3	→	_	3_	1		
4	→	A	4A	1		
4	→	U	4U	1		
4	→	I	4I	2		

Seluruh hasil perhitungan dari Proses Pemisahan Data Training dalam Implementasi HMM, akan digunakan untuk proses selanjutnya, yaitu: Proses *Decoding* dengan menggunakan Algoritma Viterbi.

4.3.5.2 Proses *Decoding* dengan Menggunakan Algoritma Viterbi

4.3.5.2.1 Algoritma Proses *Decoding* dengan Menggunakan Algoritma Viterbi

Setelah proses pemisahan data training selesai dilakukan, proses yang akan dilakukan selanjutnya adalah proses *decoding* menggunakan algoritma *Viterbi*. Proses ini akan menghitung probabilitas *state* berdasarkan data training yang sudah diproses, yang didapat dari proses-proses sebelumnya. Proses *decoding* ini akan menggunakan 2 matriks HMM, yaitu matriks A (*transition matrix*) dan matriks B (*observation matrix*). Kedua matriks ini akan diisi dengan data yang berbeda dengan HMM biasa, karena dalam proses *decoding* ini menggunakan *Trigram HMM*.

Algoritma 4.5 Proses Decoding Algoritma Viterbi

- 1: Lakukan pengambilan hasil data training dari file
- 2: Lakukan proses perhitungan pada data observasi
- 3: Ambil dan simpan nilai maksimum dari perhitungan untuk data observasi yang bersangkutan
- 4: Simpan state yang paling baik untuk data observasi
- 5: Lakukan proses 2 sampai dengan 4 hingga seluruh data observasi selesai dihitung

Pada algoritma 4.5, langkah pertama yang dilakukan tentunya adalah pengambilan hasil data training dari file yang sudah dihitung yaitu pada file *datalist.txt*. Setelah pengambilan file selesai, proses perhitungan *Viterbi* pada data observasi akan dilakukan. Proses perhitungan ini akan dilakukan sampai semua state yang mungkin sudah selesai dihitung, dan setelah itu akan diambil *state* yang paling baik dari nilai maksimum yang sudah didapat. State yang paling baik tersebut akan disimpan untuk data observasi yang bersangkutan. Proses perhitungan ini akan dilakukan sampai semua data observasi selesai dihitung dan mempunyai *state* yang paling baik.

Langkah-langkah yang dilakukan pada Algoritma *Viterbi* untuk Proses *Decoding* adalah sebagai berikut:

1. Inisialisasi matriks A, matriks B, dan *vektor* π .
2. Matriks A diisi data training dengan menggunakan persamaan 4.9.

$$a(z | x, y) = \frac{c(x, y, z)}{c(x, y)} \quad (4.9)$$

3. Matriks B diisi data training dengan menggunakan persamaan 4.10.

$$b(z | S) = \frac{c(S \rightarrow z)}{c(S)} \quad (4.10)$$

4. Hitung probabilitas dari masing-masing *state* dengan menggunakan persamaan 4.11.

$$V_{t,k} = \max_{x \in S} [P(y_t | k) \cdot a_{x,k} \cdot V_{t-1,x}] \quad (4.11)$$

5. Simpan *state* terbaik pada perhitungan state berikutnya dengan menggunakan langkah (4) di atas.
6. Ulangi langkah (4) dan (5) dari proses di atas sampai semua perhitungan telah diselesaikan.

Proses perhitungan algoritma Viterbi dimulai dengan menggunakan persamaan 4.11, dimana $V_{t,k}$ adalah perhitungan probabilitas untuk urutan state yang paling mungkin untuk observasi $ke - t$ yang mempunyai k sebagai kemungkinan statenya. Persamaan 4.11 ini, akan menghitung nilai maksimum dari setiap perhitungan, $P(Y_t/k)$ akan menghitung probabilitas untuk observasi y yang $ke - t$ saat diberikan sebuah state k . $P(Y_t | k)$ ini dapat ditemukan pada matriks B atau matriks observasi. $a_{x,k}$ akan menghitung nilai probabilitas transisi dari state x yang merupakan state sebelumnya menuju state k yang merupakan state sekarang. $a_{x,k}$ ini dapat ditemukan pada matriks A atau matriks transisi. $V_{t-1,x}$ merupakan perhitungan probabilitas dari observasi $ke - (t-1)$ atau observasi sebelumnya pada state x , dimana nantinya akan dipilih nilai maksimum dari perhitungan-perhitungan tersebut, dan akan dipilih sebuah state k sebagai state final untuk observasi $ke - t$.

Untuk dapat mencapai perhitungan seperti pada persamaan 4.11 dalam bentuk *trigram HMM*, maka diperlukan parameter yang lain yang harus dihitung khusus untuk bentuk *trigram HMM*. Parameter ini akan terdiri dari 2 macam, yaitu

matriks A dan matriks B khusus untuk *trigram HMM*. Matriks A *trigram HMM* masih merupakan matriks transisi, hanya saja transisi pada matriks ini merupakan transisi dari state bigram menuju state trigram. Perhitungan nilai matriks A trigram HMM dapat digambarkan dengan persamaan 4.9, dengan $a(z/x,y)$ merupakan perhitungan dari salah satu perhitungan nilai matriks A milik *trigram HMM* saat transisi dari *bigram state* x,y menuju trigram state x,y,z . Transisi tersebut akan dapat dihitung dengan cara membagi hasil jumlah *trigram state* dari data training yang digambarkan dengan persamaan $c(x,y,z)$ dengan hasil jumlah bigram state dari data training yang digambarkan dengan persamaan $c(x,y)$. Perhitungan ini tidak akan menghitung seluruh matriks A, akan tetapi hanya nilai yang dibutuhkan pada saat diberikan nilai observasi dan state tertentu.

Selain nilai matriks A, terdapat juga nilai matriks B yang akan dihitung. Perhitungan nilai matriks B atau matriks observasi trigram HMM dapat digambarkan dengan persamaan 4.10, dimana $b(z|S)$ merupakan perhitungan dari salah satu perhitungan nilai matriks B milik trigram HMM saat diberikan sebuah data observasi z dan sebuah state S. Perhitungan ini akan dapat dihitung dengan cara membagi hasil jumlah state terhadap data observasi yang digambarkan dengan persamaan $c(S \rightarrow z)$ dengan hasil jumlah state dari data training yang digambarkan dengan persamaan $c(S)$. Sama dengan persamaan sebelumnya, perhitungan ini tidak akan menghitung seluruh matriks B, akan tetapi hanya nilai yang dibutuhkan pada saat diberikan nilai observasi dan state tertentu.

Dengan munculnya persamaan-persamaan tersebut, maka persamaan algoritma Viterbi dapat ditulis ulang menggunakan persamaan 4.12:

$$V_{t,k} = \max_{x \in S} (a(z|x,y) \cdot b(z|S) \cdot V_{t-1,x}) \quad (4.12)$$

dimana $V_{t,k}$ adalah perhitungan probabilitas untuk urutan state yang paling mungkin untuk observasi ke - t yang mempunyai k sebagai kemungkinan statenya, seperti pada persamaan 4.11. Persamaan ini juga akan tetap menghitung nilai maksimum dari setiap state yang mungkin, dengan $\max_{x \in S}$, akan memaksimumkan setiap x data observasi dari anggota S sebagai urutan state yang mungkin. $a(z|x,y)$ merupakan persamaan perhitungan nilai parameter matriks A atau matriks transisi seperti pada persamaan 4.9. $b(z|S)$ merupakan persamaan perhitungan nilai

parameter matriks B atau matriks observasi seperti pada persamaan 4.10. $V_{t-1,x}$ merupakan perhitungan probabilitas dari observasi ke - (t-1) atau observasi sebelumnya pada state x. dimana nantinya akan dipilih nilai maksimum dari perhitungan-perhitungan tersebut, dan akan dipilih sebuah state k sebagai state final untuk observasi ke - t seperti pada persamaan 4.11.

4.3.5.2.2 Penerapan Proses Decoding dengan Menggunakan HMM Trigram dan Algoritma Viterbi

Berdasarkan penjelasan algoritma 4.5 dan Algoritma Viterbi di atas, berikut ini adalah tahap yang dilakukan dalam menerapkan proses *decoding* dengan menggunakan *Trigram* HMM dan Algoritma Viterbi. Mula-mula lakukan pengambilan hasil data training dari file sesuai dengan parameter afeksi/emosi yang diinginkan sesuai dengan pemisahan data training pada proses sebelumnya.

Dalam perhitungan penerapan proses *decoding* ini, tetap menggunakan data training dari 6 contoh kalimat training seperti pada Gambar 4.10. Data yang tertulis pada Gambar 4.10 tersebut, dipakai sebagai data training pada proses *decoding*. Hasil perhitungan pada proses pemisahan data training pada Tabel 4.7 dan Tabel 4.8 digunakan sebagai data input untuk proses *decoding*.

Untuk melakukan uji coba proses *decoding* dengan menggunakan Algoritma Viterbi, dalam bab ini akan digunakan data input teks kalimat Bahasa Indonesia “Saya sakit kepala” dengan parameter afeksi netral. Berikut ini adalah barisan informasi yang diperoleh dari data input teks kalimat Bahasa Indonesia yang memiliki viseme sejumlah 17 viseme.

Data input :	S A Y A _ S A K I T _ K E P A L A
Viseme :	S A S A _ S A G I T _ G E P A T A
State :	1 2 1 2 2 1 2 1 2 1 2 1 2 1 2

Perhitungan Algoritma Viterbi, mula-mula menggunakan persamaan seperti yang bersesuaian dengan persamaan 4.11 dan persamaan 4.12, yaitu menentukan probabilitas maksimum dari hasil perkalian probabilitas dari nilai viterbi sebelumnya (*vektor* π), probabilitas dari matriks A (matriks transisi) dan probabilitas matriks B (matriks observasi). Persamaan untuk menghitung Viterbi adalah sebagai berikut:

$$\text{Viterbi} = \max \{ P(\text{Nilai Viterbi Sebelumnya}) * P(S3 | S1, S2) * P(\text{Karakter} | S3) \}$$

Matriks A berisi data training yang menerapkan persamaan 4.9.

$$P(S3 | S1, S2) = \frac{\text{count}(S1, S2, S3)}{\text{count}(S1, S2)}$$

Matriks B berisi data training yang menerapkan persamaan 4.10.

$$P(\text{String} | S3) = \frac{\text{count}(S3 \rightarrow \text{String})}{\text{count}(S3)}$$

Dengan menggunakan hasil perhitungan pada pemisahan data training sebelumnya, di bawah ini dilakukan uji coba dengan menggunakan 17 panjang viseme untuk input teks kalimat “saya sakit kepala”. Terdapat 17 loop untuk 17 panjang viseme yang dijadikan input teks dalam uji coba untuk menghasilkan Viterbi maksimum pada setiap urutan viseme “S,A,S,A,_,S,A,G,I,T,_,G,E,P,A,T,A”.

Langkah 1: Viseme S

	LOOP 1			
State ke-1 (S1) = *	0	0	0	0
State ke-2 (S2) = *	0	0	0	0
State ke-3 (S3) = S	1	2	3	4
Kombinasi String Bigram	00	00	00	00
Kombinasi String Trigram	001	002	003	004
Count(S1, S2, S3)	5	1	0	0
Count (S1,S2)	6	6	6	6
String	S	S	S	S
Penggabungan (3S → String)	1S	2S	3S	4S
Count (S3 → String)	22	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	1	1	1	1
P(S3 S1,S2)	0,83333	0,16667	0	0
P(String S3)	0,36667	0	0	0
Perhitungan Viterbi	0,30556	0	0	0
Viterbi Maksimum	0,305556			
State yang Terpilih	1			

Langkah 2: Viseme A

	LOOP 2			
State ke-1 (S1) = *	0	0	0	0
State ke-2 (S2) = S	1	1	1	1
State ke-3 (S3) = A	1	2	3	4
Kombinasi String Bigram	01	01	01	01
Kombinasi String Trigram	011	012	013	014
Count(S1, S2, S3)	0	5	0	0
Count (S1,S2)	5	5	5	5
String	A	A	A	A
Penggabungan (3S → String)	1A	2A	3A	4A
Count (S3 → String)	0	28	1	1
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	0,30556	0,30556	0,30556	0,30556
P(S3 S1,S2)	0	1	0	0
P(String S3)	0	0,49123	0,5	0,25
Perhitungan Viterbi	0	0,1501	0	0
Viterbi Maksimum		0,1501		
State yang Terpilih		2		

Langkah 3: Viseme S

	LOOP 3			
State ke-1 (S1) = S	1	1	1	1
State ke-2 (S2) = A	2	2	2	2
State ke-3 (S3) = S	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	121	122	123	124
Count(S1, S2, S3)	41	7	0	0
Count (S1,S2)	49	49	49	49
String	S	S	S	S
Penggabungan (3S → String)	1S	2S	3S	4S
Count (S3 → String)	22	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	0,1501	0,1501	0,1501	0,1501
P(S3 S1,S2)	0,83673	0,14285	0	0
P(String S3)	0,36667	0	0	0
Perhitungan Viterbi	0,04605	0	0	0
Viterbi Maksimum		0,04605		
State yang Terpilih		1		

Langkah 4: Viseme A

	LOOP 4			
State ke-1 (S1) = A	2	2	2	2
State ke-2 (S2) = S	1	1	1	1
State ke-3 (S3) = A	1	2	3	4
Kombinasi String Bigram	21	21	21	21
Kombinasi String Trigram	211	212	213	214
Count(S1, S2, S3)	2	41	1	9
Count (S1,S2)	49	49	49	49
String	A	A	A	A
Penggabungan (3S → String)	1A	2A	3A	4A
Count (S3 → String)	0	28	1	1
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	0,04605	0,04605	0,04605	0,04605
P(S3 S1,S2)	0,04081	0,83673	0,02041	0,08163
P(String S3)	0	0,49123	0,5	0,25
Perhitungan Viterbi	0	0,01893	0,00047	0,00094
Viterbi Maksimum	0,01893			
State yang Terpilih	2			

Langkah 5: Viseme _

	LOOP 5			
State ke-1 (S1) = S	1	1	1	1
State ke-2 (S2) = A	2	2	2	2
State ke-3 (S3) = _	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	121	122	123	124
Count(S1, S2, S3)	41	7	0	0
Count (S1,S2)	49	49	49	49
String	-	-	-	-
Penggabungan (3S → String)	1_	2_	3_	4_
Count (S3 → String)	0	15	1	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	0,04605	0,04605	0,04605	0,04605
P(S3 S1,S2)	0,04081	0,83673	0,02041	0,08163
P(String S3)	0	0,26316	0,5	0
Perhitungan Viterbi	0	0,00071	0	0
Viterbi Maksimum	0,00071			
State yang Terpilih	2			

Langkah 6: Viseme S

	LOOP 6			
State ke-1 (S1) = A	2	2	2	2
State ke-2 (S2) = _	2	2	2	2
State ke-3 (S3) = S	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	221	222	223	224
Count(S1, S2, S3)	7	0	0	0
Count (S1,S2)	7	7	7	7
String	S	S	S	S
Penggabungan (3S → String)	1S	2S	3S	4S
Count (S3 → String)	22	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	0,00071	0,00071	0,00071	0,00071
P(S3 S1,S2)	1	0	0	0
P(String S3)	0,36667	0	0	0
Perhitungan Viterbi	0,00026	0	0	0
Viterbi Maksimum		0,00026		
State yang Terpilih			1	

Langkah 7: Viseme A

	LOOP 7			
State ke-1 (S1) = _	2	2	2	2
State ke-2 (S2) = S	1	1	1	1
State ke-3 (S3) = A	1	2	3	4
Kombinasi String Bigram	21	21	21	21
Kombinasi String Trigram	211	212	213	214
Count(S1, S2, S3)	2	41	1	4
Count (S1,S2)	49	49	49	49
String	A	A	A	A
Penggabungan (3S → String)	1A	2A	3A	4A
Count (S3 → String)	0	28	1	1
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	0,00026	0,00026	0,00026	0,00026
P(S3 S1,S2)	0,04082	0,83673	0,02041	0,08163
P(String S3)	0	0,49123	0,5	0,25
Perhitungan Viterbi	0	0,00011	2,66E-6	5,32E-6
Viterbi Maksimum		0,00011		
State yang Terpilih			2	

Langkah 8: Viseme G

	LOOP 8			
State ke-1 (S1) = S	1	1	1	1
State ke-2 (S2) = A	2	2	2	2
State ke-3 (S3) = G	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	121	122	123	124
Count(S1, S2, S3)	41	7	0	0
Count (S1,S2)	49	49	49	49
String	G	G	G	G
Penggabungan (3S → String)	1G	2G	3G	4G
Count (S3 → String)	11	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	0,00011	0,00011	0,00011	0,00011
P(S3 S1,S2)	0,83673	0,14286	0	0
P(String S3)	0,18333	0	0	0
Perhitungan Viterbi	1,65E-05	0	0	0
Viterbi Maksimum	1,65E-05			
State yang Terpilih	1			

Langkah 9: Viseme I

	LOOP 9			
State ke-1 (S1) = A	2	2	2	2
State ke-2 (S2) = G	1	1	1	1
State ke-3 (S3) = I	1	2	3	4
Kombinasi String Bigram	21	21	21	21
Kombinasi String Trigram	211	212	213	214
Count(S1, S2, S3)	2	41	1	4
Count (S1,S2)	49	49	49	49
String	I	I	I	I
Penggabungan (3S → String)	1I	2I	3I	4I
Count (S3 → String)	0	6	0	2
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	1,65E-05	1,65E-05	1,65E-05	1,65E-05
P(S3 S1,S2)	0,04082	0,83673	0,02041	0,08163
P(String S3)	0	0,10526	0	0,5
Perhitungan Viterbi	0	1,45E-06	0	6,72E-07
Viterbi Maksimum	1,45E-06			
State yang Terpilih	2			

Langkah 10: Viseme T

	LOOP 10			
State ke-1 (S1) = G	1	1	1	1
State ke-2 (S2) = I	2	2	2	2
State ke-3 (S3) = T	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	121	122	123	124
Count(S1, S2, S3)	41	7	0	0
Count (S1,S2)	49	49	49	49
String	T	T	T	T
Penggabungan (3S → String)	1T	2T	3T	4T
Count (S3 → String)	17	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	1,45E-06	1,45E-06	1,45E-06	1,45E-06
P(S3 S1,S2)	0,83673	0,14286	0	0
P(String S3)	0,28333	0	0	0
Perhitungan Viterbi	3,43E-07	0	0	0
Viterbi Maksimum			3,43E-07	
State yang Terpilih			1	

Langkah 11: Viseme _

	LOOP 11			
State ke-1 (S1) = I	2	2	2	2
State ke-2 (S2) = T	1	1	1	1
State ke-3 (S3) = _	1	2	3	4
Kombinasi String Bigram	21	21	21	21
Kombinasi String Trigram	211	212	213	214
Count(S1, S2, S3)	2	41	1	4
Count (S1,S2)	49	49	49	49
String	_	_	_	_
Penggabungan (3S → String)	1_	2_	3_	4_
Count (S3 → String)	0	15	1	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	3,44E-07	3,44E-07	3,44E-07	3,44E-07
P(S3 S1,S2)	0,04082	0,83673	0,02041	0,08163
P(String S3)	0	0,26316	0,5	0
Perhitungan Viterbi	0	7,56E-08	3,51E-07	0
Viterbi Maksimum			7,56E-08	
State yang Terpilih			2	

Langkah 12: Viseme G

	LOOP 12			
State ke-1 (S1) = T	1	1	1	1
State ke-2 (S2) = _	2	2	2	2
State ke-3 (S3) = G	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	121	122	123	124
Count(S1, S2, S3)	41	7	0	0
Count (S1,S2)	49	49	49	49
String	G	G	G	G
Penggabungan (3S → String)	1G	2G	3G	4G
Count (S3 → String)	11	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	7,56E-08	7,56E-08	7,56E-08	7,56E-08
P(S3 S1,S2)	0,83673	0,14286	0	0
P(String S3)	0,18333	0	0	0
Perhitungan Viterbi	1,16E-08	0	0	0
Viterbi Maksimum	116E-08			
State yang Terpilih	1			

Langkah 13: Viseme E

	LOOP 13			
State ke-1 (S1) = _	2	2	2	2
State ke-2 (S2) = G	1	1	1	1
State ke-3 (S3) = E	1	2	3	4
Kombinasi String Bigram	21	21	21	21
Kombinasi String Trigram	211	212	213	214
Count(S1, S2, S3)	2	41	1	4
Count (S1,S2)	49	49	49	49
String	E	E	E	E
Penggabungan (3S → String)	1E	2E	3E	4E
Count (S3 → String)	0	7	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	1,16E-08	1,16E-08	1,16E-08	1,16E-08
P(S3 S1,S2)	0,04082	0,8367	0,02041	0,08163
P(String S3)	0	0,12281	0	0
Perhitungan Viterbi	0	1,19E-09	0	0
Viterbi Maksimum	1,19E-09			
State yang Terpilih	2			

Langkah 14: Viseme P

	LOOP 14			
State ke-1 (S1) = G	1	1	1	1
State ke-2 (S2) = E	2	2	2	2
State ke-3 (S3) = P	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	121	122	123	124
Count(S1, S2, S3)	41	7	0	0
Count (S1,S2)	49	49	49	49
String	P	P	P	P
Penggabungan (3S → String)	1P	2P	3P	4P
Count (S3 → String)	7	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	1,19E-09	1,19E-09	1,19E-09	1,19E-09
P(S3 S1,S2)	0,83673	0,14286	0	0
P(String S3)	0,11667	0	0	0
Perhitungan Viterbi	1,16E-10	0	0	0
Viterbi Maksimum		1,16E-10		
State yang Terpilih		1		

Langkah 15: Viseme A

	LOOP 15			
State ke-1 (S1) = E	2	2	2	2
State ke-2 (S2) = P	1	1	1	1
State ke-3 (S3) = A	1	2	3	4
Kombinasi String Bigram	21	21	21	21
Kombinasi String Trigram	211	212	213	214
Count(S1, S2, S3)	2	41	1	4
Count (S1,S2)	49	49	49	49
String	A	A	A	A
Penggabungan (3S → String)	1A	2A	3A	4A
Count (S3 → String)	0	28	1	1
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	1,16E-10	1,16E-10	1,16E-10	1,16E-10
P(S3 S1,S2)	0,04082	0,83673	0,02041	0,08163
P(String S3)	0	0,49123	0,5	0,25
Perhitungan Viterbi	0	4,78E-11	1,19E-12	2,38E-12
Viterbi Maksimum		4,78E-11		
State yang Terpilih		2		

Langkah 16: Viseme T

	LOOP 16			
State ke-1 (S1) = P	1	1	1	1
State ke-2 (S2) = A	2	2	2	2
State ke-3 (S3) = T	1	2	3	4
Kombinasi String Bigram	12	12	12	12
Kombinasi String Trigram	121	122	123	124
Count(S1, S2, S3)	41	7	0	0
Count (S1,S2)	49	49	49	49
String	T	T	T	T
Penggabungan (3S → String)	1T	2T	3T	4T
Count (S3 → String)	17	0	0	0
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	4,78E-11	4,78E-11	4,78E-11	4,78E-11
P(S3 S1,S2)	0,83673	0,14286	0	0
P(String S3)	0,28333	0	0	0
Perhitungan Viterbi	1,13E-11	0	0	0
Viterbi Maksimum	1,13E-11			
State yang Terpilih	1			

Langkah 17: Viseme A

	LOOP 17			
State ke-1 (S1) = A	2	2	2	2
State ke-2 (S2) = T	1	1	1	1
State ke-3 (S3) = A	1	2	3	4
Kombinasi String Bigram	21	21	21	21
Kombinasi String Trigram	211	212	213	214
Count(S1, S2, S3)	2	41	1	4
Count (S1,S2)	49	49	49	49
String	A	A	A	A
Penggabungan (3S → String)	1A	2A	3A	4A
Count (S3 → String)	0	28	1	1
Count (S3)	60	57	2	4
P(Nilai Viterbi Sebelumnya)	1,13E-11	1,13E-11	1,13E-11	1,13E-11
P(S3 S1,S2)	0,04082	0,83673	0,02041	0,08163
P(String S3)	0	0,49123	0,5	0,25
Perhitungan Viterbi	0	4,66E-12	1,16E-13	2,32E-13
Viterbi Maksimum	4,66E-12			
State yang Terpilih	2			

Jadi ringkasan hasil perhitungan Viterbi untuk setiap data input teks kalimat Bahasa Indonesia dari Loop 1 sampai Loop 17 dapat diringkas menjadi Tabel 4.9.

Tabel 4.9 Rekapitulasi Hasil Perhitungan Viterbi

Loop	Data Input	Viseme	Viterbi Maksimum	State yang Terpilih
1	S	S	0,30556	1
2	A	A	0,15010	2
3	Y	S	0,04605	1
4	A	A	0,01892	2
5	-	-	0,00071	2
6	S	S	0,00026	1
7	A	A	0,00011	2
8	K	G	1,65E-05	1
9	I	I	1,45E-06	2
10	T	T	3,44E-07	1
11	-	-	7,56E-08	2
12	K	G	1,16E-08	1
13	E	E	1,19E-09	2
14	P	P	1,16E-10	1
15	A	A	4,78E-11	2
16	L	T	1,13E-11	1
17	A	A	4,66E-12	2

Perhitungan Akurasi dari Trigram HMM dan Viterbi ini dibahas di Bab 5, pada saat melakukan sinkronisasi visualisasi 3D wajah animasi dengan gerak bibir yang dihasilkan. Pada Bab 4 ini lebih fokus pada pemisahan data training dengan menggunakan HMM untuk mendapatkan urutan viseme yang terbaik. Hasil yang didapatkan dari tahap ini, sangat mempengaruhi hasil tahap visualisasi.

4.4 Inesh² Corpora

Corpora (definisi dari Wikipedia) adalah basis pengetahuan utama dalam linguistik korpus, yang merupakan sekumpulan besar dari teks yang terstruktur. Saat ini corpora biasanya disimpan dan diproses secara elektronik. Analisis dan pengolahan berbagai jenis corpora menjadi subyek dari banyak pekerjaan di komputasi linguistik, pengenalan suara dan mesin penterjemah, yang sering digunakan untuk membuat Hidden Markov Model, untuk penandaan kelas kata dan

keperluan lainnya. Sekumpulan teks ini digunakan untuk analisis statistik dan pengujian hipotesa, memeriksa kemunculan dari suatu kejadian atau memvalidasi aturan linguistik atau sebuah wilayah bahasa tertentu.

Untuk keperluan penelitian pembuatan HMM ini, penulis mencoba untuk membuat pengujian hipotesa yang merupakan penerapan dari kumpulan pemisahan data training dari input teks kalimat Bahasa Indonesia. Pemisahan data training ini untuk mencari urutan viseme terbaik dan alami dari pengucapan kalimat Bahasa Indonesia yang berafeksi/emosi. Untuk itu, penulis mencoba membuat data training *Inesh² Corpora*, singkatan dari *Indonesian Emotional Speech – Indonesian Endang Surya Hery Corpora*, dibaca “*inesh square corpora*”, yang memiliki makna kumpulan ucapan kalimat Bahasa Indonesia yang beremosi/afeksi yang dibuat oleh penulis (Endang Setyati) dan dua promotor penulis (Bapak Surya Sumpeno dan Bapak Mauridhi Hery Purnomo), diambil dari huruf kapital identitas nama panggilan masing-masing.

Selain untuk keperluan penelitian penulis, diharapkan *Inesh² Corpora* ini dapat bermanfaat bagi pengetahuan dan penelitian orang lain. Sehingga ketika *Inesh² Corpora* ini sudah teruji dan tervalidasi dengan baik, maka kelak dalam waktu dekat akan diletakkan di web agar bisa dimanfaatkan banyak orang. Sementara ini, karena keterbatasan waktu penulis dan ide ini baru muncul dalam beberapa bulan terakhir, maka *Inesh² Corpora* versi 1.0 masih jauh dari sempurna.

Inesh² Corpora versi 1.0 ini, memiliki karakteristik sebagai berikut:

1. Menggunakan 8 ekspresi: netral (*neutral*), senang (*happy*), sedih (*sad*), takut (*fear*), jijik (*disgust*), terkejut (*surprise*), malu (*shame*), dan marah (*angry*).
2. Menggunakan simbol panjang durasi dengan 4 *State* (satuan *frame*): 1F, 2F, 3F, 4F; yang mewakili panjang durasi pengucapan dalam satuan frame.
3. Menggunakan 12 model viseme Indonesia, yang terklasifikasi ke dalam kelas viseme P, F, G, T, R, S, A, I, U, E, O, dan diam (*silent*).
4. Model viseme Indonesia ini, merupakan representasi dari 49 fonem Indonesia, yang terdistribusi dari 5 fonem vokal tunggal, 5 fonem vokal alofon, 3 vokal ganda/diftong, 21 fonem konsonan tunggal, dan 15 konsonan ganda yang berasal dari kata serapan bahasa daerah/lokal dan bahasa asing (Arab dan Inggris).

5. Menggunakan 12 gambar bentuk mulut yang mewakili viseme dari wajah animasi, pada setiap ekspresi dan masing-masing viseme memiliki 5 model transisi inbetween = $12 \times 8 \times 5 = 480$ gambar
6. Menggunakan 50 kalimat Bahasa Indonesia berbeda pada setiap ekspresi.
7. Menggunakan 157 kata Bahasa Indonesia berbeda pada setiap ekspresi.
8. Kemunculan masing-masing viseme pada setiap ekspresi:

a. Viseme P = 100	g. Viseme A = 199
b. Viseme F = 11	h. Viseme I = 91
c. Viseme G = 119	i. Viseme U = 66
d. Viseme T = 175	j. Viseme E = 62
e. Viseme R = 53	k. Viseme O = 29
f. Viseme S = 105	l. Viseme _ (silent) = 152
9. Menggunakan panjang (jumlah) viseme bervariasi, mulai dari 14 sampai dengan 39 viseme dalam satu kalimat pada setiap ekspresi. Satu kalimat yang memiliki:

a. panjang viseme 14 = 1 kalimat	k. panjang viseme 25 = 1 kalimat
b. panjang viseme 15 = 3 kalimat	l. panjang viseme 26 = 4 kalimat
c. panjang viseme 16 = 7 kalimat	m. panjang viseme 27 = 4 kalimat
d. panjang viseme 17 = 2 kalimat	n. panjang viseme 29 = 2 kalimat
e. panjang viseme 18 = 2 kalimat	o. panjang viseme 30 = 5 kalimat
f. panjang viseme 19 = 3 kalimat	p. panjang viseme 31 = 2 kalimat
g. panjang viseme 20 = 3 kalimat	q. panjang viseme 33 = 1 kalimat
h. panjang viseme 21 = 2 kalimat	r. panjang viseme 35 = 1 kalimat
i. panjang viseme 22 = 2 kalimat	s. panjang viseme 38 = 1 kalimat
j. panjang viseme 23 = 3 kalimat	t. panjang viseme 39 = 1 kalimat
10. Akurasi 50 kalimat bahasa Indonesia pada setiap ekspresi:

a. Netral : 100%	e. Takut : 86,27%
b. Sedih : 75,43%	f. Jijik : 75,33%
c. Marah : 77,26%	g. Terkejut : 94,25%
d. Senang : 75,98%	h. Malu : 85,35%.
11. Format pembuatan data training *Inesh² Corpora* diperoleh dari urutan state viseme terbaik pada kalimat Indonesia berafeksi pada setiap ekspresi dalam bentuk *text file*. Setiap file data training ini memuat 2 baris informasi, yaitu:

baris pertama menunjukkan parameter input kalimat yang sudah menjadi urutan viseme dan baris kedua merupakan parameter afeksi/emosi yang disimpan dalam urutan state (*frame*) yang benar untuk kalimat pada baris pertama. File data training tersebut akan dibaca dan diambil barisnya satu persatu, dimana setiap baris ganjil merupakan kalimat training, dan baris genap merupakan urutan state yang benar untuk kalimat pada baris sebelumnya.

12. Sintaks penulisan urutan viseme dan *frame* di dalam file data training *Inesh² Corpora*, untuk membantu mendapatkan urutan viseme yang alami, didahului dengan pemberian 2 tanda bintang (*star marker*), “**”. *Star marker* ini membantu sebagai penanda awal untuk dimulainya gerak bibir dari posisi diam (*silent*), yang seolah-olah membutuhkan 2 *state*. Urutan state dengan satuan *frame*, merepresentasikan kebutuhan *frame* dalam setiap viseme yang muncul. Setiap *frame* yang muncul (0F, 1F, 2F, 3F atau 4F) mewakili *range* kebutuhan dari sejumlah *frame* berdasarkan panjang durasi pengucapan dari pergerakan animasi gerakan bibir (*lip movement*). Sintaksnya sebagai berikut:

Baris ganjil: **{urutan_viseme}: **<viseme-1>,<viseme-2>,...,<viseme-n>
Baris genap: *0F,*0F, <viseme-1> <urutan_state_viseme-1 dengan satuan frame F>, <viseme-2> <urutan_state_viseme-2 dengan satuan frame F>,...,<viseme-n> <urutan_state_viseme-n dengan satuan frame F>.

Contoh:

**S,A,S,A,_S,E,G,A,T (dari input teks kalimat Indonesia “saya sehat”)

*0F,*0F,S1F,A2F,S3F,A4F,_2F,S1F,E1F,G3F,A2F,T2F

**T,A,G,I,R_,P,A,T,I,T (dari input teks kalimat Indonesia “lahir batin”)

*0F,*0F,T2F,A2F,G3F,I3F,R2F,_2F,P2F,A2F,T2F,I2F,T2F

Jumlah kalimat Bahasa Indonesia yang menjadi data training dan data testing dalam *Inesh² Corpora* ini, masih sangat jauh dari cukup. Sehingga pada penelitian selanjutnya, akan terus ditingkatkan jumlah dan urutan state alaminya. Dengan sedikitnya jumlah data training yang digunakan, maka akan mempengaruhi hasil eksplorasi untuk mendapatkan pemilihan urutan viseme terbaik.

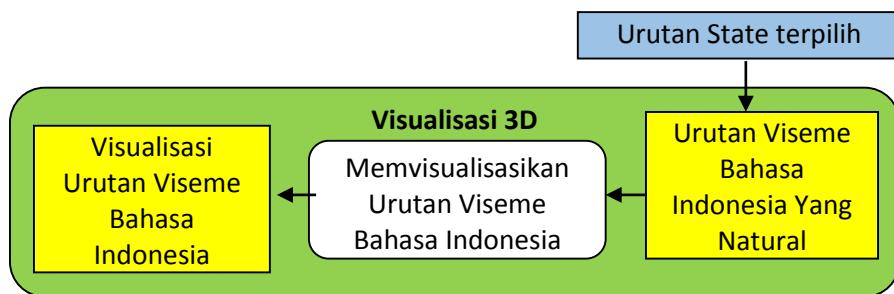
BAB V

VISUALISASI WAJAH ANIMASI PADA TALKING HEAD SYSTEM DALAM BAHASA INDONESIA

Pembahasan pada Bab 5 ini merupakan tahapan keempat atau tahapan terakhir dari penelitian ini, yaitu pembuatan visualisasi model wajah 3D Animasi sebagai representasi visual 3D pada Talking Head System dalam Bahasa Indonesia. Input dari sistem representasi visual 3D ini diperoleh dari hasil implementasi HMM pada Bab 4 yang menghasilkan pengucapan urutan viseme yang alami berafeksi.

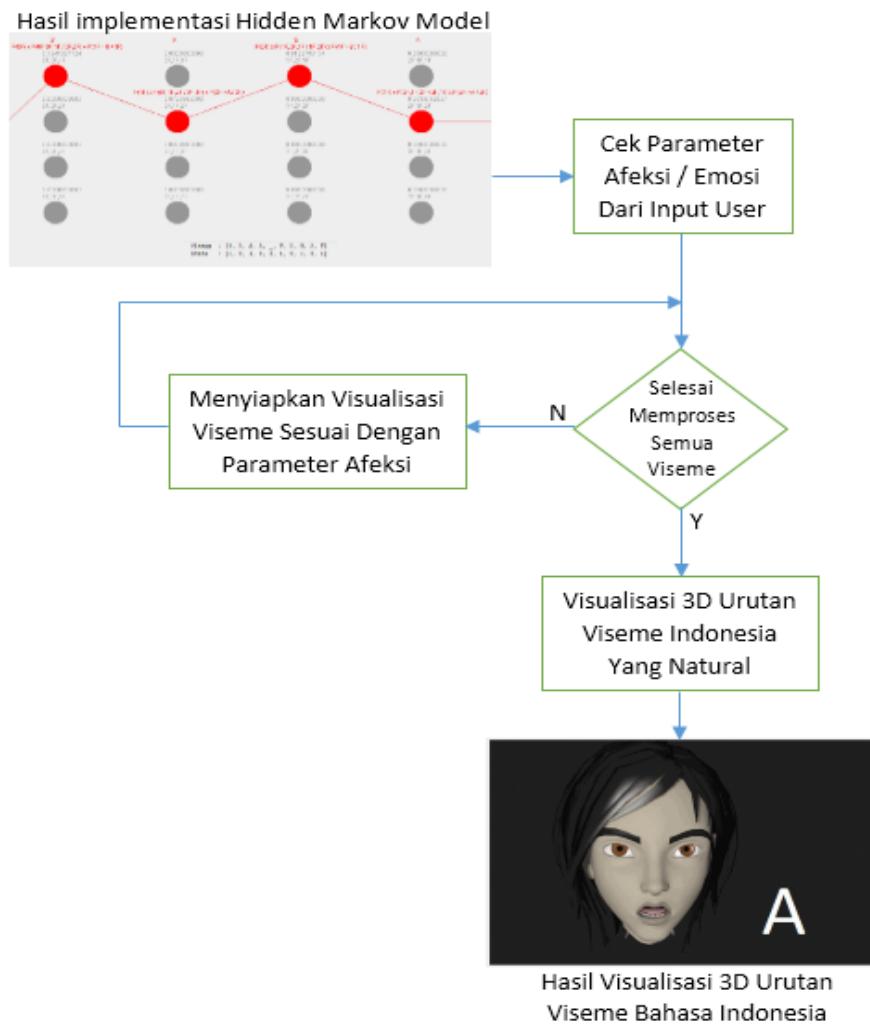
5.1 Konstruksi dari Proses Visualisasi Urutan Viseme

Pada sub-bab ini menjelaskan tentang proses visualisasi urutan viseme yang dilakukan. Proses visualisasi urutan viseme ini berfungsi untuk menampilkan urutan viseme dalam bentuk visualisasi 3D. Proses visualisasi urutan viseme ini dilakukan setelah Implementasi HMM dan algoritma Viterbi selesai dilakukan. Proses ini terdiri dari satu proses utama saja, yaitu pemvisualisasian viseme dalam bentuk 3D, yang ditunjukkan pada Gambar 5.1



Gambar 5.1 Proses Visualisasi Urutan Viseme

Sistem visualisasi 3D urutan viseme Indonesia berfungsi untuk memvisualisasi urutan viseme Indonesia dengan baik dan benar dari input kalimat yang sudah diinputkan oleh user. Visualisasi 3D ini dilakukan pada setiap viseme dari input yang sudah diproses pada sistem-sistem sebelumnya. Setiap proses yang dilakukan pada sistem visualisasi 3D urutan viseme Indonesia ditunjukkan pada Gambar 5.2.



Gambar 5.2 Desain Sistem Visualisasi 3D Urutan Viseme Indonesia

Sistem visualisasi 3D urutan viseme Indonesia ini diawali dengan melakukan pengecekan terhadap input parameter afeksi dari user. Visualisasi 3D yang dilakukan bergantung dari afeksi/emosi yang sudah diinputkan. Selanjutnya dilakukan proses menyiapkan visualisasi 3D terhadap setiap viseme hingga setiap viseme mempunyai bentuk visualisasinya sendiri.

Setelah semua viseme selesai disiapkan visualisasinya, maka dilakukan penggabungan visualisasi setiap viseme menjadi sebuah visualisasi 3D urutan viseme Indonesia. Teknik visualisasi 3D urutan viseme Indonesia ini dijelaskan satu per satu dalam sub-bab berikutnya.

5.2 Algoritma Visualisasi Urutan Viseme

Pemvisualisasian ini bergantung pada panjang frame viseme yang sudah dihitung, ada 4 macam panjang frame, yaitu: 1 Frame (1F), 2 Frame (2F), 3 Frame (3F), dan 4 Frame (4F). Untuk setiap panjang tersebut, akan dilakukan penyesuaian pada pemvisualisasian dengan cara mengalikan 4 gambar viseme penuh pada setiap panjang frame yang bersangkutan. Untuk lebih jelasnya akan dilanjutkan pada Algoritma 5.1.

Algoritma 5.1 Pemvisualisasian Viseme dalam Bentuk 3D

- 1: Lakukan pengambilan urutan viseme yang sudah dihitung menggunakan algoritma Viterbi
- 2: Lakukan pemilihan folder jenis gambar sesuai dengan emosi yang sudah diinputkan sejak awal
- 3: Ambil jenis dan urutan viseme
- 4: Lakukan pengambilan dan penyimpanan gambar yang akan divisualisasikan pada sebuah vektor
- 5: Ulangi langkah 3 dan 4 sampai semua viseme selesai disimpan gambarnya
- 6: Lakukan pembuatan .gif dari vektor gambar yang sudah disimpan

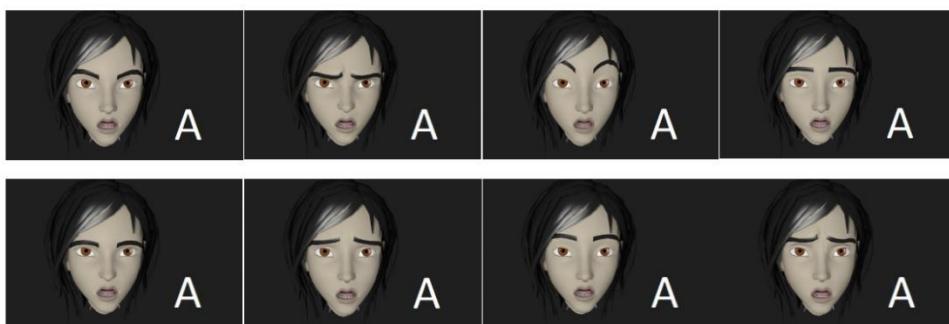
Pada algoritma 5.1 ini, langkah pertama yang harus dilakukan adalah melakukan pengambilan urutan viseme yang sudah dihitung menggunakan algoritma Viterbi sebelumnya. Setelah itu, akan dilakukan pemilihan folder jenis gambar yang akan digunakan sesuai dengan emosi yang sudah diinputkan oleh user sejak awal. Setelah pemilihan folder, langkah selanjutnya adalah melakukan pengambilan viseme dari urutan viseme yang sudah dihitung, dan kemudian dilakukan pengambilan dan penyimpanan gambar yang akan divisualisasikan, nama file gambar tersebut akan disimpan pada semua vektor. Proses pengambilan viseme dan penyimpanan gambar tersebut dilakukan berulang-ulang sampai semua viseme dari urutan viseme yang ada sudah selesai disimpan gambarnya.

Setelah semua urutan viseme selesai diproses dan disimpan gambarnya, maka dilakukan pembuatan .gif sebagai bentuk visualisasi 3D dari vektor file gambar yang sudah disimpan. Pembuatan .gif ini ditampilkan pada tampilan program utama. Gambar (*image*) seluruh hasil 3D Animasi yang telah dibuat untuk setiap 12 model viseme Indonesia dari 8 jenis afeksi dapat dilihat pada Lampiran 2.

5.3 Penerapan Visualisasi Urutan Viseme

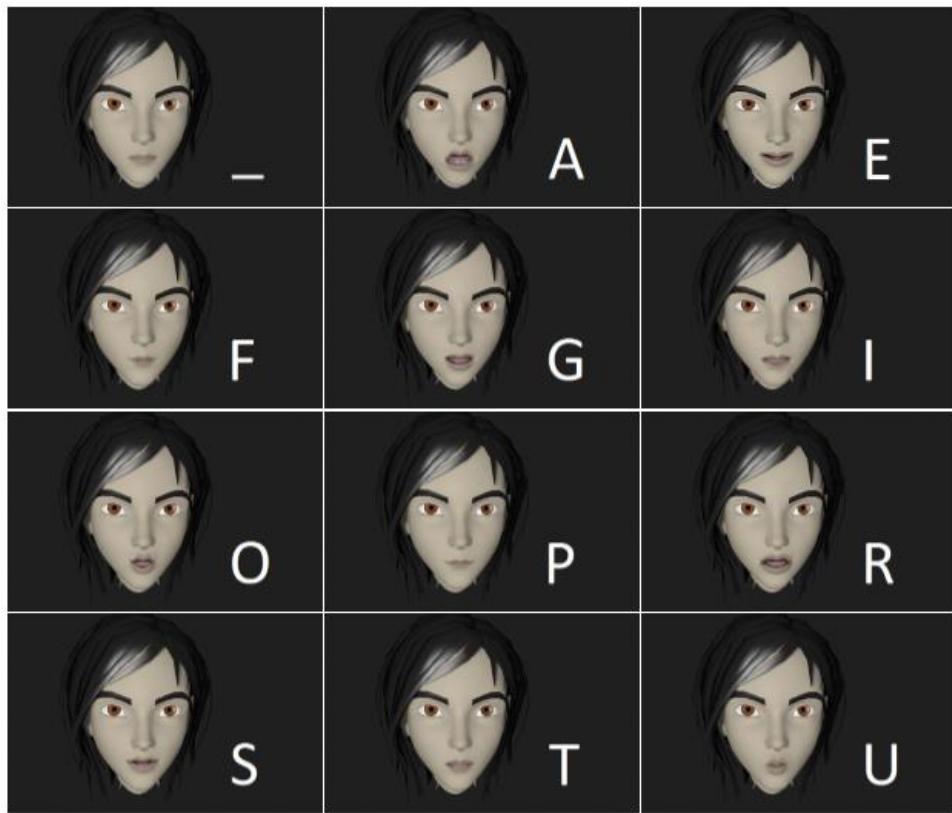
Pada proses visualisasi urutan viseme ini, terdapat 8 macam jenis afeksi/emosi yang digunakan, yaitu sedih, senang, marah, malu, takut, terkejut, jijik, dan netral. Dari 8 emosi yang ada dipisah ke dalam bentuk gambar dalam folder tersendiri yang divisualisasikan pada proses selanjutnya.

Pada setiap gambar ini disertakan pula jenis viseme yang bersangkutan. Viseme yang digunakan adalah 12 model viseme Indonesia seperti yang ada pada Tabel 2.12. Jadi total gambar hasil rendering yang dibuat seperti pada Gambar 5.3 dalam penelitian ini adalah adalah 96 gambar atau 12 gambar untuk masing-masing 8 jenis afeksi.



Gambar 5.3 Contoh Gambar Viseme A Hasil Rendering dari 8 Visualisasi Afeksi, dengan Ekspresi Wajah Mulai dari Baris Pertama Kiri ke Kanan: Netral, Marah, Terkejut, Malu, Sedih, Jijik, Senang, Takut

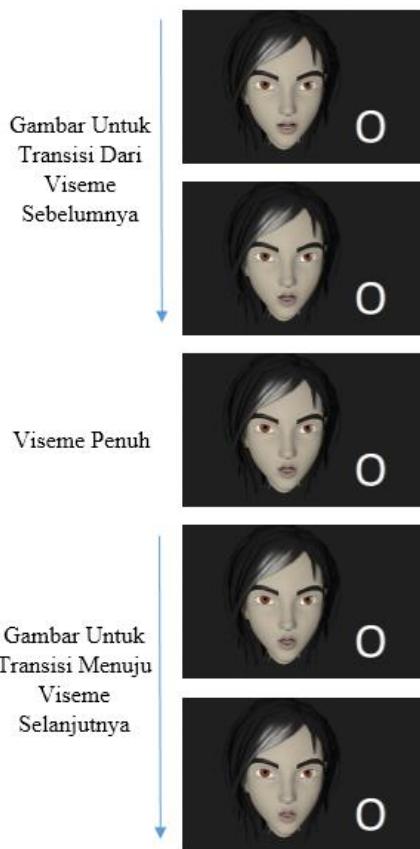
Pada gambar-gambar wajah 3D Animasi yang akan divisualisasikan juga terdapat jenis-jenis viseme yang berfungsi untuk membandingkan bentuk bibir dengan viseme yang sedang divisualisasikan. Tulisan huruf yang menempel pada gambar visualisasi di ujung kanan bawah adalah jenis kelas viseme yang berfungsi agar user dapat membandingkan bentuk bibir dengan ucapan yang ada, seperti terlihat pada Gambar 5.4. Dalam hal ini jenis viseme yang dimaksud di atas direpresentasikan sebagai pengganti suatu *running text* pada sebuah video. Pembuatan gambar seperti pada Gambar 5.3 dan Gambar 5.4 sebagai visualisasi 3D animasi, diperoleh dari hasil rendering wajah 3D animasi pada Bab 3, yang merupakan hasil akuisisi *IFFE-3D-Database*.



Gambar 5.4 Contoh Visualisasi Afeksi Netral dari 12 Macam Jenis Bentuk Bibir Menurut Model Viseme Indonesia

Setelah terdapat 8 jenis afeksi/emosi dan 12 jenis viseme, seperti pada Gambar 5.3 dan Gambar 5.4, akan dilakukan penghalusan bentuk ucapan untuk setiap viseme. Penghalusan ini dilakukan agar tampilan visualisasi ucapan wajah terlihat alami pada saat transisi ucapan antar viseme.

Penghalusan ini akan dilakukan dengan cara mengkombinasikan 5 gambar untuk setiap viseme, 2 gambar untuk transisi viseme dari viseme sebelumnya, 1 gambar transisi penuh, dan 2 gambar untuk transisi viseme ke viseme selanjutnya. Gambar ini nantinya akan digabung menjadi sebuah .gif sehingga dapat terlihat sebuah bentuk ucapan viseme dari kalimat input yang sudah pertama kali diinputkan dengan bentuk afeksi/emosi yang benar, dapat terbaca viseme yang sedang divisualisasikan, serta terlihat pula bentuk ucapan yang halus dan natural, lihat Gambar 5.5.



Gambar 5.5 Contoh Gambar Transisi Viseme Untuk Viseme O

5.4 Pengujian Pengucapan dari Hasil Eksperimen

Hasil dari eksperimen ini untuk menjelaskan validasi pengujian dari pengucapan urutan viseme secara alami dengan menggunakan *Trigram HMM* dan Algoritma *Viterbi*. Uji validasi pengucapan urutan viseme secara alami ini telah dilakukan pada 50 kalimat Bahasa Indonesia yang digunakan sebagai data testing untuk masing-masing afeksi. Ada 8 afeksi, berarti jumlah data testing sebanyak 400 data. Dalam 50 kalimat Bahasa Indonesia sebagai data testing, telah merepresentasikan penggunaan seluruh 12 klasifikasi viseme Indonesia (termasuk *silent*), yang dihasilkan dari 49 fonem Indonesia.

Pada uji validasi dari input teks kalimat Bahasa Indonesia, kemudian input teks ini dibandingkan dengan *state* yang telah diperoleh dari hasil pelaksanaan implementasi HMM dan Algoritma Viterbi pada *state* yang tersedia dalam data training pada proses sebelumnya. Setiap *state* pada masing-masing viseme dari input teks kalimat Bahasa Indonesia yang digunakan, akan dihitung akurasinya.

Jika ada lebih dari 1 data training untuk sebuah kalimat yang diinputkan, maka setiap data training akan dibandingkan dengan *state-state* yang telah diperoleh. Kemudian, rata-rata dari semua data training yang tersedia akan dihitung. Perhitungan akurasi dilakukan dengan menggunakan kalimat bahasa Indonesia yang sama di setiap afeksinya.

5.4.1 Pengujian Pengucapan Urutan Viseme secara Alami

Akurasi perhitungan pada data testing ini dapat diilustrasikan dengan menggunakan persamaan 5.1.

$$\text{II.} \quad \text{Akurasi} = \left(\frac{V - s}{V} \right) \times 100\% \quad \dots \dots \dots \quad (5.1)$$

dengan V sebagai banyaknya viseme hasil konstruksi dan s sebagai banyaknya prediksi dari state viseme yang tidak tepat. Dari perhitungan ini, akan diperoleh prosentase akurasi dari hasil uji coba pada setiap state. Rata-rata hasil dari perhitungan akurasi ini dihitung berdasarkan pada setiap afeksi yang ada.

Untuk memudahkan ilustrasi perhitungan, dengan menggunakan contoh yang sama pada hasil uji coba data testing pada proses sebelumnya, yaitu diberikan data input teks “Saya sakit kepala”, maka diperoleh hasil pengucapan pada urutan viseme dan urutan *state* sebagai berikut:

Data Input Teks : *Saya Sakit Kepala*

Urutan Viseme : S, A, S, A, _, S, A, G, I, T, _, G, E, P, A, T, A

Urutan State : 2, 3, 2, 3, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3

Panjang Viseme : 17

Misalkan diberikan sebuah kalimat yang memiliki panjang viseme yang sama dalam 6 data training yang berbeda, yaitu: data training 1 (DT1), data training 2 (DT2), data training 3 (DT3), data training 4 (DT4), data training 5 (DT5), data training 6 (DT6) dengan output dari urutan viseme dan *state* sebagai berikut:

DT1 : S, A, S, A, _, S, A, G, I, T, _, G, E, P, A, T, A

State : 2, 3, 2, 3, 3, 2, 3, 2, 3, 2, 3, 2, 2, 2, 3, 2, 4

DT2 : S, A, S, A, _, S, A, G, I, T, _, G, E, P, A, T, A
State : 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 4

DT3 : S, A, S, A, _, S, A, G, I, T, _, G, E, P, A, T, A
State : 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 2

DT4 : S, A, S, A, _, S, A, G, I, T, _, G, E, P, A, T, A
State : 2, 3, 2, 2, 3, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 3

DT5 : S, A, S, A, _, S, A, G, I, T, _, G, E, P, A, T, A
State : 3, 2, 2, 3, 3, 2, 2, 2, 3, 2, 3, 2, 2, 2, 3, 3, 3

DT6 : S, A, S, A, _, S, A, G, I, T, _, G, E, P, A, T, A
State : 3, 3, 2, 3, 3, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 4

Ilustrasi dari perhitungan untuk setiap output dari urutan pengucapan alami ditunjukkan pada persamaan 5.1 dengan masing-masing data training, dapat dilihat hasil akurasinya pada langkah-langkah berikut:

State dari DT1 : 2, 3, 2, 3, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3

Output_DT1 : 2, 3, 2, 3, 3, 2, 3, 2, 3, 2, 3, 2, 2, 3, 2, 4

Diperoleh Akurasi Output dari DT1: ($V = 17$ and $s_1 = 2$)

$$Akurasi_DT1 = \left(\frac{17 - 2}{17} \right) \times 100\% = 88.23\%$$

State of DT2 : 2, 3, 2, 3, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3

Output_DT2 : 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 4

Diperoleh Akurasi Output dari DT2: ($V = 17$ and $s_2 = 7$)

$$Akurasi_DT2 = \left(\frac{17 - 7}{17} \right) \times 100\% = 58.82\%$$

Perhitungan akurasi dari output untuk DT3, DT4, DT5, dan DT6 menggunakan persamaan 5.1 dengan panjang viseme $V = 17$ yang sama, serta memiliki perbedaan berturut-turut untuk DT3 hingga DT6, yaitu: $s_3 = 6$, $s_4 = 4$, $s_5 = 5$ and $s_6 = 3$, maka akurasi output dari masing-masing DT3 hingga DT6 adalah sebagai berikut:

$$Akurasi_DT3 = \left(\frac{17 - 6}{17} \right) \times 100\% = 64.70\%$$

$$Akurasi_DT4 = \left(\frac{17 - 4}{17} \right) \times 100\% = 76.47\%$$

$$Akurasi_DT5 = \left(\frac{17 - 5}{17} \right) \times 100\% = 70.58\%$$

$$Akurasi_DT6 = \left(\frac{17 - 3}{17} \right) \times 100\% = 82.35\%$$

Rata-rata gabungan akurasi yang diperoleh adalah sebagai berikut:

$$\begin{aligned} Akurasi &= \frac{88.23\% + 58.82\% + 64.70\% + 76.47\% + 70.58\% + 82.35\%}{6} \\ &= 73.51\% \end{aligned}$$

5.4.2 Pengujian dengan Mean Square Error

Untuk menghitung reliabilitas dari pendekatan yang telah diusulkan, maka diperlukan pembanding akurasi dari metode lain, misalkan Mean Square Error (MSE). MSE digunakan untuk mengukur kinerja dari State, yaitu antara State dari Data Output dan State dari Data Training. MSE ditentukan berdasarkan pada perbedaan antara nilai *observation state* dan *expectation state*. Nilai MSE yang lebih rendah berarti memiliki nilai kinerja yang lebih baik.

Tabel 5.1 merupakan hasil perhitungan nilai MSE dari data yang diberikan pada sebuah kalimat yang sama dengan 6 data training berbeda, yaitu: DT1 hingga DT6 dengan contoh yang sama dari kalimat yang digunakan pada proses data training sebelumnya.

MSE dapat diukur dengan menggunakan persamaan 5.2.

$$MSE = \frac{1}{MN} \sum_{j=1}^M \sum_{i=1}^N (S_{i_obs} - S_{j_exp})^2 \quad (5.2)$$

dengan S_{i_obs} merepresentasikan *output state* dari *observation state*, dan S_{j_exp} merepresentasikan *data training state* atau *expectation state*, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, N$. N adalah panjang state dan M adalah banyaknya data training.

Tabel 5.1 Mean Square Error dari DT1 hingga DT6

Input Teks : SAYA SAKIT KEPALA Output Viseme : SASA_SAGIT_GEPATA Length of State : 17						
Output State	Training Data of State					
	DT1	DT2	DT3	DT4	DT5	DT6
2	2	2	2	2	3	3
3	3	2	2	3	2	3
2	2	2	2	2	2	2
3	3	3	2	2	3	3
3	3	2	2	3	3	3
2	2	2	2	2	2	2
3	3	2	2	2	2	2
2	2	2	2	2	2	2
3	3	2	2	2	3	2
2	2	2	2	2	2	2
3	3	3	3	3	3	3
2	2	2	2	2	2	2
3	2	2	2	2	2	2
2	3	2	2	2	2	2
3	3	3	3	3	3	3
2	2	2	2	2	3	2
3	4	4	2	3	3	4
MSE	0.1765	0.3530	0.4118	0.2353	0.2941	0.2941
Rata-rata MSE = 0.294117647						

Hasil dari uji validasi pengucapan urutan viseme secara alami pada 50 kalimat Bahasa Indonesia yang sudah tersedia dalam data testing dan membandingkannya dengan 50 kalimat Bahasa Indonesia dalam data training pada *Inesh² Corpora* untuk masing-masing afeksi (ada 8 afeksi), dapat dilihat pada Tabel 5.2. Data testing dari kalimat Bahasa Indonesia pada Tabel 5.3 direpresentasikan

dari 12 klasifikasi viseme Indonesia (termasuk diam), yang berasal dari pemetaan 49 fonem Indonesia.

Akurasi tertinggi ditemukan dalam afeksi netral (*neu*) dengan rata-rata akurasi 100%, sementara akurasi terendah ditemukan pada akurasi jijik (*dis*) dengan rata-rata akurasi 75,33%. Total dari 50 data testing, diperoleh nilai akurasi rata-rata adalah 83,73%.

Tabel 5.2 Hasil Perhitungan Uji Validasi Urutan Viseme Alami

I	Kalimat Bahasa Indonesia	Panjang Viseme	Akurasi untuk Setiap Afeksi (%)							
			NEU	SAD	ANG	HAP	FEA	DIS	SUR	SHA
1	<i>Fotocopi Xerox</i>	14	100.00	80,71	77,14	76,43	86,43	82,14	97,14	87,86
2	<i>Saya Minum Susu</i>	15	100.00	70	75,56	73,33	83,33	70,00	96,67	80,00
3	<i>Kucing Saya Mati</i>	15	100.00	73,33	71,11	66,67	83,33	70,00	96,67	80,00
4	<i>Anjing Saya Mati</i>	15	100.00	63,33	73,33	63,33	83,33	70,00	96,67	80,00
5	<i>Aku Cinta Padamu</i>	16	100.00	62,5	72,92	62,50	84,38	68,75	96,88	81,25
6	<i>Saya Sakit Perut</i>	16	100.00	75	77,08	75,00	84,38	71,88	96,88	81,25
7	<i>Dia Suka Bermain</i>	16	100.00	71,88	62,50	71,88	84,38	68,75	96,88	81,25
8	<i>Mereka Bawa Kayu</i>	16	100.00	68,75	72,92	71,88	84,38	68,75	96,88	81,25
9	<i>Foto hitam putih</i>	16	100.00	81,88	80	80,63	88,13	78,75	96,25	91,15
10	<i>Sayur mayur hijau</i>	16	100.00	76,25	81,88	76,88	85,63	79,38	95,63	83,75
11	<i>Aku Capek Sekali</i>	17	100.00	71,88	72,92	62,50	84,38	68,75	96,88	81,25
12	<i>Rumah Makan Padan</i>	17	100.00	74,71	79,41	79,41	88,82	80,59	97,06	86,47
13	<i>Saya Sakit Kepala</i>	17	100.00	73,53	74,51	73,53	85,29	70,59	97,06	82,35
14	<i>Jeihan sudah insyaf</i>	18	100.00	80,56	81,67	78,33	83,33	76,67	96,67	85
15	<i>Pulang Pergi Sekolah</i>	19	100.00	76,32	75,44	73,68	86,84	73,68	97,37	92,11
16	<i>Saya Tidak Mau Tahu</i>	19	100.00	68,42	75,44	71,05	86,84	68,42	97,37	84,21
17	<i>Muadzin adzan di surau</i>	19	100.00	75,79	78,42	78,42	82,63	78,42	95,26	84,74
18	<i>Hari Ini Malam Minggu</i>	20	100.00	72,5	70,00	72,50	87,50	70,00	97,50	85,00
19	<i>Segeralah Kamu Pulang</i>	20	100.00	75	70,00	72,50	87,50	70,00	97,50	85,00
20	<i>Pasar itu sangat ramai</i>	20	100.00	77,5	82,5	81	85,5	79,5	99	85,5
21	<i>Amboi, pulau itu indah</i>	21	100.00	77,62	78,57	82,38	81,35	78,57	97,14	85,32

I	Kalimat Bahasa Indonesia	Panjang Viseme	Akurasi untuk Setiap Afeksi (%)							
			NEU	SAD	ANG	HAP	FEA	DIS	SUR	SHA
22	Bersyukur kepada Tuhan	21	100.00	80,95	79,52	78,1	83,81	76,19	95,24	88,1
23	Saya Tidak Naik Kelas	21	100.00	71,43	73,02	73,81	88,10	71,43	97,62	85,71
24	Pelangi itu sangat indah	22	100.00	79,09	79,09	75,91	83,71	78,64	97,27	84,47
25	Kebun binatang Surabaya	22	100.00	81,36	80,91	86,36	88,64	76,82	98,18	83,64
26	Singa Itu Berlari Kencang	23	100.00	73,91	75,36	73,91	89,13	71,74	97,83	86,96
27	Nyonya Endang sedang menyanyi	23	100.00	80,71	77,14	76,43	86,43	82,14	97,14	87,86
28	Tips disayang oleh orang tua	25	100.00	70	75,56	73,33	83,33	70,00	96,67	80,00
29	Antar Saya Pergi Ke Kantor	26	100.00	73,08	78,21	73,08	90,38	73,08	98,08	61,54
30	Kami Suka Makan Nasi Padang	26	100.00	71,15	78,21	73,08	90,38	71,15	98,08	88,46
31	Jangan Ajak Aku Pergi Dahulu	26	100.00	72,22	76,54	70,37	90,74	70,37	98,15	88,89
32	Bintang kecil di atas langit	26	100.00	79,23	76,54	79,23	89,74	80,77	97,69	83,33
33	Kelelawar Masuk Universitas	27	100.00	74,07	72,84	74,07	90,74	74,07	98,15	88,89
34	Rujak Cingur Ini Sangat Pedas	27	100.00	72,22	74,07	68,52	90,74	72,22	98,15	88,89
35	Vas bunga itu berwarna merah	27	100.00	78,89	82,59	81,11	86,11	78,89	96,3	88,15
36	Kaum dhuafa berdzikir di masjid	28	100.00	81,07	84,29	79,64	84,29	77,5	97,5	85,00
37	Monyet bergelantungan di pohon	28	100.00	78,57	76,43	81,43	88,69	77,5	97,14	87,14
38	Ilmu anthropologi dan psikologi	29	100.00	80,34	86,55	83,45	86,21	80,34	96,69	88,28
39	Hari Ini Saya Tidak Bisa Hadir	30	100.00	70,00	75,56	70,00	91,67	70,00	98,33	90,00
40	Gelombang tsunami sedang menerjang	30	100.00	82,33	79,67	79,00	80,67	80,33	99,33	86,67
41	Tolong Buatkan Minum Untuk Tamu	30	100.00	73,33	77,78	70,00	91,67	73,33	98,33	90,00

I	Kalimat Bahasa Indonesia	Panjang Viseme	Akurasi untuk Setiap Afeksi (%)							
			NEU	SAD	ANG	HAP	FEA	DIS	SUR	SHA
42	<i>Puisi ini karangan Chairil Anwar</i>	30	100.00	79.67	77.33	86.33	80.33	78.33	76.00	86.06
43	<i>Marsya sedang belajar teori graph</i>	30	100.00	76.67	78.33	79.33	81.67	82.00	76.67	83.61
44	<i>Sholat Maghrib pada bulan ramadlan</i>	31	100.00	77.10	82.9	83.55	83.87	77.10	77.1	87.1
45	<i>Jangan Kau Pecundangi Diriku Sayang</i>	32	100.00	71.88	73.96	71.88	92.19	71.88	98.44	90.63
46	<i>Jeni sedang bermain pasir di pantai</i>	33	100.00	78.18	78.79	81.52	81.82	81.82	70.91	88.18
47	<i>Sang Khaliq adalah penentu akhir zaman</i>	36	100.00	74.44	77.78	80.56	84.26	81.67	74.44	85.42
48	<i>Membaca sebuah buku teks</i>	38	100.00	78.42	81.58	83.42	89.47	83.16	75.53	82.11
49	<i>Saya Sangat Menyesal Dengan Perbuatan Saya</i>	38	100.00	78.21	79.49	74.36	91.03	73.08	98.72	91.03
50	<i>Gadis cantik itu mempunyai rambut keriting</i>	39	100.00	78.97	88.21	85.13	87.69	79.74	77.69	87.95
<i>Rata-rata Akurasi</i>			100.00	75.43	77.26	75.98	86.27	75.33	94.25	85.35
<i>Rata-rata Akurasi dari Seluruh Ekspresi</i>			83,73375							

5.5 Pengujian Representasi Visual

Untuk pengujian representasi visual, telah dilakukan survei secara *online* terhadap hasil representasi visual ekspresi dan gerak bibir dari 94 orang responden. Survei *online* yang diberikan terdiri dari 11 pertanyaan, setiap pertanyaan akan terdapat sebuah video yang menunjukkan representasi gerak bibir dan emosi. Sebuah kalimat akan diucapkan oleh model pada video tersebut sebanyak 3 (tiga) kali tanpa suara. Form pertanyaan pada setiap interface video yang ditayangkan, ditunjukkan pada Gambar 5.6.

Dalam survei yang dilakukan secara online ini, mula-mula responden akan diminta data berupa usia, pekerjaan, dan pendidikan terakhir. Selanjutnya, responden akan diminta untuk memilih jenis parameter afeksi/emosi dan gerak bibir

yang diucapkan oleh model pada video dari pilihan jawaban yang ada dengan mengabaikan huruf yang muncul pada setiap video.

Video 4



Gerakan bibir pada Video 4 menunjukkan kalimat apa ?*

- Saya sakit perut
- Saya belum tahu

Ekspresi pada Video 4 menunjukkan afeksi/emosi apa ?*

- Netral
- Takut

Gambar 5.6 Contoh Sebuah Pertanyaan dalam Survei Online

Contoh sebagian pertanyaan pada form survei online seperti yang terlihat di Gambar 5.6. Survei online ini dengan memanfaatkan *Google Form*, serta sebuah pertanyaan akan terdiri dari 2 pertanyaan lagi, dimana setiap pertanyaan tersebut akan terdapat 2 jawaban yang harus dipilih oleh responden. Sebanyak minimum 94 respon telah dicatat, dan dihitung akurasinya untuk menentukan akurasi representasi visualnya. Dari survei online yang sudah dibuat, didapat sebanyak 94

respon dari umur, jenis kelamin, dan pekerjaan yang berbeda seperti yang terlihat pada Tabel 5.3 berikut.

Tabel 5.3 Profil dari Responden pada Survei Online

Jenis Kelamin	:	Pria: 62 (66%) Wanita: 32 (34%)
Usia (tahun)	:	Rata-rata: 30,77 Tahun Standar Deviasi: 8,39 Tahun
Status Pendidikan	:	S0: 1 (1,10%) S1: 58 (61,70%) S2: 32 (34,00%) S3: 3 (3,20%)
Pekerjaan	:	Masih mahasiswa: 14 (14,90%) Dosen: 26 (27,70%) Wiraswasta: 40 (42,60%) Guru: 14 (14,90%)

Pada Tabel 5.3 dapat dilihat bahwa usia rata-rata responden adalah 30,77 tahun dengan simpangan baku 8,39 tahun atau range usia responden sekitar 21 – 39 tahun. Responden dengan jenis kelamin pria mendominasi 66% dibanding wanita yang hanya 34%. Responden dengan status pendidikan S1 paling banyak, yaitu 61,70% dan pekerjaan wiraswasta merupakan responden terbesar dari profil pengisi survei online. Hasil survei dari 94 responden tersebut dicatat dan dihitung akurasinya secara otomatis berdasarkan banyaknya jumlah jawaban Benar (B) dan Salah (S) dari pilihan responden untuk emosi dan gerak bibir masing-masing pertanyaan. Hasil perhitungan akurasi gerak bibir dan emosi menjadi ukuran akurasi representasi visual untuk visualisasi gerak bibir serta afeksi. Hasil survei representasi visual secara online dapat dilihat pada Tabel 5.4.

Pada Tabel 5.4 ini menunjukkan sebuah ringkasan dari Form Survei, yang merepresentasikan tayangan dari 11 video yang berisi visualisasi gambar wajah 3D animasi. Setiap tayangan video yang ditampilkan, terdapat 2 pertanyaan, yaitu: (1) Gerakan bibir pada Video ini menunjukkan kalimat apa? dan (2) Ekspresi pada Video ini menunjukkan afeksi/emosi apa? Masing-masing pertanyaan tersedia 2 jawaban, dimana salah satu jawaban adalah Benar (B) dan satu jawaban lagi adalah Salah (S). Responden diminta untuk memberikan isiannya, dalam bentuk tekan “klik” pada jawaban yang dianggap benar. Hasil survei online dari 94 responden

yang mengisi kuesioner online secara lengkap, diperoleh pilihan jawaban benar untuk kalimat yang diucapkan memiliki akurasi tertinggi untuk visualisasi afeksi / emosi “sedih” sebesar 96,81% dan akurasi terendah ada pada emosi “jijik dan terkejut”, masing-masing sebesar 60,64%. Sedangkan untuk gerak bibir yang menunjukkan pilihan jawaban kalimat yang benar tertinggi adalah pada ucapan “Saya tidak mau tahu” sebesar 93,62% dan pilihan jawaban kalimat yang benar terendah adalah pada ucapan “Singa itu berlari kencang” sebesar 57,45%. Sehingga secara keseluruhan, dari hasil survei online ini, diperoleh rata-rata akurasi untuk hasil survei parameter afeksi/emosi sebesar 80,56% dan untuk gerak bibir, mempunyai rata-rata akurasi sebesar 81,04%.

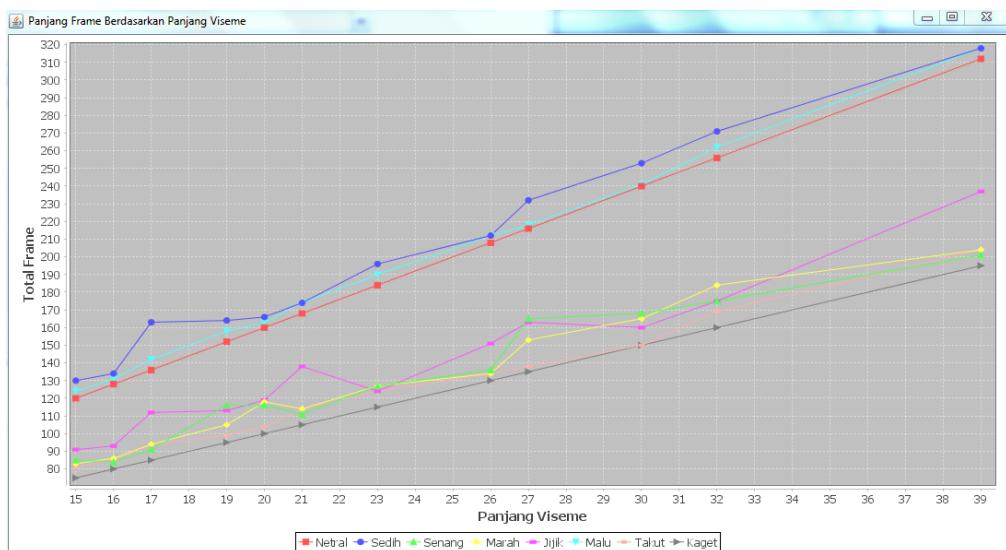
Tabel 5.4 Hasil Survei Online dari Representasi Visual

Video ke-	Kalimat Benar	Paramet er Afeksi	Hasil Survei Emosi		Hasil survei Gerak bibir		Akurasi %	
			B	S	B	S	Emosi	Gerak Bibir
1	Pulang Pergi Sekolah	Netral	87	7	84	10	92,55	89,36
2	Saya Melihat Serangga	Takut	88	6	78	16	93,62	82,98
3	Saya Minum Jus Pepaya	Marah	87	7	71	23	92,55	75,53
4	Saya sakit perut	Takut	69	25	78	16	73,40	82,98
5	Saya tidak mau tahu	Marah	86	8	88	6	91,49	93,62
6	Anjing saya mati	Jijik	57	37	65	29	60,64	69,15
7	Kamu suka makan nasi padang	Senang	82	12	84	10	87,23	89,36
8	saya belum belanja	Terkejut	57	37	76	18	60,64	80,85
9	Aku suka kamu	Malu	63	31	84	10	67,02	89,36
10	Kucing saya mati	Sedih	91	3	76	18	96,81	80,85
11	Singa itu berlari kencang	Netral	66	28	54	40	70,21	57,45
			Rata-rata Akurasi (%) =					80,56
								81,04

5.6 Perbandingan Jumlah Frame

Pada perbandingan jumlah frame ini, ditunjukkan dengan membandingkan panjang jumlah frame dari tiap afeksi/emosi terhadap panjang viseme dari data yang sudah diinputkan. Terdapat beberapa panjang viseme dari kalimat yang sudah diinputkan dan panjang framenya dihitung untuk melihat perbandingan jumlah

frame antar afeksi/emosi. Panjang frame yang telah diuji cobakan mulai dari 15 hingga 39 frame.



Gambar 5.7 Grafik Panjang Frame Berdasarkan Panjang Viseme

Tabel 5.5 Hasil Perbandingan Jumlah Frame

Panjang Viseme	Rata- Rata Panjang Frame							
	Netral	Sedih	Marah	Senang	Takut	Jijik	Kaget	Malu
15	120	130	85	83	81	91	75	124
16	128	134	86	84	85	93	80	131
17	136	163	94	90	94	112	84	142
19	152	164	105	116	99	113	95	158
20	160	166	118	116	104	119	100	163
21	168	174	114	111	114	138	105	174
23	184	196	127	127	124	124	115	190
26	208	212	134	136	134	151	130	212
27	216	232	153	165	138	163	135	218
30	240	253	165	168	150	160	150	241
32	256	271	184	175	169	175	160	262
39	312	318	204	201	204	237	165	318
Total Rata-Rata Panjang Frame	2280	2413	1569	1572	1496	1676	1394	2333

Gambar 5.7 menampilkan grafik perbandingan jumlah frame antar emosi dengan panjang visemanya. Dari gambar tersebut terlihat bahwa emosi sedih memiliki prosentase jumlah frame tertinggi dibandingkan parameter afeksi lainnya untuk setiap panjang viseme. Hal ini disebabkan emosi sedih, biasanya memiliki durasi lambat dalam pengucapan kalimat, sehingga membutuhkan banyak waktu dan meningkatkan pemakaian jumlah frame.

Data dari Gambar 5.7 ini ditunjukkan pada Tabel 5.5 yang menampilkan data hasil perbandingan jumlah frame antar emosi dengan panjang viseme dari kalimat-kalimat yang sudah diinputkan di dalam implementasi HMM dan Algoritma Viterbi.

Pada Tabel 5.5 dapat dilihat rata-rata panjang frame terhadap panjang viseme tersebut didapatkan bahwa total-rata-rata panjang frame terbanyak adalah pada emosi sedih dengan total rata-rata panjang frame sebanyak 2413 frame, sedangkan total rata-rata panjang frame terpendek adalah untuk emosi terkejut dengan total rata-rata panjang frame sebanyak 1394 frame.

5.7 Tampilan Urutan Viseme Alami Berafeksi

Pada sub ini, memberikan suatu tampilan sebuah barisan urutan viseme alami berafeksi dari sebuah kalimat bahasa Indonesia yang diinputkan. Agar memudahkan pengamatan, akan diberikan sebuah contoh uji coba visualisasi 3D animasi dari input kalimat bahasa Indonesia “SAYA” dengan afeksi marah.

Contoh:

Parameter Afeksi Marah	:	Marah
Input Kalimat	:	sayang ibu
Kata	:	“sayang” “ibu”
Fonem	:	/s//a//y//a//ng//_//i//b//u/
Model	:	/s//a//y//a//G//_//i//b//u/
Kalimat	:	sayaG_ibu
Viseme	:	SASAG_IPU

Data Training HMM Saya (marah) :

*0F,*0F,S1F,A2F,S1F,A2F,G3F_1F,I1F,P3F,U1F

Output Urutan State Viseme berafeksi marah:

*0F,*0F,S1F,A2F,S2F,A3F,G3F_1F,I3F,P3F,U4F

Program akan menghitung Urutan State Viseme HMM untuk hasil pengucapan output urutan viseme yang alami. Kemudian program akan menampilkan video visualisasinya, yang diilustrasikan pada Tabel 5.6 sebagai berikut:

Tabel 5.6 Ilustrasi Barisan Gambar dari Output “SASAG_IPU”

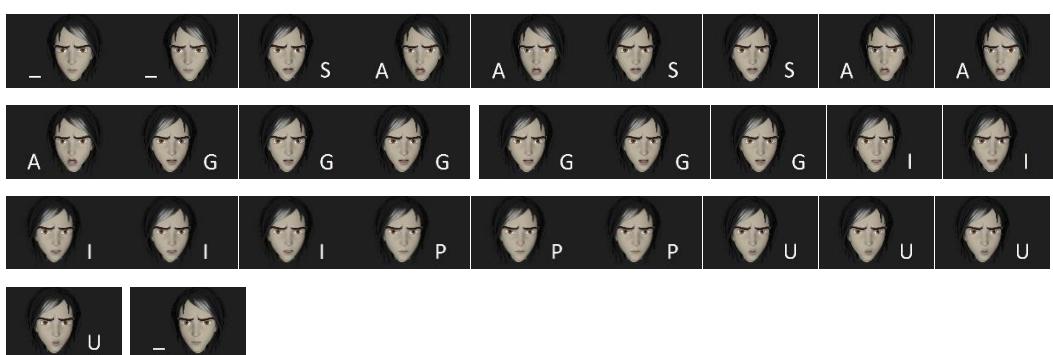
dengan Parameter Afeksi Marah

Urutan Viseme	Frame	Visualisasi 3D ANimasi Urutan Viseme yang alamai	Keterangan Ilustrasi
*	*0F		Silent (mulut masih tertutup)
*	*0F		Silent (mulut mulai akan bergerak)
S	S1F		S (panjang durasi S dibaca 1 frame)
A	A2F		AA (panjang durasi A dibaca 2 frame)
S	S2F		SS (panjang durasi S dibaca 2 frame)
A	A3F		AAA (panjang durasi A dibaca 3 frame)
G	G3F		GGG (panjang durasi G dibaca 3 frame)
—	G-1		Inbetween G ke-satu
—	G-2		Inbetween G ke-dua
—	G-3		Inbetween G penuh

Urutan Viseme	Frame	Visualisasi 3D Animasi Urutan Viseme yang alamai	Keterangan Ilustrasi
—	I-4		Inbetween I ke-empat
—	I-5		Inbetween I ke-lima
I	I3F		III (panjang durasi I dibaca 3 frame)
P	P3F		PPP (panjang durasi P dibaca 3 frame)
U	U4F		UUUU (panjang durasi U dibaca 4 frame)
*	*0F		Silent (mulut tertutup kembali)

Apabila dalam ilustrasi model wajah nyata manusia, maka kalimat yang diucapkan dari hasil urutan state yang alami adalah seolah-olah seseorang itu menggerakkan bentuk mulut dengan urutan viseme “SAASSAAAGGG_IIPPPUUUU” yang artinya SAAYYAAANNNGGG_IIBBBUUUU dengan aksi marah.

Urutan wajah animasinya adalah:



BAB VI

PENUTUP

Pada bab penutup ini akan membahas Kesimpulan dari Hasil Penelitian dan Usulan Penelitian Selanjutnya.

6.1 Kesimpulan Hasil Penelitian

Berdasarkan sintesa atas hasil penelitian, dapat disimpulkan sebagai berikut:

1. Pengembangan sistem afektif berwujud karakter virtual khas Indonesia dapat dicapai dengan menggunakan 12 model viseme Indonesia berdasarkan 9 parameter geometrik *blend shape animation* dari bentuk mulut di sekitar bibir model wajah animasi. Ke-12 model viseme dihasilkan dari pemetaan fonem ke viseme Bahasa Indonesia sebagai representasi dari 49 fonem Bahasa Indonesia termasuk diam (*silent*). Hasil pengukuran menunjukkan bahwa:
 - Visualisasi bentuk mulut model animasi yang mewakili masing-masing model viseme Indonesia memiliki persentase kesesuaian dengan bentuk mulut model nyata manusia sebesar 80,48% (berasal dari jumlah respon responden 57,74% sangat setuju + 22,74% setuju), lihat Tabel 2.9. Akan tetapi masih terdapat 19,52% responden yang mengatakan representasi visualisasi bentuk mulut model animasi tidak mewakili bentuk mulut model nyata manusia, terutama untuk 6 kelas viseme V2, V3, V5, V6, V8, dan V9. Dari observasi sebaran jumlah respon yang dilakukan, ada dugaan 13,54% responden merasa ragu bahwa 1 viseme bisa mewakili lebih dari 8-12 jumlah fonem. Sedangkan sisa 5,98% kurang setuju karena faktor lainnya.
 - Berdasarkan hasil perhitungan Mean Rating Indeks (MRI), yang bisa dipakai untuk mengukur indeks keberhasilan visualisasi bentuk mulut antara model animasi dan model nyata manusia, memiliki indeks keberhasilan 3,33, lihat Tabel 2.10. Artinya visualisasi bentuk mulut animasi berhasil merepresentasikan viseme sebesar 83,25% atau hanya 16,75% yang mengatakan viseme dari model animasi tidak sesuai dengan

model nyata manusia. Setelah dilakukan observasi pada sebaran jumlah respon responden, kelas viseme V8 memiliki nilai MRI terendah, yaitu 2,98. Penyebab rendahnya nilai MRI dari kelas viseme V8 (/i/, /I/, /oi/) ini, dugaan yang bisa diberikan adalah responden ragu bahwa fonem vokal /i/ dan /I/ memiliki bentuk mulut yang sama dengan fonem diftong /oi/.

- Berdasarkan hasil nilai rata-rata Mean Square Error (MSE), visualisasi model animasi memiliki rata-rata kesalahan pengukuran sebesar 0,01045 atau 1,045%, lihat Tabel 2.11. Artinya, kesalahan dalam pengukuran yang telah dilakukan dalam perhitungan MSE relatif rendah.
2. Visualisasi 3D model wajah animasi ucapan berafeksi ditentukan melalui klasifikasi 8 parameter afeksi, yaitu: senang, sedih, marah, takut, terkejut, jijik, malu dan netral. Proses klasifikasi didasarkan pada 37 marker (penanda) pada *facial motion capture*. Prosentase terbesar dari hasil survei pada masing-masing parameter afeksi yang terpilih pada posisi frame tertentu, dipakai sebagai titik-titik referensi untuk menentukan pose afeksi yang akan ditampilkan. Hasil pengukuran menunjukkan bahwa:
- Pemilihan afeksi untuk masing-masing pose dipilih dari salah satu pose dengan parameter afeksi terbaik yang dijadikan *IFFE-3D Database*. Data hasil survei secara keseluruhan untuk masing-masing pose dari *IFFE-3D Database* memiliki prosentase pemilih rata-rata sebesar 86,5%, lihat Tabel 3.8, dengan rincian sebagai berikut:
 - afeksi terkejut, yang terpilih frame ke-597 dari model ke-2, pada pengambilan rekaman ke-4, dengan prosentase pemilih 91%,
 - afeksi takut, yang terpilih frame ke-440 dari model ke-1, pada pengambilan rekaman ke-1, dengan prosentase pemilih 91%,
 - afeksi sedih, yang terpilih frame ke-581 dari model ke-2, pada pengambilan rekaman ke-4, dengan prosentase pemilih 91%,
 - afeksi senang, yang terpilih frame ke-934 dari model ke-6, pada pengambilan rekaman ke-3, dengan prosentase pemilih 73%,
 - afeksi jijik, yang terpilih frame ke-398 dari model ke-2, pada pengambilan rekaman ke-3, dengan prosentase pemilih 91%,

- afeksi marah, yang terpilih frame ke-406 dari model ke-4, pada pengambilan rekaman ke-1, dengan prosentase pemilih 82%,
 - afeksi netral, yang terpilih frame ke-800 dari model ke-1, pada pengambilan rekaman ke-4, dengan prosentase pemilih 73%,
 - afeksi malu, yang terpilih frame ke-122 dari model ke-9, pada pengambilan rekaman ke-2, dengan prosentase pemilih 100%.
- Wajah animasi yang dibuat berdasarkan titik-titik kontrol dari IFFE-3D Database yang terpilih, memiliki tampilan wajah 3D yang lebih ekspresif, karena terdapat gerakan alis dan gerakan bibir sesuai dengan afeksi yang diinputkan. Akan tetapi bentuk mulut yang merupakan representasi dari viseme ikut berubah seiring dengan perubahan emosinya.
 - Wajah animasi yang dibuat dengan blend shape animation, tidak bisa digunakan untuk representasi visualisasi dengan menggunakan input teks kalimat Bahasa Indonesia secara *real time*, karena *blend shape animation* bersifat statis dan hanya menampilkan pergerakan bentuk mulut dengan wajah berafeksi dari kalimat yang sudah disimpan dalam *script*.
 - Visualisasi 3D wajah animasi berafeksi, diperoleh dari hasil rendering pada wajah animasi blend shape. Akan tetapi hasil rendering ini, sedikit memberikan kesan perubahan bentuk wajah yang semula tampak 3D cenderung terlihat seperti gambar 2D.
3. Pengembangan *talking head system* dengan input teks kalimat Bahasa Indonesia berafeksi, berbasis teknik cerdas, pemodelan emosi dan visualisasi, dilakukan dalam dua tahap yang berbeda. Tahap pertama adalah pembangkitan urutan viseme berbasis teknik cerdas dengan menggunakan Hidden Markov Model, untuk mendapatkan ucapan urutan viseme Indonesia yang alami dan berafeksi. Sedangkan tahap kedua adalah pembuatan model wajah emosi berbasis visualisasi dengan menggunakan *blend shape animation*. Hasil pengukuran dari kedua tahap ini menunjukkan bahwa:
- a. Pembangkitan urutan viseme alami berafeksi, dengan menggunakan HMM pada pemisahan data training *Inesh² Corpora* yang terdiri dari 50 kalimat

bahasa Indonesia pada setiap ekspresi yang berbeda, hasil pengukurannya menunjukkan bahwa:

- Akurasi rata-rata keberhasilan dalam pembangkitan urutan viseme berafeksi sebesar 83,73%, lihat Tabel 4.29, artinya masih terdapat kesalahan 16,27%. Faktor penyebabnya adalah jumlah data training kalimat bahasa Indonesia yang ada pada *Inesh² Corpora* masih sangat kurang dari segi jumlah, maupun dari segi variasi kemunculan untuk masing-masing viseme. Sehingga masih harus banyak data training yang dilatihkan, sehingga menjadi cerdas.
 - Tiga proses pemotongan kalimat menjadi kata, kata menjadi fonem, dan fonem menjadi frame pada tahap segmentasi input teks kalimat Bahasa Indonesia, semua proses yang dilakukan berhasil 100%, sesuai dengan model klasifikasi yang ditentukan. Akan tetapi ketika melakukan proses decoding dengan algoritma Viterbi, untuk menebak state mana yang muncul pada urutan viseme berikutnya pada setiap ekspresi, menampilkan akurasi rata-rata yang berbeda-beda, untuk emosi netral 100% (akurasi tertinggi), sedih 75,43%, marah 77,26%, senang 75,98%, takut 86,27%, jijik 75,33% (akurasi terendah), kaget 94,25% dan emosi malu 85,35%. Faktor penyebabnya adalah belum memiliki standarisasi parameter afeksi yang baku pada saat diinputkan, sehingga menyebabkan terjadinya perbedaan total kemunculan untuk setiap urutan viseme yang dibangkitkan.
- b. Pembuatan model wajah emosi/afeksi berbasis visualisasi dengan menggunakan *blend shape animation*, telah menerapkan transisi viseme (inbetween), agar perubahan bentuk mulut pada pengucapan viseme dari input kalimat Bahasa Indonesia berafeksi terlihat halus dan alami. Hasil pengukuran menunjukkan:
- Tingkat akurasi berdasarkan penilaian visualisasi afeksi/emosi yang tampak pada wajah animasi adalah 80,56%. Akan tetapi 19,44% dinilai responden visualisasi wajah emosinya kurang akurat. Faktor penyebabnya adalah visualisasi 3D yang dibuat hanya menampilkan

perubahan gerakan alis yang statis pada setiap emosi, tidak ada perubahan gerakan pada mata, sedangkan gerakan bibir dan bentuk mulutnya tetap mempertahankan bentuk mulut sesuai viseme, tidak dipengaruhi oleh perubahan emosi.

- Tingkat akurasi berdasarkan visualisasi yang tampak pada gerakan bibir dalam mengucapkan kalimat Bahasa Indonesia yang diinputkan memiliki tingkat akurasi 81,04%. Terdapat 18,96% visualisasi gerakan bibir tidak terlihat jelas oleh responden. Faktor penyebabnya adalah responden tidak terbiasa mengamati wajah animasi berbicara tanpa suara, hanya menggerakkan bibir saja. Untuk itu diperlukan latihan berulang-ulang agar terbiasa dengan kondisi melihat gerak bibir tanpa suara seperti kaum deaf people.
- Tingkat pengukuran akurasi pembentukan visualisasi 3D animasi yang diukur berdasarkan survei ini belum merepresentasikan hasil penilaian akurasi visualisasi 3D animasi yang baik, subyektifitasnya cukup tinggi, karena dalam setiap tayangan video yang ditampilkan hanya memuat 2 pertanyaan untuk melihat hasil ekspresinya dan 2 pertanyaan untuk melihat hasil gerakan bibir dari ucapan yang disampaikan.

6.2 Usulan Penelitian Selanjutnya

Berdasarkan hasil-hasil yang diperoleh pada penelitian ini, maka usulan penelitian selanjutnya perlu diperhatikan hal-hal sebagai berikut:

1. Pengembangan sistem afektif berwujud karakter virtual khas Indonesia yang dicapai dengan menggunakan 12 model viseme Indonesia lebih mengeplorasi jumlah parameter geometrik *blend shape animation* yang tidak hanya 9 titik kontrol pada bentuk mulut di sekitar bibir model wajah animasi, akan tetapi menggunakan tambahan 10 titik-titik kontrol lainnya di sekitar mata dan alis, untuk mengembangkan kelas viseme Indonesia dengan lebih baik dan realistik.
2. Dalam pembuatan *IFFE-3D Database* ekspresi wajah yang berbasis pose 3D dapat dikembangkan menjadi ekspresi wajah yang mengacu tidak hanya pada bentuk parameter afeksi, akan tetapi juga memperhatikan parameter afeksi

pada gerakan kepala. Hal ini dikarenakan, gerakan kepala merupakan salah satu fitur yang bisa digunakan sebagai fitur penting dalam ekspresi wajah berbasis 3D, sehingga proses pengklasifikasian emosi lebih alami dan realistik.

3. Untuk mengoptimalkan akurasi dari hasil penelitian yang telah dilakukan ini, akan dilakukan eksplorasi perhitungan akurasi secara riil menggunakan metode penelitian lainnya dengan teknik-teknik tertentu dan baru. Jadi tidak hanya berdasarkan hasil survei saja, karena perhitungan hasil survei sering kali bersifat subjektif untuk banyak kasus penelitian di Indonesia.
4. Dengan menggunakan hasil gerakan bibir dari 12 model viseme Indonesia, sangat memungkinkan untuk dilakukan penelitian lebih lanjut dalam mengenali dan memahami kalimat yang diucapkan dengan cara mencocokkan sinyal digital dengan suatu pola tertentu (*speech recognition*). Kalimat Bahasa Indonesia yang diucapkan, diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi sekumpulan digit angka yang kemudian diekstrak dan disesuaikan dengan kode dan pola tertentu untuk mengidentifikasi kalimat tersebut, sehingga hasil dari identifikasi kalimat yang diucapkan disesuaikan dengan *pattern* dan ditampilkan dalam bentuk mimik dan gerak bibir, agar visualisasi yang dihasilkan lebih realistik.

DAFTAR PUSTAKA

- ____ “Arena Expression, Seven Step Quick Start Guide”, Natural Point Inc. 2008.
- A. Kappas, 2003. “ *What facial activity can and cannot tell us about emotions* ”. The human face: Measurement and meaning, M. Katsikitis, Ed. Kluwer Academic Publisher, pp. 215-234.
- A. Nogueiras, A. Moreno, A. Bonafonte, and Mario, 2001. "Speech Emotion Recognition Using Hidden Markov Models", in *Proceedings Eurospeech*, Scandinavia, pp. 2679-2682.
- Alfan Farizki Wicaksono, Ayu Purwarianti, 2010. “HMM Based Part-of-Speech Tagger for Bahasa Indonesia”, On Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop, 2nd August 2010.
- Anh Tuan Do, 2011. “*Creating Emotional Speech for Conversational Agents* ”. Proceeding of the 2011 IEEE Workshop on Digital Media and Digital Content Management, IEEE Computer Society, pp. 107-110.
- Anis, Muhammad dan Imam K, Achmad, 2012. *Penentuan Jarak Objek dari Kamera dengan Menggunakan Stereo Vision pada Mobile Phone*. Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika. Volume 1, Number 2, Juli 2012.
- Arifin, Mulyono, Surya Sumpeno, Mochamad Hariadi, 2013. “Towards Building Indonesian Viseme : A Clustering-Based Approach”, CYBERNETICSCOM 2013 IEEE International Conference on Computational Intelligence and Cybernetics, pp 57-61, December 2013.
- Arifin, Surya Sumpeno, Mochamad Hariadi, Hanny Haryanto, 2015. “*A Text-to-Audiovisual Synthesizer for Indonesian by Morphing Viseme*”, International Review on Computers and Software (IRECOS), Vol.10, No.11, November 2015.
- Bianca Aschenberner and Christian Weiss, 2005. “Phoneme-Viseme Mapping for German Video-Realistic Audio-Visual-Speech-Synthesis,” IKP-Working Paper NF 11, Institut für Kommunikations for Schung und Phonetik, Universität Bonn, 2005.
- C. Cullen, B. Vaughan, S. Kousidis, Wang Yi, C. McDonnell, and D. Campbell, 2015. “Generation of High Quality Audio Natural Emotional Speech Corpus using Task Based Mood Induction,” Digital Media Centre, Dublin Institute of Technology, Ireland, pp. 1-6.

Dailey, Mathew N.; Joyce, Carrie; dkk. 2010. "Evidence and a Computational Explanation of Cultural DJIFFERences in Facial Expression Recognition. Emotion, 2010, Volume 10(6), halaman 874-893.

David Alfter, Caroline Sporleder, Sven Naumann, 2015, Dissertation of Computerlinguistic & Digital Humanities, Language Segmentation, pp. 1-149.

Daniel Jurafsky & James H. Martin, 2015. "Part of Speech Tagging", Speech and Language Processing, pp. 1-39.

Dong-Mei Yu and Jian-An Fang, 2007. "Research on A Methodology to Model Speech Emotion". Proceeding of the 2007 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR 2007), Beijing, China, November 2-4, 2007, pp. 825-830.

Elif Bozkurt, Cigdem Eroglu Erdem, Engin Erzin, Tanju Erdem, Mehmet Ozkan, 2007. "Comparison of Phoneme and Viseme Based Acoustic Units for Speech Driven Realistic Lip Animation". Proceeding of the 2007 IEEE International Conference.

Endang Setyati, Surya Sumpeno, Mauridhi Hery Purnomo, Koji Mikami, Masanori Kakimoto, and Kunio Kondo, 2015. "Phoneme- Viseme Mapping for Indonesian Language Based on Blend Shape Animation," IAENG International Journal of Computer Science, Vol. 42, No. 3, pp. 233-244.

Endang Setyati, Yoyon K. Suprapto, and Mauridhi Hery Purnomo, 2012. "Facial Emotional Expressions Recognition Based on Active Shape Model and Radial Basis Function Network," in Proceeding of 2012 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA 2012), July, 2-4, 2012, pp. 41-46.

Endang Setyati, Surya Sumpeno, Mauridhi Hery Purnomo, 2016. "Hidden Markov Models based Indonesian Viseme Model for Natural Speech with Affection", Kursor, Scientific Journal on Information Technology, University of Trunojoyo Madura, ISSN 0216 - 0544, e-ISSN: 2301-6914, Vol.8, No.3, July 2016, pp. 215–222.

Etienne B. Roesch, Lucas Tamarit, Lionel Reveret, Didier Grandjean, David Sander, Klaus R. Scherer, 2010. "FACSGen: A Tool to Synthesize Emotional Facial Expressions Through Systematic Manipulation of Facial Action Units". Original Paper, Journal Nonverbal Behav, Springer Science+Business Media, LLC 2010, Publised online: 19 November 2010. DOI 10.1007/s10919-010-0095-9.

- G. Jonghwa Kim, 2008. “*Emotion Recognition Based on Physiological Changes in Music Listening*”. IEEE Transactions on Pattern Analysis and Machine Intelligence, December 2008, Vol. 30, No. 12.
- G. Potamianos, *et al.*, 2003. “Recent advances in the automatic recognition of audio-visual speech,” *Proceeding of the IEEE*, vol. 91, no. 9, 2003.
- Goranka Zorich, 2005. “Automatic Lip Synchronization by Speech Signal Analysis”, Master Thesis, Department of Telecommunications, Faculty of Electrical Engineering and Computing, pp. 1-87.
- Hao Tang, Yuxiao Hu, Yun Fu, et all., 2008. “*Real-Time Conversion from A Single 2D Face Image to A 3D Text-Driven Emotive Audio-Visual Avatar*”. Proceeding of the 2008 IEEE International Conference on Multimedia and Expo (ICME 2008), June 23-26, 2008, Hannover, Germany, pp. 1205-1208.
- Hazen, Saenko, La, and Glass, 2004. “A segment-based audio-visual speech recognizer: data collection, development, and initial experiments,” In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 235–242, State College, PA, USA. ACM, 2004.
- Jacey-Lynn Minoi, S. Hassan Amin, Carlos Eduardo Thomaz and Duncan Fyfe Gillies, 2008. “*Synthesizing Realistic Expressions in 3D Face Data Sets*”. Proceeding of the 2008 2nd IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2008), September 29-October 1, 2008, pp. 1-6.
- Jerome R. Bellegarda, 2011. “*A Data Driven Affective Analysis Framework Toward Naturally Expressive Speech Synthesis*”. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 19, No. 5, July 2011, pp. 1113-1122.
- Jianhua Tao and Le Xin, 2004. “*Mutual Features Controlling of Dynamic Visemes for Emotional Talking Head System*”. Proceeding of The 14th International Conference on Artificial Reality and Telexistence (ICAT) 2004, November 30-December 2, 2004, Coex-Seoul-Korea.
- Jianhua Tao, Yongguo Kang and Aijun Li, 2006. “*Prosody Conversion from Neutral Speech to Emotional Speech*”. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, July 2006, pp. 1145-1154.
- Jian-Qing Wang, Ka-Ho Wong, Pheng-Ann Heng, Helen M. Meng, and Tien-Tsin Wong, 2004. “*A Real Time Cantonese Text-To-Audiovisual Speech Synthesizer*”. Proceeding of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), May 17-21, 2004, Montreal, Quebec, Canada, pp. I-653-656.

Jui-Chen Wu, Yung-Sheng Chen, and I-Cheng Chang, 2007. “An Automatic Approach to Facial Feature Extraction for 3-D Face Modeling,” IAEENG International Journal of Computer Science, vol. 33, no. 2, pp1-7.

Kang Liu and Joern Ostermann, 2009. “*Optimization of an Image-Based Talking Head System*”, EURASIP Journal on Audio, Speech, and Music Processing, Volume 2009, Article ID 174192, pp. 1-13

Luca Cappelletta and Naomi Harte, 2012. “*Phoneme-to Viseme Mapping for Visual Speech Recognition*”. Proceeding of the 2012 International Conference on Pattern Recognition Applications and Methods (ICPRAM 2012), February 7, 2012.

Luca Cappelletta and Naomi Harte, 2011. “*Viseme Definitions Comparison for Visual-Only Speech Recognition*”. Proceeding of the 2011 European Signal Processing Conference (Eusipco), Barcelona, Spain. September 3, 2011.

Luca Cappelletta and Naomi Harte, 2010. “*Nostril Detection for Robust Mouth Tracking*”. Proceeding of the 2010 Irish Signals and Systems Conference (ISSC 2010), UCC, Cork, June 23–24, 2010.

M. Schroeder, 2014. “Speech and Emotion Research An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis, in Faculty of Philoshopy, 2014, Universit”at des Saarlandes, pp. 288-298.

Menache, Alberto. 2000. *Understanding Motion Caption for Computer Animation and Video Games*. London: Academic Press.

Meng Qingmei, Wu Weiguo, Zhong Yusheng, and Song Ce, 2008. “*Research and Experiment of Lip Coordination with Speech for the Humanoid Head Robot-H&Frobot-III*”. Proceeding of the 2008 8th IEEE-RAS International Conference on Humanoid Robots, December 1-3, 2008, Daejeon, Korea, pp. 603-608.

Ming Xu and Ruimin Hu, 2006. “*Mouth Shape Sequence Recognition Based on Speech Phoneme Recognition*”. Proceeding of the 2006 IEEE First International Conference on Communications and Networking in China (ChinaCom 2006), October 25-27, 2006, pp. 1-5.

Nadia Mana and Fabio Pianesi, 2006. “*HMM-based Synthesis of Emotional Facial Expressions during Speech in Synthetic Talking Heads*”. Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI 2006), Banff, Alberta, Canada, November 2-4, 2006, pp. 380-387.

Pascal Damien, Nagi Wakim, and Marcel Egéa, 2009. “*Phoneme-Viseme Mapping for Modern, Classical Arabic Language*,” *Proceeding of the The 2009 IEEE International Conference on Advances in Computational Tools for*

Engineering Applications (ACTEA 2009), Zouk Mosbeh, Lebanon, July 14-17, 2009, pp. 547-552.

Patrick Lucey, Terrence Martin and Sridha Sridharan, 2004. “*Confusability of Phonemes Grouped According to Their Viseme Classes in Noisy Environments*”. Proceedings of the 10th Australian International Conference on Speech Science and Technology, Sydney, December 8-10, 2004, pp. 265-270.

Pushkar Joshi, Wen C. Tien, Mathieu Desbrun and Frédéric Pighin, 2003. “Learning Controls for Blend Shape Based Realistic Facial Animation,” *Eurographics/SIGGRAPH Symposium on Computer Animation*, D. Breen, M. Lin (Editors).

R. Cowie, E. Douglas-Cowie, K. Karpouzis, G. Caridakis, M. Wallace and S. Kollias, 2008. “Recognition of Emotional States in Natural Human-Computer Interaction”, pdfs.semanticscholar.org, Department of Computer Science, University of Indianapolis, Athens Campus, pp. 1-39

Rainer Reisenzein, 2010. “*Broadening the Scope of Affect Detection Research*”. IEEE Transaction On Affective Computing, Vol. 1, No. 1, January-June 2010, pp. 42-45.

Rong Chuanzhen, Yue Zhenjun, wang Yuan, Yang Yu, 2012. “*Research on Chinese Viseme Based on Fuzzy Clustering and Grey Relation Analysis*”, Proceeding of the 2012 IEEE 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012), pp. 321-324.

Sakriani Sakti, Eka Kelana, Hammam Riza, Shinsuke Sakai, Konstantin Markov, Satoshi Nakamura, 2008. “Development of Indonesian Large Vocabulary Continuous Speech Recognition System within A-STAR Project,” In *Proceeding TCAST*, Hyderabad, India, January 2008, pp. 19-24.

Salah Werda, Walid Mahdi and Abdelmajid Ben Hamadou, 2007. “*Lip Localization and Viseme Classification for Visual Speech Recognition*”. International Journal of Computing and Information Sciences, Vol. 5, No. 1, April 2007, On-Line, pp. 62-75.

Savran, Arman; dkk. 2012. *Comparative Evaluation of 3D vs. 2D Modality for Automatic Detection of Facial Action Units*. Pattern Recognition, volume 45(2012), halaman 767-782.

Shen Zhang, Zhiyong Wu, Helen M. Meng, and Liannhong Cai, 2007. “*Facial Expression Synthesis Using PAD Emotional Parameters for a Chinese Expressive Avatar*”. Proceeding of the Second International Conference on Affective Computing and Intelligent Interaction,, Lisbon, Portugal

September 12-14, 2007. ACII 2007, LNCS 4738, pp. 24-35. Springer-Verlag Berlin Heidelberg 2007.

Surya Sumpeno, Mochamad Hariadi, and Mauridhi Hery Purnomo, 2011. "Facial Emotional Expressions of Life-like Character Based on Text Classifier and Fuzzy Logic," *IAENG International Journal of Computer Science*, Vol. 38, no. 2, pp. 122-133, 2011.

Tekalp, A. Murat; Ostermann, Jorn. 2000. *Face and 2D Mesh Animation in MPEG-4*. Signal Processing: Image Communication 15, pp. 387-421.

Tie Yun, Ling Guan, 2010. "*Human Emotion Recognition Using Real 3D Visual Features from Gabor Library*". Proceeding of the 2010 IEEE International Workshop on Multimedia Signal Processing (MMSP 2010), October 4-6, 2010, Saint-Malo, France, pp. 505-510.

Tie Yun, Ling Guan, 2012. *Human Emotion State Recognition Using Real 3D Visual Features from Gabor Library*, 2012 Elsevier International Journal of Pattern Recognition 46 (2013), pp. 529-538.

Wang Zhiming, CAI Lianhong, AI Haizhou, 2002. "*A Dynamic Viseme Model for Personalizing A Talking Head*". Proceeding of IEEE the 6th International Conference on Signal Processing (ICSP 2002), August, 26-30, 2002, pp. 1015-1018.

Yingli Tian, Takeo Kanade, and Jeffrey F. Cohn, 2001. "*Recognizing Action Units for Facial Expression Analysis*". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, pp. 97-116.

Yui, Lijun; Xiazhou, Wei, dkk. 2006. A 3D Facial Expression Database for Facial Behavioral Research. The 7th International Conference on Automatic Face and Gesture Recognition (FG 2006), IEEE Computer Society TC PAMI. Southampton, UK.



Biodata Penulis

Nama	:	Endang Setyati
NRP	:	2211 301 005
Program	:	Doktor
Program Studi	:	Teknik Elektro
Fakultas	:	Fakultas Teknologi Industri Institut Teknologi Sepuluh Noverember Surabaya (ITS)
Tempat/tanggal lahir	:	Probolinggo, 20 Nopember 1967
Agama	:	Islam
Pekerjaan	:	Dosen Tetap Sekolah Tinggi Teknik Surabaya (STTS)
Alamat Kantor	:	Kampus STTS Jl. Ngagel Jaya Tengah 73-77 Surabaya – 60293
Email	:	endang@stts.edu , esetyati@gmail.com
Nama Orang Tua	:	K. Sastro Soemarno (almarhum) Oemijati K. Tjiptodihardjo (almarhumah)
Nama Suami	:	Didik Wahyudi
Nama Anak	:	1. Anggraita Arindra Febrianti, S.H. (<i>mahasiswa S2 Magister Kenotariatan, Fakultas Hukum, Universitas Airlangga, Surabaya</i>) 2. Nadia Laksmita Sari (<i>mahasiswa S1 Ekonomi Islam, Fakultas Ekonomi dan Bisnis, Universitas Airlangga, Surabaya</i>) 3. Amadhea Trizza Audina (<i>siswa kelas 7, SMP Negeri 6 Surabaya</i>)

Riwayat Pendidikan:

A. Pendidikan Dasar dan Menengah :

1. TK Siwi Peni I, Probolinggo – 1972
2. SD Negeri Sukabumi II, Probolinggo – 1974
3. SMP Negeri I, Probolinggo – 1980
4. SMA Negeri I, Probolinggo – 1983

B. Pendidikan Tinggi:

1. Strata-1 : Sarjana Matematika – 1986
Institut Teknologi Bandung (ITB)
Skripsi : Analisa Komponen Utama Pengaruh Perlakuan
Pembimbing : Prof. Dr. Sutawanir Darwis.

2. Strata-2 : Magister Teknik Informatika – 1998
Institut Teknologi Sepuluh November (ITS)
Tesis : Pengenalan Wajah dengan Menggunakan Analisa Komponen Utama (Principal Component Analysis)
Pembimbing : Prof. Ir. Handayani Tjandrasa, M.Sc., Ph.D.

3. Strata-3 : Doktor Teknik Elektro – 2011
Institut Teknologi Sepuluh November (ITS)
Disertasi : Talking Head System in Indonesian Language with Affective Facial Expressions Synthesis
Promotor : Prof. Ir. Mauridhi Hery Purnomo, M.Sc., Ph.D.
Co-Promotor : Dr. Surya Sumpeno, S.T., M.Sc.

Riwayat Pekerjaan / Jabatan:

1. Dosen Tetap STTS (1993 – sekarang)
2. Dosen Luar Biasa UK Petra (2003 – sekarang)
3. Ketua Program Studi S2 Magister Teknologi Informasi STTS (2002 – sekarang)
4. Direktur Pengembangan Kelembagaan STTS (2008 – sekarang)

Daftar Publikasi selama studi Program Doktor:**A. Jurnal Internasional**

1. **Endang Setyati**, Surya Sumpeno, Mauridhi Hery Purnomo, Koji Mikami, Masanori Kakimoto, Kunio Kondo, “*Phoneme-Viseme Mapping for Indonesian Language Based on Blend Shape Animation*”, IAEENG International Journal of Computer Science, (terindex scopus), ISSN: 1819-9224 (online version); 1819-656X (print version), Vol. 42, No. 3, pp. 233-244, September 2015, Online: 25 July 2015.

B. Jurnal Nasional Terakreditasi

1. **Endang Setyati**, Surya Sumpeno, Mauridhi Hery Purnomo, ”*Hidden Markov Models based Indonesian Viseme Model for Natural Speech with Affection*”, Kursor, Scientific Journal on Information Technology, University of Trunojoyo Madura, ISSN 0216 - 0544, e-ISSN: 2301-6914, Vol.8, No.3, July 2016, pp. 215–222.

C. Seminar Internasional

1. **Endang Setyati**, Yoyon K. Suprapto, Mauridhi Hery Purnomo, "*Facial Emotional Expressions Recognition Based on Active Shape Model and Radial Basis Function Network*", 2012 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA 2012), Tianjin, China, 2-4 July 2012.
2. **Endang Setyati**, Yoyon K. Suprapto, Mauridhi Hery Purnomo, "*Facial Emotional Expressions Synthesis using Radial Basis Function Network*", 3rd International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT 2012), ISBN: 978-602-97124-1-4, UK Petra, Denpasar, 24-25 Mei 2012.

Daftar Publikasi Hibah Penelitian:

A. Jurnal Nasional Terakreditasi:

1. Raymond Sutjiadi, **Endang Setyati**, Resmana Lim, "*Adaptive Background Extraction for Video Based Traffic Counter Application Using Gaussian Mixture Models Algorithm*", Telkomnika, Indonesian Journal of Electrical Engineering and Computer Science, ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v13i3.1772, Vol.13, No.3, September 2015, pp. 1006–1013.

B. Seminar Internasional

1. **Endang Setyati**, Daniel Widjaja, David Alexandre, "*Face Tracking Implementation with Pose Estimation Algorithm in Augmented Reality Technology*", The 2012 International Summer Conference on Asia Pacific Business Innovation & Technology Management, (APBITMS 2012), Kuala Lumpur, Malaysia, July 01-03, 2012.
2. **Endang Setyati**, Yosi Kristian, Yuliana Melita Pranoto, David Alexandre, "*Augmented Reality 3D Eyeglasses Frame Simulator Using Active Shape Model and Real Time Face Tracking*", 11th IEEE International Conference on ICT and Knowledge Engineering (ICT & KE 2013), Siam University, Bangkok, Thailand, November 20-22, 2013.
3. Yosi Kristian, **Endang Setyati**, Yuliana Melita Pranoto, Arya Tandy Hermawan, "*Human Facial Age Classification Using Active Shape Model, Geometrical Feature, And Support Vector Machine On Early Growth Stage*", ISICO 2015, Institut Teknologi Sepuluh Nopember, Surabaya, ISSN 2477-5541, ISBN 978-602-73429-0-3, November 5, 2015.
4. Yuliana Melita Pranoto, **Endang Setyati**, Edwin Pramana, Yosi Kristian, Renato Budiman, "*Real Time Handwriting Recognition For Matematic Expressions Using Hidden Markov Model*", International Seminar on Intelligent Technology and Its Application (ISITIA 2016), Institut

Teknologi Sepuluh November, Surabaya, ISSN/ISBN 9781509017089, July 28-30, 2016.

C. Seminar Nasional

1. **Endang Setyati**, David Alexandre, "Pemanfaatan Real-Time Face Tracking dalam Aplikasi Augmented Reality Frame Kacamata", 1st Seminar Nasional Inovation Desain and Technology (Ideatech 2011), STTS, Surabaya, ISSN: 2089-1121, 11 Nopember 2011.
2. **Endang Setyati**, David Alexandre, Daniel Widjaja, "Pengembangan Program Simulator Frame Kacamata Secara Real-Time 3D Face Tracking dengan Menggunakan Augmented Reality", 13th Seminar of Intelligent Technology and Its Applications (SITIA 2012), ITS Surabaya, 23 Mei 2012.
3. **Endang Setyati**, Joan Santoso, Sonny Budiarso, "Particle Swarm Optimization Untuk Contrast Stretching Citra Grayscale", Mataram NTB, 14-15 Februari 2013. ISBN 978-602-17488-0-0.
4. Ronny Makhfuddin Akbar, **Endang Setyati**, "Penilaian Otomatis Lembar Jawaban Komputer (LJK) Secara Real Time dengan Memanfaatkan Webcam", 2nd Seminar Nasional Innovation Desain and Technology (Ideatech 2015), STTS, Surabaya, ISSN: 2089-1121, 19 Maret 2015.

Pengalaman Hibah Penelitian:

1. Ketua Peneliti dalam Program Penelitian Hibah Bersaing, "*Pengembangan Program Simulator Kacamata dengan Menggunakan Augmented Reality*", Kemdikbud, Direktorat Jenderal Pendidikan Tinggi, Direktorat Penelitian dan Pengabdian kepada Masyarakat, **tahun 2011 dan tahun 2012 (dalam 2 tahun pendanaan)**.
2. Program Penelitian Hibah Disertasi Doktor, "*Pemetaan Fonem Bahasa Indonesia ke Viseme Indonesia untuk Pengembangan Sistem Afektif Berwujud Karakter Virtual*", Kemdikbud, Direktorat Jenderal Pendidikan Tinggi, Direktorat Penelitian dan Pengabdian kepada Masyarakat, **tahun 2014**.
3. Anggota Peneliti dalam Program Penelitian Hibah Bersaing, "*Pengembangan Sistem Klasifikasi Usia Manusia Berdasarkan Fitur Wajah Menggunakan Active Shape Model dan Multiple Classifier*", Kemdikbud, Direktorat Jenderal Pendidikan Tinggi, Direktorat Penelitian dan Pengabdian kepada Masyarakat, **tahun 2015 dan tahun 2016 (2 tahun pendanaan)**.
4. Ketua Peneliti, dalam Program Penelitian Produk Terapan, "*Pembentukan Urutan Viseme (Visual Phoneme) Indonesia untuk Pengembangan Animasi*

3D dengan Menggunakan Hidden Markov Model”, Kemristek Dikti, Direktorat Penelitian dan Pengabdian kepada Masyarakat, tahun 2017.

5. Ketua Peneliti dalam Program Hibah Iptek Bagi Kewirausahaan, “Kewirausahaan di Sekolah Tinggi Teknik Surabaya”, Kemristek Dikti, Direktorat Penelitian dan Pengabdian kepada Masyarakat, **tahun 2017**.

Penerima Beasiswa dari Ditjen Dikti:

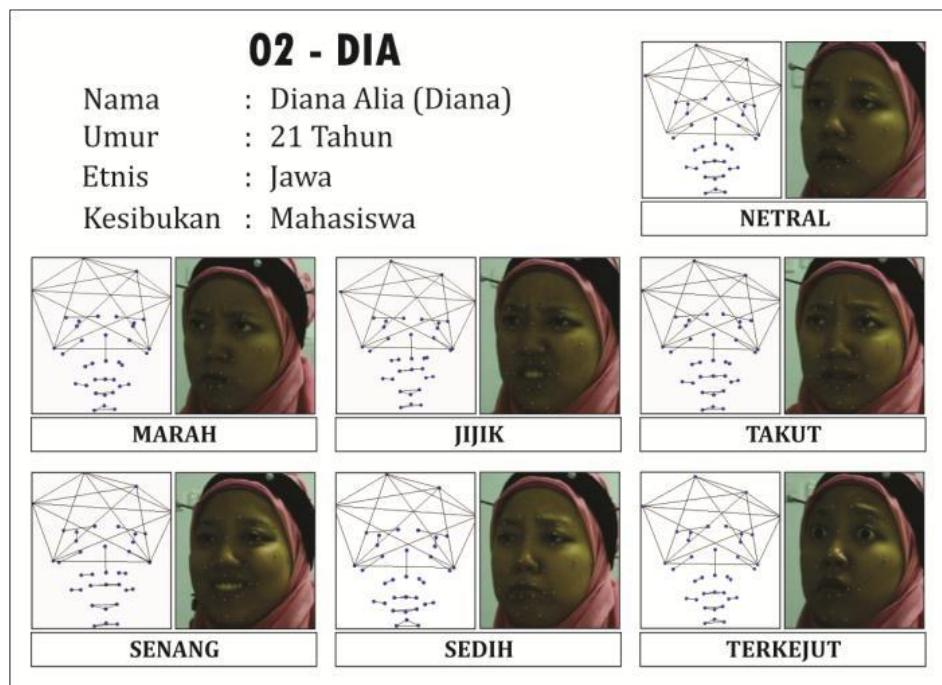
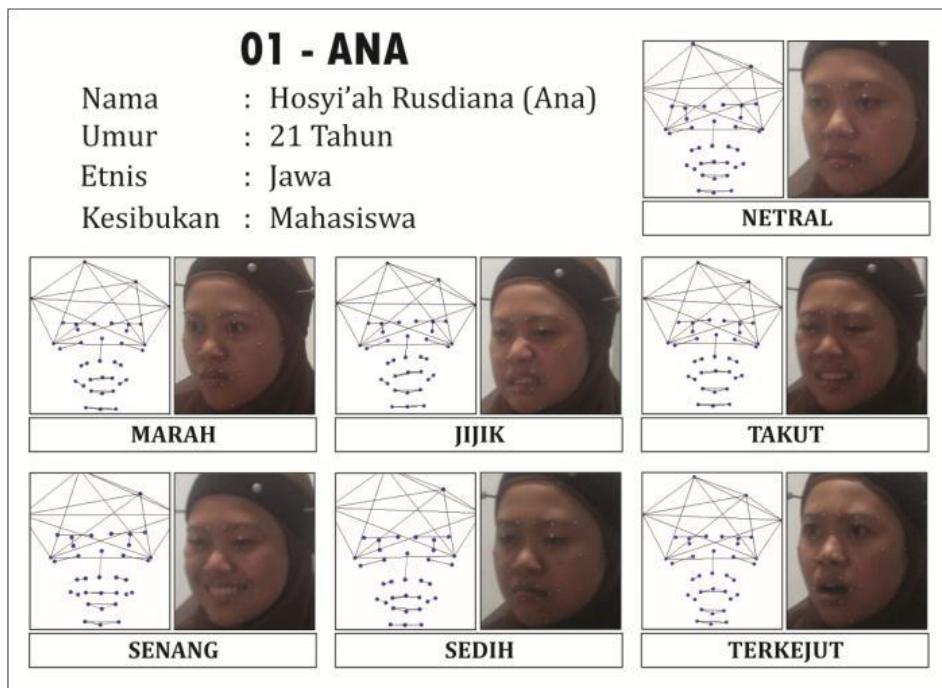
1. Beasiswa Program Pascasarjana Dalam Negeri (BPPDN), 2011-2015, Kemdikbud, Direktorat Jenderal Pendidikan Tinggi, Direktorat Pendidikan dan Tenaga Kependidikan.
2. Beasiswa Program Sandwich, Peningkatan Kualitas Publikasi Internasional (PKPI), Agustus-Desember 2013, di Tokyo University of Technology, Tokyo, Japan, disponsori oleh Kemdikbud, Direktorat Jenderal Pendidikan Tinggi, Direktorat Pendidikan dan Tenaga Kependidikan.

Keanggotaan Organisasi Ilmiah Internasional

1. Anggota IEEE (The Institute of Electrical and Electronics Engineer)
2. Anggota IAENG (International Association of Engineers)

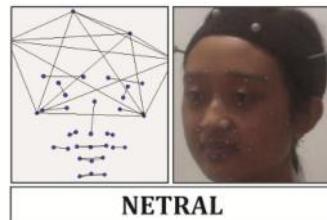
[halaman ini sengaja dikosongkan]

LAMPIRAN 1.
DATA DAN BIODATA MODEL JIFFE-3D DATABASE
(Hosyiah Rusdiana, 2015)

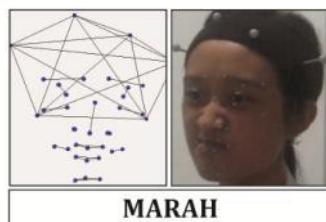


03 - FEB

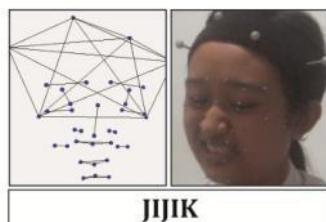
Nama : Febriyanti A. P. (Febri)
Umur : 18 Tahun
Etnis : Jawa
Kesibukan : Pelajar SMA



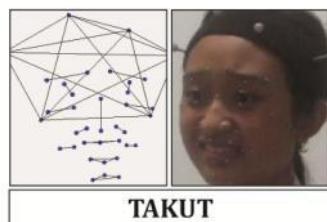
NETRAL



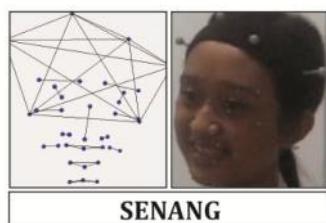
MARAH



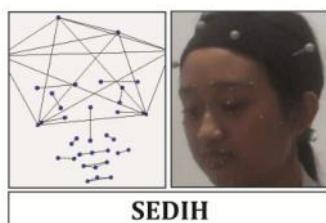
JIJIK



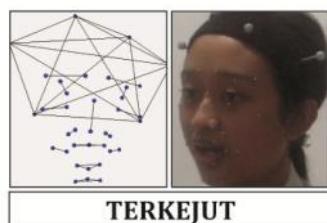
TAKUT



SENANG



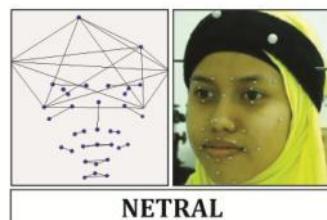
SEDIH



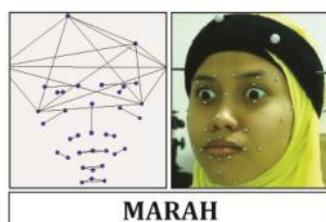
TERKEJUT

04 - NES

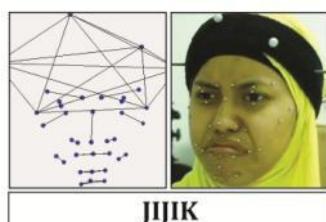
Nama : Annisa Pujiantara (Nesya)
Umur : 20 Tahun
Etnis : Jawa
Kesibukan : Mahasiswa



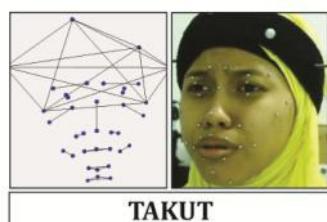
NETRAL



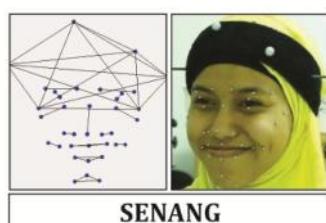
MARAH



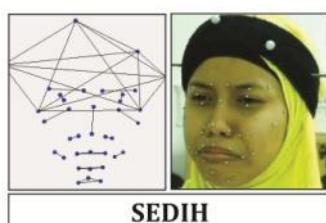
JIJIK



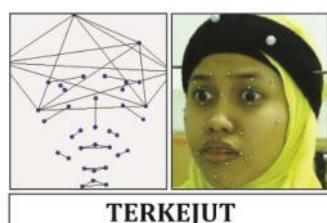
TAKUT



SENANG



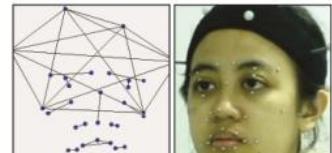
SEDIH



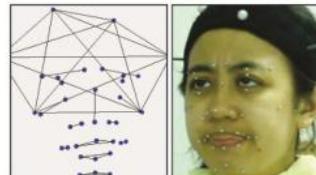
TERKEJUT

05 - NIN

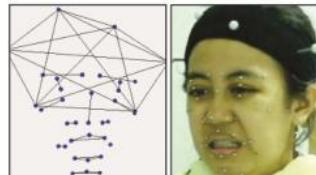
Nama : Nindita S. Putri (Nindita)
Umur : 22Tahun
Etnis : Jawa
Kesibukan : Mahasiswa



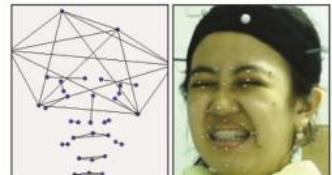
NETRAL



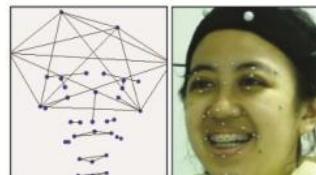
MARAH



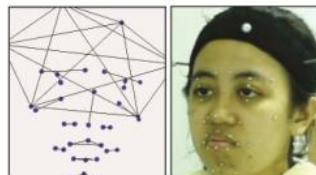
JIIK



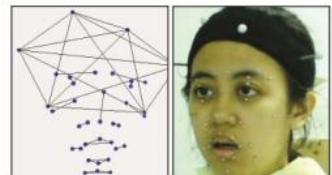
TAKUT



SENANG



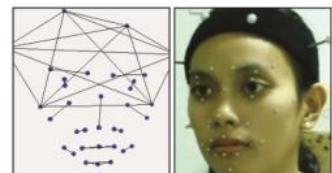
SEDIH



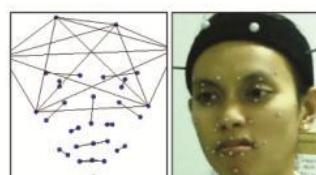
TERKEJUT

06 - NIT

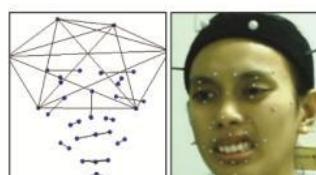
Nama : Nita Indriyani P. (Nita)
Umur : 21Tahun
Etnis : Jawa
Kesibukan : Mahasiswa



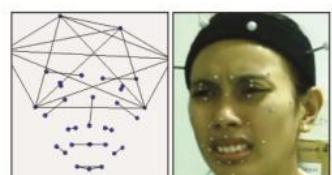
NETRAL



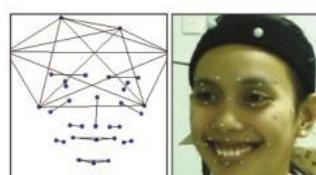
MARAH



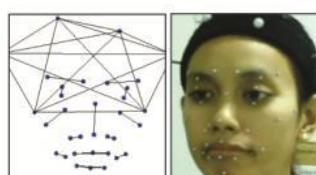
JIIK



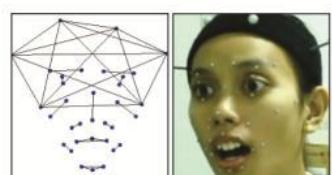
TAKUT



SENANG



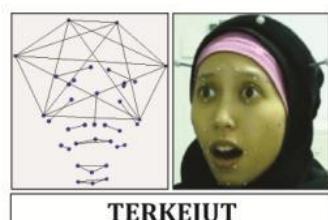
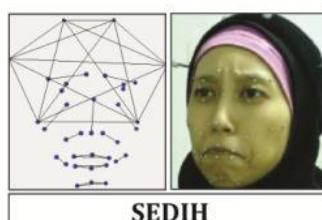
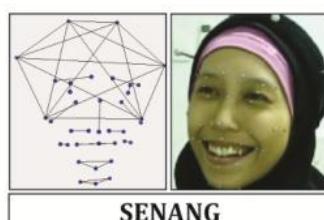
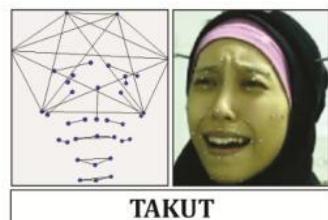
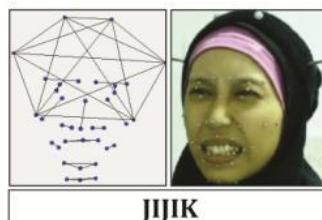
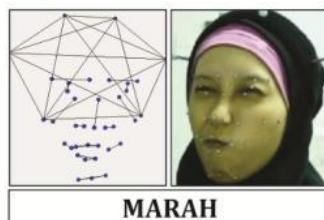
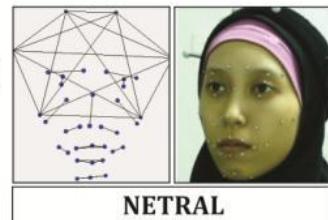
SEDIH



TERKEJUT

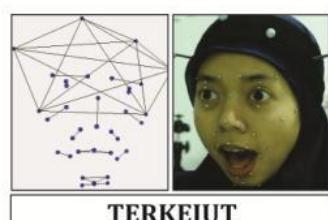
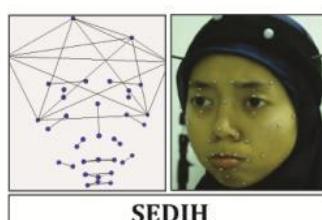
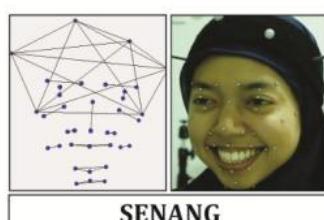
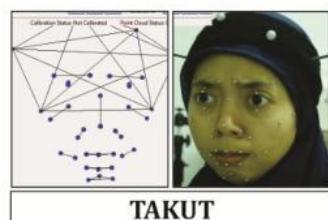
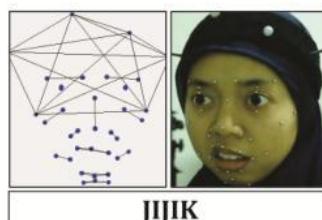
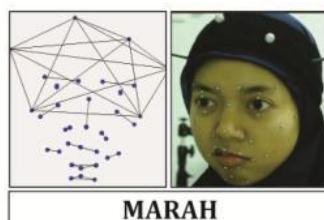
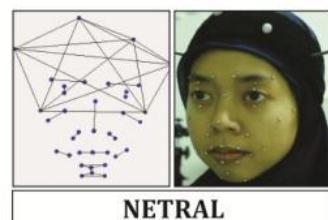
07 - NOO

Nama : Noora Miftakhul R. (Noora)
Umur : 22Tahun
Etnis : Jawa
Kesibukan : Mahasiswa



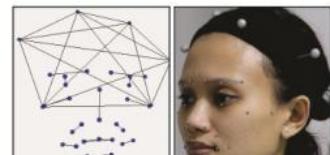
08 - RIA

Nama : Annisa Riani S. (Riani)
Umur : 21Tahun
Etnis : Jawa
Kesibukan : Mahasiswa

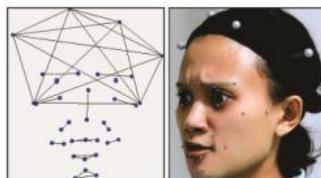


09 - TIA

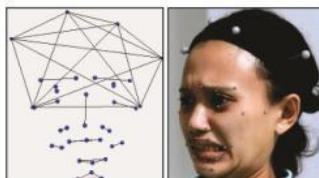
Nama : Artiarini K. N. (Tia)
Umur : 24 Tahun
Etnis : Jawa
Kesibukan : Mahasiswa



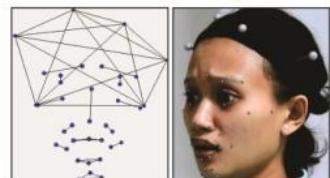
NETRAL



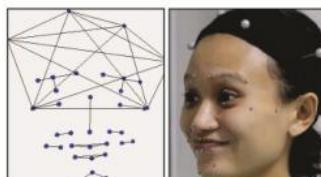
MARAH



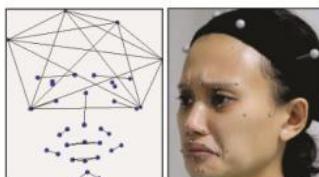
JIJIK



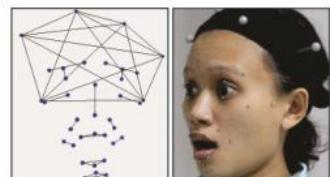
TAKUT



SENANG



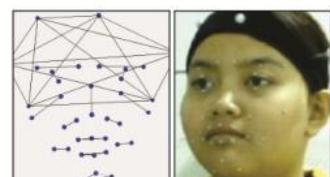
SEDIH



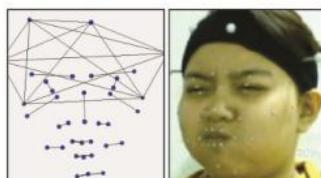
TERKEJUT

10 - YOK

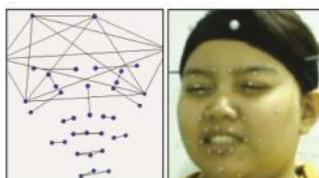
Nama : Yoka Cininta P. (Yoka)
Umur : 22 Tahun
Etnis : Jawa
Kesibukan : Mahasiswa



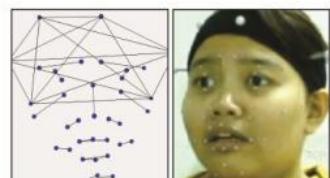
NETRAL



MARAH



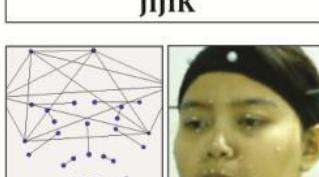
JIJIK



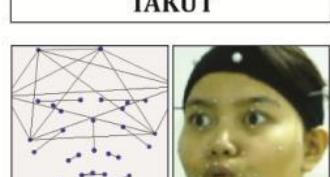
TAKUT



SENANG



SEDIH



TERKEJUT

[halaman ini sengaja dikosongkan]