



TESIS TE 142599

PEMODELAN VISEME DINAMIS 3D UNTUK BAHASA INDONESIA

NURUL FADILLAH.
2213205005

Dosen Pembimbing:
Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D.
Dr. Surya Sumpeno, ST, M.Sc.

PROGRAM MAGISTER
BIDANG KEAHLIAN JARINGAN CERDAS MULTIMEDIA
JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI INDUSTRI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2015

[Halaman ini sengaja dikosongkan.]



TESIS TE 142599

3D VISEME DYNAMIC MODELLING FOR INDONESIAN

NURUL FADILLAH.
2213205005

Supervisors:

Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D.

Dr. Surya Sumpeno, ST, M.Sc.

MAGISTER PROGRAM
AREAS OF EXPERTISE NETWORK INTELLIGENT MULTIMEDIA
DEPARTMENT OF ELECTRICAL ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
SEPULUH NOPEMBER INTITUTE OF TECHNOLOGY
SURABAYA
2015

[Halaman ini sengaja dikosongkan.]

Tesis ini disusun untuk memenuhi salah satu syarat untuk memperoleh gelar
Magister Teknik(MT)

di

Institut Teknologi Sepuluh Nopember Surabaya

Oleh :

Nurul Fadillah.

NRP. 2213205005

Tanggal Ujian : 17 Juni 2015
Periode Wisuda : September 2015

Disetujui oleh:


Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D.

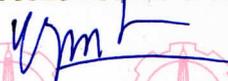
NIP. 19580916 198601 1 001

(Pembimbing I)


Dr. Surya Sumpeno, ST, M.Sc.

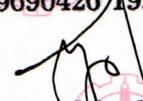
NIP. 19690613 199702 1 003

(Pembimbing II)


Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc.

NIP. 19690426 199403 1 003

(Penguji I)


Dr. Supeno Mardj Susiki Nugroho, ST, MT.

NIP. 19700313 199512 1 001

(Penguji II)

Mengetahui,
Direktur Program Pascasarjana ITS


Prof. Dr. Ir. Adi Soeprijanto, MT

NIP. 19640405 199002 1 001



PEMODELAN VISEME DINAMIS 3D UNTUK BAHASA INDONESIA

Nama Mahasiswa : NURUL FADILLAH.
NRP : 2213205005
Pembimbing I : Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D.
Pembimbing II : Dr. Surya Sumpeno, ST, M.Sc.

Abstrak

Interaksi Manusia-Komputer yang natural memerlukan integrasi audio dan visual berbicara yang realistik untuk persepsi dan tampilan, sistem ekspresi yang multimodal membuat komunikasi yang user-friendly antara manusia dan komputer. Hal ini mencakup beberapa hal, seperti : pengolahan suara , wajah manusia 3D, dan lain-lain. Pemodelan 3D adalah objek atau benda yang di desain sebaik mungkin sehingga objek tersebut terlihat hidup. Pemodelannya berupa wajah manusia, Wajah merupakan penutur menyampaikan berbagai jenis informasi, baik mengenai penutur maupun apa yang dikatakan. Viseme adalah visual fonem yaitu bentuk bibir ketika mengucapkan suatu fonem atau bunyi bahasa. Penelitian mengenai viseme ini telah banyak dilakukan, misalnya untuk bahasa English, Swedish, Spanish, Persian, dan lain-lain. Sayangnya, penelitian viseme bahasa Indonesia belum banyak dilakukan. Untuk menghasilkan animasi berbicara yang realistik, diperlukan viseme yang merepresentasikan fonem tertentu dengan memperhitungkan fonem yang mengikutinya. Viseme seperti ini disebut dengan viseme dinamis. Animasi berbicara dengan menggunakan viseme dinamis dapat menghasilkan animasi berbicara yang lebih natural dan masuk akal dibandingkan menggunakan viseme statis . Penelitian ini bertujuan untuk pemodelan viseme dinamis 3D untuk Bahasa Indonesia yang prosesnya hanya fokus pada gerakan mulutnya saja dari model tersebut. Metode yang diusulkan dalam penelitian ini terdiri dari pembuatan model viseme dinamis 3D bahasa Indonesia dengan pendekatan data-driven , Segmentasi, Normalisasi,clustering menggunakan Algoritma K-mean.Dari hasil cluster k- mean didapatkan hasil cluster yang baik dengan K=20 dari 22 cluster berdasarkan hasil penelitian cluster yang dilakukan, hasil cluster yang baik ini berdasarkan dari matrik kalkulasi untuk SSE dan rasio untuk BCV dan WCV . Penelitian ini diharapkan dapat menghasilkan animasi berbicara Bahasa Indonesia yang lebih natural. Salah satu bidang yang akan dikembangkan dalam penelitian adalah penerapan viseme dinamis pada animasi berbicara . Hasil penelitian ini juga diharapkan dapat memberikan kontribusi pada perkembangan berbagai bidang, misalnya produksi film animasi, game, interaksi manusia dan komputer serta pada linguistik Bahasa Indonesia.

Kata-kunci: 3D, fonem, viseme , Viseme Dinamis

[Halaman ini sengaja dikosongkan.]

3D VISEME DYNAMIC MODELLING FOR INDONESIAN

Name : NURUL FADILLAH.
NRP : 2213205005
Promotor : Prof. Ir. Mauridhi Hery Purnomo, M.Eng., Ph.D.
Co-Promotor : Dr. Surya Sumpeno, ST, M.Sc.

Abstract

Natural Human-Computer Interface requires integration of realistic audio and visual information for perception and display, Multimodal expression system makes the communication easy and friendly between human and computer. It covers several areas, such as speech processing , 3D human face, etc. 3D modeling is the object or objects that are designed the best possible so that the object seen alive. The face of a speaker conveys many kinds of information about both the speaker and the content of what is being said. Viseme is visual phonemes that form lip when to say something phonemes or sounds of language. Viseme associated with a particular language. Therefore, each language has a structure different viseme classes. Research on this viseme have been conducted, for example, for English, Swedish, Spanish, persian, etc. Unfortunately, Indonesian viseme research has not been done. To generate realistic speech animation, necessary of viseme represent a particular phoneme with to account phoneme that follow which is called the dynamic viseme. Lip Animation by using the dynamic viseme is more natural and realistic be compared using static viseme. This study aims to 3D viseme dynamic modelling for Indonesian that the process is only focused on her mouth movements of the model. The method proposed in this study consisted of 3D viseme dynamic modeling Indonesian with data - driven approach, segmentation, normalization, clustering using C-mean algorithm. the results of cluster k-mean cluster obtained good results with $K = 20$ of 22 clusters based on the results of research conducted cluster, cluster good results is based on matrix calculations for SSE and ratios for BCV and WCV. Dari C-mean cluster results obtained cluster good results with $K = 20$ of 22 clusters based on the results of research conducted cluster, cluster good results is based on matrix calculations for SSE and ratios for BCV and WCV. The results of this study are expected to contribute to the development of various fields, such as the production of animated films, games, and interaction human and computer as well as the linguistic Indonesian.

Key-words: *3D, Viseme, Fonem, Viseme dynamic*

[Halaman ini sengaja dikosongkan.]

CONTENTS

TITLE PAGE	i
STATEMENT OF AUTHENTICITY THESIS	vii
ABSTRAK	ix
ABSTRACT	xi
ACKNOWLEDGEMENTS	xiii
CONTENTS	xvii
LIST OF FIGURES	xix
LIST OF TABLES	xxiii
NOMENCLATURE	xxv
Chapter I INTRODUCTION	1
1.1 Background	1
1.2 Problem Formulation	2
1.3 Research Purposes	3
1.4 Benefit	3
Chapter II LITERATURE AND BASIC THEORY	5
2.1 <i>State of the art</i>	5
2.2 Basic Concepts 3D Modelling	6
2.3 Basic 3D Modelling Methods	7
2.4 Marker	11
2.5 <i>Facial Motion Capture</i>	12
2.6 C3D	12
2.7 Basic Definition phoneme	13
2.8 Various phoneme	14
2.9 Pattern Syllable	15
2.10 Viseme	17
2.11 Viseme Dynamic	17
2.12 Viseme Dynamic Modeling Indonesian	17
2.13 Normalisationi 3D	17
2.14 Segmentation	20
2.15 C-Means	20
2.16 Measuring Cluster Quality	25
2.17 Visualization Speech Animation	26

Chapter III	RESEARCH METHODS	29
3.1	Materials research	29
3.2	Research Methods	30
3.3	Preparation Model	31
3.4	Area Capture / Position Camera	32
3.5	Recording process in marker	33
3.6	Formation of Face Template	33
3.7	Model 3D	34
3.8	The Process of the World Coordinate System	35
3.9	The Process of Normalisation Position 3D	35
3.10	The Process of Segmentation	37
3.11	Clustering	38
Chapter IV	RESULTS AND DISCUSSION	41
4.1	Data Research	41
4.2	Results of Face Recording	42
4.3	Results of formation face Template	42
4.4	Normalisation Position 3D	43
4.5	Results of Segmentation	43
4.6	Clustering with Algorithm C - mean	44
4.7	Measuring Quality of Clustering	45
Chapter V	CONCLUSION	71
5.1	Conclusion	71
5.2	Further work	71
	APPENDIX	75
	BIOGRAPHY	81

DAFTAR TABEL

Tabel 2.1	Pemilihan Fonem Bahasa Indonesia	15
Tabel 2.2	Pemilihan Fonem Bahasa Indonesia	16
Tabel 2.3	Pola Suku Kata	17
Tabel 4.1	Kalimat yang digunakan sebagai data	41
Tabel 4.2	Kalkulasi SSE dan rasio untuk BCV dan WCV	56
Tabel 4.3	Deskripsi Hasil Klasterisasi K=20	59
Tabel 4.4	Deskripsi Hasil Klasterisasi K=20	60
Tabel 4.5	Struktur Kelas Viseme	61

[Halaman ini sengaja dikosongkan.]

DAFTAR GAMBAR

Gambar 2.1	Sumber: (Hui Zhao dan Chaojing Tang, 2011)	5
Gambar 2.2	Proses <i>Grafting</i> , input geometri (a), <i>triangles</i> yang tumpang tindih dihapus (c), Batas model bibir dan batas dari segitiga dihapus (d), <i>triangles</i> baru ditambahkan (e), akhirnya, model geometri bibir ditambahkan ke input geometri (f)(sumber: Robert Anderson, 2012).	7
Gambar 2.3	Otot-otot Kulit Wajah(Sumber : (R. Kent and F. Mini e, 1977))	8
Gambar 2.4	Gambar yang menunjukkan tampilan berbagai atribut layar dan gerakan dari model bibir Untuk menganimasikan visualisasi wicara perlu dipahami mengenai(Sumber :Thbaut weise, 2005)	8
Gambar 2.5	<i>Vertex</i> , <i>edge</i> dan wajah serta <i>polygon mesh</i> yang membentuk objek 3D(Sumber :(M. E. Hennecke, K. V, dkk ,1988).	9
Gambar 2.6	<i>Vertex-vertex</i> yang terkait dengan gerakan bibir atas (a) dan bibir bawah (b) (Sumber : (P. Delmas, 1999.))	9
Gambar 2.7	Contoh model <i>viseme</i> Bahasa Indonesia dengan nilai FAP tertentu(Sumber: Mazonaviciute dkk, 2011)	10
Gambar 2.8	Titik-Titik Fidusia(Sumber : Jose Mario De Martino.dkk, 2006)	10
Gambar 2.9	Penanda aktif(Sumber:(Zign Creation, 2013))	12
Gambar 2.10	Model kamera lubang jarum(Sumber:(Eka Prasetyo, 2010))	18
Gambar 2.11	Model kamera lubang jarum(Sumber:(Eka Prasetyo, 2010))	18
Gambar 2.12	Beberapa <i>cluster</i> dari <i>clustering criterion</i> . Gambar (a) kelompok cara binatang membawa keturunannya, Gambar (b) kelompok paru-paru binatang(sumber: Kardi , 2007))	21
Gambar 2.13	Identifikasi Kelompok(Sumber:(Kardi, 2007))	21
Gambar 2.14	Contoh Kasus(sumber: Kardi , 2007))	23
Gambar 2.15	<i>iterasi</i> (sumber: Kardi , 2007))	24
Gambar 2.16	Contoh Sinkronisasi Sinyal Wicara, Model Viseme Dinamis(sumber: Sarah L. Taylor , 2012))	27
Gambar 3.1	Kamera V100:R2	29
Gambar 3.2	Penanda aktif yang digunakan pada model	30
Gambar 3.3	Metode Penelitian	31

Gambar 3.4	Ilustrasi peletakan penanda aktif di wajah yang terdapat di <i>OptiTrack</i> , b peletakan penanda aktif pada Model	32
Gambar 3.5	Ilustrasi pemasangan kamera(Sumber : (Ana, 2010)) . . .	33
Gambar 3.6	Bone	33
Gambar 3.7	contoh model <i>Edge Planning</i> dan Topologi	35
Gambar 3.8	Bentuk Bidang Sebagai Acuan	37
Gambar 3.9	Penanda aktif yang digunakan untuk normalisasi dibagian titik penanda aktif pada area mulut	37
Gambar 3.10	Titik penanda aktif pada area kepala	38
Gambar 3.11	Flowchart K-means	39
Gambar 4.1	hasil bone	42
Gambar 4.2	hasil mapping.	42
Gambar 4.3	hasil mapping.	43
Gambar 4.4	Hasil C3D dan Viseme animasi saat mengucapkan awal dan akhir suku kata	44
Gambar 4.5	Hasil <i>cluster</i> K=3	46
Gambar 4.6	Hasil <i>cluster</i> K=4	46
Gambar 4.7	Hasil <i>cluster</i> K=5	47
Gambar 4.8	Hasil <i>cluster</i> K=6	47
Gambar 4.9	Hasil <i>cluster</i> K=7	48
Gambar 4.10	Hasil <i>cluster</i> K=8	48
Gambar 4.11	Hasil <i>cluster</i> K=9	49
Gambar 4.12	Hasil <i>cluster</i> K=10	49
Gambar 4.13	Hasil <i>cluster</i> K=11	50
Gambar 4.14	Hasil <i>cluster</i> K=12	50
Gambar 4.15	Hasil <i>cluster</i> K=13	51
Gambar 4.16	Hasil <i>cluster</i> K=14	51
Gambar 4.17	Hasil <i>cluster</i> K=15	52
Gambar 4.18	Hasil <i>cluster</i> K=16	52
Gambar 4.19	Hasil <i>cluster</i> K=17	53
Gambar 4.20	Hasil <i>cluster</i> K=18	53
Gambar 4.21	Hasil <i>cluster</i> K=19	54
Gambar 4.22	Hasil <i>cluster</i> K=20	54
Gambar 4.23	Hasil <i>cluster</i> K=21	55
Gambar 4.24	Hasil <i>cluster</i> K=22	55
Gambar 4.25	hasil dari <i>cluster viseme</i> berupa kelas- kelas <i>viseme</i> yang dimodelkan oleh model <i>viseme</i> (area mulut model).	58
Gambar 4.26	hasil dari <i>cluster viseme</i> berupa kelas- kelas <i>viseme</i> yang dimodelkan oleh model <i>viseme</i> (area mulut model).	62
Gambar 4.27	hasil dari <i>cluster viseme</i> berupa kelas- kelas <i>viseme</i> yang dimodelkan oleh model <i>viseme</i> (area mulut model).	63
Gambar 4.28	hasil dari <i>cluster viseme</i> berupa kelas- kelas <i>viseme</i> yang dimodelkan oleh model <i>viseme</i> (area mulut model).	64

- Gambar 4.29 hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model). 65
- Gambar 4.30 hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model). 66
- Gambar 4.31 hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model). 67
- Gambar 4.32 hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model). 68
- Gambar 4.33 hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model). 69

[Halaman ini sengaja dikosongkan.]

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi komputer yang sangat cepat, kebutuhan untuk peningkatan fleksibilitas dan kemampuan futuristik pada interaksi manusia komputer meningkat. Hal ini juga terjadi pada animasi karakter, *game*, analisa biomekanik dan yang lain. Di bidang animasi, tuntutan penyajian animasi yang realistis dan pantas serta menarik semakin tinggi. Animasi harus dapat menampilkan karakter yang sangat mirip dengan karakter di dunia nyata. Pada industri film animasi 3D yang diproduksi dari studio animasi di Indonesia dan dibuat versi bahasa Indonesia . Beberapa contoh adegan film animasi yang memperlihatkan animasi wajah dan animasi berbicara . Animasi-animasi tersebut diproduksi dengan masih memerlukan animasi manual yang besar, yaitu memerlukan waktu yang lama dan mahal, serta memerlukan skill animator yang sangat tinggi. Animasi wajah dan animasi berbicara pada beberapa film animasi tersebut, juga belum ditangani dengan baik sehingga animasi-animasi bibir khususnya belum terlihat alami. Produksi animasi yang lebih canggih dengan animasi wajah dan khususnya animasi berbicara yang kelihatan sangat alami masih merupakan suatu tantangan.

Wajah manusia merupakan salah satu bagian tubuh yang sangat spesifik dan istimewa. Sebuah wajah merupakan bagian penting dari tubuh agar seseorang dapat mengenal secara visual dan komunikasi, sehingga peran gerak wajah sangatlah di perlukan untuk visual dan komunikasi.

Gerakan wajah berperan penting bagi manusia dalam berkomunikasi. terutama gerakan bibir yang memudahkan berinteraksi antara manusia dengan manusia yang lain. Dengan demikian gerakan bibir adalah hal penting untuk Pemodelan viseme dinamis 3D .

Pemodelan *viseme* dinamis 3D prosesnya mencocokkan antara suara dan gerakan bibir merupakan gabungan teknologi dalam bidang ilmu interaksi manusia komputer dan grafika komputer. Pemodelan 3D merupakan rangkaian titik yang membentuk garis sehingga memunculkan sebuah objek. Pemodelan 3D biasanya digunakan dalam dunia maya seperti pada aplikasi kartun, acara TV, dan video game. Penelitian Hui Zhao et al (Hui Zhao dan Chaojing

Tang, 2008) melakukan penelitian dengan judul *Visual Speech Synthesis based on Chinese Dynamic Visemes* yang mana pada penelitian ini menggunakan pemodelan 2D dengan bahasa cina. Tentu penelitian ini menjadi tantangan untuk bahasa lainnya, terutama salah satunya adalah bahasa Indonesia. Hal ini didasarkan bahwa belum ada penelitian tentang pemodelan 3D untuk *viseme* yang dinamis berbahasa Indonesia sehingga penelitian yang dilakukan Hui Zhao (Hui Zhao dan Chaojing Tang, 2008) dapat dijadikan rujukan bagi penelitian tersebut.

Penelitian yang diajukan adalah pemrosesan kecocokan antara suara dan gerakan bibir dengan tepat dan benar, sehingga pemodelan 3D seolah-olah dapat berbicara sesuai dengan kata yang diucapkan. Manfaat dari penelitian ini adalah untuk tujuan pembelajaran dalam pengucapan Bahasa Indonesia karena pada Bahasa Indonesia ucapan fonem sama dengan yang ditulis sedangkan bahasa inggris berbeda dengan bahasa indonesia ucapan fonem dengan ucapan berbeda. Manfaat lain adalah dapat membantu orang bisu atau tuli untuk berkomunikasi satu sama lain melalui sintesis pembicaraan.

1.2 Perumusan Masalah

1. Fonem yang dibutuhkan untuk *viseme* Bahasa Indonesia adalah Bahasa Indonesia, sedangkan model yang akan digunakan adalah orang Indonesia untuk virtual wicara bahasa Indonesia di karenakan untuk viseme dinamis bahasa Indonesia.
2. *Viseme* dinamis bahasa Indonesia harus memiliki kualitas tampilan yang baik sehingga *virtual* yang di inginkan seolah-olah berbicara seperti manusia aslinya dan secara natural , sehingga diperlukan Pemodelan 3D agar tampilnya seperti manusia yang berbicara secara natural untuk bahasa Indonesia .
3. *Viseme* dinamis telah diteliti sebelumnya oleh Hui Zhao et dengan judul *Visual Speech Synthesis based on Chinese Dynamic Visemes* yang mana pada penelitian ini menggunakan pemodelan 2D dengan bahasa cina. Sehingga belum ada di Indonesia yang teliti tentang Pemodelan *viseme* dinamis 3D Bahasa Indonesia.
4. Pemodelannya hanya fokus pada bentuk mulut. .
5. Pengambilan nilai 3D bentuk mulut dari hasil data C3D

1.3 Tujuan Penelitian

Menghasilkan model 3D *viseme* dinamis untuk Bahasa Indonesia.

1.4 Manfaat

Dapat membantu orang bisu atau tuli untuk berkomunikasi satu sama lain melalui sintesis pembicaraan dan *virtual* untuk berita.

[Halaman ini sengaja dikosongkan.]

BAB II

KAJIAN PUSTAKA DAN DASAR TEORI

2.1 *State of the art*

Penelitian yang berkaitan dengan *viseme* dinamis belum banyak dilakukan. Berdasarkan paper yang menjadi referensi, penelitian yang berkaitan dengan *viseme* dinamis yang sudah dilakukan adalah Bahasa Inggris dan bahasa Cina. Jumlah *viseme* dinamis yang digunakan dalam penelitian ini, lebih banyak dibandingkan dengan jumlah *viseme* statis yang sudah dilakukan sebelumnya. Hal ini disebabkan dalam *viseme* dinamis, koartikulasi yang menyertai setiap artikulasi diperhitungkan. Pada penelitian yang dilakukan oleh Hui Zhao dan Chaojing Tang dengan judul paper “*Visual Speech Synthesis based on Chinese Dynamic Visemes*”, jumlah *viseme* dinamis yang digunakan adalah 40 *viseme* dinamis. Berdasarkan hasil pemetaan fonem-ke-*viseme* statis, dilakukan kombinasi dengan vokal dan konsonan masing-masing, sehingga diperoleh kombinasi fonem yang berbeda yang sesuai dengan *viseme-viseme* dinamis seperti yang terlihat pada Tabel 2.5.

Tabel 2.5. Kombinasi Fonem Yang Sesuai Dengan *Viseme* Dinamis Bahasa China

	b type	d type	j type	z type	zh type	a type	o type	e type	i type	u type
a type	ba (爸)	da (大)	jia (加)	za (咋)	zha (扎)	a (阿)	Ao (奥)	×	ai (爱)	×
o type	po (迫)	dou (斗)	jie (节)	zou (走)	zhou (周)	×	×	×	×	ou (欧)
e type	bei (北)	Le (乐)	jiong (窘)	ce (策)	zhe (折)	×	×	e (恶)	×	×
i type	bi (逼)	di (弟)	ji (鸡)	zi (字)	zhi (纸)	ya (牙)	you (由)	ye (也)	yi (一)	yu (鱼)
u type	bu (补)	du (杜)	ju (居)	zu (组)	zhu (住)	wa (挖)	wo (我)	wei (为)	×	wu (五)

Gambar 2.1: Sumber: (Hui Zhao dan Chaojing Tang, 2011)

Sementara penelitian *viseme* dinamis untuk Bahasa Inggris yang dilakukan oleh Sarah L. Taylor, Moshe Mahler, Barry-John Theobald dan Iain Matthews pada tahun 2012 (Iain Matthews, 2012) dengan judul “*Dynamic Units of Visual Speech*”, dijelaskan bahwa *viseme* dinamis yang digunakan dalam penelitian ini berjumlah sekitar 150 *viseme*. Untuk menentukan jumlah

viseme dinamis pengukuran *two goodness-of-fit* yang mencakup seluruh visualisasi wicara dengan menghitung masing-masing $k = 40, 45, 50, \dots, 600$. D_m adalah jarak rata-rata fitur super dengan median *cluster* masing-masing. D_n adalah jarak rata-rata fitur super dengan data terdekat yang tidak memiliki *cluster* yang sama. D_m akan menjadi besar untuk *cluster* yang bentuknya kurang baik karena data yang berbeda secara *visual* akan dimasukkan ke *viseme* yang sama. Sebaliknya, D_n akan menjadi kecil ketika ada banyak *viseme* yang mirip secara *visual* dimasukkan ke *viseme* yang berbeda.

Pada penelitian ini akan dibuat penelitian tentang pemodelan *viseme* dinamis 3D untuk Bahasa Indonesia yang sinkron dengan wicara sehingga menghasilkan animasi berbicara Bahasa Indonesia yang lebih alami (netral). *Viseme* yang di gunakan dalam penelitian ini adalah *viseme* dinamis, yaitu *viseme* fonem dapat ditampilkan dengan beberapa bentuk *viseme* yang berbeda sesuai dengan koartikulasi yang menyertainya .oleh karena itu animasi berbicara Bahasa Indonesia yang dihasilkan belum pernah diteliti sebelumnya. Penelitian sebelumnya yang pernah meneliti ini adalah Hui Zhao dan Chaojing Tang pada tahun 2008 (Hui Zhao dan Chaojing Tang, 2008) melakukan penelitian dengan judul *Visual Speech Synthesis based on Chinese Dynamic Visemes* yang mana pada penelitian ini menggunakan pemodelan 2D dengan bahasa cina .

2.2 Konsep Dasar Pemodelan 3D

Pemodelan adalah membentuk suatu benda-benda atau obyek. Pembuatan dan pendesainan suatu objek harus dilakukan sebaik mungkin sehingga objek tersebut terlihat seolah-olah hidup. Sesuai dengan obyek dan basisnya, proses ini secara keseluruhan dikerjakan secara komputerisasi. Melalui konsep dan proses desain, keseluruhan obyek bisa diperlihatkan secara 3D sehingga banyak yang menyebut hasil ini sebagai pemodelan 3D (3D modelling) (Nalwan, 1998).

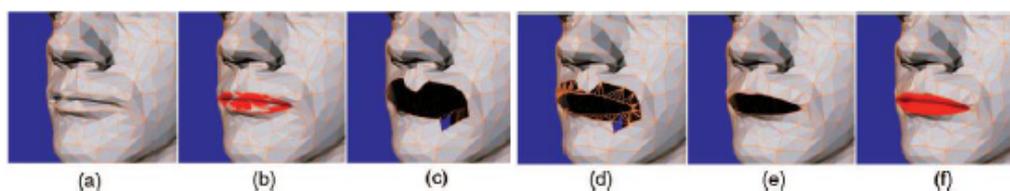
Ada beberapa aspek yang harus dipertimbangkan bila membangun model obyek, seluruh aspek harus memberi kontribusi pada kualitas hasil akhir. Hal-hal tersebut meliputi metode untuk mendapatkan atau membuat data yang mendeskripsikan obyek, tujuan dari model, tingkat kerumitan, perhitungan biaya, kesesuaian dan kenyamanan, serta kemudahan manipulasi model. Proses pemodelan 3D membutuhkan perancangan yang dibagi dengan beberapa tahapan untuk pembentukannya. Seperti obyek apa yang ingin dibentuk sebagai obyek dasar, metoda pemodelan obyek 3D, pencahayaan dan

animasi gerakan obyek sesuai dengan urutan proses yang akan dilakukan.

2.3 Dasar Metode Modeling 3D

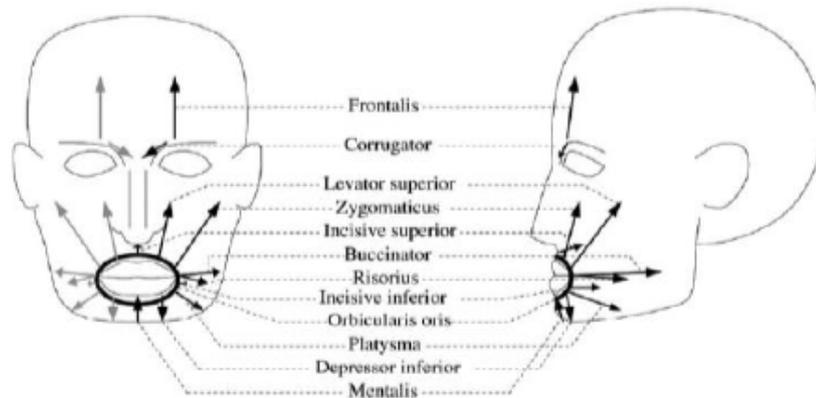
Ada beberapa metode yang digunakan untuk pemodelan 3D seperti metode pemodelan obyek yang disesuaikan dengan kebutuhannya seperti dengan nurbs dan polygon ataupun subdivision . Modeling polygon merupakan bentuk segitiga dan segiempat yang menentukan area dari permukaan sebuah karakter. Setiap polygon menentukan sebuah bidang datar dengan meletakkan sebuah jajaran polygon sehingga kita bisa menciptakan bentuk-bentuk permukaan. Untuk mendapatkan permukaan yang halus, dibutuhkan banyak bidang polygon. Bila hanya menggunakan sedikit polygon, maka object yang didapat akan terbagi menjadi sejumlah pecahan polygon (Nalwan, 1998).

Dalam tahap ini, kami melakukan *grafting* model bibir dengan input geometri untuk mendapatkan keuntungan baik input geometri penuh maupun mulut dengan resolusi yang tinggi. *Grafting* geometri bibir merupakan proses interaktif seperti yang terlihat pada Gambar 2.2. Ditambahkan juga *deformable* lidah, gigi, gusi, langit-langit atas, dan mata. Penambahan geometri mulut bagian dalam dan mata menghasilkan animasi yang lebih baik secara signifikan (Robert Anderson, 2012).



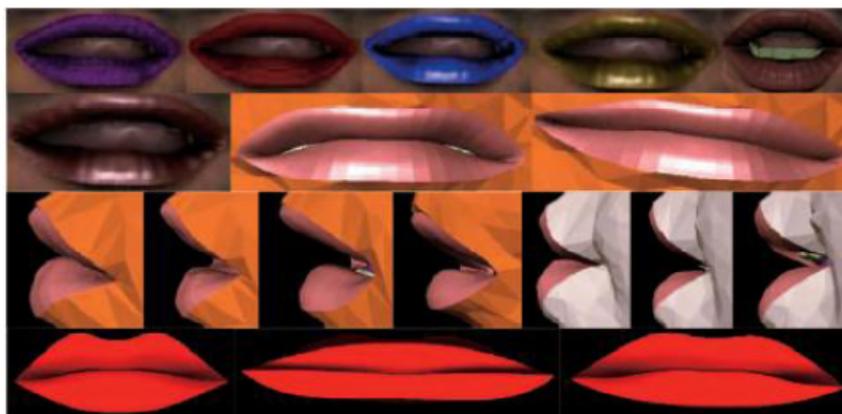
Gambar 2.2: Proses *Grafting*, input geometri (a), *triangles* yang tumpang tindih dihapus (c), Batas model bibir dan batas dari segitiga dihapus (d), *triangles* baru ditambahkan (e), akhirnya, model geometri bibir ditambahkan ke input geometri (f) (sumber: Robert Anderson, 2012).

Skin merupakan parameter model wajah yang mendefinisikan permukaan wajah yang berbasis *muscle*. Dalam hal ini digunakan parameter untuk masing-masing otot-otot wajah yang menempel pada kulit. Seperti yang terlihat pada Gambar 2.3 Didefinisikan karakteristik poin pada permukaan kulit yang bergerak sebagai artikulator (rahang dan bibir) bergerak. Poin-poin karakteristik juga dapat berubah bentuk karena kontraksi otot. Karakteristik poin menyediakan metode untuk pemetaan perubahan parameter untuk deformasi permukaan (R. Kent and F. Mini e, 1977).



Gambar 2.3: Otot-otot Kulit Wajah(Sumber : (R. Kent and F. Mini e, 1977))

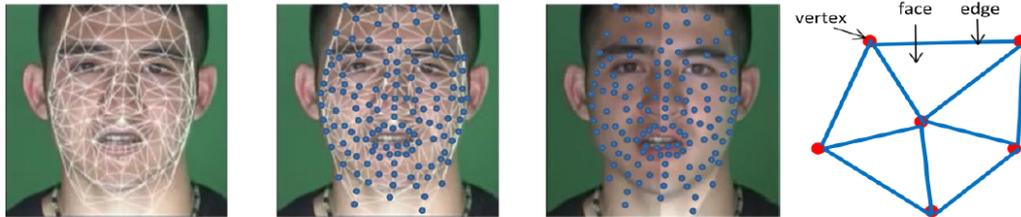
Dari bagian visualisasi wicara, bibir menyampaikan informasi yang paling penting. Hal ini sangat penting bagi untuk menghasilkan animasi berbicara yang kelihatan alami dan juga mampu menghasilkan bentuk yang realistis. Selain itu, bagian dalam dari bibir harus ada. Kami mengembangkan *deformable* model bibir dengan parameterisasi berbasis otot dan geometri bibir secara penuh. Model bibir memiliki pemetaan yang *embedded* antara parameter dan *deformasi* yang akan memberikan dampak besar pada ekspresi wajah dan visualisasi wicara(Thbaut wise, 2005).



Gambar 2.4: Gambar yang menunjukkan tampilan berbagai atribut layar dan gerakan dari model bibir Untuk menganimasikan visualisasi wicara perlu dipahami mengenai(Sumber :Thbaut wise, 2005)

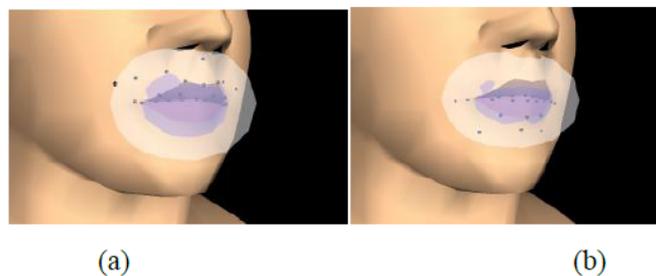
Untuk menganimasikan visualisasi wicara perlu dipahami mengenai *polygon mesh* yang merupakan sekumpulan *vertex* dan *edge* yang tersusun

sedemikian rupa sehingga membentuk suatu bentuk objek wajah 3D. *Vertex-vertex* dihubungkan oleh *edge-edge* sehingga membentuk *polygon-polygon convex* yang disebut wajah. Rangkaian wajah akan membentuk suatu bentuk objek 3D yang disebut *polygon mesh*. Gambar 2.6 menjelaskan mengenai *vertex*, *edge* dan wajah serta *polygon mesh* yang membentuk objek wajah 3D (M. E. Hennecke, K. V, dkk ,1988).



Gambar 2.5: *Vertex*, *edge* dan wajah serta *polygon mesh* yang membentuk objek 3D(Sumber :(M. E. Hennecke, K. V, dkk ,1988)

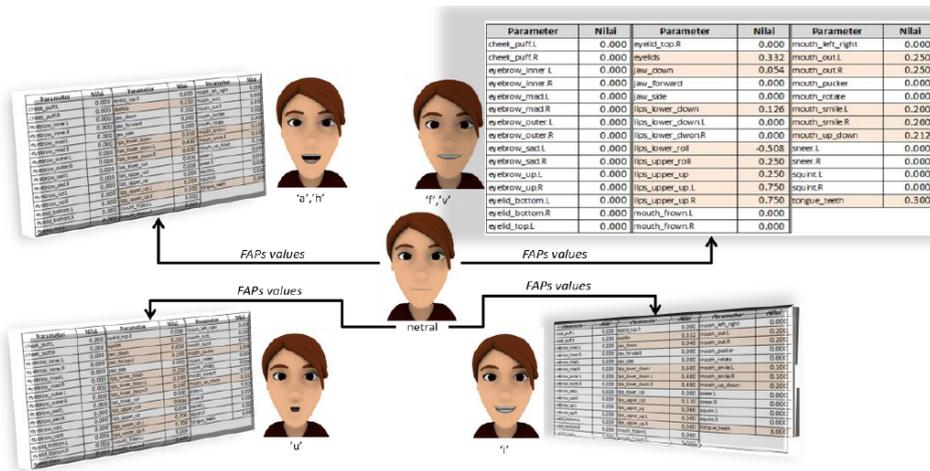
Bentuk objek wajah 3D dipengaruhi oleh nilai parameter-parameter animasi wajah (Facial Animation Parameter / FAP) yang merupakan parameter-parameter untuk menentukan nilai masing-masing *vertex*. Tidak semua nilai FAP digunakan untuk merepresentasikan bentuk mulut saat mengucapkan suatu fonem, hanya beberapa nilai-nilai parameter yang mempengaruhi *vertex-vertex* yang membentuk objek bentuk mulut 3D. Gambar 2.7 menjelaskan beberapa *vertex-vertex* yang mempengaruhi bentuk mulut saat mengucapkan suatu fonem (P. Delmas, 1999.).



Gambar 2.6: *Vertex-vertex* yang terkait dengan gerakan bibir atas (a) dan bibir bawah (b) (Sumber : (P. Delmas, 1999.))

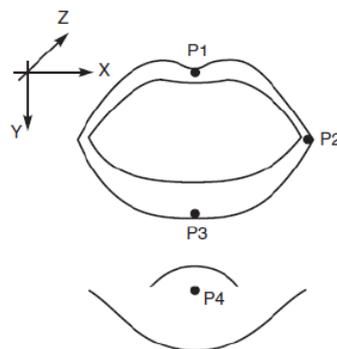
Ketika satu titik kontrol dipindahkan, maka *vertex-vertex* objek yang terpengaruh titik kontrol tersebut akan ikut berpindah. Setiap *viseme* didefin-

isikan oleh kombinasi FAP. Gambar 2.7 menjelaskan contoh-contoh model *viseme* Bahasa Indonesia yang dibentuk dengan memberikan nilai FAP yang berbeda-beda sehingga membentuk *viseme* tertentu. Pada kondisi *viseme netral* (silence) semua nilai FAP adalah 0 (I Mazonaviciute dkk, 2011).



Gambar 2.7: Contoh model *viseme* Bahasa Indonesia dengan nilai FAP tertentu(Sumber: Mazonaviciute dkk, 2011)

Karakteristik *viseme* merupakan *transisi* antara target-target artikulasi pengucapan fonem yang diukur dan dianalisis menggunakan titik-titik Fidusia (Fiduciary Points) seperti Gambar 2.8. Gerakan titik-titik Fidusia dibagi menjadi tiga komponen, yaitu : komponen *rigid* yang terkait dengan gerakan *mandibula* dan komponen *non-rigid* yang terkait dengan gerakan bibir atas dan bibir bawah. Selama berbicara, rotasi, maju dan mundur *mandibula*.(Jose Mario De Martino.dkk, 2006)

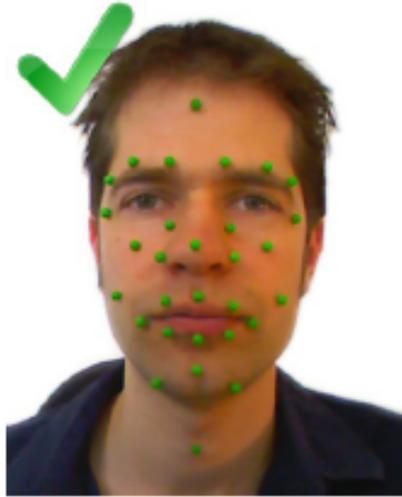


Gambar 2.8: Titik-Titik Fidusia.(Sumber : Jose Mario De Martino.dkk, 2006)

2.4 Marker(Penanda Aktif)

Anda dapat menggunakan warna penanda berikut dengan Zign Jalur Pro: merah, hijau, biru, kuning dan merah muda. Anda akan melihat pilihan di jendela konfigurasi untuk memilih 'putih' sebagai warna penanda , tapi ini untuk kasus saja. Anda hanya dapat menggunakan penanda putih jika mereka memiliki titik hitam di tengah, atau jika Anda menggunakan kamera inframerah khusus dengan spidol terang yang akan menghasilkan video hitam putih(Zign Creation, 2013).

Pilihan untuk menggunakan spidol dengan titik hitam ditambahkan ke Zign Pro untuk Melacak karena beberapa pengguna dari versi 1 telah meminta fitur ini. Meskipun demikian kita dianjurkan untuk menggunakan spidol yang berwarna normal karena dalam praktek ini telah terbukti lebih handal. Mengapa saya tidak dapat menggunakan spidol putih dalam keadaan normal Anda bertanya? Berikut ini alasannya: karena Zign Pro menggunakan algoritma deteksi penanda Zign Pro akan melacak secara otomatis mendeteksi penanda. Mendeteksi penanda berdasarkan warna dan ukuran yang dipilih. Jika Anda menggunakan spidol putih sehingga mata dan gigi yang terlihat akan memiliki masalah untuk pelacakan dari awal karena algoritma pelacakan dapat mendeteksi penanda di mata Anda dan mulut yang akan menyebabkan solusi yang salah dari wajah. Zign pro merekomendasikan menggunakan spidol hijau seperti gambar 2.9 . Warna hijau telah terbukti menjadi yang paling dapat diandalkan dalam deteksi Warna dikarenakan warna hijau biasanya tidak terdapat pada kulit. Jika memilih menggunakan spidol merah, kulit akan memiliki warna yang hampir sama dengan warna merah di kulit. Pelacak akan mendeteksi penanda warna di wajah . Jika memilih untuk menggunakan warna lain gunakan warna hijau sebagai penanda fungsi *preview* untuk menguji mana warna terbaik untuk yang akan tempelkan ke wajah (Zign Creation, 2013).



Gambar 2.9: Penanda aktif(Sumber:(Zign Creation, 2013))

2.5 *Facial Motion Capture*

Karena sulit untuk membuat animasi wajah secara *manual*, sejumlah metode yang diusulkan untuk menangkap gerak wajah secara otomatis. Ada dua jenis pendekatan untuk menangkap gerakan wajah: berbasis penanda-pendekatan dan pendekatan penanda-kurang.

Pendekatan berbasis penanda memiliki keuntungan bahwa itu dapat lebih kokoh mendeteksi pergerakan penanda daripada pendekatan penanda-kurang. (Huang et al). diusulkan metode untuk menangkap gerakan wajah dengan menggunakan seratus spidol . Metode ini dapat menangkap keriput yang realistis dan dinamis dan skala halus secara rinci di wajah. Namun, pendekatan berbasis penanda membutuhkan waktu untuk mengatur penanda dan sulit untuk menggunakan untuk menangkap adegan, karena objek harus berada dalam area wajah. (Bickel et al). diusulkan metode yang menggunakan spidol lukisan untuk menangkap *deformasi* rinci wajah seperti keriput . pendekatan ini berbasis penanda yang memiliki masalah umum untuk menempatkan penanda dan sulit untuk menangkap tekstur yang alami pada saat yang sama dengan gerak wajah(Zign Creation, 2013).

2.6 C3D

C3D adalah data hasil dari *motion capture* yaitu *sampling* dan gerak manusia, hewan,benda mati sebagai data 3D (Midori Kitagawa, Brian Windsor,181, 2008).

2.7 Pengertian Dasar Fonem

Sebelum ditemukan sejumlah fonem dalam Bahasa Indonesia. Terlebih dahulu akan dirumuskan mengenai pengertian tentang fonem. Fonem adalah bahasa yang terkecil dan dapat membedakan arti atau makna (Gleason, H.A.2011). Berdasarkan definisi diatas maka setiap bunyi bahasa, baik *segmental* maupun *suprasegmental* 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 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. Berdasarkan kenyataan, ternyata di dalam Bahasa Indonesia hanya ditemukan fonem *segmental* saja, dan bunyi *suprasegmental* tidak terbukti dapat membedakan arti. Oleh karena itu, dalam Bahasa Indonesia tidak ditemukannya fonem *suprasegmental*. Itulah sebabnya dalam kajian berikut ini hanya dibicarakan fonem *segmental* Bahasa Indonesia yang meliputi fonem ealias, fonem konsonan, dan fonem semi konsonan. Dalam mengenali fonem terdapat beberapa pokok pikiran umum yang disebut premis-premis fonologis. Berdasarkan sifat umumnya premis-premis bahasa tersebut adalah sebagai berikut:

1. Bunyi bahasa mempunyai kecenderungan untuk dipengaruhi oleh lingkungannya.
2. Sistem bunyi mempunyai kecenderungan bersifat simetris.
3. Bunyi-bunyi bahasa yang secara fonetis mirip harus digolongkan ke dalam kelas-kelas bunyi (fonem) yang berbeda, apabila terdapat pertentangan di dalam lingkungan yang sama.
4. 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/.

2.8 Macam-macam Fonem

Setelah menganalisis dari banyak literatur dan mengamati frekuensi penggunaan dalam kalimat bahasa Indonesia secara umum, jumlah fonem bahasa Indonesia yang digunakan pada penelitian ini adalah 48 fonem. Fonem yang digunakan meliputi vokal (V) dan konsonan (K), baik *monoftong* (huruf tunggal) maupun *diftong* (huruf rangkap / ganda). Fonem vokal yang digunakan dalam penelitian ini terdiri dari 10 vokal tunggal (a, i, I, u, U, o, O) dan 3 vokal ganda (ai, au, oi). Sedangkan 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 14 konsonan ganda (ch, dh, dl, dz, gh, kh, ks, ky, 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, ky, q, sh, sy, th, ts, v, w, x, ny, y, dan z fonem vokal I, dan U. Dari dua puluh satu akhiran fonem konsonan yang tidak (jarang) terpakai ini, seringkali cara membacanya diganti menjadi fonem lainnya seperti yang tertulis dalam penjelasan berikut ini : o Akhiran /c/, /j/ dan /th/ sering kali dibaca menjadi /t/. o Akhiran /ch/ sering kali dibaca menjadi /kh/. o Akhiran /dh/, /dl/ dan /dz/ sering kali dibaca menjadi /d/. o Akhiran /q/ sering kali dibaca menjadi /k/. o Akhiran /sh/, /sy/, /ts/ dan /z/ sering kali dibaca menjadi /s/. o Akhiran /v/ sering kali dibaca menjadi /f/. o Akhiran /U/ dan /w/ sering kali dibaca menjadi /u/. o Akhiran /x/ sering kali dibaca menjadi /ks/. o Akhiran /I/, /ky/ dan /y/ sering kali dibaca menjadi /i/

Pada tabel 2.1 berikut, 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 I, O dan U, juga tidak pernah digunakan di akhir kata dalam bahasa Indonesia, akan tetapi sering dipakai sebagai pengganti ucapan seperti pulau (dibaca pulO), sedangkan diftong oi sangat jarang dipakai dalam bahasa Indonesia.

Tabel 2.1: Pemilihan Fonem Bahasa Indonesia

No	V	K	Fonem	Diawal	Ditengah	Diakhir
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/	chairul	ahmad	-
7		d	/d/	dua	ada	abad
8		dh	/dh/	dhuafa	baghdhad	mudlaradh
9		dl	/dl/	dluhur	ramadlan	faridl
10		dz	/dz/	dzikir	muzdzin	mahfuzd
11				emas	internasional	kode
12	e		/e/	eja	faedah	-
13			//	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	I		/I/	-	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		ky	/ky/	kyai	tokyo	-
25		l	/l/	lain	kalau	halal
26		m	/m/	makan	kamu	salam
27		n	/n/	nama	pintar	kawan
28		ng	/ng/	ngilu	angkasa	siang
29		ny	/ny/	nyonya	senyap	-
30	o		/o/	obat	saolin	bakso

2.9 Pola Suku Kata

Suku kata (syllable) adalah unit pembentuk kata yang tersusun dari satu fonem atau urutan fonem. Suku kata sering dianggap sebagai unit pembangun fonologis kata, karena dapat mempengaruhi ritme dan artikulasi suatu kata. Sementara ritme atau irama adalah suatu ukuran gerakan yang simetris dan aksentuasi dari suatu suara yang teratur. Ilmu tentang ritme, penekanan, dan laras / nada dalam percakapan disebut prosodi (prosody), yang merupakan bagian dari ilmu linguistik. Suku kata ialah unit penyusun aturan bunyi percakapan,

Tabel 2.2: Pemilihan Fonem Bahasa Indonesia

No	V	K	Fonem	Diawal	Ditengah	Diakhir
31	O		/O/	orang	Borang	-
32		oi	/oy/	-	-	-
33		p	/p/	paman	sopan	gelap
34		q	/q/	qur'an	wuquf	baliq
35		r	/r/	ramah	hari	pudar
36		s	/s/	suka	kasih	kasus
37		sh	/sh/	shalat	nashab	-
38		sy	/sy/	syukur	termasyur	arsy
39		t	/t/	tuhan	hutan	kabut
40		th	/th/	therapi	-	-
41		ts	/ts/	tsunami	batsuni	hadits
42	u		/u/	udang	pusat	tabu
43	U		/U/	-	bertarung	-
44		v	/v/	vitam	lava	-
45		w	/w/	wanita	hawa	-
46		x	/x/	xerox	-	xerox
47		y	/y/	yang	payah	-
48		z	/z/	zaman	lezat	lafaz
Total	13	35	/48			

yang boleh mempengaruhi irama, prosodi, puitis, corak tekanan, dsb. dalam sebuah bahasa. Contohnya, perkataan sayang terdiri dari dua suku kata sa dan yang. Satu suku kata biasanya terdiri dari nukleus suku kata (selalu vokal) dengan pinggir awalan dan akhiran biasanya konsonan. Satu kata yang terdiri dari satu suku kata saja (seperti yang) disebut eka-suku. Kata yang terdiri dari dua suku kata (seperti mana) disebut dwi-suku. Kata yang terdiri dari tiga suku kata (seperti belalang) disebut tri-suku, dan kata yang memiliki lebih dari tiga suku kata (seperti kebaikan) disebut banyak suku. Suku kata dalam bahasa Indonesia memiliki beberapa pola. Pola suku kata tersebut terdiri dari kombinasi vokal dan konsonan. Pemenggalan atau pemisahan suku kata dari sebuah kata, harus didasarkan pada kata dasar dan pola suku katanya. Tabel 2.3 berikut ini adalah daftar pola suku kata dalam bahasa Indonesia.

Tabel 2.3: Pola Suku Kata

No	Pola Suku Kata	Didepan	Di belakang
1	V	i-bu	di-a
2	VK	er-na	ma-in
3	KV	sa-ya	ban-tu
4	KVK	pin-tu	ban-tu
5	KKV	dra-ma	in-dus-tri
6	KKVK	trak-tor	e-lek-trik
7	VKK	eks-port	eks
8	KKVK	sank-si	kon-teks
9	KKKV	groos	kom-leks
10	KKKV	stra-tegi	de-mon-stra-si
11	KKKVK	struk-tur	in-do-spring

2.10 *Viseme*

Viseme merupakan representasi visual dari fonetik Wicara (I. Mazonavi-ciute, 2011). *Viseme* yang disinkronisasi dengan fonem yang diucapkan dapat menghasilkan animasi berbicara yang lebih *ealistic*.

2.11 *Viseme* dinamis

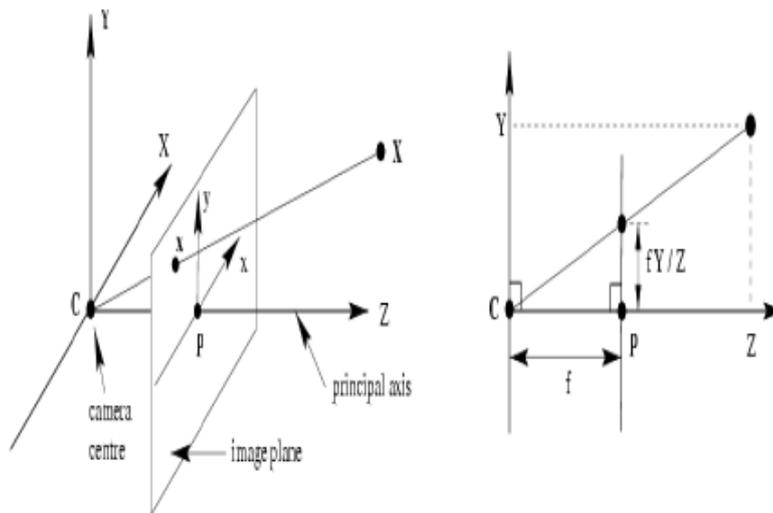
Viseme dinamis lebih merepresentasikan sinyal visualisasi wicara dalam setiap *viseme* yang melayani fungsi tertentu dan juga mengganti satu *viseme* dinamis untuk mengubah tampilan visual ucapan lainnya, yang merupakan analog visual yang benar untuk fonem. Sifat dinamis ini berarti bahwa efek koartikulasi eksplisit dalam model ini dan batas-batas antara *viseme* yang tidak terikat dengan bunyi yang mendasari (Ka-Ho WONG. dkk, 2012).

2.12 **Pemodelan *Viseme* Dinamis Bahasa Indonesia**

Pendekatan yang digunakan untuk menghasilkan model *viseme* dinamis adalah pendekatan data-driven terhadap data hasil ekstraksi *fitur* untuk *frame-frame* visual gerakan bibir dari video yang berisi orang berbicara dalam kalimat-kalimat berbahasa Indonesia. (Arifin, 2013)

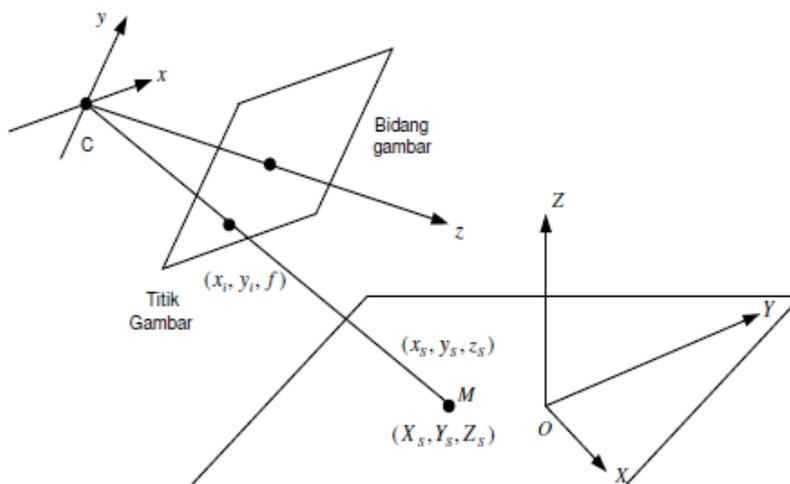
2.13 **Normalisasi 3D**

Model kamera adalah model kamera lubang jarum, model ini dipilih karena sudah umum digunakan penelitian-penelitian sebelumnya dan berhasil dengan baik (Eka prasyono, 2010). Apabila diasumsikan sebuah kamera dengan pusat titik kamera berhimpit dengan pusat sumbu koordinat dunia (Eka prasyono, 2010), maka dapat di gambarkan seperti nampak pada gambar



Gambar 2.10: Model kamera lubang jarum(Sumber:(Eka Prasetyo, 2010))

Apabila suatu kamera dengan pusat titik kamera tidak berhimpit dengan pusat sumbu koordinat dunia seperti digambarkan pada gambar 2.10. Untuk memperoleh informasi titik gambar dalam koordinat dunia maka diperlukan suatu matrik transformasi dari koordiant 2D gambar ke koordinat 3D dunia (Eka prasetyono, 2010).



Gambar 2.11: Model kamera lubang jarum(Sumber:(Eka Prasetyo, 2010))

Untuk dapat melakukan suatu rekonstruksi 3D diperlukan parameter-parameter intrinsik dan ekstrinsik dari kamera, nilai tersebut dapat diperoleh dari gambar 2.1

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \begin{bmatrix} a_x & 0 & X_0 & 0 \\ 0 & a_y & Y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_s \\ Y_s \\ 1 \end{bmatrix} \quad (2.1)$$

Keterangan :

a_x = fokus kamera

a_y = fokus kamera

X_0 = Distorsi radial

Y_0 = Distorsi radial

X_s = Koordinat image pada sumbu x

Y_s = Koordinat image pada sumbu y

Persamaan 2.2 merupakan persamaan umum untuk mendapatkan parameter ekstrinsik kamera sampai $R33$ merupakan nilai rotasi dari kamera sedangkan T_x, T_y dan T_z merupakan nilai dari kamera, X_s, Y_s, Z_s dan x_s, y_s, z_s dunia

$$\begin{bmatrix} X_3 \\ Y_3 \\ Z_3 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_3 \\ Y_3 \\ Z^3 \\ 1 \end{bmatrix} \quad (2.2)$$

Keterangan :

R = Matriks Rotasi ($R11-R33$)

T = Matriks Translasi (T_x, T_y, T_z)

X_3 = Koordinat dunia pada sumbu x

Y_3 = Koordinat dunia pada sumbu y

Z_3 = Koordinat dunia pada sumbu z

Proses triangulasi dari dua citra pada gambar5, titik pusat kamera $PC1$ diperoleh dengan mengalikan matrik parameter eksternal kamera 1 $M1$ dengan titik $(0,0,0)$ seperti pada persamaan 2.3.

$$PC1 = M1 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (2.3)$$

Keterangan :

PCI = titik Pusat kamera

$M1$ = parameter *Eksternal* kamera dengan titik (0,0,0)

2.14 Segmentasi

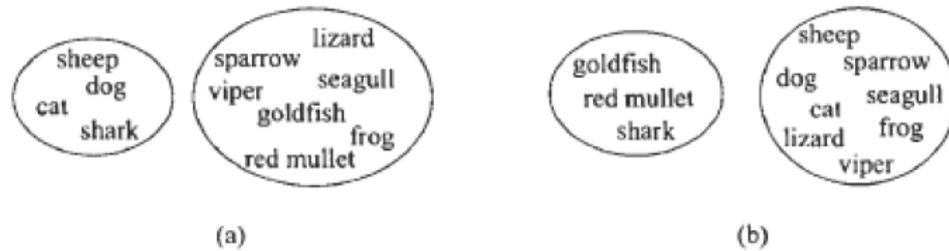
Motion segmentation bertujuan untuk memisahkan suatu gerakan ke dalam beberapa gerakan - gerakan dengan suatu kriteria tertentu. Jenis operasi ini berkaitan dengan pengenalan pola. Kebanyakan segmentasi gambar memilih daerah dengan kualitas gambar yang bagus dan setelah itu mencari daerah yang terbaik dan cocok untuk gambar. Segmentasi mengacu pada label piksel-piksel yang dihubungkan dengan daerah atau gerakan objek dengan perbedaan yang jelas. Segmentasi mempunyai hubungan erat dengan dua masalah lain antara *motion detection* dan *motion estimation*. *Motion detection* adalah panggilan khusus untuk *motion segment* dengan hanya dua segmen yang sesuai dengan gerakan kemudian dibandingkan dengan daerah gambar yang tidak bergerak (untuk kamera yang tidak bergerak) atau global dibandingkan dengan daerah gerakan lokal (untuk kamera yang bergerak) (Miftahur Rohman, 2012).

2.15 K-Means

Dalam *system* klasifikasi terdapat 2 jenis yaitu *supervised classification* dan *unsupervised classification*. Pada *unsupervised classification*, dimana pembelajaran pola tentang pembagian *class* tidak diberikan, sehingga lebih banyak fokus untuk memahami pola dalam *cluster* yang dapat dimengerti untuk menemukan persamaan dan perbedaan antar pola dan untuk memperoleh kesimpulan bermanfaat. Ide tersebut dapat dijumpai pada banyak bidang, seperti ilmu yang mempelajari hidup (biologi, zoology), ilmu pengetahuan medis (psychiatry, pathology), ilmu-ilmu sosial (sociology, archeology), ilmu pengetahuan bumi (geography, geology), dan rancangbangun (Kardi, 2007).

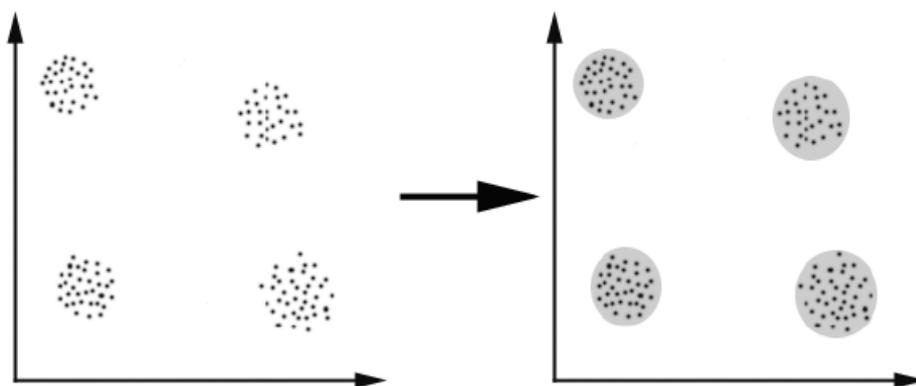
Misal binatang seperti domba, anjing, kucing termasuk *cluster* binatang menyusui; burung pipit, burung camar termasuk *cluster* burung; ular, kadal termasuk *cluster* binatang melata; ikan mas, ikan mullet merah, ikan hiu biru termasuk *cluster* ikan; dan kodok termasuk *cluster* binatang amfibi. Dalam mengelompokkan binatang-binatang tersebut ke dalam suatu *cluster* dibutuhkan penggambaran *clustering criterion*, hal ini sama halnya jika kita akan mengelompokkan cara binatang membawa keturunan mereka ke dalam sebuah *cluster*. Sebagai contoh domba, anjing, kucing, dan ikan hiu biru

dapat dikelompokkan dalam satu *cluster* sedangkan binatang yang lain dapat dibentuk ke dalam *cluster* yang lain (Kardi, 2007). Untuk jelasnya dapat dilihat pada gambar berikut.



Gambar 2.12: Beberapa *cluster* dari *clustering criterion*. Gambar (a) kelompok cara binatang membawa keturunannya, Gambar (b) kelompok paru-paru binatang(sumber: Kardi , 2007))

Clustering dapat dianggap yang paling penting dalam masalah *unsupervised learning*, karena setiap masalah semacam ini, ia berurusan dengan mencari struktur dalam kumpulan yang tidak diketahui datanya. Sehingga dapat didefinisikan bahwa *clustering* merupakan "proses mengatur objek menjadi anggota kelompok yang hampir sama dalam beberapa cara". Sebuah *cluster* merupakan kumpulan objek-objek yang "sama" di antara mereka dan "berbeda" pada objek dari *cluster* lainnya (Kardi, 2007).



Gambar 2.13: Identifikasi Kelompok(Sumber:(Kardi, 2007))

Dengan memperhatikan gambar di atas, dengan mudah mengidenti-

fikasikan 4 kelompok menjadi data yang dapat dibagi yaitu kesamaan dengan kriteria jarak antara dua atau lebih benda dalam *cluster* yang sama jika mereka dekat dan sesuai dengan jarak yang diberikan. Hal ini disebut *distance-based clustering*. Lain halnya untuk jenis pengelompokan konseptual *clustering*, dimana dua atau lebih benda dalam *cluster* yang sama dengan mendefinisikan konsep secara umum untuk semua benda, dengan kata lain objek dikelompokkan menurut konsep deskriptif (Kardi, 2007).

Tujuan dari *clustering* adalah untuk mengklasifikasikan data, dengan cara menentukan pengelompokan dalam satu set data yang tidak diketahui. Tetapi bagaimana untuk menentukan *clustering* yang baik? Dapat menunjukkan tidak ada *criteria absolut* "terbaik" yang akan bergantung pada tujuan akhir dari *clustering*. Akibatnya, pengguna yang harus menyertakan kriteria ini, sehingga hasil *clustering* akan memenuhi kebutuhan mereka. Syarat yang harus dipenuhi dalam *clustering* algoritma adalah skalabilitas; berhadapan dengan berbagai jenis atribut; menemukan bentuk kelompok persyaratan minimal adalah domain pengetahuan untuk menentukan parameter masukan; kemampuan untuk menangani gangguan; dimensi tinggi; serta *interpretability* dan *usability* (Kardi, 2007). *Clustering* algoritma dapat diklasifikasikan sebagai berikut:

1. *Exclusive Clustering* merupakan Data dikelompokkan ke dalam suatu cara yang eksklusif, sehingga jika suatu fakta milik suatu *cluster* maka tidak dapat dipakai (menjadi anggota) di *cluster* lain.
2. Menggunakan fuzzy set untuk *cluster* data sehingga titik kemungkinan memiliki dua atau lebih kelompok yang berbeda sesuai derajat keanggotaannya. Dalam hal ini data akan dihubungkan dengan nilai keanggotaannya.
3. *Hierarchical Clustering* merupakan Didasarkan pada kesatuan antara dua kelompok terdekat. Permulaan kondisi diwujudkan dengan menetapkan setiap datum sebagai *cluster*. Setelah beberapa *iterasi* mencapai final kelompok yang diinginkan.
4. *Probabilistic Clustering* merupakan Sepenuhnya menggunakan pendekatan *probabilistic* Terdapat empat algoritma yang paling sering digunakan dalam *clustering*, yaitu:

- (a) K-means (*exclusive clustering*)
- (b) *Fuzzy C-means* (*overlapping clustering*)
- (c) *Hierarchical clustering*
- (d) *Mixture of Gaussians* (*probabilistic clustering*)

K-mean adalah sebuah algoritma yang digunakan untuk mengklasifikasikan atau kelompok objek berdasarkan atribut / fitur ke K (cluster) jumlah kelompok. K adalah jumlah bilangan bulat positif. Pengelompokan ini dilakukan dengan meminimalkan jumlah kuadrat dari jarak antara data dan centroid klaster yang sesuai. Dengan demikian, tujuan dari K-rata pengelompokan adalah untuk mengklasifikasikan data (Kardi, 2007).

contoh kasus K-Mean Misalnya kita memiliki 4 objek sebagai titik data pelatihan dan setiap obyek memiliki 2 atribut. Tiap atribut mewakili koordinat dari objek, yaitu:

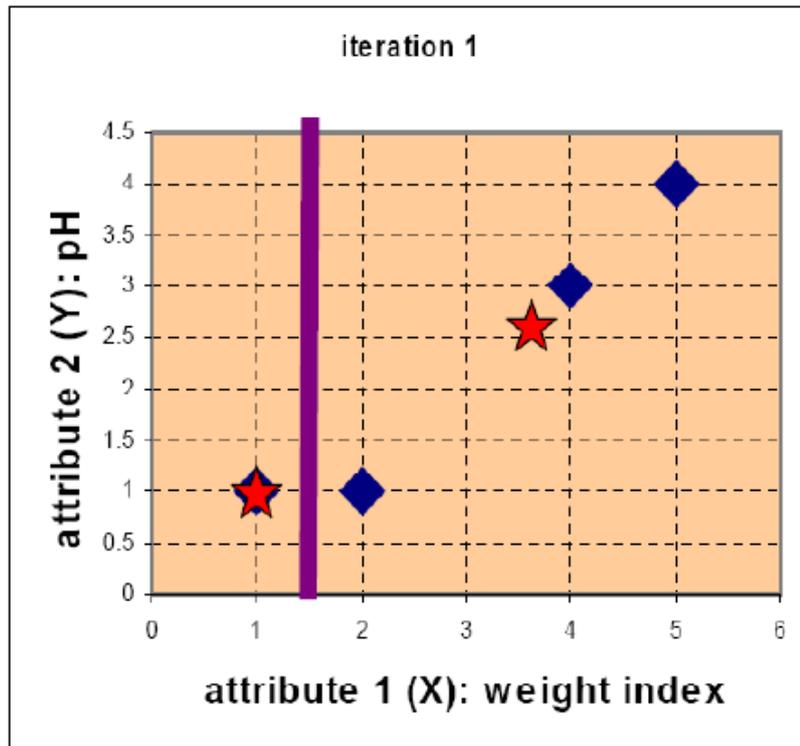
Objek Atribut 1 (X): bobot indeks

Objek Atribut 2 (Y): pH

	Medicine A	Medicine B	Medicine C	Medicine D
X	1	2	4	5
Y	1	1	3	4

Gambar 2.14: Contoh Kasus(sumber: Kardi , 2007))

Hasil *cluter* seperti gambar 2.6



Gambar 2.15: *iterasi*(sumber: Kardi , 2007))

K-Means merupakan algoritma untuk *cluster* n objek berdasarkan atribut menjadi k partisi, dimana k lebih kecil dari n dengan menggunakan algoritma k-means sebagai berikut (Arifin, 2013).

Langkah-langkah dalam algoritma K-means Klasterisasi adalah :

1. Menentukan jumlah *cluster*
2. Menentukan nilai *centroid*, dalam menentukan nilai *centroid* untuk awal *iterasi*, nilai awal *centroid* dilakukan secara acak. Sedangkan jika menentukan nilai *centroid* yang merupakan tahap dari *iterasi*, maka digunakan rumus sebagai berikut :

$$\bar{V}_{i,j} = \frac{1}{N_i} \sum_{K=0}^{N_i} X_{kj} \quad (2.4)$$

Keterangan :

V_{ij} = *centroid*/ rata-rata *cluster* ke-I untuk variable ke-j

N_i = jumlah data yang menjadi anggota *cluster* ke-i

$i, k = \text{indeks dari cluster}$

$j = \text{indeks dari variabel}$

$X_{kj} = \text{nilai data ke-}k \text{ yang ada di dalam cluster tersebut untuk variable ke-}j$

3. Menghitung jarak antara titik *centroid* dengan titik tiap objek Untuk menghitung jarak tersebut dapat menggunakan *Euclidean Distance*, yaitu:

rumus

$$d_{x,y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_n - y_n)^2} \quad (2.5)$$

Keterangan:

$$d_{x,y} \quad (2.6)$$

$d_{xy} = \text{Euclidean Distance}$

$$x = (x_1, x_2, \dots, x_n) \quad (2.7)$$

$x = \text{data point}$

$$y = (y_1, y_2, \dots, y_n) \quad (2.8)$$

$y = \text{point centroid}$

4. Pengelompokan objek untuk menentukan anggota *cluster* adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada *cluster* matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke klaster dan nilai 0 untuk data yang dialokasikan ke *cluster* yang lain.
5. Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke klaster lain (Arifin, 2013).

2.16 Measuring Cluster Quality

Salah satu metode untuk menentukan *clustering* terdefinisi dengan baik adalah dengan menggunakan kriteria fungsi yang mengukur kualitas *clustering*. Ada metode yang digunakan secara luas, yaitu *Sum of*

Squared Error (SSE), yang dihitung dengan menggunakan Persamaan. (2.1). Semakin kecil nilai *SSE* adalah, kualitas pengelompokan yang lebih baik akan.

$$SSE = \sum_{i=1}^k \sum_{p \in i} d(p, m_i)^2 \quad (2.9)$$

Keterangan:

SSE = *Sum of Squared Error*

k = Jumlah kluster

p = Data *point* untuk setiap *cluster* .

m_i = *node* terdekat .

d(p, m_i) = *Distance* untuk setiap data *point* *p* ke *m_i* untuk *cluster* *i* atau *iterasi* *i* .

Kualitas *cluster* juga dapat dievaluasi dengan menggunakan *between-class variation* (BCV) dan *within-class variation* (WCV). BCV adalah rata-rata jarak antar*centroid* dan WCV adalah *Sum of Square Error* (arifin, 2013). Nilai rasio yang lebih besar menunjukkan kualitas pengelompokan yang lebih baik. Rasio BCV dan WCV diformulasikan dengan menggunakan Persamaan. (arifin, 2013).

$$Q = \frac{BCV}{WCV} = \frac{\frac{1}{n_k} \sum_{1=i}^k d(m_i m_i)}{SSE} \quad (2.10)$$

$$BCV = \frac{1}{n_k} \sum_{1=i}^k d(m_i m_i) \quad (2.11)$$

Keterangan:

Q = Hasil kualitas *Clustering*

BCV = rata-rata jarak antara *centroid*

WCV = *Sum Square Error*

SSE = *Sum of Squared Kesalahan*

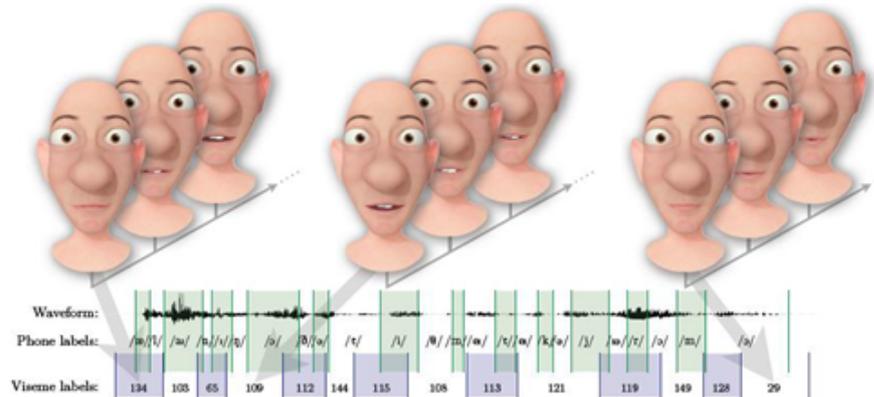
k = jumlah kluster

m_i = *node* terdekat

2.17 Menganimasikan Visualisasi Wicara

Model 3D *viseme* yang dinamis dapat dianimasikan dengan menggunakan perangkat lunak pemodelan 3D seperti Blender. Visualisasi wicara

dapat dianimasikan ketika urutan *viseme* dinamis diketahui, yang dilakukan kemudian hanya merangkai *viseme* dinamis yang diperlukan dalam urutan. Kemudian rangkaian animasi visualisasi wicara tersebut dibaurkan sehingga didapatkan animasi yang lebih *smooth*. Gambar 2.3 memperlihatkan contoh *frame-frame* dari urutan yang dianimasikan. Setiap unit *viseme* dinamis dianimasikan oleh rata-rata 4 *frame* (Sarah L. Taylor , 2012).



Gambar 2.16: Contoh Sinkronisasi Sinyal Wicara, Model Viseme Dinamis(sumber: Sarah L. Taylor , 2012))

[Halaman ini sengaja dikosongkan.]

BAB III

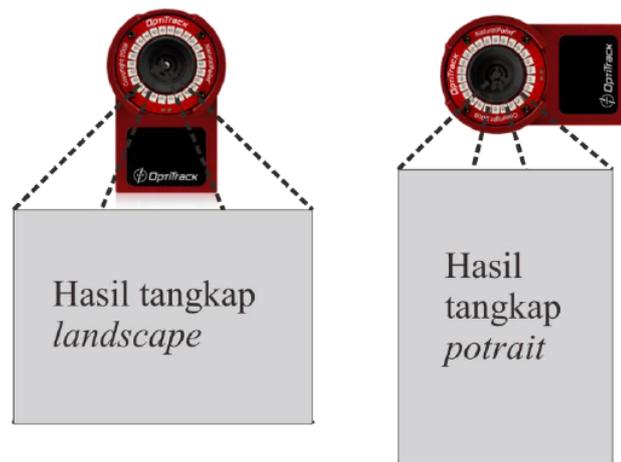
METODE PENELITIAN

3.1 Material dan bahan

Alat dan bahan yang akan digunakan pada penelitian ini:

- *Opti Track*

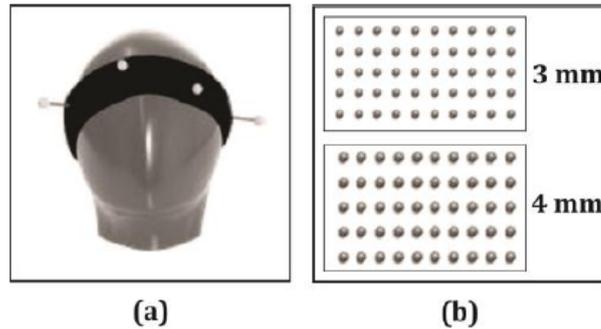
Kamera yang digunakan pada pengambilan data yang digunakan adalah kamera bertipe VR100:R2, kamera *motion capture Opti Track* yang memiliki resolusi sebesar 480 x 640 dan memiliki kecepatan tangkap sebesar 100 *frame per second (fps)* , seperti gambar 3.1



Gambar 3.1: Kamera V100:R2

- Penanda aktif (*marker*)

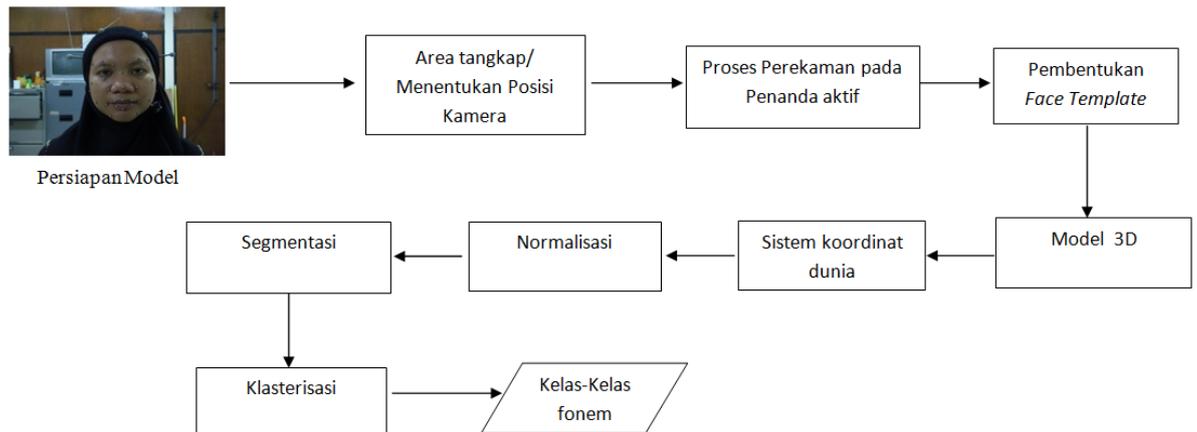
Penanda aktif yang digunakan telah disediakan secara khusus oleh *OptiTrackTM*. Pada *facial motion capture*, penanda yang dibutuhkan adalah *reflective* penanda aktif *headband* set gambar dan *reflective* Penanda aktif set gambar . Untuk *reflective* penanda aktif set dapat digunakan *reflective* penanda aktif dengan ukuran 4 mm atau 3 mm.



Gambar 3.2: Penanda aktif yang digunakan pada model

3.2 Metode Penelitian

Metode penelitian ini secara garis besar digambarkan pada gambar 3.4. Ada beberapa proses yang akan dilakukan dalam penelitian ini. Pertama melakukan proses persiapan model dengan menggunakan 37 penanda aktif yang diletakan pada wajah model. Kedua mengatur tata letak kamera yang akan digunakan untuk mendeteksi penanda aktif pada wajah model. Lalu, dilakukan proses perekaman penanda aktif pada wajah model yang akan menghasilkan *bone*. Hasil *bone* diproses untuk membentuk *face template* yang digunakan untuk menangkap gerakan pada wajah model. Selanjutnya proses mencari data koordinat dunia yang terdapat pada data *motion capture*. Data dari koordinat dunia yang didapat akan digunakan untuk proses normalisasi agar data yang besar menjadi yang lebih kecil. Data normalisasi digunakan untuk proses segmentasi data. Setelah didapatkan hasil segmentasi, data hasil segmentasi akan digunakan untuk proses klasterisasi sehingga didapatkan hasil kelas - kelas fonem berupa suku kata.



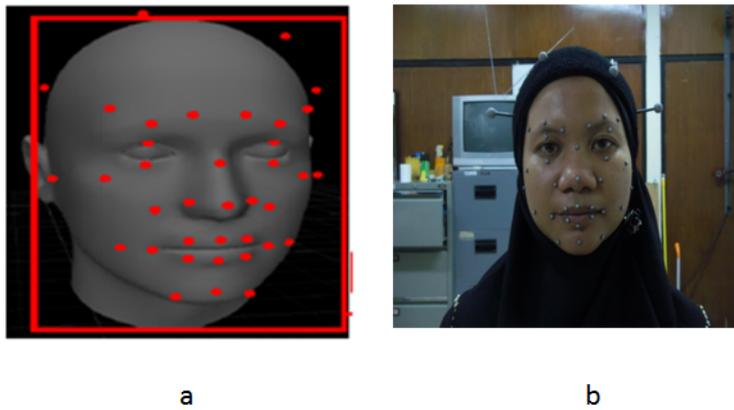
Gambar 3.3: Metode Penelitian

3.3 Persiapan Model

Model untuk pengambilan data adalah model wajah manusia. Wajah model akan diletakan penanda aktif sebanyak 37 yang terdiri dari 33 + 4. 33 Penanda aktif akan diletakan di area wajah model sedangkan 4 penanda aktif diletakan diatas kepala. Ilustrasi tersebut dapat dilihat pada gambar 3.3 b. Sebagai contoh peletakan penanda aktif pada model wajah manusia dapat dilihat pada model penanda aktif yang disediakan khusus oleh *OptiTrack TM* seperti pada gambar 3.3 a. Pada *OptiTrack TM* terdapat konfigurasi untuk peletakan penanda aktif di wajah model. Konfigurasi *OptiTrack TM* untuk penanda aktif di wajah terdiri dari 3 peletakan penanda aktif yaitu:

1. 23 di wajah + 4 di atas kepala
2. 33 di wajah + 4 di atas kepala
3. 37 di wajah + 4 di atas kepala

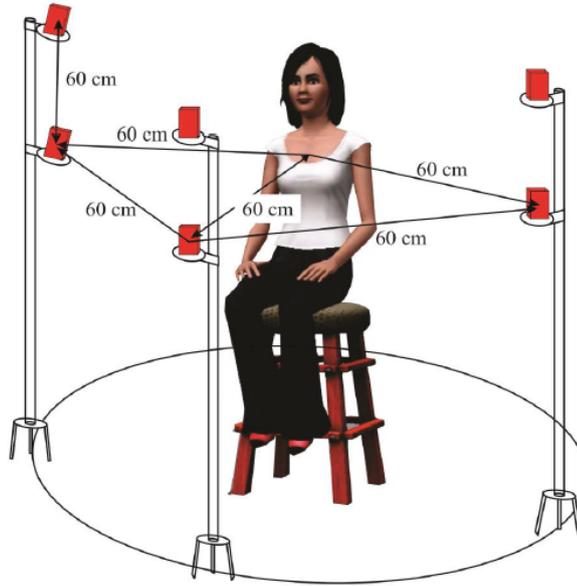
Pada penelitian ini untuk penggunaan penanda aktif di wajah model menggunakan 33 + 4 penanda aktif dikarenakan hasil yang diperoleh untuk membentuk *template facial motion* sangat baik dan hasil gerakan bibir sesuai dengan model. Sedangkan untuk 23 + 4 hasil yang diperoleh untuk membentuk *template facial motion* dan gerakan bibir kurang baik, sehingga sulit untuk dilakukan proses segmentasi.



Gambar 3.4: Ilustrasi peletakan penanda aktif di wajah yang terdapat di *OptiTrack*, b peletakan penanda aktif pada Model

3.4 Area Tangkap/Posisi Kamera

Kamera *motion capture OptiTrackTM* yang berjumlah enam buah disusun menyerupai busur lingkaran dengan kisaran sudut 120 derajat. Masing-masing kamera dipasang secara *orientasi* dan *landscape*. Tiga kamera disusun diatas kepala dan tiga kamera disusun setinggi dada. Jarak kamera dengan model sepanjang 60 cm.



Gambar 3.5: Ilustrasi pemasangan kamera(Sumber : (Ana, 2010))

3.5 Proses Perekaman pada Penanda Aktif

Proses perekaman penanda aktif bertujuan untuk membentuk letak penanda aktif yang terdapat pada wajah model. Proses ini dilakukan untuk menghasilkan tampilan *facial capture*. Sebelum proses perekaman dimulai. Terlebih dahulu menentukan waktu sebelum perekaman. kemudian menentukan lama waktu yang digunakan untuk perekaman dan menyalakan kamera, sehingga kamera akan mulai merekam. Hasil perekaman penanda aktif seperti pada Gambar 3.6.



Gambar 3.6: Bone

3.6 Pembentukan *Face Template*

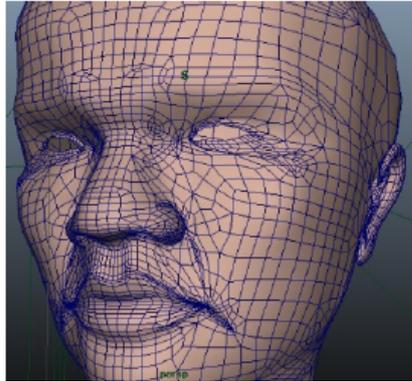
Sebelum proses pembentukan *face Template* dilakukan terlebih dahulu model duduk di depan sistem kamera seperti ilustrasi pada Gambar 3.5.

Sehingga gerak wajah model dapat dilihat oleh beberapa kamera. Setiap *frame* gerakan pada wajah model akan dilacak oleh *Optik Track*. Sehingga kamera mulai berkerja secara *real time* melacak penanda aktif yang terdapat pada wajah model. *Software Optik Track* akan secara *real time* membentuk *face templete*

3.7 Model 3D

Model 3D di sini digunakan untuk memperlihatkan gerakan mulut yang di modelkan oleh model di dunia nyata yaitu manusia, proses Model 3D di sini yaitu persiapan Model wajah animasi. yaitu prosesnya:

1. siapkan *image* model
2. *Polygon* Modelan dengan jenis yang sering digunakan untuk pemodelan.
3. *Smoothing* merupakan proses yang di gunakan untuk menghaluskan permukaan yang spesifikasi bentuk kasar dari permukaan dasar. sehingga proses ini dapat digunakan untuk mesh dan dapat menciptakan *vertex-vertex* baru dan bentuk baru. proses ini dilakukan lebih dari sekali, posisi lama juga akan berubah.
4. normal proses disini adalah proses teknik yang digunakan model agar dapat memalsukan cahaya sehingga menghasilkan lubang *polygon* lebih banyak. Normal map biasanya merupakan gambar RGB yang menggunakan koordinat dunia X, Y, Z dari permukaan normal.
5. Pemetaan UV merupakan proses pembentukan tekstur pada objek 3D dengan mempresentasikan 2D dari model 3D .Huruf U dan V digunakan untuk menggambarkan *mesh* 2D karena X, Y dan Z yang di gunakan untuk menggambar objek 3D dalam ruang model.
6. *Edge Planning* dan Topologi proses ini merupakan bagian penting untuk membangun bentuk kepala serta perhitungan bagaimana bagian kepala dapat di animasikan yang serupa dengan model yang diperankan manusia. sehingga di dapatkan model 3D



Gambar 3.7: contoh model *Edge Planning* dan Topologi

3.8 Proses Sistem Koordinat Dunia

Proses sistem koordinat dunia merupakan proses yang dihasilkan dari *facial motion capture* yang bersifat relatif terhadap gerakan kepala. Untuk mendapatkan data dari gerakan mulut pada saat model mengucapkan beberapa kalimat yang akan menghasilkan suku kata

3.9 Proses Normalisasi Posisi 3D

Sistem koordinat yang dihasilkan dari *facial motion capture* adalah sistem koordinat dunia yang bersifat relatif terhadap gerakan kepala. Data-data koordinat tiap *frame* akan mudah berubah seiring dengan gerakan kepala. Oleh karena itu, diperlukan transformasi dari sistem koordinat dunia ke sistem koordinat lokal.

Pada proses transformasi ini diperlukan sebuah bidang yang digunakan sebagai acuan terhadap data-data koordinat dari penanda aktif yang lain. Bidang ini disusun dari titik-titik penanda aktif yang mempunyai sifat relatif tetap terhadap gerakan kepala.

Kami memilih tiga titik penanda aktif, yaitu titik penanda aktif yang akan digunakan terdiri dari *head1*, *head2* dan *head4* (lihat gambar 3.9) yang masing-masing disebut sebagai p_1 , p_2 dan p_3 sehingga sebuah bidang seperti terlihat pada gambar 3.7. Sumbu Z tegak lurus terhadap bidang $P_1P_2P_3$, maka :

$$V_Z = \frac{\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}}{p_1p_2 \times p_1p_3} = (Z_1, Z_2, Z_3) \quad (3.1)$$

$$V_X = \frac{\overrightarrow{P_1P_2}}{p_1p_2} = (X_1, X_2, X_3) \quad (3.2)$$

$$V_Y = V_Z \times V_X = (Y_1, Y_2, Y_3) \quad (3.3)$$

Sehingga terbentuk matriks M :

$$M = \begin{bmatrix} X_1 & Y_1 & Z_1 & P_{1_1} \\ X_2 & Y_2 & Z_2 & P_{1_2} \\ X_1 & Y_3 & Z_3 & P_{1_3} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.4)$$

Keterangan :

$p1$ = Bibir penanda aktif kiri ketiga bibir atas (LLipUpperBend)

$p2$ = Bibir penanda aktif kanan kelima bibir atas (RLipUpperBend)

$p3$ = Bibir penanda aktif kanan keenam bibir atas (LMouthCorner)

$P1$ = nilai kepala atas bagian kanan yang menggunakan penanda aktif kepala ($head_1$)

$P2$ = nilai kepala atas tengah kedua yang menggunakan penanda aktif kepala ($head_2$)

$P3$ = nilai kepala atas tengah pertama yang menggunakan penanda aktif kepala ($head_4$)

Vz = Hasil dari gerakan yang tetap ($P1, P2, P3$) dan tegak lurus ($p1, p2, p3$) yang berupa matrik ($Z1, Z2, Z3$)

Vx = Hasil dari gerakan yang tetap ($P1, P2$) dan tegak lurus ($p1, p2$) yang berupa matrik ($X1, X2, X3$)

Vz = Hasil dari perkalian hasil (Vz) dengan (Vx) yang berupa matrik ($Y1, Y2, Y3$)

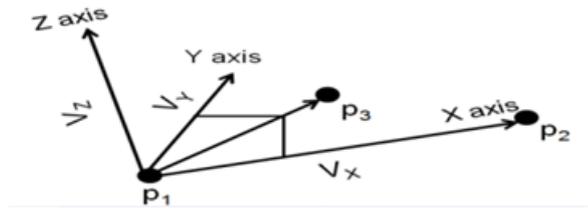
M = Hasil matrik (Vz, Vx, Vy) dan nilai dari ($P1$) serta nilai matrik koordinat dunia $[0, 0, 1]$

M_i = Hasil normalisasi (M) yang di invers

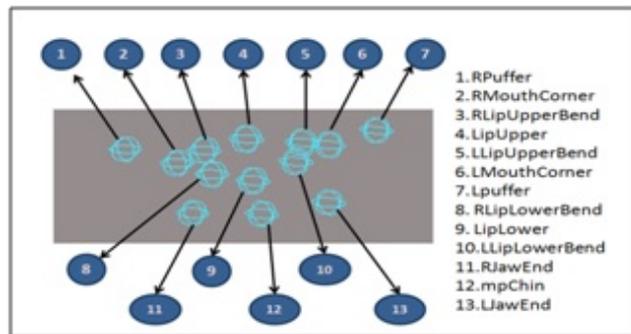
$$M_i = Inv(M) \quad (3.5)$$

Selanjutnya, sistem koordinat dari seluruh titik penanda aktif di area mulut (lihat gambar 3.8) dikalikan dengan dengan matriks M_i . Sistem koordinat yang dihasilkan ini yang digunakan pada segmentasi.

Hasilnya hanya diambil pada bagian area mulut saja untuk digunakan pada proses segmentasi



Gambar 3.8: Bentuk Bidang Sebagai Acuan



Gambar 3.9: Penanda aktif yang digunakan untuk normalisasi dibagian titik penanda aktif pada area mulut

3.10 Proses Segmentasi

Proses segmentasi *motion capture* merupakan proses yang berdasarkan pencarian pada suku kata. Kami mencatat jumlah *frame* pada awal dan akhir setiap pengucapan kalimat untuk mendapatkan suku kata. Data koordinat X, Y, Z masing-masing penanda aktif dari kumpulan *frame* dihitung nilai rata-ratanya. Nilai rata-rata ini yang selanjutnya digunakan sebagai data *fitur* untuk masing-masing penanda aktif. Nilai rata-rata tersebut menjadi hasil dari segmentasi *motion capture*, digunakan rumus sebagai berikut :

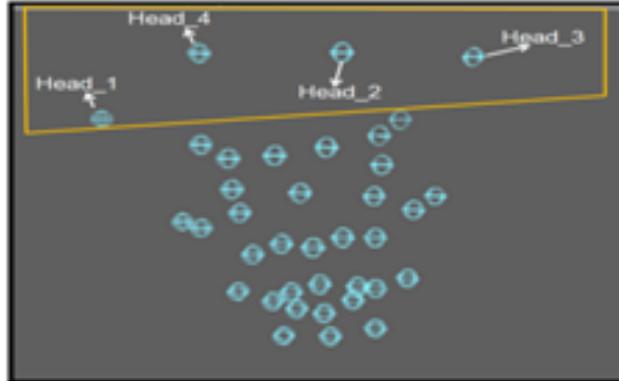
$$S = \frac{L}{N} \quad (3.6)$$

Keterangan:

S =Nilai rata-rata

L =Jumlah seluruh frame awal per suku kata dan akhir persuku kata

N = Jumlah banyak frame per suku kata



Gambar 3.10: Titik penanda aktif pada area kepala

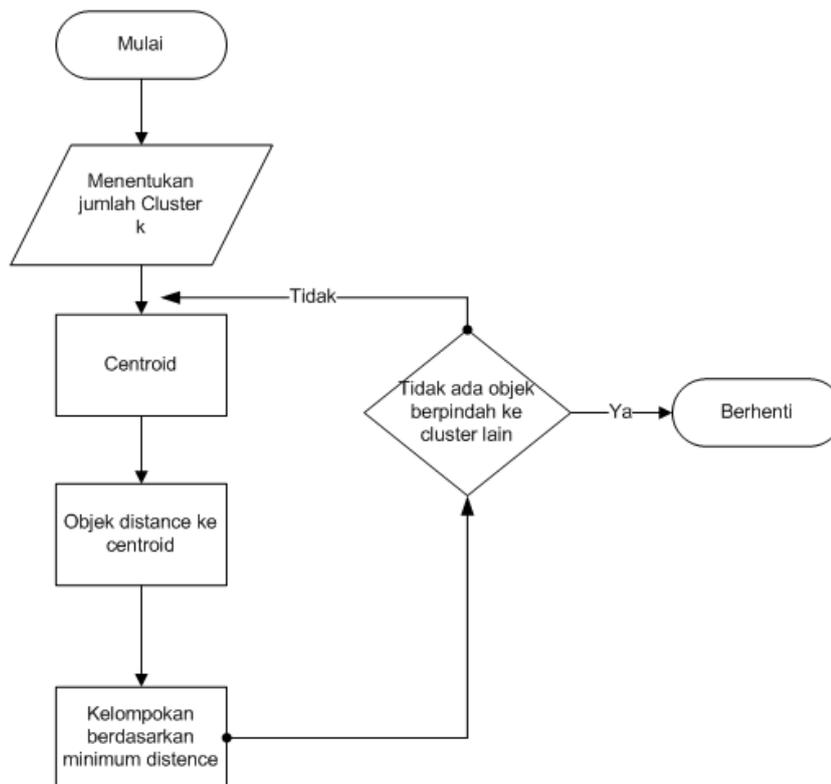
3.11 Klasterisasi

Proses klasterisasi di penelitian ini menggunakan metode K-means. K-Means merupakan algoritma untuk *cluster* n objek berdasarkan atribut menjadi k partisi, dimana k lebih kecil dari n dengan menggunakan algoritma k-means sebagai berikut.

Langkah-langkah dalam algoritma K-means Klasterisasi adalah :

1. Menentukan jumlah *cluster*
2. Menentukan nilai *centroid*, dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak. Sedangkan jika menentukan nilai *centroid* yang merupakan tahap dari iterasi, maka digunakan rumus 2.4
3. Menghitung jarak antara titik *centroid* dengan titik tiap objek untuk menghitung jarak tersebut dapat menggunakan *Euclidean Distance* dengan menggunakan rumus 2.5
 y adalah *point centroid*
4. Pengelompokan objek untuk menentukan anggota *cluster* adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada *distance* matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke *cluster* dan nilai 0 untuk data yang dialokasikan ke *cluster* yang lain.
5. Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.

Berikut penggambaran algoritma K-mean klasterisasi menggunakan flowchart:



Gambar 3.11: Flowchart K-means

[Halaman ini sengaja dikosongkan.]

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menjelaskan proses dan hasil percobaan Normalisasi dan Segmentasi serta klasterisasi untuk menghitung jumlah kelas- kelas fonem untuk suku kata dengan terlebih dahulu menggunakan proses segmentasi yang selanjutnya akan menggunakan Algoritma K-means untuk cluster fitur suku kata yang diperoleh dari kalimat Bahasa Indonesia yang diucapkan oleh seorang model

4.1 Data Percobaan

Percobaan ini menggunakan data C3D yang didapatkan dari hasil rekaman *motion capture* pada model wajah manusia di dunia nyata pada saat model manusia membacakan beberapa kalimat . C3D yang diperoleh akan dikoneksikan dengan model 3D wajah manusia. Model 3D tersebut digunakan untuk visualisasi pada saat model mengucapkan beberapa kalimat. kalimat digunakan sebanyak 10 seperti tabel 4.1 kalimat yang akan dibacakan oleh model. Setelah model 3D mengucapkan beberapa kalimat maka akan diperoleh gerakan bibir per *frame* , data per *frame* ini yang akan diproses. Dari 10 kalimat yang diucapkan oleh model akan diperoleh 1764 frame selanjutnya diproses segmentasi menjadi 103 fitur suku kata. Hasil segmentasi berupa fitur suku kata tersebut yang akan di klasterisasi menjadi 20 kelas *cluster* dengan menggunakan Algoritma K-means.

Tabel 4.1: Kalimat yang digunakan sebagai data

No	Kalimat
1	Saya suka baju yang berwarna merah tua
2	Boneka beruang di toko itu lucu sekali
3	Jalan itu ramai sekali setiap pagi hari
4	Lusa aku akan pergi ke rumah paman
5	Lusa aku akan pergi ke koto baju bersama ibu
6	Besok saya ke toko baju bersama ibu
7	Toko itu buka mulai jam enam pagi
8	ibu menyirami bunga di halaman setiap hari
9	Buanglah sampah ditempat yang telah disediakan
10	Masakan nenek paling enak di rumah kami

4.2 Hasil Rekaman Wajah

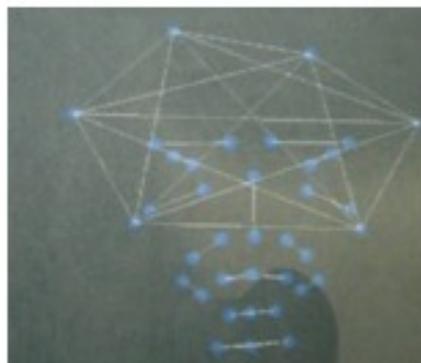
Data dari visualisasi pada wajah manusia menggunakan penanda aktif diwajah model yang direkam melalui komputer sehingga akan menghasilkan *bone* seperti gambar 4.1



Gambar 4.1: hasil bone

4.3 Hasil Pembentukan *face Template*

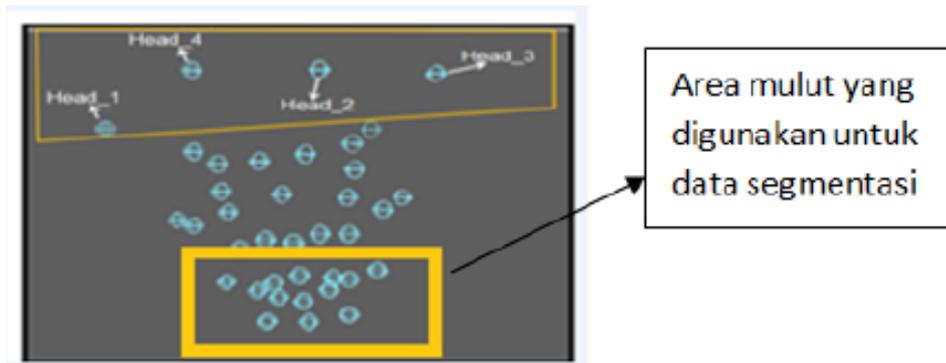
hasil *bone* ini digunakan untuk pembentukan *face template* pada proses perekaman yang akan dijadikan file C3D. Sebelum proses pembentukan *face template* dilakukan terlebih dahulu model duduk di depan sistem kamera seperti ilustrasi pada Gambar 3.5. Sehingga gerak wajah model dapat dilihat oleh beberapa kamera dan hasil dari *bone* yang akan di satukan antara titik penanda aktif ke penanda aktif lainnya sehingga membentuk *face template* dari setiap *frame* gerakan pada wajah model akan dilacak oleh *OptiTrack*. Sehingga kamera mulai berkerja secara *real time* melacak penanda aktif yang terdapat pada wajah model. *Software OptiTrack* akan secara *real time* membentuk *face template* yang akan dihasilkan seperti pada gambar 4.2 .



Gambar 4.2: hasil mapping

4.4 Normalisasi Posisi 3D

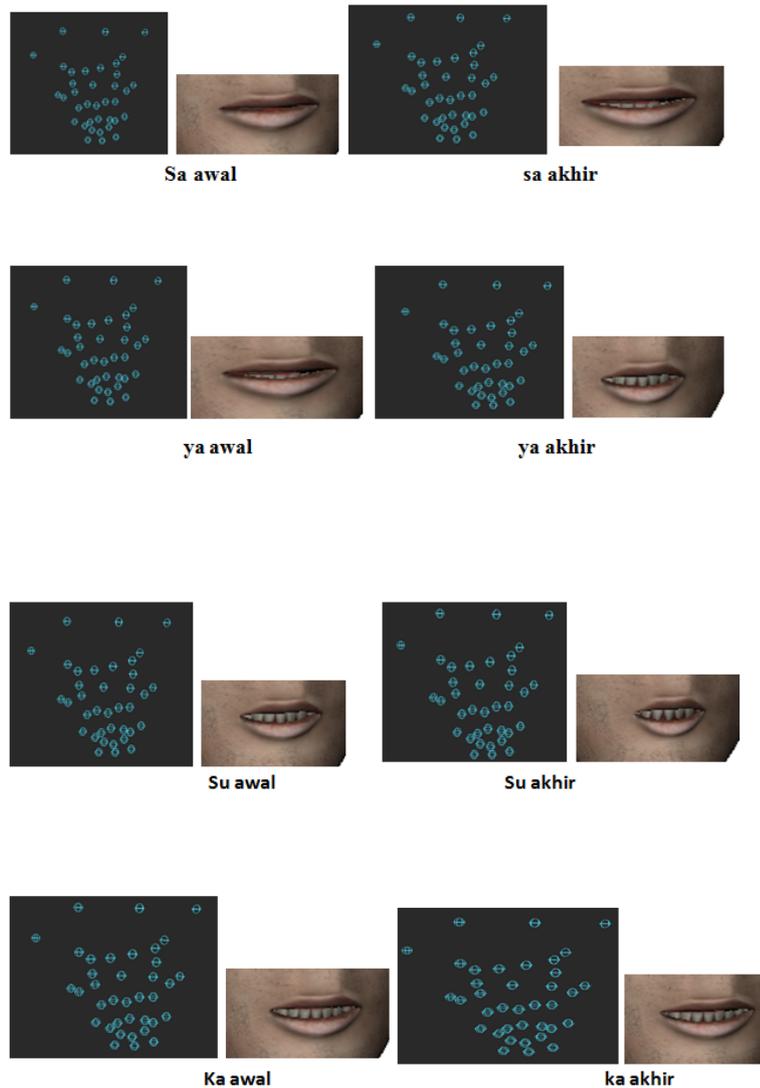
Setelah didapatkan *face template* maka akan dicari data setiap gerakan yang disinkronkan dengan gerakan wajah dari model tetapi datanya masih besar, sehingga dilakukan proses normalisasi yang hanya diambil di bidang tetap dengan bergerak di area mulut dan di kepala dengan menggunakan persamaan 3.1, 3.2, 3.4, 3.5. Sehingga didapatkan nilai hanya pada bagian area mulut saja. Seperti yang diberi tanda pada gambar 4.3.



Gambar 4.3: hasil mapping

4.5 Hasil Segmentasi

Visualisasi pada wajah animasi pada saat mengucapkan 10 kalimat yang diperankan oleh model seperti gambar 4.2 dan gambar 4.3, hasil dari rekaman *motion* data berupa file C3D yang akan dikoneksikan dengan wajah animasi, sehingga didapatkan gerakan bibir animasi yang sama dengan model yang diperankan oleh manusia di dunia nyata pada saat mengucapkan 10 kalimat dan diambil awal dan akhir dari suku kata untuk mendapatkan nilai rata-rata dari setiap fitur suku kata.



Gambar 4.4: Hasil C3D dan Viseme animasi saat mengucapkan awal dan akhir suku kata

4.6 KLASTERISASI DENGAN ALGORITMA K - MEAN

Setelah hasil diperoleh dari proses segmentasi berupa data fitur suku kata seperti tabel 1, selanjutnya pada penelitian ini melakukan uji coba hasil segmentasi. Hasil Segmentasi tersebut di proses klasterisasi dengan menggunakan Algoritma K-mean. Sehingga hasil segmentasi akan di kelompokkan dalam beberapa *cluster*.

Dimana langkah klasterisasi untuk melakukan uji coba hasil penelitian menggunakan Algoritma K-mean adalah berikut:

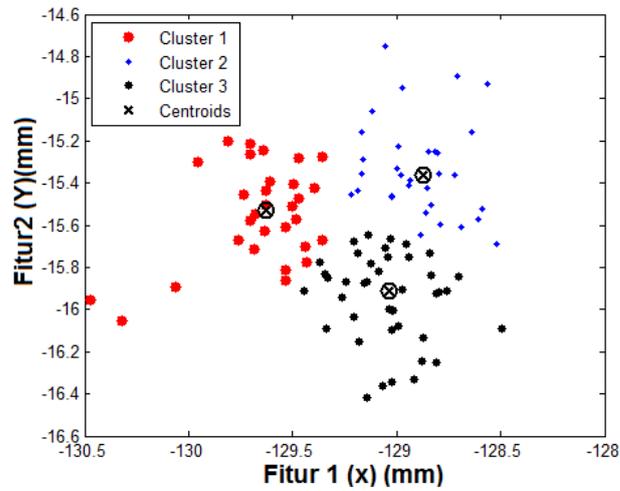
1. Menentukan nilai k secara acak.

2. Menentukan nilai pusat massa. Pada awal iterasi, nilai-nilai *centroid* yang ditentukan secara acak. Pada langkah iterasi berikutnya, nilai massa ditentukan dengan menghitung rata-rata setiap cluster.
3. Menghitung jarak *centroid* dan masing-masing yang memiliki data .
4. Pengelompokan data berdasarkan minimum *Euclidean Distance*.
5. Selanjutnya akan kembali ke langkah 2, mengulangi langkah-langkah sampai nilai *centroid* tetap dan anggota cluster tidak pindah ke *cluster* lain. Salah satu metode untuk menentukan *cluster* terdefinisi dengan baik adalah dengan menggunakan kriteria fungsi yang mengukur kualitas Klasterisasi. Ada metode yang digunakan secara luas, yaitu *Sum of Squared Error (SSE)*.

4.7 Mengukur kualitas Klasterisasi

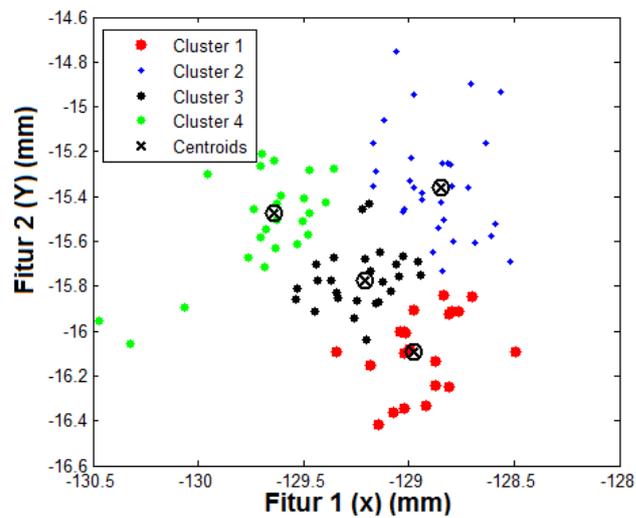
Untuk menentukan *cluster* terdefinisi dengan baik adalah menggunakan kriteria fungsi yang mengukur kualitas yaitu menggunakan *Sum of Squared Error(SSE)*, dengan menggunakan rumus 2.9

Kualitas *cluster* juga dapat *evaluated* yang menggunakan variasi kelas (BCV) dan dalam variasi kelas.(WCV). BCV adalah nilai rata-rata yang dihasilkan dari antara *centroid* dan WCV adalah *Sum of Squire Error*. Nilai rasio untuk BCV dan WCV menggunakan rumus 2.10 dari hasil setiap *cluster*. Dengan demikian kita dapat menguji kualitas dari *cluster* untuk pengelompokan yang digunakan dalam penelitian ini. Dari gambar 4.5 mendapatkan nilai rata-rata dari *Distance Centroid Between-Class Variation* (BCV) = 0.45 sedangkan $SSE/WCV = 34.74$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10



Gambar 4.5: Hasil *cluster* K=3

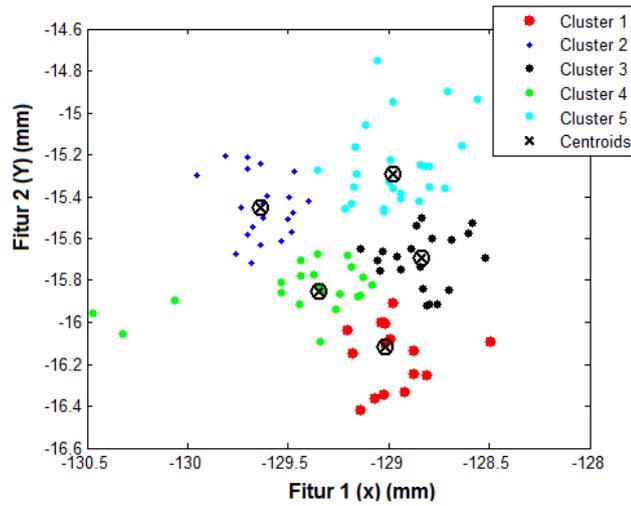
$$Q = \frac{0.45}{34.74} = 0.013 \quad (4.1)$$



Gambar 4.6: Hasil *cluster* K=4

Dari gambar 4.6 mendapatkan nilai rata-rata dari *Distance Centroid Between-Class Variation*(BCV) = 0.49 sedangkan $SSE/WCV = 30.99$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

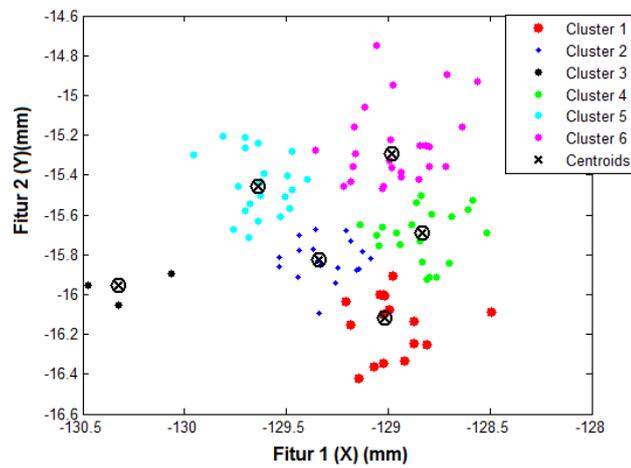
$$Q = \frac{0.49}{30.99} = 0.016 \quad (4.2)$$



Gambar 4.7: Hasil *cluster* K=5

Dari gambar 4.7 mendapatkan nilai rata-rata dari *Distance Centroid Between-Class Variation* (BCV) = 0.50 sedangkan $SSE/WCV = 27.29$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV menggunakan rumus 2.10

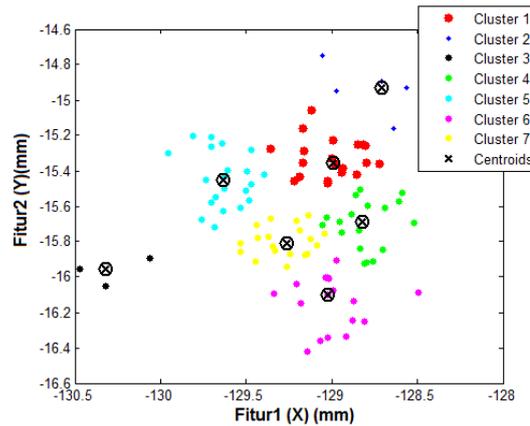
$$Q = \frac{0.50}{27.29} = 0.018 \quad (4.3)$$



Gambar 4.8: Hasil *cluster* K=6

Dari gambar 4.8 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.69 sedangkan $SSE/WCV = 24.64$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan persamaan 2.10

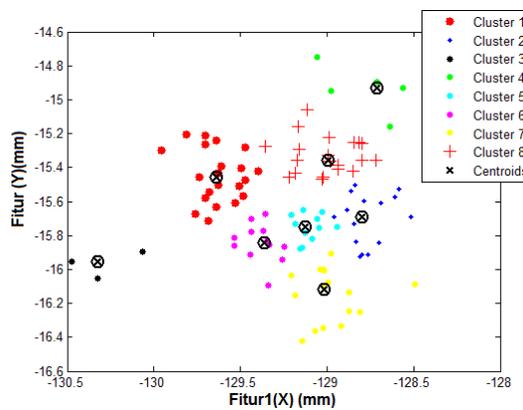
$$Q = \frac{0.69}{24.64} = 0.028 \quad (4.4)$$



Gambar 4.9: Hasil *cluster* K=7

Dari gambar 4.9 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.69 sedangkan $SSE/WCV = 22.81$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan persamaan 2.10

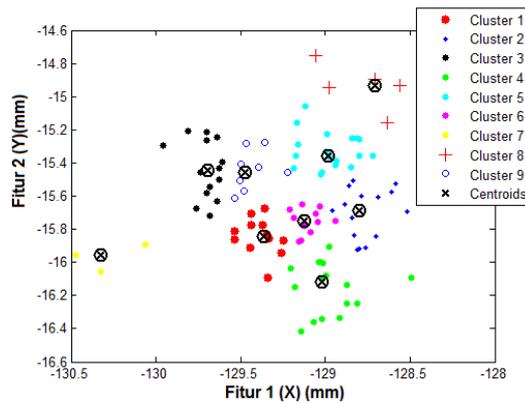
$$Q = \frac{0.69}{22.81} = 0.03 \quad (4.5)$$



Gambar 4.10: Hasil *cluster* K=8

Dari gambar 4.10 mendapatkan nilai rata-rata dari *Distance Centroid Between-Class Variation* (BCV) = 0.76 sedangkan $SSE/WCV = 21.33$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan persamaan 2.10

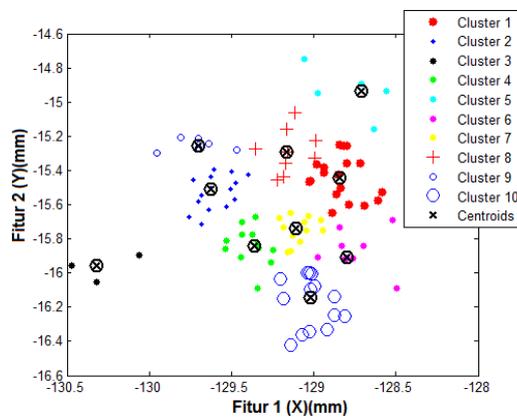
$$Q = \frac{0.76}{21.33} = 0.036 \quad (4.6)$$



Gambar 4.11: Hasil *cluster* K=9

Dari gambar 4.11 mendapatkan nilai rata-rata dari *Between-Class Variation*(BCV) = 0.6678 sedangkan $SSE/WCV = 19.8215$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

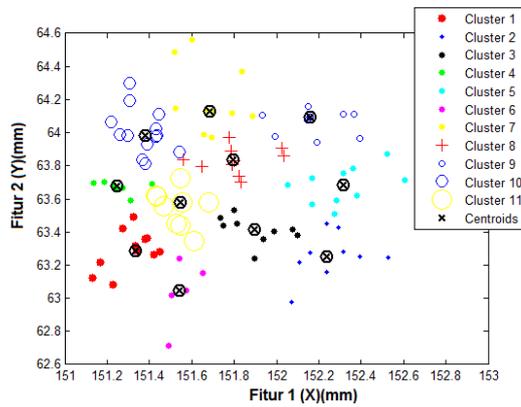
$$Q = \frac{0.67}{19.82} = 0.034 \quad (4.7)$$



Gambar 4.12: Hasil *cluster* K=10

Dari gambar 4.12 mendapatkan nilai rata-rata dari *Distance Centroid*(BCV) = 0.69 sedangkan $SSE/WCV = 18.71$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

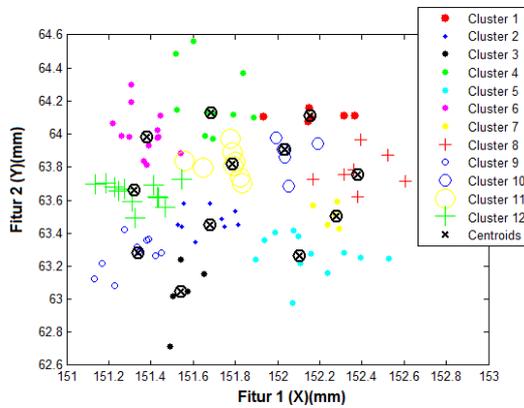
$$Q = \frac{0.69}{18.71} = 0.037 \quad (4.8)$$



Gambar 4.13: Hasil *cluster* K=11

Dari gambar 4.13 mendapatkan nilai rata-rata dari *Distance Centroid*(BCV) = 0.6241 sedangkan $SSE/WCV = 18.3157$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

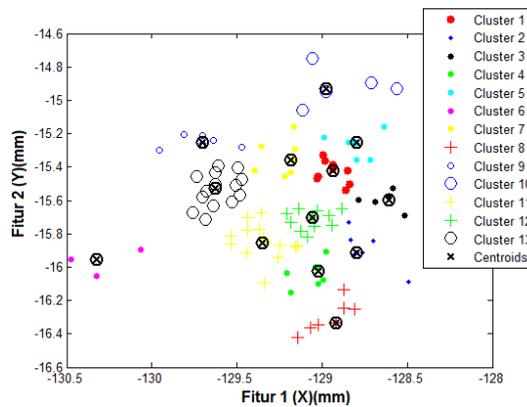
$$Q = \frac{0.62}{18.31} = 0.034 \quad (4.9)$$



Gambar 4.14: Hasil *cluster* K=12

Dari gambar 4.14 mendapatkan nilai rata-rata dari *Distance Centroid*(BCV) = 0.6655 sedangkan SSE/WCV = 17.3665 , Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

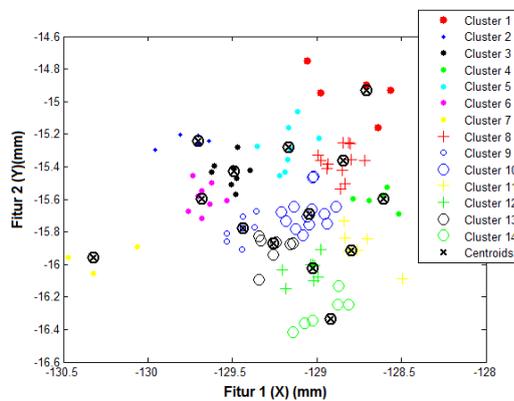
$$Q = \frac{0.67}{17.37} = 0.038 \quad (4.10)$$



Gambar 4.15: Hasil *cluster* K=13

Dari gambar 4.15 mendapatkan nilai rata-rata dari *Distance Centroid*(BCV) = 0.7078 sedangkan SSE/WCV = 15.1617 , Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

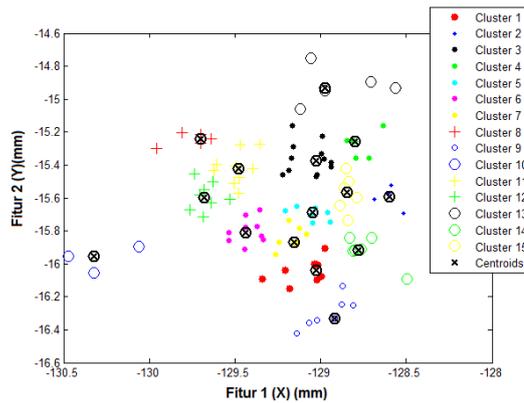
$$Q = \frac{0.71}{15.16} = 0.047 \quad (4.11)$$



Gambar 4.16: Hasil *cluster* K=14

Dari gambar 4.16 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.70 sedangkan $SSE/WCV = 14.23$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

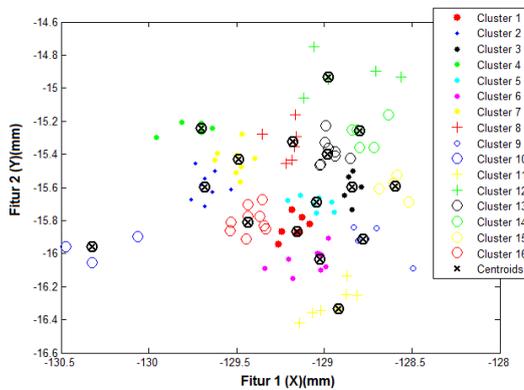
$$Q = \frac{0.70}{14.23} = 0.049 \quad (4.12)$$



Gambar 4.17: Hasil *cluster* K=15

Dari gambar 4.17 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.69 sedangkan $SSE/WCV = 13.55$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

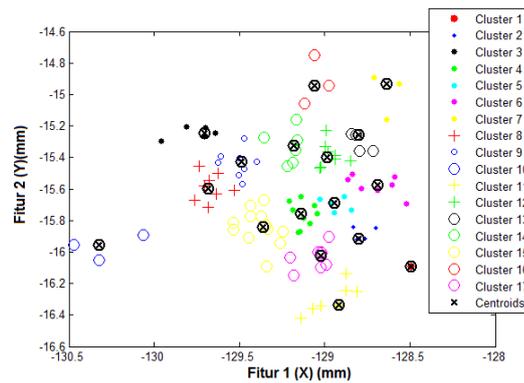
$$Q = \frac{0.69}{13.55} = 0.051 \quad (4.13)$$



Gambar 4.18: Hasil *cluster* K=16

Dari gambar 4.18 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.67 sedangkan $SSE/WCV = 12.96$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

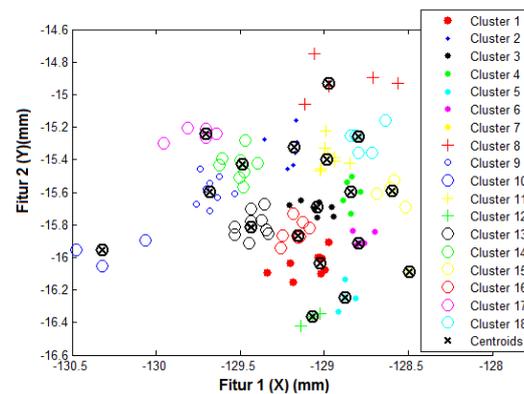
$$Q = \frac{0.67}{12.96} = 0.052 \quad (4.14)$$



Gambar 4.19: Hasil *cluster* K=17

Dari gambar 4.19 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.70 sedangkan $SSE/WCV = 12.48$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

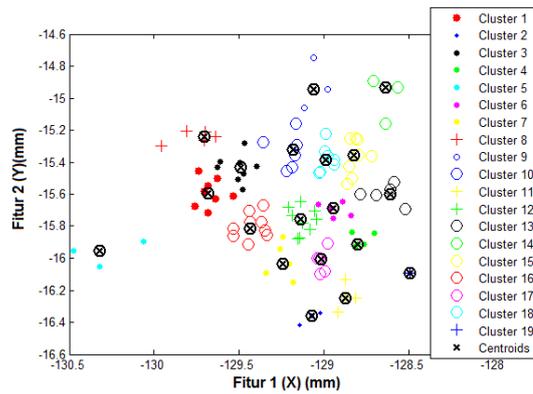
$$Q = \frac{0.70}{12.48} = 0.056 \quad (4.15)$$



Gambar 4.20: Hasil *cluster* K=18

Dari gambar 4.20 mendapatkan nilai rata-rata dari *Distance Centroid Between-Class Variation* (BCV) = 0.72 sedangkan $SSE/WCV = 11.10$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

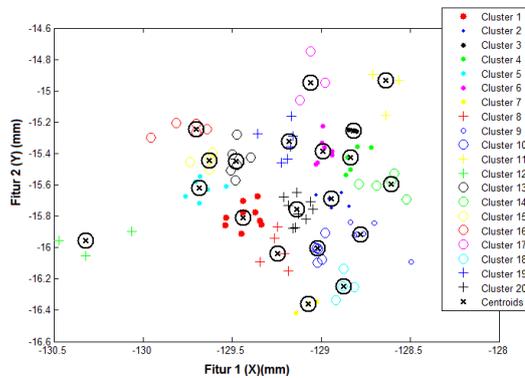
$$Q = \frac{0.72}{11.81} = 0.061 \quad (4.16)$$



Gambar 4.21: Hasil *cluster* K=19

Dari gambar 4.21 mendapatkan nilai rata-rata dari *Distance Centroid Between-Class Variation* (BCV) = 0.74 sedangkan $SSE/WCV = 11.10$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

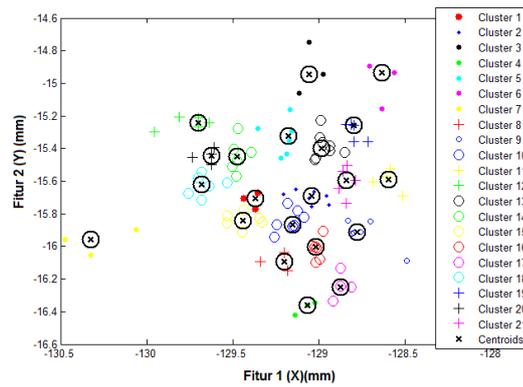
$$Q = \frac{0.74}{11.10} = 0.067 \quad (4.17)$$



Gambar 4.22: Hasil *cluster* K=20

Dari gambar 4.22 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.69 sedangkan $SSE/WCV = 10.67$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

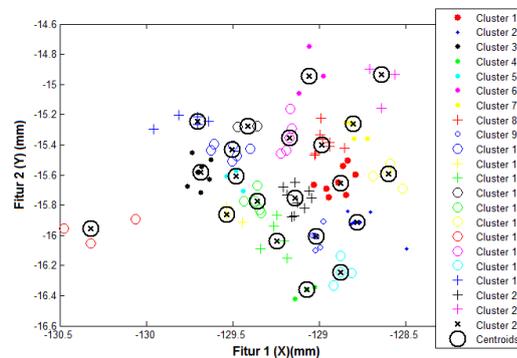
$$Q = \frac{0.69}{10.67} = 0.065 \quad (4.18)$$



Gambar 4.23: Hasil *cluster* K=21

Dari gambar 4.23 mendapatkan nilai rata-rata dari *Distance Centroid* *Between-Class Variation* (BCV) = 0.75 sedangkan $SSE/WCV = 10.36$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

$$Q = \frac{0.75}{10.36} = 0.072 \quad (4.19)$$



Gambar 4.24: Hasil *cluster* K=22

Dari gambar 4.24 mendapatkan nilai rata-rata dari *Distance Centroid Between-Class Variation* (BCV) = 0.72 sedangkan $SSE/WCV = 10.31$, Q adalah hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dengan menggunakan rumus 2.10

$$Q = \frac{0.72}{10.31} = 0.07 \quad (4.20)$$

Sehingga hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV dapat dilihat di tabel 4.1, hasil tersebut menggunakan formula untuk kalkulasi SSE dan rasio untuk BCV dan WCV yaitu persamaan 2.10

Tabel 4.2: Kalkulasi SSE dan rasio untuk BCV dan WCV

K=value	(BCV)	SSE/WCV	BCV/SSE (WCV)
K=3	0.45	34.74	0.013
K=4	0.49	30.99	0.016
K=5	0.50	27.29	0.018
K=6	0.69	24.64	0.028
K=7	0.69	22.80	0.03
K=8	0.76	21.33	0.036
K=9	0.67	19.82	0.034
K=10	0.69	18.71	0.037
K=11	0.62	18.31	0.034
K=12	0.67	17.37	0.038
K=13	0.71	15.16	0.047
K=14	0.70	14.23	0.049
K=15	0.69	13.55	0.051
K=16	0.67	12.96	0.052
K=17	0.70	12.48	0.056
K=18	0.72	11.81	0.061
K=19	0.74	11.10	0.067
K=20	0.69	10.67	0.065
K=21	0.75	10.36	0.072
K=22	0.72	10.31	0.07

Keterangan:

K = Jumlah *cluster*

BCV = rata-rata jarak antara *centroid*

WCV = *Sum of Squared Error*

SSE = *Sum of Squared Error*

BCV/SSE (WCV) = hasil rasio

Setelah didapatkan hasil dari kalkulasi SSE dan rasio untuk BCV dan WCV yang digunakan sebagai kualitas klusterisasi maka kita bisa menentukan

kualitas klasterisasi yang baik untuk menentukan kelas- kelas dari fonem berupa suku kata KV (konsonan vokal), dimana jika hasil dari kalkulasi *SSE* dan rasio untuk *BCV* lebih kecil dari hasil *WCV* lebih besar maka kualitas klasterisasi baik. sehingga diambillah $K=20$ sebagai kualitas klasterisasi lebih baik yang akan digunakan untuk *cluster* dan menentukan persentasi fonem suku kata. dengan membandingkan jumlah suku kata per jumlah per kelas dikalikan 100 persen. sehingga diketahui *viseme* suku kata yang muncul.seperti yang pada tabel 4.2, 4.3, 4.4. dan visual yang lebih jelas dari masing-masing kelas viseme ditunjukkan oleh *viseme* model pada Gambar 4.25,4.26, 4.27,4.28,4.29 dan 4.30.

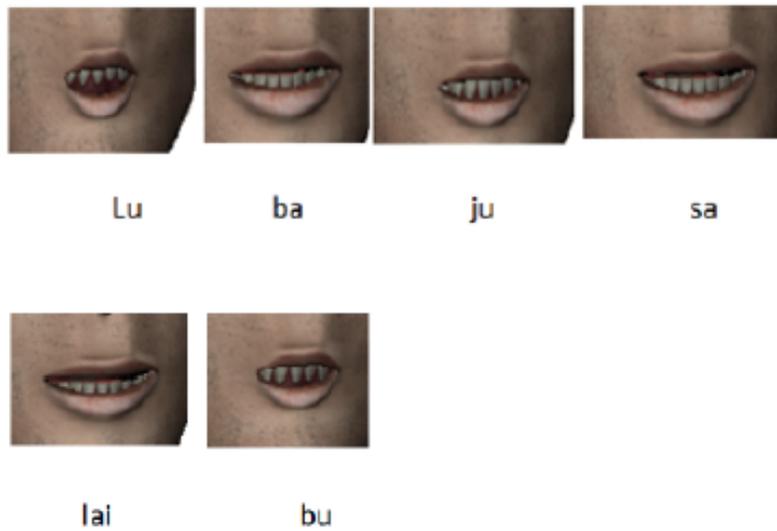
Tabel 4.5. menampilkan persentase keanggotaan setiap *viseme* yang bervariasi. Persentase keanggotaan ini dapat menentukan dengan tingkat keanggotaan dari *viseme* untuk suku kata. anggota *viseme* dengan persentase keanggotaan yang tinggi ini berarti dari tingkat keanggotaan untuk *cluster* yang sangat kuat. *viseme* dengan tingkat keanggotaan yang lemah akan dihapus dari keanggotaan *cluster*, namun keanggotaannya ke *cluster* lain yang tingkat keanggotaan yang tinggi dipertahankan atau ditempatkan pada persentase kelompok yang lebih besar. Tingkat keanggotaan setiap *cluster* ini kemudian digunakan sebagai dasar pemetaan ke dalam kelas *viseme*. *visemes* yang kondisi 'diam' tidak dimasukkan dalam proses *clustering*.

Setelah didapatkan hasil struktur kelas dari *viseme* dapat di aplikasikan oleh model pada gambar 4.23, 4.24 dan gambar 4.25 yang dijadikan sebagai model *viseme*

Kelas 1



Kelas2



Gambar 4.25: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model)

Tabel 4.3: Deskripsi Hasil Klasterisasi K=20

Cluster	Anggota Viseme Untuk Suku Kata	%	Cluster	Anggota Viseme Untuk Suku Kata	%
Cluster1	se	25	Cluster 7	ma	16.67
	ti	25		to	16.67
	di	25		ko	16.67
	ru	25		tu	16.67
Cluster 2	lu	12.5	Cluster 8	bu	16.67
	sa	12.5		nyi	16.67
	ba	12.5		na	12.5
	ju	12.5		se	12.5
	sa	12.5		cu	12.5
	lai	12.5		ci	12.5
	bu	12.5		ja	12.5
	ru	12.5		tu	12.5
Cluster3	ke	16.67	Cluster 9	ma	12.5
	ka	16.67		pa	12.5
	ya	16.67		sa	20
	gi	16.67		ya	20
	ke	16.67		su	20
	di	16.67		li	20
Cluster4	sa	20	Cluster 10	ne	20
	gi	20		ko	25
	to	20		cu	12.5
	ko	20		se	12.5
	mi	20		ku	25
Cluster5	bo	33.3	Cluster 11	da	12.5
	ne	33.3		ra	12.5
	ka	33.3		tu	25
Cluster6	ka	20	Cluster 12	ke	25
	ba	20		pa	25
	ju	20		ru	25
	me	20		be	25
	tu	20		bu	25
				ra	25
				mi	25

Tabel 4.4: Deskripsi Hasil Klasterisasi K=20

Cluster	Anggota Viseme Untuk Suku Kata	%	Cluster	Anggota Viseme Untuk Suku Kata	%
Cluster11	tu	25	Cluster16	ka	33.3
	ke	25		di	33.3
	pa	25		re	33.3
	ru	25	Cluster 17	to	16.67
Cluster 12	be	25		lu	16.67
	bu	25		be	16.67
	ra	25		pa	16.67
	mi	25		so	16.67
Cluster13	gi	33.3	Cluster18	bu	16.67
	ha	33.3		ri	12.5
	ri	33.3		ma	12.5
Cluster14	se	33.3		se	12.5
	ti	33.3		li	12.5
	pa	33.3	be	12.5	
Cluster15	me	25	so	12.5	
	nga	25	Cluster19	ru	14.29
	di	25		mu	14.29
	ka	25		bu	14.29
			Cluster 20	ha	20
				te	20
				di	40
				se	20

Tabel 4.5: Struktur Kelas Viseme

Kelas Viseme	Anggota fonem untuk Suku kata	Viseme	Contoh	Kalimat
Kelas 0	Diam	-	-	
Kelas 1	Se, ti, di, ru	se	setiap, dirumah	4 , 10
Kelas 2	lu, ba, ju,	sa	lusa, baju, sama	5, 6, 7
	sa, lai, bu		mulai, ibu	
Kelas 3	ke, ka, ya	ke	ke, sekali, saya	3, 4, 6
	gi, ke, di		pergi, di, ke	
Kelas 4	sa, gi, to, ko, mi	gi	lusa, pergi, toko, kami	5, 6, 10
Kelas 5	bo, ne, ka	bo	boneka	2
Kelas 6	ka, ba,ju, me, tu	ka	suka, baju, merah, tua	1
Kelas 7	ma, to, ko,	tu	bersama, toko, itu, ibu	6, 7
	tu, bu, nyi		ibu, menyiram	8
Kelas 8	na, se, cu,	na	berwarna, sekali, cuci	1, 2, 3
	ci, ja, tu, ma, pa		jalan, itu, masakan, paling	4,10
Kelas 9	sa, ya, su, li, ne	ne	saya, suka, sekali, nenek	1, 2, 10
	li,ne			
Kelas 10	ko, cu, se, ku,	ku	toko, lucu, sekali, sepatuku	2, 3
	ko,ku, da, ra		kotor, belumku, dari ,ramai	4
Kelas 11	tu, ke, pa, ru	pa	itu, ke, paman, rumah	2, 5
Kelas 12	be, bu, ra, mi	be	beruang, ibu	2, 6
			menyiram	8
Kelas 13	gi, ha, ri	ha	pagi, hari	4
Kelas 14	se, ti, pa	ti	setiap, pagi	4
Kelas 15	me, nga	nga	menyiram, bunga	8, 9
	di, ka		di, tempat, kami	10
Kelas 16	ka, di, re	re	buka, di, sore	2, 7, 8
kelas 17	to, lu, be, pa	to	toko, lucu, belum	2, 3, 4
	bu, so		pagi, sore, buanglah	8,9
Kelas 18	ri, ma, se,	ma	dari, kemarin, sekali,	3, 4,
	li, be, so			6
Kelas 19	ru, mu, bu	ru	besok,beruang	2,7
			mulai, buka	
Kelas 20	ha, te, di	di	halaman, tempat,	8, 9
	se, di		di, sediakan	

Kelas 3



Ke

ka

ya



gi

ke

di

Kelas 4



Sa

gi

to



ko

mi

Gambar 4.26: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimod-
elkan oleh model *viseme*(area mulut model)

Kelas 5



Kelas 6



Gambar 4.27: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimod- elkan oleh model *viseme*(area mulut model)

Kelas 7



Ma

to

ko



tu

bu

nyi

Kelas 8



Na

se

cu



Ja

Ma

pa



ci

Gambar 4.28: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimod-
elkan oleh model *viseme*(area mulut model)

Kelas9



Sa

ya

su



Li

ne

Kelas 10



Ko

cu

se



Ku

ko

ku

ra

Gambar 4.29: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimod- elkan oleh model *viseme*(area mulut model)

Kelas11



Tu

ke

ru



pa

Kelas 12



Be

bu

ra



mi

Gambar 4.30: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model)

Kelas13



Gi

ha

ri

Kelas 14



Se

ti

pa

Kelas 15



Me

nga

di

ka

K16



ka

di

re

Gambar 4.31: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimod- elkan oleh model *viseme*(area mulut model)

Kelas 17



To

lu

so



Pa



be



bu

K18



Ri

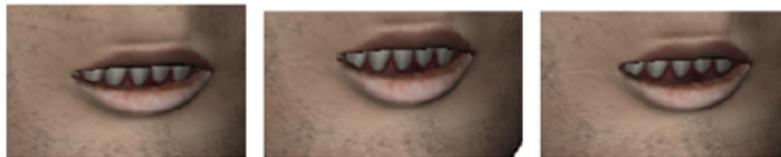
ma

se

li

Gambar 4.32: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model)

Kelas 19



ru

mu

bu

K20



Ha

te

di



se



di

Gambar 4.33: hasil dari *cluster viseme* berupa kelas- kelas *viseme* yang dimodelkan oleh model *viseme*(area mulut model)

[Halaman ini sengaja dikosongkan.]

BAB V

KESIMPULAN

Dalam bab terakhir penelitian ini dimuat suatu kesimpulan dan saran hasil pengujian dan pembahasan yang telah dilakukan sebelumnya. Adapun kesimpulan dan saran ini bermanfaat untuk pengembangan penelitian selanjutnya.

5.1 Kesimpulan

Penulis dapat menyimpulkan dari beberapa proses yaitu proses pada saat mulut animasi digerakan, pada saat mulut animasi digerakan terdapat perbedaan viseme pada setiap suku kata, perbedaannya antara lain yaitu viseme yang paling dekat dengan yang paling jauh di saat animasi berbicara secara netral. Adapun yang paling dekat yaitu suku kata sa dengan ya dikarenakan gerakan mulutnya hampir sama sedangkan yang paling jauh yaitu suku kata su dan suku kata ka, sedangkan pada proses *cluster* K-mean untuk viseme bahasa Indonesia berdasarkan gerakan mulut pada saat mengucapkan 10 kalimat bahasa Indonesia , 10 kalimat bahasa Indonesia yang diucapkan akan menghasilkan gerakan mulut yang sama dan berbeda, gerakan mulut yang sama akan di jadikan dalam 1 kelas, sedangkan gerakan mulut yang berbeda akan dijadikan dalam kelas-kelas yang lainnya sesuai dengan jarak gerakan mulut animasi, dari gerakan mulut animasi didapatkan hasil kelas-kelas viseme dengan menggunakan *cluster* K-mean. Dari hasil *cluster* k-mean didapatkan hasil *cluster* yang baik dengan $K=20$ dari 22 klaster berdasarkan hasil penelitian *cluster* yang dilakukan, hasil *cluster* yang baik ini berdasarkan dari matrik kalkulasi untuk SSE dan rasio untuk BCV dan WCV, dengan demikian sehingga dapat di cari anggota viseme pada suku kata untuk dapat digunakan pada ligustik bahasa Indonesia.

5.2 Penelitian Selanjutnya

Penelitian lanjutan dapat dikembangkan menggunakan metode klasterisasi atau *clustering* lain yang dapat meningkatkan kualitas *cluster* untuk viseme Bahasa Indonesia Penelitian lanjutan dapat menggunakan data suara yang berupa sinyal yang dapat di sinkronkan dengan gerakan mulut dan dapat mengklaster viseme bukan hanya berupa konsonan vokal tapi juga suku kata

lainnya yang dinamis sehingga didapatkan viseme yang lebih dinamis untuk Bahasa Indonesia.

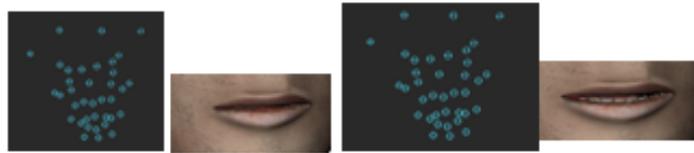
DAFTAR PUSTAKA

- 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, Yogyakarta.
- Eka Prasetyono, Mochamad Hariadi. 2010. *Penangkapan Gerak 3D Pada Manusia Menggunakan Penanda Aktif* .Jurusan Teknik Elektro, Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember (ITS).
- Gleason, H.A. 2011. *Introduction to Descriptive Linguistics* .New York: Rinehart and Winston.
- Hui Zhao and Chaojing Tang. 2008. *Visual Speech Synthesis based on Chinese Dynamic Visemes* . International Conference on Information and Automation, IEEE, June 20 -23, Zhangjiajie, China.
- I. Mazonaviciute, R. Bausys. *Translingual Visemes Mapping for Lithuanian Speech Animation* .Department of Graphical Systems, Vilnius Gediminas Technical University, ISSN 1392-1215, pp. 95-98.
- Ioannis A. Ypsilos, Adrian Hilton, Aseel Turkmani and Philip J. B. Jackson. 2011. *Speech-driven Face Synthesis from 3D Video*. Centre for Vision Speech and Signal Processing University of Surrey, Guildford, GU2 7XH, U, Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04) , IEEE ,hal. 1-8.
- Jiyong Ma, Member, IEEE, Ron Cole, Member, IEEE, Bryan Pellom, Member, IEEE, Wayne Ward, and Barbara Wise. *Accurate Visible Speech Synthesis Based on Concatenating Variable Length Motion Capture Data* . IEEE Transactions on Visualization and Computer Graphics, hal.1-12. Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04) , IEEE ,hal. 1-8.
- Jose Mario De Martino, Leo Pini Magalhaes, Fabio Violaro. *Facial Animation based on Context-Dependent Visemes* . Elsevier Computer Graphics.

- Ka-Ho WONG, Wai-Kim LEUNG, Wai-kit LO and Helen MENG. 2012. *Development of an Articulatory Visual Speech Synthesizer to Support Language Learning*. IEEE Transaction on Speech and Audio Processing, pp. 139-143.
- M. Barnard, E. Holden, and R. Owens. *Lip tracking using pattern matching snakes* .In proceedings of the 5th Asian Conference on Computer Vision.
- M. E. Hennecke, K. V. Prasad, and D. G. Stork. *Using deformable templates to infer visual speech dynamics* . Elsevier Computer Graphics.
- Midori Kitagawa, Brian Windsor. 2008. *Mocap for Artists*. Publish by Elsevier Inc, United States of America, pp. 181.
- Miftahur Rohman, Wirawan. 2012. *Analisa Gerakan Manusia Pada Video Digital*. Jurusan Teknik Elektro, Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember (ITS).
- Sarah L. Taylor 1, Moshe Mahler 2, Barry-John Theobald 1 and Iain Matthews. 2012. *Dynamic Units of Visual Speech*. 1 University of East Anglia, Norwich, England , Disney Research, Pittsburgh, USA, Eurographics/ ACM SIGGRAPH Symposium on Computer Animation.
- Ren. C. Luo,IEEE, Chien-Chieh Huang, Shu-Ruei Chang, and Yi-Jeng Tsai. 2011. *Speech Synchronization between Speech and Lip Shape Movements for Service Robotics Applications*. IEEE ,hal. 1-6,
- R. Kent and F. Mini e. 1977. *Coarticulation in recent speech production models*. Journal of Phonetics.
- Yasuhiro Akagi dkk. 2013. *Marker-less Facial Motion Capture based on the Parts Recognition*. Kagoshima University Korimoto, 1-21-24 Kagoshima 890-8580, JAPAN
- Yudi Prayudi, Iwan Aprizal. 2004. *Pemodelan Wajah 3D Berbasis Foto diri Menggunakan Maya Embedded Language (MEL) Script* .Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia , Media Informatika, Vol. 2,No. 2, Desember
- Zign Creations. 2013. *Zign Track Real-Time Facial Motion Capture Pro* . Zign Creations.

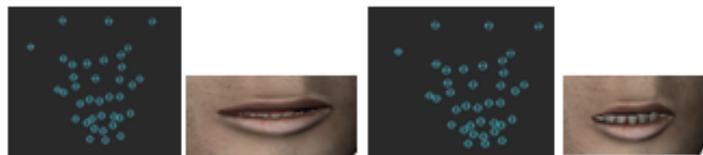
LAMPIRAN I

Hasil segmentasi yang merupakan data yang akan digunakan untuk proses klasterisa (clustering). pemisahan setiap suku kata yaitu dari awal dan akhir dari sepuluh kalimat disini hanya melampirkan 3 kalimat dari 10 kalimat yang di ucapkan secara alami (netral).sehingga kita dapat menganalisa pada saat Clustering. Kalimat yang di pisahkan pada lampiran adalah sa,ya, su,ka



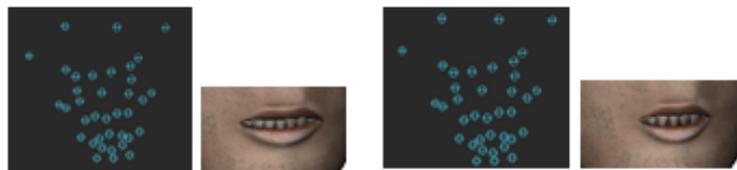
Sa awal

sa akhir



ya awal

ya akhir



Su awal

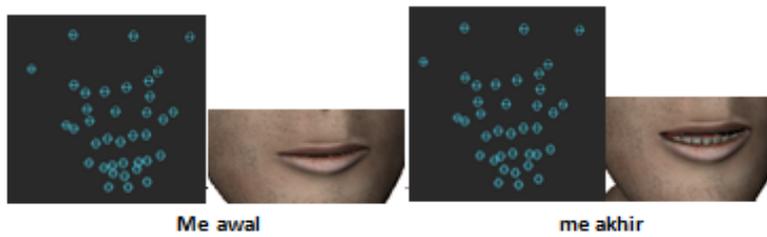
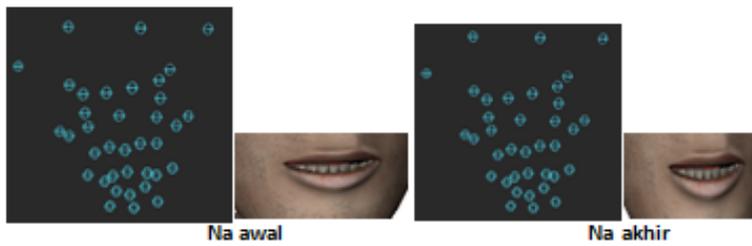
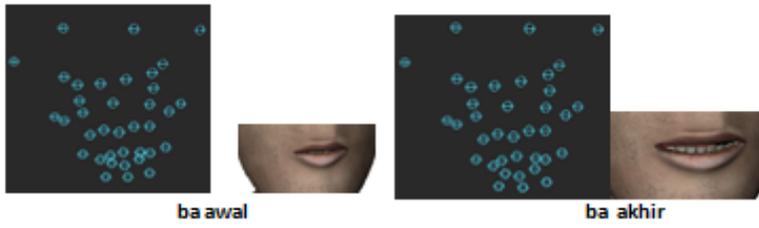
Su akhir



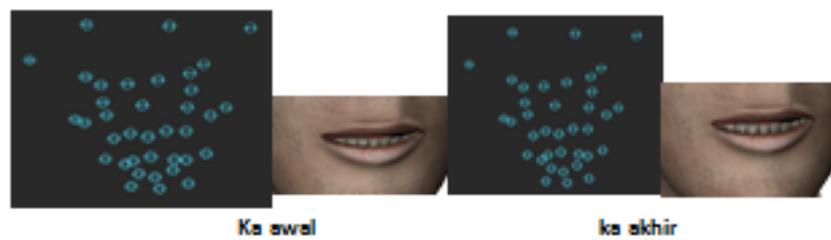
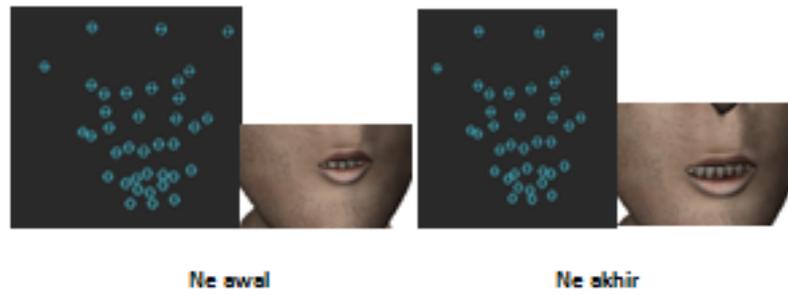
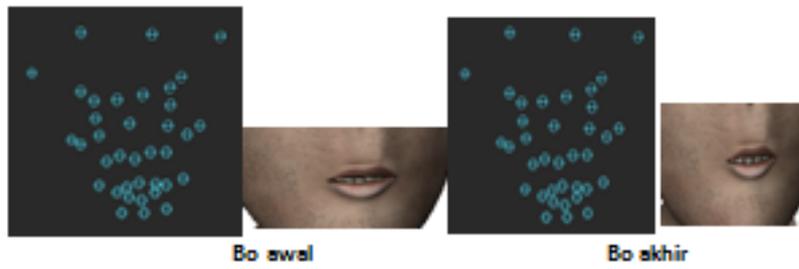
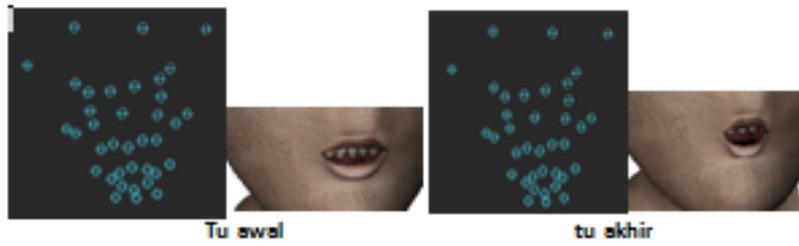
Ka awal

ka akhir

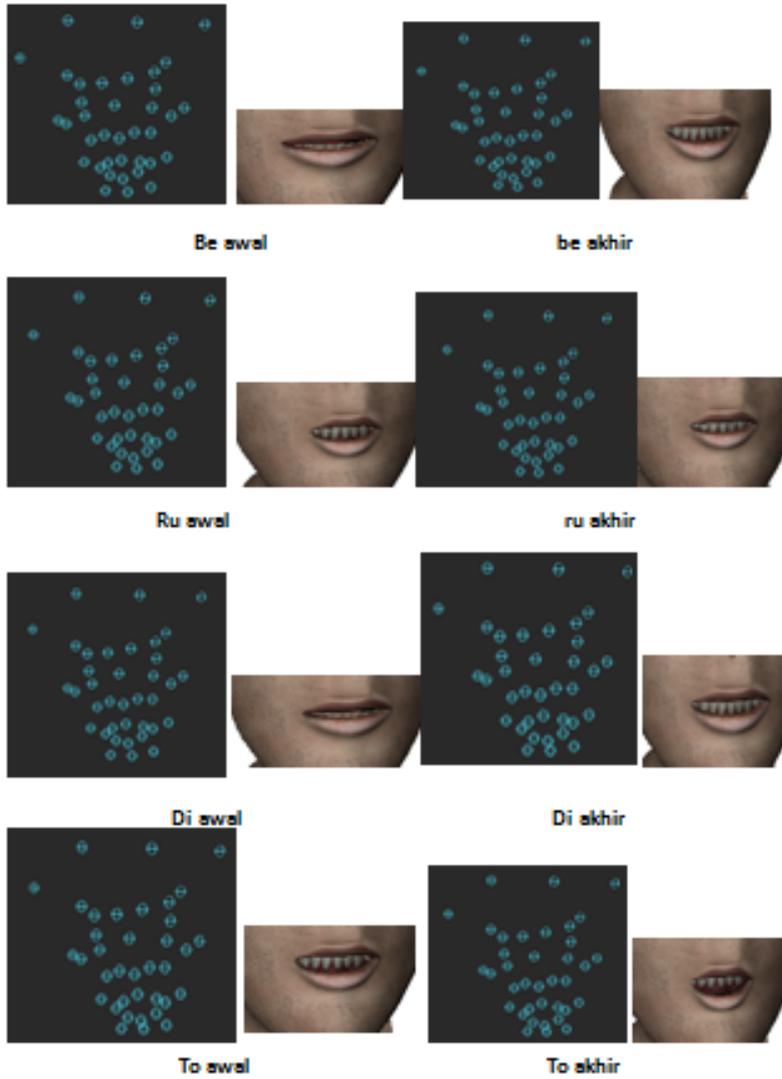
Kalimat yang di pisahkan pada lampiran adalah ba, ju, na, me



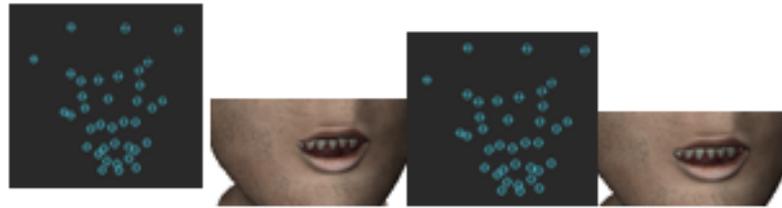
Kalimat yang di pisahkan pada lampiran adalah tu, bo, ne, ka



Kalimat yang di pisahkan pada lampiran adalah be, ru, di, to



Kalimat yang di pisahkan pada lampiran adalah ko,tu, lu, cu



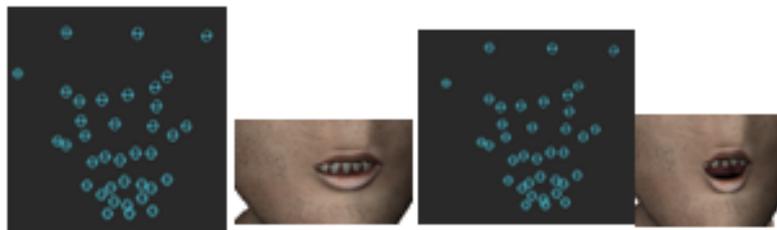
Ko awal

ko akhir



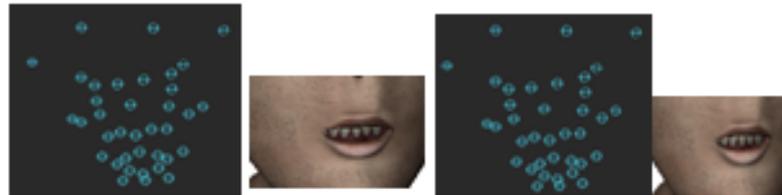
Tu awal

tu akhir



Lu awal

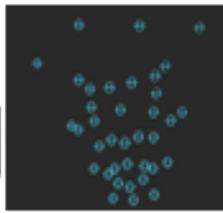
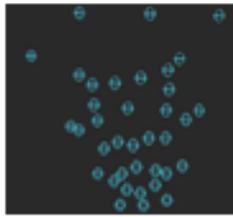
lu akhir



Cu awal

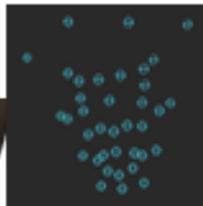
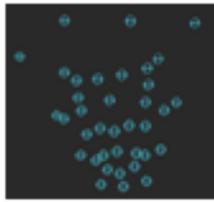
Cu akhir

Kalimat yang di pisahkan pada lampiran adalah se,ka, li, se



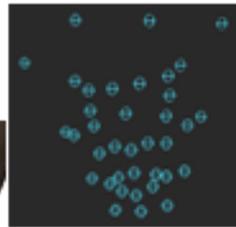
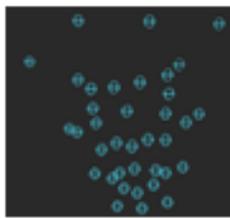
Se awal

se akhir



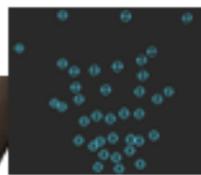
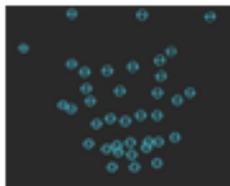
Ka awal

ka akhir



Li awal

Li akhir



Se awal

se akhir

DAFTAR RIWAYAT HIDUP



Nama : Nurul Fadillah
Tempat, tanggal lahir : Pulau Tiga, 1 Oktober 1989
Agama : Islam
Alamat :Jln. Arief rahman hakim,
keputih Gang 3C no 1
,sukolilo surabaya,
Dusun simpang tiga , desa
kaloy ,kecamatan tamiang
hulu kabupaten aceh
Tamiang, Nanggroe Aceh
Darussalam, Aceh
Email : nurulf479@gmail.com /
Fadillah_tite@yahoo.co.id
Facebook / Path : nurul fadillah
Twitter / Instagram : @nurultamy

Riwayat Pendidikan :

1. SD Negeri 4 Tamiang hulu, Aceh tamiang (1996 - 2003)
2. MTs Negeri 1 Tamiang hulu, Aceh tamiang (2003 - 2005)
3. SMA Negeri 1 Tamiang hulu, Aceh tamiang (2005 - 2006)
4. SMA Negeri 5 Bulu Blang Ara, Lhokseumawe (2006 - 2008)
5. D4 Teknik Informatika Politeknik Negeri Lhokseumawe (2008 - 2012)
6. S2 Teknik Elektro ITS Surabaya,
bidang keahlian Jaringan Cerdas Multimedia, angkatan 2013 - ganjil

Riwayat Pekerjaan :

1. Mengajar di Politeknik Negeri Lhokseumawe Nanggroe Aceh Darussalam (2012).
2. Mengajar di STMIK Bina Bangsa Cunda Lhokseumawe Nanggroe Aceh Darussalam (2012).
3. Mengajar di STIKES buket rata lhokseumawe Nanggroe Aceh Darussalam (2013 Febuari)