

TEXT TO AUDIOVISUAL MENGGUNAKAN METODE MORPHING VISEMES UNTUK PENGUCAPAN BAHASA INDONESIA BAGI PENUTUR ASING

Galih Muji Nugroho¹, Arifin²

^{1,2}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Jl. Nakula 5 – 11, Semarang, 50131, 024 – 3517261

E-mail : 111201105950@mhs.dinus.ac.id¹, arifin@dsn.dinus.ac.id²

Abstrak

Dewasa ini bahasa Indonesia dianggap penting dan menarik untuk dipelajari. Terbukti dari program BIPA (Bahasa Indonesia bagi Penutur Asing) yang tersebar di 22 negara dan meningkatnya wisatawan asing berkunjung ke Indonesia setiap tahun. Sehingga pemahaman terhadap bahasa Indonesia sangat dibutuhkan bagi warga asing. Berdasarkan latar belakang masalah tersebut, penulis membuat aplikasi text to audiovisual menggunakan metode morphing visemes untuk pengucapan bahasa Indonesia bagi penutur asing. Morphing visemes adalah perubahan bentuk viseme pertama hingga viseme terakhir. Diperlukan correspondence untuk mempertahankan bentuk objek. Forward warping, mengirim setiap pixel pada gambar pertama untuk menyesuaikan pixel pada gambar kedua. Cross-dissolve untuk mencampur dua gambar menjadi gambar peralihan. Selanjutnya proses sinkronisasi untuk menyesuaikan audio, teks, dan viseme. Hasilnya adalah aplikasi Text to Audiovisual Indonesia (TTAVI) dengan suara dan viseme Bahasa Indonesia. Dari pengujian Mean Opinion Score terhadap 30 responden untuk mendengarkan dan melihat kesesuaian viseme pada aplikasi TTAVI diperoleh rata-rata nilai MOS adalah 4,4196 dengan kategori nilai 1(buruk), 2(kurang), 3(cukup), 4(bagus), dan 5(sangat bagus) maka hasilnya adalah bagus. Dari penelitian ini penulis mengambil kesimpulan bahwa aplikasi TTAVI dapat membantu para penutur asing untuk belajar Bahasa Indonesia dengan lebih mudah dan interaktif.

Kata Kunci: fonem, viseme, text to audiovisual, morphing viseme

Abstract

Nowadays, Indonesian considered important and interesting to learn. Evident from the program of BIPA (Bahasa Indonesia bagi Penutur Asing) which is spread over 22 countries and the increasing foreign tourists to visit Indonesia every year. So the comprehension of Indonesian is necessary for foreigners. Based on the background of the problem, the author makes an application text to audiovisual using morphing visemes method for Indonesian speech toward foreign speakers. Morphing visemes is a change of the first viseme shape until the last viseme. correspondence is necessary to maintain the shape of the object. Forward warping, is sending each pixel in the first image to adjust pixel in the second image. Cross-dissolve for mixing two images into an image transition. Furthermore, the synchronization process to costumize audio, text, and viseme. The result is an application TTAVI (Text to Audiovisual Indonesia) with sound and Indonesian viseme. From the test of Mean Opinion Score against 30 respondents to listen and look at the viseme suitability of TTAVI application obtained an average value of MOS is 4.4196 with the value categories of 1(poor), 2(less), 3(enough), 4(good), and 5(very good) then the result is good. From this study, the author concludes that TTAVI application can help foreign speakers to learn Indonesian with easier and interactive.

Keyword: phoneme, viseme, text to audiovisual, morphing viseme

1. PENDAHULUAN

Indonesia merupakan negara dengan sumber daya alam dan sumber daya manusia yang melimpah. Selain itu Indonesia memiliki kekayaan budaya tersendiri, salah satunya adalah bahasa. Hal tersebut menjadi salah satu faktor bagi negara lain untuk mempelajari Indonesia terutama dalam berbahasa. Dewasa ini antusiasme bangsa asing memberikan gambaran bahwa bahasa Indonesia dianggap penting dan menarik untuk dipelajari. Terbukti adanya program Bahasa Indonesia bagi Penutur Asing (BIPA) yang tersebar di 22 negara. Menjadikan kemungkinan suatu saat bahasa Indonesia bisa menjadi bahasa Internasional kedua setelah bahasa Inggris. Mohammad Nuh mengatakan, “Bahasa Indonesia tidak hanya sekedar digunakan sebagai bahasa lokal, tetapi untuk berinteraksi dengan bahasa dunia [1].”

Pada saat ini kunjungan wisatawan asing oleh Badan Pusat Statistik (BPS) mengalami peningkatan setiap tahunnya, pada Agustus 2014 telah mencapai 826,8 ribu kunjungan. Berdasarkan data tersebut, minat wisatawan asing untuk berkunjung ke Indonesia cukup tinggi. Oleh karena itu pemahaman terhadap bahasa Indonesia sangat dibutuhkan bagi warga asing, salah satu contohnya dengan belajar pengucapan bahasa Indonesia. Namun di Indonesia masih banyak ditemukan kesalahan pelafalan fonem seperti penggunaan fonem /e/ pada pelafalan /beri/ yang bermakna “menyerahkan” dan /ber/ yang bermakna “buah”. Apabila kata tersebut diucapkan tidak sesuai yang dimaksudkan, maka akan membuat salah paham. fonem adalah bunyi bahasa yang minimal yang membedakan bentuk dan makna kata [2].

Dari penelitian sebelumnya Indonesia memiliki 33 set pelafalan fonem yang terdiri dari konsonan, vokal, dan *natural* kemudian di kelompokkan menjadi sepuluh viseme [3]. viseme atau *visual phoneme* merupakan penggambaran dari sekumpulan fonem dengan pengucapan melalui gerak mulut yang hampir sama. Pada penerapannya, *visual* menggunakan algoritma morphing yaitu perubahan gambar dari suatu objek ke objek lain. Dengan *morphing* bersama korespondensi ini, transisi yang halus antara gambar viseme dapat dihasilkan [4]. Morphing artinya mengubah satu bentuk menjadi bentuk lain dengan menampilkan serangkaian frame yang menciptakan gerakan halus begitu bentuk pertama mengubah dirinya menjadi bentuk lain [5]. Banyak perangkat lunak yang menggunakan algoritma *morphing* dalam *text to audiovisual* terutama dalam bahasa Inggris, namun versi bahasa Indonesia masih jarang.

Dengan demikian, dalam rangka membantu para penutur asing untuk belajar Bahasa Indonesia dengan lebih mudah dan interaktif diperlukan media pembelajaran dengan menggunakan alat bantu aplikasi. Sehingga perlu dibuat *software text to audiovisual* menggunakan metode *morphing visemes* untuk pengucapan bahasa Indonesia bagi penutur asing.

2. METODE

2.1 Instrumen Penelitian

2.1.1 Software

Dalam melakukan penelitian ini dibutuhkan beberapa *software* sebagai alat untuk membantu berlangsungnya proses penelitian yaitu sistem operasi *Windows* edisi *Windows 8 Pro With*

Media Center 32-bit. CorelDRAW X6 versi 16.0.0.707 sebagai aplikasi untuk membuat animasi wajah dan *Software Audacity* 2.0.6 sebagai pengolah suara. Kemudian digunakan aplikasi *wxFormBuilder* versi 3.5.0-beta yaitu sebuah *software* untuk mengembangkan atau membuat *software* baru. Bahasa pemrograman yang digunakan adalah *python* versi 2.7.8.

2.1.2 Hardware

Selain menggunakan *software* tentu dibutuhkan *hardware* sebagai wadah *software* tersebut untuk dapat dioperasikan. Dalam penelitian ini menggunakan *hardware* berupa *microphone Bear Mountain Audio (BMA)* versi SR-370 untuk pengambilan suara dan laptop BenQ versi joybookR46 dengan spesifikasi sebagai berikut :
 Prosesor : Pentium(R) Dual-Core CPU T4200 @ 2.00 GHz 2.00 GHz.
 RAM : 1 GB + 2 GB, DDR II.
 GPU : Mobile Intel(R) 4 Series Express Chipset.

2.2 Prosedur Pengambilan Data

Data pada penelitian ini menggunakan data kualitatif yaitu data yang berbentuk kata, bukan dalam bentuk angka. Dalam pengambilan sumber data didapatkan data primer dan sekunder. Data primer diperoleh dari sumber langsung berupa rekaman suara berupa kata kemudian dipotong sehingga menjadi fonem bahasa Indonesia. Sedangkan data sekunder didapat dari dokumen berupa jurnal, informasi pada internet, media massa, buku, dan laporan karya tulis ilmiah yang berhubungan tentang *text to audiovisual*. Salah satu contoh data sekunder yang didapatkan adalah kumpulan fonem dan kelas viseme bahasa Indonesia dari [3].

2.2.1 Audio

Salah satu data yang digunakan dalam penelitian ini adalah berkas *audio* dimana berisi suara yang telah direkam. Suara pada penelitian ini diambil atau direkam menggunakan *microphone Bear Mountain Audio (BMA)* yang dihubungkan ke *port microphone* pada laptop BenQ joybookR46. Pengambilan suara dibantu dengan perangkat lunak *audacity* versi 2.0.6.

Proses perekaman dilakukan seperti yang sudah dijelaskan menghasilkan 32 berkas *audio* dengan ekstensi *.wav*. Berkas tersebut yang akan digunakan sebagai kumpulan fonem yang akan dikelompokkan berdasarkan viseme Indonesia. selain itu juga sebagai bahan untuk mengeluarkan suara dalam percobaan dan pengujian pada perangkat lunak *text to audiovisual* Indonesia (TTAVI).

No	File	Ukuran(KB)	No	File	Ukuran(KB)
1	a.wav	26.3	17	t.wav	17.5
2	h.wav	17	18	ai.wav	42.2
3	b.wav	31.7	19	E.wav	17.7
4	m.wav	18	20	oi.wav	60.2
5	p.wav	15.7	21	y.wav	17.7
6	c.wav	18	22	f.wav	17.2
7	i.wav	26	23	v.wav	18.2
8	j.wav	21	24	g.wav	17
9	ny.wav	25.7	25	k.wav	17.5
10	s.wav	27.2	26	kh.wav	27
11	sy.wav	34.5	27	e1.wav	25.5
12	z.wav	26.2	28	ng.wav	34.5
13	d.wav	18	29	au.wav	34.5
14	l.wav	18	30	o.wav	16.5
15	n.wav	18.5	31	u.wav	28
16	r.wav	17.5	32	w.wav	26

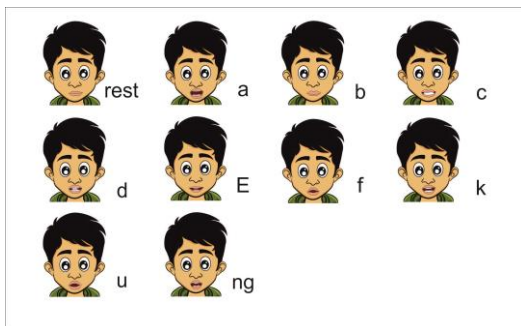
Gambar 1. Data Suara Fonem

2.2.2 Viseme

Dalam data gambar didapatkan dengan membuat karakter wajah menggunakan perangkat lunak *CorelDRAW X6*. Pembuatan bentuk mulut didasarkan pada penelitian berjudul “*Towards Building Indonesian Viseme: A Clustering-Based Approach*”, pada penelitian tersebut membahas bahwa

viseme Indonesia didapatkan dari beberapa kumpulan fonem Indonesia yang terdiri atas 33 simbol fonemis yang mana berisi sepuluh vokal termasuk diftong, 22 konsonan dan satu diam [3]. Kemudian dari acuan jurnal tersebut penulis membuat model karakter bentuk mulut sesuai [3].

Proses pembuatan karakter dilakukan seperti membuat karakter pada umumnya hingga membentuk wajah. Namun pembuatan dilakukan dengan satu acuan wajah yang telah jadi, yaitu wajah dengan bentuk mulut *rest* dan bentuk mulut lain dibedakan menyesuaikan viseme Indonesia sehingga tidak perlu membuat bentuk wajah dari awal. Kemudian *export* menjadi berkas gambar berekstensi .jpg dan atur dengan pixel 300*300, hasilnya ditunjukkan pada gambar berikut:



Gambar 2. Viseme Indonesia

2.3 Teknik Analisis Data

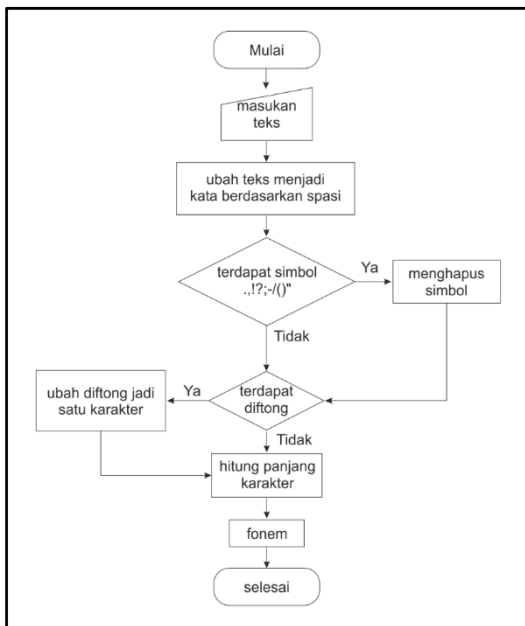
Pada data yang berupa suara atau *audio* dilakukan rekaman suara berupa kata kemudian dipotong menjadi fonem dan disimpan pada komputer dengan format WAV (*waveform*). Data yang ada digunakan untuk memberikan efek pada masukan kata, sehingga dihasilkan keluaran suara yang sesuai dengan fonem pada pengucapan kata. Langkah tersebut dilakukan dengan menggunakan *hardware* berupa laptop dan *microphone*, sedangkan *software* yang digunakan adalah *audacity*.

Data diperoleh berupa fonem bahasa Indonesia yang terdiri dari kelompok fonem konsonan berjumlah 23, fonem vokal berjumlah sembilan, dan fonem diam berjumlah satu. Fonem tersebut akan menjadi dasar pembentukan kata yang memiliki sifat atau kegunaan untuk membedakan makna apabila terdapat kata yang sama. Dari kumpulan fonem tersebut didapat viseme dengan beberapa kelas. Setiap kelas memiliki kumpulan fonem yang pengucapannya hampir sama dan didapatkan viseme dari masing-masing kelas yaitu *rest*, a, b, d, u, k, c, E, f, dan ng. Kelas viseme tersebut akan menjadi acuan gambar bentuk mulut yang dapat diterapkan pada animasi sehingga apabila suatu karakter mengucapkan kata, maka kata yang berupa fonem tersebut dapat diwakilkan pada kelas viseme.

2.4 Algoritma Split Text

Split text yaitu proses dimana suatu masukan berupa teks atau kalimat yang dipecah menjadi potongan kata dan fonem. Hal ini untuk mempermudah menggolongkan fonem kedalam kelas viseme. Pada proses *split* teks sebagai contoh, apabila suatu kalimat “sampai jumpa” dilakukan *split* menjadi “sampai” “jumpa”, sedangkan pada fonem menjadi “/s//a//m//p//a//i//j//u//m//p//a/”. Ada beberapa kondisi dimana inputan kata terdapat diftong didalamnya. Apabila terdapat diftong pada suatu kata seperti /au/, /oi/, /ai/, /kh/, /sy/, /ny/, /ng/ harus diberikan kondisi agar tidak terbaca sebagai monoftong. Contohnya kata “sampai” jika diucapkan dengan fonem monoftong menjadi /s//a//m//p//a//i/, untuk menghindari hal tersebut dibuat kondisi berupa konversi fonem diftong seperti pada kelas-kelas viseme sehingga fonem /s//a//m//p//ai/ menjadi “cabbE”.

Proses *split text* dimulai dengan memasukan teks pada *keyboard* yang akan dibaca tiap karakter apakah terdapat spasi. Kemudian dilakukan *split* teks berdasarkan spasi untuk mengubahnya menjadi kata. Setelah itu dilakukan pengecekan apakah terdapat simbol, jika iya maka simbol akan dihapus untuk memudahkan pembacaan karakter. Pengecekan dilakukan lagi apakah terdapat diftong, jika iya maka dua karakter akan dijadikan satu karakter, misalnya karakter ‘n’ yang diikuti ‘g’ akan diubah menjadi ‘ng’. Lalu menghitung panjang karakter untuk mengetahui urutan karakter yang akan diambil menjadi tiap-tiap fonem.



Gambar 3. Flowchart Split Text

2.5 Metode Morphing Viseme

Dari pengertian *morphing* dan viseme yang telah dijelaskan dapat diambil kesimpulan bahwa *morphing visemes* adalah perubahan bentuk dari viseme satu menuju viseme kedua dan seterusnya hingga mencapai bentuk viseme terakhir. Penggunaan metode *morphing visemes* telah banyak digunakan pada animasi gerakan mulut

yang membutuhkan kesesuaian dengan pengucapan fonem.

2.5.1 Correspondence

Correspondence memiliki sinonim dengan *match*, *equivalence*, *congruity*, dan sebagainya yang berarti kesesuaian atau kecocokan. Langkah pertama, metode *morphing* memerlukan spesifikasi *correspondence maps* atau peta kesesuaian dengan $C_0 : I_0 \Rightarrow I_1$ dan $C_1 : I_1 \Rightarrow I_0$ berhubungan dengan gambar I_0 dan I_1 . Peta ini berfungsi untuk memastikan bahwa selanjutnya mempertahankan kesesuaian proses *warping* yang diinginkan antara sifat geometris objek yang akan berubah.

$$C_0(P_0) = \{d_x^{0 \rightarrow 1}(P_0), d_y^{0 \rightarrow 1}(P_0)\} \quad (1)$$

$$C_1(P_1) = \{d_x^{1 \rightarrow 0}(P_1), d_y^{1 \rightarrow 0}(P_1)\} \quad (2)$$

Pixel dalam gambar I_0 pada posisi $P_0 = (x, y)$ menyesuaikan menuju pixel dalam gambar I_1 pada posisi $(x + d_x^{0 \rightarrow 1}(x, y), y + d_y^{0 \rightarrow 1}(x, y))$. Begitu pula pixel dalam gambar I_1 pada posisi $P_1 = (x, y)$ menyesuaikan menuju pixel dalam gambar I_0 pada posisi $(x + d_x^{1 \rightarrow 0}(x, y), y + d_y^{1 \rightarrow 0}(x, y))$.

2.5.2 Forward Warping

Langkah selanjutnya adalah *forward warping* atau melengkungkan kedepan dengan diberikan gambar I_0 dan I_1 , dengan menghitung kesesuaian $C_0(P_0) = \{d_x^{0 \rightarrow 1}(P_0), d_y^{0 \rightarrow 1}(P_0)\}$. skala aliran vektor yang seragam dengan parameter α antara 0 dan 1, dapat menghasilkan serangkaian gambar peralihan yang dilengkungkan dengan $I_0^{warped}(\alpha)$ yang mendekati perubahan antara I_0 dan I_1 .

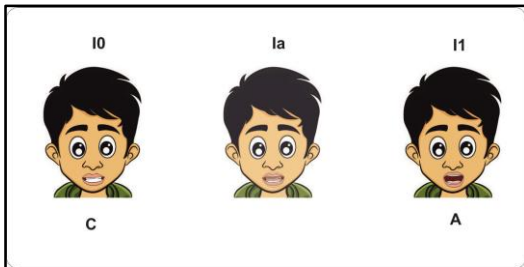
$$W_0(P_0, \alpha) = P_0 + \alpha C_0(P_0) \quad (3)$$

$$I_0^{warped}(W_0(P_0, \alpha)) = I_0(P_0) \quad (4)$$

Dimana pada persamaan (3) W_0 merupakan *forward warping*, P_0 adalah pixel pada gambar I_0 dengan posisi $P_0 = (x, y)$, α parameter berkisar 0 sampai 1, C_0 merupakan *correspondence maps* dan pada persamaan (4) I_0^{warped} sebagai gambar melengkung.

2.5.3 Morphing Visemes

Diberikan dua buah gambar I_0 pada gambar paling kiri dan I_1 pada gambar paling kanan, hasil gambar peralihan I_α pada gambar tengah, dimana α adalah parameter berkisar dari 0 sampai 1. Gambar peralihan dihasilkan oleh *warping* I_0 terhadap I_1 , *warping* I_1 terhadap I_0 , dan *cross-dissolving* gambar yang dilengkungkan untuk menghasilkan gambar akhir yang diinginkan.



Gambar 4. *Morphing* dengan *cross-dissolve*, gambar paling kiri adalah I_0 , paling kanan I_1 , dan ditengah merupakan I_α

Karena pada proses *forward warping* hanya bisa memindahkan pixel disekitarnya, hal itu tidak dapat memodelkan munculnya tekstur pixel yang baru. *Forward warp* dalam viseme I_0 disepanjang aliran vektor pada C_0 bisa tidak menghasilkan gambar akhir seperti yang terlihat pada viseme I_1 , karena viseme I_1 sendiri mengandung sejumlah besar tekstur dari dalam mulut.

Morphing mengatasi tekstur pixel baru ini dengan menggabungkan tekstur yang ditemukan di kedua *forward warp*. Kombinasi ini dilakukan dengan *scaling* atau meningkatkan lengkungan gambar peralihan dengan *cross-dissolve* atau parameter *blending* (perpaduan), dan kemudian menambahkan hasil akhir perubahan gambar $I^{morph}(\alpha)$.

$$I^{morph}(P, \alpha) =$$

$$(1 - \alpha)I_0^{warped}(P, \alpha) + \alpha I_1^{warped}(P, (1 - \alpha)) \quad (5)$$

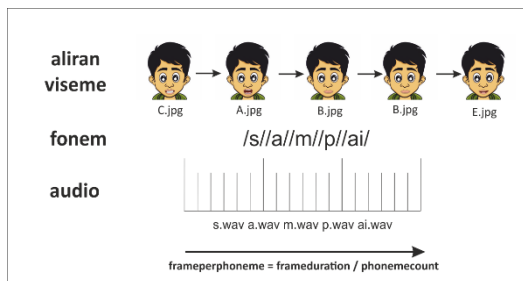
Pada persamaan (5), $I^{morph}(P, \alpha)$ merupakan hasil perubahan (*morph*) dimana P adalah pixel dan α adalah parameter 0 sampai 1, I_0^{warped} merupakan gambar pertama yang akan dilengkungkan dan I_1^{warped} merupakan gambar tujuan yang dilengkungkan. Dengan interpolasi (penyisipan) parameter perpaduan perubahan, “memudar keluar” versi pelengkungan dari viseme awal dan “memudar kedalam” versi pelengkungan dari viseme akhir. Proses pencampuran tersebut memungkinkan dua lengkungan secara efektif digabungkan dan pixel baru dari viseme kedua yang terlibat mentransisikan dirinya sendiri [4].

2.6 Proses Sinkronisasi

Setelah dilakukan proses *split text* untuk menentukan fonem dari suatu kata yang diucapkan, selanjutnya adalah perangkaian fonem dengan suara. Untuk membangun aliran visual dimana kata atau kalimat diucapkan, kita perlu mensinkronisasi gambar viseme dengan kelas viseme secara bersamaan dimana teks sudah dinyatakan sebagai fonem. Untuk melengkapi viseme diperlukan *audio* dan teks sebagai masukan dalam perangkat lunak. Proses yang dilakukan

dalam *audio* yaitu perekaman suara fonem sampai disimpan dalam bentuk *.wav*, sedangkan masukan teks akan diubah menjadi kata dan fonem dengan *split text*. Setelah itu masuk dalam tahap sinkronisasi atau penyelarasan, pada tahap ini akan disamakan antara *audio*, teks, dan viseme.

Proses sinkronisasi bekerja dengan cara menyambungkan tiap-tiap data sehingga data tersebut akan terkait dan ketika dijalankan akan bekerja secara bersama-sama. Pertama adalah memasukkan data berupa teks dengan mengetik pada *keyboard*, melalui proses *split text* dimana data berupa fonem diperoleh. Fonem yang telah dikonversi menjadi kumpulan viseme Indonesia seperti yang telah dijelaskan diatas dihubungkan dengan direktori yang berisi kumpulan gambar viseme yang telah diberi nama viseme sesuai bentuk mulut. Sehingga kumpulan konversi viseme Indonesia akan diwakilkan oleh kumpulan gambar viseme Indonesia.



Gambar 5. Proses Sinkronisasi

Gambar 5 diatas adalah proses sinkronisasi dimana fonem dihubungkan dengan viseme Indonesia dan *audio* sehingga menciptakan aliran viseme yang sesuai dengan fonem dan suara. Sebagai contoh proses sinkronisasi adalah pernyataan berupa kalimat “sampai jumpa” yang terdiri dari fonem /s//a//m//p//ai//j//u//m//p//a/, sedangkan pada kelas viseme Indonesia menjadi bentuk ‘rest’ ‘c’ ‘a’ ‘b’ ‘b’ ‘E’ ‘rest’ ‘c’ ‘u’ ‘b’ ‘b’ ‘a’ ‘rest’. Bentuk kelas viseme

tersebut disinkronkan dengan gambar viseme Indonesia untuk membentuk rangkaian viseme dari viseme pertama hingga terakhir.

3. HASIL DAN PEMBAHASAN

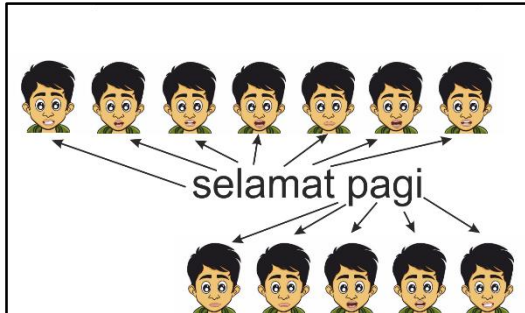
3.1 Hasil Percobaan

Percobaan pada perangkat lunak TTA VI dilakukan untuk mengetahui apakah proses memasukkan data seperti suara dan teks serta pergerakan gambar viseme bisa berjalan sesuai dengan yang diharapkan. Percobaan dilakukan dengan memasukkan atau mengetik pada *keyboard*, misalnya kata atau kalimat. Tekan *play* untuk memulai proses sinkronisasi sehingga pergerakan gambar viseme dapat terlihat. Sebagai contoh adalah kalimat “selamat pagi.” dengan viseme yang akan disesuaikan sesuai fonem dari kalimat tersebut. Hasil keluaran adalah suara “selamat pagi” dengan viseme “CNGDABAD BACK”. Setelah melalui proses percobaan memasukkan data seperti yang telah dijelaskan, maka didapatkan hasil dari percobaan tersebut dengan sampel audio sebanyak 10 berkas dapat dilihat pada tabel berikut:

Tabel 1: Hasil Percobaan

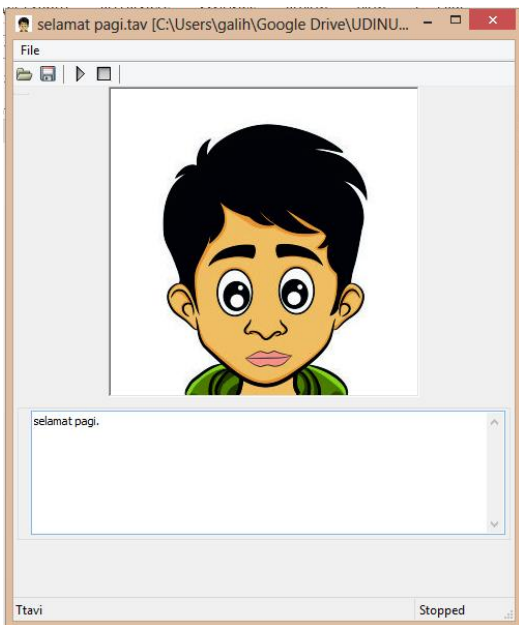
No	Suara	Viseme
1	akhir pekan	AKCD BNGKAD
2	apa kabar	ABA KABAD
3	bahasa indonesia	BAAACA CDDUDNGCCA
4	sampai jumpa	CABBE CUBBA
5	sekarang hari apa	CNGKADANG AADC ABA
6	sekarang hari minggu	CNGKADANG AADC BCNGKU
7	selamat datang	CNGDABAD DADANG
8	selamat pagi	CNGDABAD BACK
9	terima kasih	DNGDCBA KACCA
10	warna biru	UADDA BCDU

Gambar 6 merupakan sinkronisasi gambar viseme dari awal hingga akhir membentuk aliran viseme “selamat pagi”.



Gambar 6. Sinkronisasi Gambar Viseme

Hasil percobaan yang telah dilakukan menggunakan aplikasi TTAVI yang telah dibuat seperti terlihat pada gambar 7.



Gambar 7. Aplikasi TTAVI

3.2 Pengujian Mean Opinion Score

Mean Opinion Score atau disingkat *MOS* merupakan metode yang digunakan untuk mengukur seberapa tinggi kesesuaian pengucapan pada perangkat lunak TTAVI terhadap pemahaman pengguna. Penilaian *MOS* didapatkan dengan melakukan survei pada pengguna TTAVI melalui kesesuaian

gambar viseme terhadap suara yang didengarkan. Kemudian pengguna akan mengisikan penilaian pada tabel kesesuaian viseme dengan 5 kriteria kesesuaian, yaitu:

1. Buruk: suara dan pergerakan mulut tidak sesuai.
2. Kurang: suara dan pergerakan mulut kurang sesuai.
3. Cukup: suara dan pergerakan mulut cukup sesuai.
4. Bagus: suara dan pergerakan mulut sesuai.
5. Sangat bagus: suara dan pergerakan mulut sangat sesuai.

No	Suara	Kesesuaian viseme				
		Buruk (1)	Kurang (2)	Cukup (3)	Bagus (4)	Sangat bagus (5)
1.	akhir pekan	-	-	2	17	11
2.	apa kabar	-	2	6	6	16
3.	bahasa indonesia	-	1	4	8	17
4.	sampai jumpa	-	-	3	11	16
5.	sekarang hari apa	-	1	4	5	20
6.	sekarang hari minggu	-	-	2	11	17
7.	selamat datang	-	-	4	8	18
8.	selamat pagi	-	-	2	6	22
9.	terima kasih	-	3	4	5	18
10.	warna biru	-	-	1	12	17

Gambar 8. Kesesuaian Viseme

$$MOS = \sum_{i=1}^n \frac{x(i).k}{N} \quad (6)$$

Dimana:

- $x(i)$ = nilai sampel ke i
- k = jumlah bobot
- N = jumlah responden

Hasil yang diperoleh dari survei pengamatan kesesuaian viseme oleh 30 responden dengan menggunakan persamaan 6 diperoleh *Mean Opinion Score* yang ditunjukkan seperti pada tabel 3.

Tabel 2: Mean Opinion Score

No	Suara	MOS
1.	akhir pekan	4,3
2.	apa kabar	4,2
3.	bahasa indonesia	4,366
4.	sampai jumpa	4,433

5.	sekarang hari apa	4,466
6.	sekarang hari minggu	4,5
7.	selamat datang	4,466
8.	selamat pagi	4,666
9.	terima kasih	4,266
10.	warna biru	4,533
	Jumlah	44,196
	Rata-rata	4,4196

Dari perolehan nilai rata-rata *MOS* tersebut, apabila ditarik nilai antara 1 sampai 5 dengan kategori nilai 1(buruk), 2(kurang), 3(cukup), 4(bagus), dan 5(sangat bagus) maka masuk dalam kategori bagus. Yang mana dapat diambil kesimpulan bahwa perangkat lunak *Text to Audiovisual* Indonesia (TTAVI) dapat membantu para penutur asing untuk belajar Bahasa Indonesia dengan lebih mudah dan interaktif.

4. KESIMPULAN DAN SARAN

Dari proses penelitian dan pembuatan perangkat lunak *Text to Audiovisual* Indonesia (TTAVI), dapat disimpulkan bahwa pengujian kebutuhan fungsional pada perangkat lunak dapat melakukan masukan berupa teks dan suara. Dari hasil percobaan yang telah dilakukan, proses sinkronisasi antara teks, suara, dan gambar viseme menghasilkan pergerakan gambar yang sesuai pengucapan suara. Berdasarkan *Mean Opinion Score*, kesesuaian viseme terhadap suara dengan rentang nilai dari 1 sampai 5 dikategorikan bagus dengan nilai rata-rata 4,4196. Sehingga dapat membantu pengguna dalam pemahaman terhadap pengucapan bahasa Indonesia. Kedepan perlu pengembangan lebih lanjut pada proses perekaman di tempat yang sunyi atau kedap suara untuk mengurangi gangguan rekaman. Menggunakan suku kata dan pola vokal konsonan, sehingga proses pengucapan oleh gambar viseme lebih jelas.

5. DAFTAR PUSTAKA

- [1] kompas.com, 2013. *Mendikbud: Perkuat Bahasa Indonesia di Percaturan Internasioanl!* [Online] (Update 28 Okt 2013) URL: http://edukasi.kompas.com/read/2013/10/28/1716436/Mendikbud.Perkuat.Bahasa.Indonesia.di.Percaturan.Internasional.?utm_campaign=related&utm_medium=bp-kompas&utm_source=news& [Diakses pada 25 Oktober 2014].
- [2] Alwi, H., Dardjowidjojo, S., Lapoliwa, H., dan Moeliono, A.M., 2003. *Tata Bahasa Baku Bahasa Indonesia*. 3rd ed. Jakarta: Balai Pustaka.
- [3] Arifin, Muljono, Surya Sumpeno dan Mochamad Hariadi, 2013. 2013 IEEE International Conference. In: Computational Intelligence and Cybernetics (CYBERNETICSCOM), *Towards Building Indonesian Viseme: A Clustering-Based Approach*. Yogyakarta, Desember 2013, IEEE: Yogyakarta.
- [4] Ezzat, T., and Poggio, T., 2000. International Journal of Computer Vision. *Visual Speech Synthesis by Morphing Visemes*, 38 (1), pp.45-57.
- [5] Suyanto, M., 2003. *MULTIMEDIA Alat untuk Meningkatkan Keunggulan Bersaing*. 1st ed. Yogyakarta: ANDI.