
Identifikasi Viseme Untuk Fonem Bahasa Madura Berbasis *Clustering* Berdasarkan *Facial Landmark Point*

Pyepit Rinekso Andriyanto^{1*}, Joan Santoso², Endang Setyati³

^{1,2,3} Institut Sains dan Teknologi Terpadu Surabaya, Fakultas Sains dan Teknologi, Magister Teknologi Informasi, Jl. Ngagel Jaya Tengah No. 73-77 Baratajaya - Surabaya, Indonesia

Informasi Artikel

Diterima: 12-05-2023

Direvisi: 30-05-2023

Diterbitkan: 30-06-2023

Kata Kunci

Viseme; Clustering; Fonem; Bahasa Madura

*Email Korespondensi:

pyepitrineksoa@gmail.com

Abstrak

Bentuk bahasa yang paling efektif dalam berkomunikasi adalah bahasa lisan atau bicara. Ketika berbicara manusia akan menggerakkan mulut dan bibirnya untuk mengucapkan kata tertentu. Model gerakan mulut dan bibir ini menggambarkan suatu viseme (*visual-phonem*), yaitu sekelompok fonem yang memiliki visual atau tampilan yang hampir sama. Bahasa Madura merupakan bahasa yang unik dan memiliki beberapa ciri tertentu. Selain mempunyai tingkat bahasa, bahasa Madura mempunyai fonem-fonem yang beraspirasi atau pengucapan kata dengan cara dihembuskan seperti: /bh/, /dh/, /Dh/, /gh/ dan /jh/ yang tidak ada pada bahasa lainnya. Penelitian ini membahas tentang identifikasi kelas-kelas viseme untuk fonem bahasa Madura berbasis *clustering* berdasarkan *facial landmark point*. Dari 47 fonem bahasa Madura diperoleh 9 viseme bahasa Madura yang dihasilkan dari proses *K-Means clustering*. Proses *clustering* menggunakan ekstraksi fitur berdasarkan *facial landmark point* sehingga diperoleh perhitungan jarak pada setiap fitur. Fitur-fitur yang digunakan adalah fitur geometri. Model viseme bahasa Madura digunakan untuk membangun animasi mulut 2D dalam mengucapkan kata atau kalimat bahasa Madura berdasarkan inputan berupa teks. Manfaat dari penelitian ini adalah untuk tujuan pembelajaran dalam pengucapan kata atau kalimat bahasa Madura, karena bahasa Madura memiliki tulisan dan cara pengucapan yang berbeda.

Abstract

The most effective form of language in communicating is spoken or spoken language. When speaking humans will move their mouth and lips to say certain words. This mouth and lip movement model describes a viseme (visual-phonem), namely a group of phonemes that have a visual or almost the same appearance. Madurese language is a unique language and has certain characteristics. In addition to having a language level, Madurese has aspirated phonemes or exhaled word pronunciation such as: /bh/, /dh/, /Dh/, /gh/ and /jh/ which do not exist in other languages. This research discusses the identification of viseme classes for Madurese phonemes based on clustering based on facial landmark points. Of the 47 Madurese language phonemes, 9 Madurese language visemes were obtained from the K-Means clustering process. The clustering process uses feature extraction based on facial landmark points so that the distance calculation for each feature is obtained. The features used are geometric features. The Madurese viseme model is used to build 2D mouth animations in uttering Madurese words or sentences based on input in the form of text. The benefit of this research is for learning purposes in pronouncing Madurese words or sentences, because Madurese has different writing and pronunciation.

1. Pendahuluan

Bentuk bahasa yang paling efektif dalam berkomunikasi adalah bahasa lisan atau bicara. Ketika berbicara manusia akan menggerakkan mulut dan bibirnya untuk mengucapkan kata tertentu. Model gerakan mulut dan bibir ini menggambarkan suatu viseme (*visual-phonem*). Viseme merupakan sekelompok fonem yang memiliki tampilan visual yang hampir sama (Bear & Harvey, 2016; Bozkurt et al, 2007). Fonem adalah satuan bunyi terkecil dari sebuah bahasa yang berfungsi untuk membedakan makna kata. Fonem juga merupakan unit terkecil dari suara yang berfungsi untuk membangun ucapan manusia. Fonem dan viseme memiliki korelasi yang tinggi (Cappeletta & Harte, 2012). Viseme dapat diturunkan menggunakan pemetaan fonem ke viseme, dan pemetaan tersebut dilakukan dari banyak ke satu (*many-to-one*), karena banyak fonem yang tidak dapat dibedakan hanya dengan menggunakan isyarat visual (Setyati et al, 2015).

Penelitian tentang pemetaan fonem ke viseme bahasa Indonesia telah pernah diteliti sebelumnya dengan menggunakan pendekatan linguistik (Setyati et al, 2015). dan dengan pendekatan *data driven* (Arifin, 2013). Selain bahasa Indonesia pemetaan fonem ke viseme bahasa Malaysia juga pernah diteliti menggunakan pendekatan linguistik dan *data driven* (Kumar et al, 2019), bahasa Marathi dengan pendekatan linguistik (Brahme & Bhadade, 2017), bahasa Arab dengan pendekatan linguistik (Damien et al, 2009) dan bahasa Jerman dengan pendekatan linguistik (Bianca & Christian, 2005). Tentu penelitian tersebut menjadi tantangan untuk bahasa lainnya, salah satunya adalah bahasa Madura. Hingga saat ini belum ada penelitian yang membahas tentang identifikasi kelas viseme untuk fonem bahasa Madura.

Bahasa Madura merupakan bahasa yang unik dan memiliki beberapa ciri tertentu. Selain mempunyai tingkat bahasa yaitu tingkat bahasa umum (*iyâ- ənjà'*), tingkat bahasa menengah (*əngghi- əntən*), dan tingkat bahasa tinggi atau halus (*əngghi-bhuntən*), bahasa Madura mempunyai fonem-fonem yang beraspirasi atau pengucapan kata dengan cara dihembuskan seperti: /bh/, /dh/, /Dh/, /gh/ dan /jh/ yang tidak ada pada bahasa lainnya (Pawitra, 2009). Bahasa Madura mengenal 6 huruf vokal yaitu: a, i, u, ə, è, ò dan 32 huruf konsonan.

Fonem merupakan satuan terkecil dari bahasa yang dapat membedakan suatu arti atau makna (Gleason, 1962). Fonem tidak akan mengandung sebuah makna apabila berdiri sendiri. Seperti pada fonem /b/ dan /d/. Keduanya tidak akan memiliki makna jika hanya berdiri sendiri. Namun apabila digabungkan dengan fonem lainnya seperti /a/ dan /r/ maka keduanya dapat membentuk makna "bârâ" (bengkak) dan "dârâ" (darah). Contoh lainnya yaitu "rowa" (itu) "towa" (tua), "powa" (lunak), "lowa" (luap) jika salah satu unsur diganti dengan unsur lainnya maka akan dapat merubah maknanya. Setelah menganalisa dan mengamati beberapa literatur dalam penggunaan bahasa Madura secara umum, fonem bahasa Madura terbagi menjadi fonem vokal dan fonem konsonan, baik dalam bentuk tunggal maupun ganda. Fonem vokal dalam bahasa Madura terdiri dari 6 abjad yaitu: /a/, /i/, /u/, /è/, /ə/, /ò/. Namun suara yang dihasilkan adalah 11 suara yang berbeda. Lima lainnya adalah alofon : "â" (/a/), "I" (/i/), "U" (/u/), "e" (/è/), "o" (/ò/). Selain itu bahasa Madura juga memiliki 3 vokal ganda (diftong), yaitu: /ay/, /uy/, /òy/ dan 1 alofon diftong /ây/. Sehingga jumlah fonem vokal bahasa Madura terdapat 15 buah. Bahasa Madura memiliki 32 fonem konsonan yang terdiri dari 24 konsonan tunggal yaitu: /b/, /c/, /d/, /D/, /f/, /g/, /h/, /j/, /k/, /l/, /m/, /n/, /p/, /q/, /r/, /s/, /t/, /T/, /v/, /w/, /x/, /y/, /z/, /' / dan 8 konsonan ganda yaitu: /bh/, /dh/, /Dh/, /jh/, /gh/, /ng/, /ny/, /sy/.

Viseme merupakan representasi visual dari fonetik wicara (Gleason, 1962). Definisi lain dari viseme merupakan sekelompok fonem yang memiliki tampilan visual yang hampir sama (Bear & Harvey, 2016; Bozkurt et al, 2007). Viseme dan fonem dapat dikorelasikan melalui pemetaan fonem ke viseme sehingga apabila keduanya disinkronkan maka dapat menghasilkan animasi berbicara.

Facial landmark adalah sebuah metode dalam proses deteksi wajah yang bekerja dengan cara mengalokasikan titik-titik pada wajah untuk menentukan bentuk biologis dari wajah manusia (Han, 2019). Metode tersebut dapat mengenali dan mendeteksi bagian-bagian dari wajah seperti bagian mata, alis, mulut dan hidung. Saat ini banyak penelitian yang memanfaatkan metode ini pada proses pengenalan wajah. Metode *facial landmark* juga dapat bekerja secara *real-time* yang dapat memprediksi *landmark* wajah dengan

tingkat akurasi yang tinggi. Pada penelitian ini metode *facial landmark* akan digunakan untuk mendeteksi bagian mulut atau bibir saja.

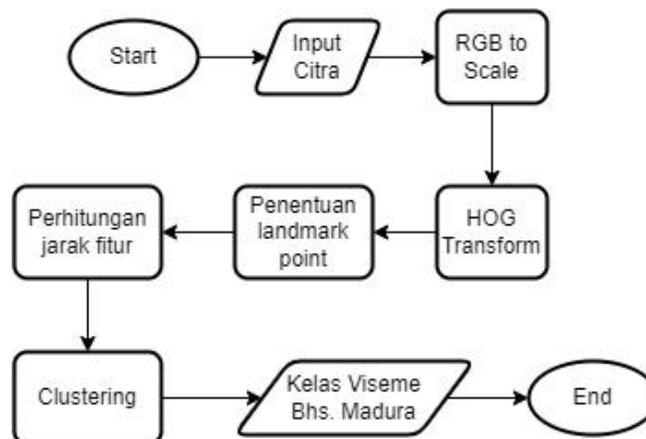
Fitur geometri yang banyak digunakan adalah jarak antara dua titik penting pada wajah (Z. Zhang, 1998). Fitur-fitur geometri yang diperoleh dari deteksi mulut atau bibir diproses dengan mengukur jarak antar fitur dan kemudian akan dilakukan proses *clustering* untuk mendapatkan kelas-kelas viseme bahasa Madura. Salah satu contoh fitur geometri yang digunakan pada pengenalan ekspresi wajah pada bagian mulut adalah fitur panjang mulut, lebar mulut bagian dalam, sudut antara ujung kiri dengan tengah kiri atas luar mulut, sudut antara ujung kanan dengan tengah kanan atas luar mulut, sudut antara ujung kiri dengan tengah kiri bawah luar mulut, sudut antara ujung kanan dengan tengah kanan bawah luar mulut (Z. Zhang, 1998). Fitur geometri akan bekerja baik dalam deteksi ucapan mulut manusia karena pada saat berbicara titik-titik tertentu pada mulut akan berpindah-pindah.

Penelitian ini membahas tentang identifikasi kelas-kelas model viseme untuk fonem bahasa Madura. Kelas-kelas viseme tersebut dapat membantu menciptakan suatu gerakan atau animasi mulut 2D untuk mengucapkan kata atau kalimat dalam bahasa Madura. Identifikasi kelas-kelas viseme pada penelitian ini menggunakan pendekatan *data driven* yaitu dengan mengelompokkan fonem berdasarkan ekstraksi fitur dari data visual hasil rekaman. Manfaat dari penelitian adalah untuk tujuan pembelajaran dalam pengucapan bahasa Madura, karena bahasa Madura memiliki tulisan dan cara pengucapan yang berbeda.

2. Metode Penelitian

Metode penelitian digunakan untuk menjelaskan langkah-langkah untuk mendapatkan kelas model viseme bahasa Madura berdasarkan *facial landmark point*. Adapun rancangan sistem pada penelitian ini ditunjukkan pada gambar 1.

Gambar 1 merupakan *flowchart sistem* yang menjelaskan bagaimana tahap-tahap untuk mendapatkan kelas model viseme bahasa Madura berdasarkan *facial landmark point*. Yang pertama adalah pengambilan *frame* pada perekaman video, kemudian setiap gambar *frame* tersebut akan dirubah dari gambar RGB menjadi gambar *grayscale* dan dilanjutkan proses deteksi wajah menggunakan algoritma *Histogram Oriented Gradient* (HOG) untuk memperkecil area komputasi. Proses HOG akan menghasilkan *bounding box* wajah dari rekaman video untuk kemudian digunakan pada proses penandaan *facial landmark point*. Selanjutnya dilakukan proses ekstraksi fitur berdasarkan *landmark point* pada bagian mulut dan dilanjutkan perhitungan jarak pada setiap fitur. Fitur-fitur yang digunakan adalah fitur geometri. Nilai ukuran jarak antar fitur yang dihasilkan akan digunakan pada proses *clustering* untuk mendapatkan kelas-kelas viseme bahasa Madura.



Gambar 1. Flowchart Sistem

2.1 Data Input

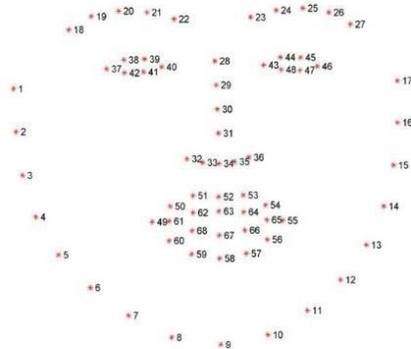
Data input yang digunakan adalah rekaman video sejumlah pembicara yang sedang berbicara atau mengucapkan suku kata dalam bahasa Madura. Pembicara duduk tegak di depan sebuah kamera video dengan jarak sekitar 1 (satu) meter yang dipasang tripod. Posisi kamera disesuaikan dengan tingkat wajah pembicara. Dari rekaman video tersebut akan diambil *frame* gambar yang mewakili setiap fonem bahasa Madura. Hasil setiap gambar akan dirubah dari gambar RGB menjadi gambar *grayscale* menggunakan persamaan "(1)".

$$G = \frac{r + g + b}{3} \quad (1)$$

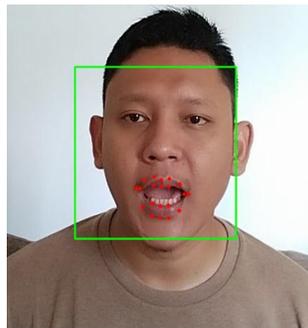
Dimana G adalah nilai *grayscale*, r nilai *red* pada setiap piksel, g nilai *green* pada setiap piksel dan b nilai *blue* pada setiap piksel. Gambar *grayscale* akan dilanjutkan pada proses *Histogram Oriented Gradient* (HOG) untuk memperkecil area komputasi karena obyek yang digunakan hanya pada area wajah. Metode HOG adalah salah satu metode yang digunakan untuk pendeteksian objek pada *computer vision* dan *image processing*. HOG akan menghitung kemunculan orientasi gradien pada bagian-bagian lokal dari sebuah gambar. Pada penelitian ini objek yang digunakan adalah pada bagian area wajah saja khususnya pada bagian mulut. Tujuan dari proses HOG adalah untuk memberikan *bounding box* wajah pada hasil rekaman video fonem bahasa Madura.

2.2 Facial Landmark Point

Setelah mendapatkan gambar pada area wajah maka tahap berikutnya adalah proses pemberian tanda menggunakan 68 penanda koordinat bentuk wajah. Namun sistem hanya mengambil koordinat pada bagian mulut saja yaitu mulai dari node ke 49 hingga node ke 68. Untuk menentukan kooordinat-koordinat tersebut digunakan *dlib regression tree*. *Dlib regression tree* adalah sebuah *library* yang digunakan untuk membantu menentukan titik *landmark* pada wajah seperti yang terlihat pada gambar 2.

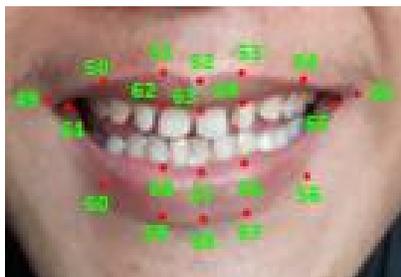


Gambar 2. Facial Landmark Point



Gambar 3. Facial Landmark Point bagian Mulut

Pada gambar 3 terlihat bahwa dengan algoritma HOG sistem akan membentuk *bounding box* pada area wajah kemudian *facial landmark point* akan memberikan tanda pada area mulut yaitu mulai dari node 49 hingga node ke 68 seperti terlihat pada gambar 4.

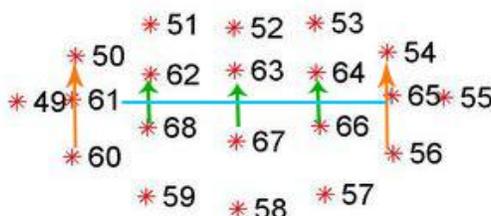


Gambar 4. Node bagian Mulut

Node-node pada gambar 4 adalah node pada bagian mulut hasil rekaman. Node-node tersebut akan menghasilkan nilai koordinat mulut yang akan digunakan pada proses perhitungan jarak antar fitur dan dilanjutkan pada proses *clustering*.

2.3 Perhitungan Jarak Antar Fitur

Setelah diperoleh nilai koordinat-koordinat pada node mulut maka tahap berikutnya yaitu proses perhitungan jarak antar fitur. Jarak antar fitur adalah nilai jarak dari setiap koordinat dari node-node yang menandai masing-masing fitur. Nilai-nilai tersebut nantinya akan digunakan pada proses *clustering*. Fitur yang dihitung jaraknya yaitu fitur panjang mulut, fitur lebar mulut bagian dalam dan fitur lebar ujung mulut bagian luar.



Gambar 5. Jarak Antara Titik Acuan pada Mulut

Gambar 5 adalah jarak antara titik acuan pada mulut yang akan dihitung pada proses perhitungan jarak antar fitur. Fitur-fitur yang digunakan pada perhitungan jarak dapat dilihat pada tabel 1 berikut :

Tabel 1. Fitur dan Node Mulut

Nama	Fitur	Node
a	Panjang Mulut	61 dan 65
b	Tinggi Ujung Mulut Kanan	54 dan 56
c	Tinggi Ujung Mulut Kiri	50 dan 60
d	Lebar Mulut Dalam Tengah	63 dan 67
e	Lebar Mulut Dalam Kanan	64 dan 66
f	Lebar Mulut Dalam Kiri	62 dan 68

Proses perhitungan jarak antar fitur menggunakan formula "(2)".

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2)$$

Dimana x dan y adalah nilai koordinat masing-masing node pada setiap fitur.

2.4 Clustering

Setelah nilai jarak setiap fitur diperoleh maka nilai-nilai tersebut akan digunakan pada proses *clustering* untuk mendapatkan kelas-kelas viseme bahasa Madura. Kelas-kelas tersebut akan berisi sekumpulan atau kelompok viseme bahasa Madura yang memiliki bentuk yang hampir sama. Metode *clustering* yang digunakan pada penelitian ini yaitu metode *K-Means clustering*. Algoritma *K-Means clustering* melakukan dua tugas utama yaitu:

- Menentukan titik pusat k atau *centroid*.
- Mencari anggota dari setiap *cluster*.

Cara kerja algoritma *K-Means clustering* adalah sebagai berikut:

- Menentukan jumlah k untuk menentukan jumlah *cluster*.
- Menentukan titik k secara acak.
- Menghitung jarak dari setiap data ke masing-masing *centroid*.
- Alokasikan setiap data ke *centroid* terdekat.
- Menentukan *centroid* yang baru dengan cara menghitung nilai rata-rata dari data yang terletak pada *centroid* yang sama.
- Kembali ke langkah 3 apabila masih terdapat data yang berpindah kelompok.

Setelah proses *clustering* dilakukan, untuk menentukan jumlah *cluster* yang terbaik, maka digunakan formula "(3)".

$$I = \sum_i (d(i, cr)) \quad (3)$$

$$W_k = \sum_{r=1}^k \frac{1}{(2 * n_r)} D_r$$

Dimana cr adalah *centroid* dari *cluster* dan d adalah kuadrat jarak. D_r adalah jumlah kuadrat jarak antara semua titik pada *cluster* r .

3. Hasil dan Pembahasan

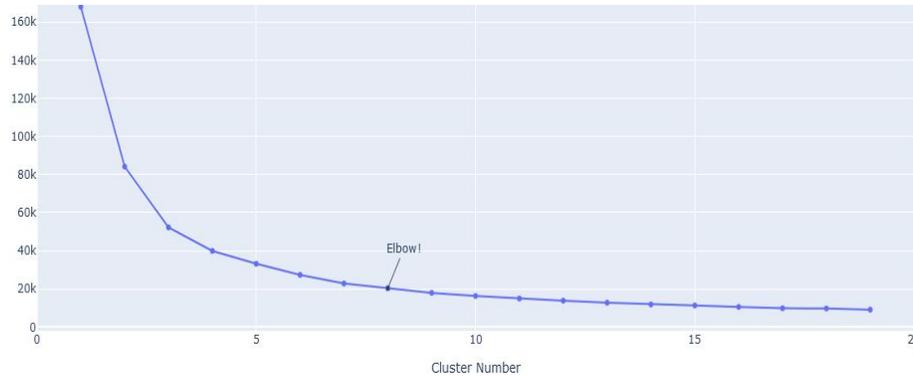
Data input yang digunakan pada penelitian ini adalah rekaman video sejumlah pembicara yang berbicara atau mengucapkan suku kata dalam bahasa Madura. Rekaman pembacaan suku kata bahasa Madura dilakukan oleh 5 (lima) orang bersuku Madura dengan variasi gender dan usia. Jumlah data yang digunakan dalam penelitian ini adalah sebanyak 1.161 frame gambar mulut yang mewakili bentuk setiap fonem bahasa Madura. Dari rekaman tersebut maka sistem akan mengambil setiap frame gambar fonem bahasa Madura untuk dirubah dari gambar RGB menjadi gambar *grayscale* dan dilanjutkan proses HOG untuk memperkecil area komputasi. Selanjutnya sistem akan mendeteksi area wajah menggunakan 68 penanda koordinat dengan *dlib regression tree*. Dari 68 node, hanya 20 node yang digunakan untuk mengukur nilai jarak antar fitur yaitu hanya pada node-node pada bagian mulut saja. Nilai jarak dari setiap fitur akan digunakan pada proses *clustering* untuk mendapatkan kelas viseme bahasa Madura. Hasil proses *clustering* dapat dilihat pada tabel 2.

Tabel 2. Hasil Proses Clustering

Gambar	Fitur						Hasil Kelas
	a	b	c	d	e	f	
1	64	14	0	1	1	13	3
2	63	14	1	1	1	14	3
3	64	37	21	22	22	38	1
4	57	39	15	18	16	40	1
...
1157	78	25	1	1	0	25	3
1158	86	36	21	20	20	33	7
1159	79	25	12	13	13	26	6
1160	72	24	0	0	0	24	3

Tabel 2 merupakan hasil perhitungan proses *clustering* dimana data yang digunakan pada tahap *clustering* adalah nilai jarak masing-masing fitur pada bagian mulut.

Untuk menentukan nilai *k* atau jumlah *cluster* yang terbaik maka digunakan metode *elbow* dengan nilai yang dihasilkan seperti pada grafik gambar 6.



Gambar 6. Nilai Elbow Clustering

Nilai *elbow* pada gambar 6 menunjukkan bahwa nilai *k* atau jumlah *cluster* yang terbaik sejumlah 8 (delapan) *cluster* ditambah dengan 1 (satu) kelas diam menjadi 9 *cluster*. 9 (sembilan) *cluster* tersebut mewakili 47 fonem dalam bahasa Madura. Hasil *cluster viseme* bahasa Madura ditunjukkan pada tabel 3.

Tabel 3. Model Viseme Bahasa Madura

Kelas	Fonem	Viseme
1	ây, ə, òy, sy, T, uy	
2	f, v	
3	a, ò	
4	c, d, D, dh, Dh, g, gh, j, jh, k, ny, q, s, t, x, z	
5	o, O, u, U, w	
6	â, è, e, h, l, r	
7	ay, i, n, ng, y	
8	b, bh, m, p	
9	-	

3.1 Uji Validitas dan Reliabilitas

Sifat-sifat instrumen pada penelitian ini menggunakan konsistensi internal yaitu validitas dan reliabilitas dimana validitas mengacu pada akurasi penilaian apakah sudah valid atau tidak dan nilai-nilai koefisien reliabilitas berkisar 0 sampai 1.0. Nilai koefisien 0 dikatakan tidak handal dan jika semua tes memiliki beberapa kesalahan maka koefisien reliabilitas tidak akan pernah mencapai 1.0 (Aschenberner dan Weiss, 2005). Instrumen validitas dan reliabilitas diukur dengan persamaan "(4)" dimana r adalah korelasi antara variabel x dan y , x_k dan y_k merupakan variabel diantara fonem-fonem, $k = 1, 2, \dots, N$, dan N adalah jumlah fonem.

$$r = \frac{N \sum_{k=1}^N x_k y_k - \left(\sum_{k=1}^N x_k \right) \left(\sum_{k=1}^N y_k \right)}{\sqrt{N \sum_{k=1}^N x_k^2 - \left(\sum_{k=1}^N x_k \right)^2} \sqrt{N \sum_{k=1}^N y_k^2 - \left(\sum_{k=1}^N y_k \right)^2}} \quad (4)$$

Pengujian model dari hasil implementasi maka dibuat kuisioner dengan menggunakan metode skala pengukuran sebagai alat statistik. Didalam kuisioner akan memuat beberapa pertanyaan dan setiap pertanyaan akan mewakili satu kelas viseme bahasa Madura. Responden akan diminta untuk mencocokkan semua himpunan fonem bahasa Madura dengan bentuk model gambar 2D. Dimana setiap himpunan fonem diukur oleh 4 (empat) poin skala Likert antara lain : Sangat tidak setuju, tidak setuju, setuju dan sangat setuju terhadap pemetaan fonem ke viseme bahasa Madura. Data responden diperoleh sebanyak 107 responden yang terdiri dari berbagai profil dan usia. Hasil persentase nilai responden ditunjukkan pada tabel 3.

Tabel 3. Skor Responden

Kelas Viseme	% Nilai Responden				Total
	1	2	3	4	
V1	6.17	18.17	52	30.67	107
V2	4.5	20.5	46.5	35.5	107
V3	4.5	7.5	47.5	47.5	107
V4	4	19.25	49	34.75	107
V5	4.75	13.25	45.75	43.25	107
V6	2.86	15.57	52.29	36.29	107
V7	4.2	18.4	49.4	35	107
V8	4.5	14.25	48	40.25	107
V9	0	3	41	63	107
Total	35.47	129.89	431.44	366.20	963
Rata-rata	3.94	14.43	47.94	40.69	107

Pada tabel 4 nilai 0,666 pada r tabel adalah nilai uji signifikansi koefisien korelasi pada taraf signifikansi 0,05 untuk $N = 9$ dimana N adalah jumlah kelas viseme bahasa Madura. Dari data tersebut dapat disimpulkan bahwa hasil dari nilai responden dinyatakan valid.

Tabel 4. Hasil Uji Validitas

Nilai Responden	r hitung	r tabel	Keterangan
1	0,954	0,666	Valid
2	0,993	0,666	Valid
3	0,999	0,666	Valid
4	0,994	0,666	Valid

Sedangkan hasil uji reliabilitas pada tabel 5 diperoleh nilai koefisien alfa untuk 4 nilai responden adalah 0,865 menunjukkan bahwa item tersebut juga memiliki konsistensi internal yang relatif tinggi.

Cronbach's Alpha	N of item
0,865	4

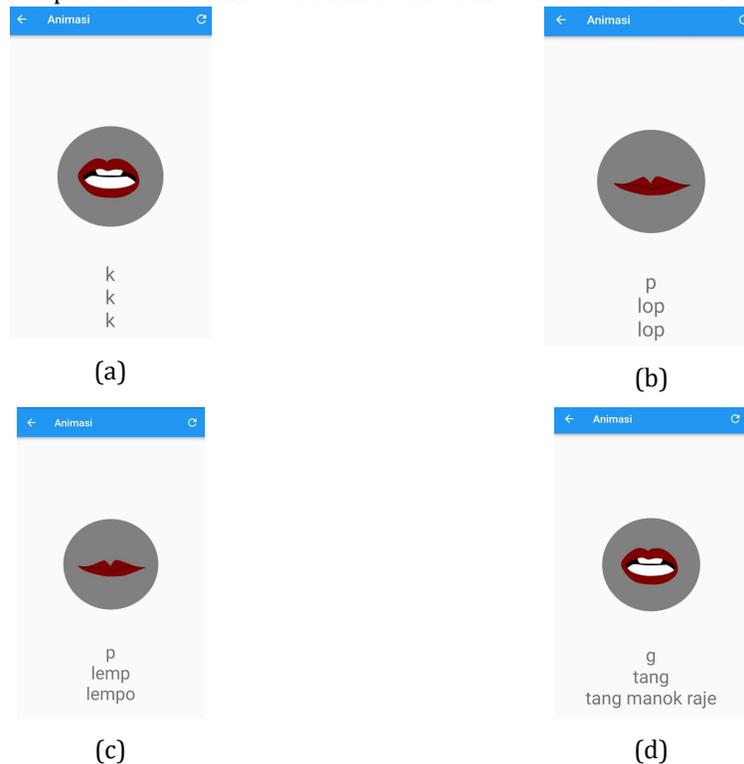
3.2 Model Animasi

Output dari sistem penelitian ini adalah animasi 2D sederhana dari mulut atau bibir dalam mengucapkan kata atau kalimat bahasa Madura. Data input berupa teks yang berisi huruf, suku kata, kata ataupun kalimat bahasa Madura. Gambar animasi output ditunjukkan pada gambar 7.

4. Kesimpulan

Berdasarkan pembahasan yang telah dijelaskan maka dapat diambil kesimpulan diantaranya :

- Metode *clustering* berdasarkan *facial landmark point* mampu menghasilkan identifikasi kelas viseme bahasa Madura.
- Jumlah kelas model viseme bahasa Madura yang dihasilkan adalah 9 kelas termasuk diam.
- Hasil dari uji validitas adalah valid.
- Hasil uji reliabilitas sebesar 0,865 menunjukkan konsistensi internal yang relatif tinggi.
- Sistem dapat menampilkan animasi 2D sederhana dari viseme bahasa Madura berdasarkan inputan teks.



Gambar 7
Output Animasi (a) Output Huruf. (b) Output Suku Kata. (c) Output Kata. (d) Output Kalimat

5. Referensi

Bear, H. L., & Harvey, R. (2016). Decoding visemes: Improving machine lip-reading Helen L. Bear and Richard Harvey. In *Icassp 2016*, 2009–2013.

- Bozkurt, E., Erdem, Ç. E., Erzin, E., Erdem, T., & Özkan, M. (2007). Comparison of phoneme and viseme based acoustic units for speech driven realistic lip animation. In *Proceedings of 3DTVCON*. <https://doi.org/10.1109/3DTV.2007.4379417>
- Luca Cappelletta and Naomi Harte, "Phoneme-to Viseme Mapping for Visual Speech Recognition," *Proceeding of the 2012 International Conference on Pattern Recognition Applications and Methods (ICPRAM 2012)*, February 7, 2012.
- Setyati, E., Sumpeno, S., Purnomo, M. H., Mikami, K., Kakimoto, M., & Kondo, K. (2015). Phoneme-viseme mapping for Indonesian language based on blend shape animation. *IAENG International Journal of Computer Science*, 42(3), 1–12.
- Arifin, Muljono, Sumpeno, S., & Hariadi, M. 2013. Towards Building Indonesian Viseme: A Clustering-Based Approach. *IEEE*. 978-1-4673-6053-1/ 3/\$31.00.
- Kumar, K. T. Bibish, Kumar, R. K. S., Sandesh, E. P. A., Sourabh, S., & Lajish, V. L. 2019. Viseme set identification from Malayalam phonemes and allophones. *International Journal of Speech Technology*. <https://doi.org/10.1007/s10772-019-09655-0>
- Brahme, A., & Bhadade, U. (2017). Phoneme visem mapping for Marathi language using linguistic approach. In *Proceedings International conference on global trends in signal processing, information computing and communication, ICGTSPICC 2016* (pp. 152–157).
- Damien, P., Wakim, N., & Egéa, M. (2009). Phoneme-viseme mapping for modern, classical arabic language. In *2009 international conference on advances in computational tools for engineering applications, ACTEA 2009* (Vol. 2(1), pp. 547–552)
- Bianca Aschenberner and Christian Weiss, "Phoneme-Viseme Mapping for German Video-Realistic Audio-Visual-Speech- Synthesis," IKP-Working Paper NF 11, Institut für Kommunikatons for Schung und Phonetik, Universität Bonn, 2005.
- Pawitra, A. Kamus Lengkap Bahasa Madura Indonesia. Jakarta: Dian Rakyat. 2009.
- Gleason, H. A. 1962. *An Introduction to Descriptive Linguistics*. New York: Holt, Reinehart and Winston
- Han, S., Yang, Z., Li, Q., & Chen, Y. (2019). Deformed landmark fitting for sequential faces. *Journal of Visual Communication and Image Representation*, 381 - 393.
- Z. Zhang, M. J. Lyons, M. Schuster, dan S. Akamatus, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," *IEEE International Conference on Automatic Face & Gesture Recognition*, no. November 2012, 1998.