

# **LIP-SYNC KARAKTER ANIMASI MENGGUNAKAN *HIDDEN MARKOV* MODEL**

**Mursyidah**

Dosen Teknik Multimedia dan Jaringan Politeknik Negeri  
Lhokseumawe

E-mail : [mursyidahpoli@gmail.com](mailto:mursyidahpoli@gmail.com)

## **ABSTRAK**

Ketika berbicara manusia menggerakkan bibir untuk mengucapkan huruf atau kata. Model bibir ini menggambarkan suatu *viseme* (*visual phoneme*). *Visual phoneme* atau visual fonem adalah bentuk bibir ketika mengucapkan suatu fonem. Bentuk bibir tersebut dapat digolongkan berdasarkan model bibir yang mirip, meskipun suara yang ditimbulkan berbeda. Konsonan konsonan tersebut terdapat pada satu golongan yang sama. Gerakan bibir juga dipengaruhi oleh bentuk koartikulasi dalam suatu kata dan kalimat. Koartikulasi muncul dari pertukaran bentuk bibir ketika mengucapkan fonem yang berurutan. Gerakan mulut /bibir selama berbicara merupakan komponen penting dari animasi wajah. Banyak dijumpai gerak mulut model karakter animasi ketika berbicara sering terlihat tidak sesuai dengan suara yang dikeluarkan. Biasanya gerak mulut relatif lebih cepat sehingga hasil gerakan bibir sebuah model karakter saat berbicara menjadi tidak realistis. Sangat dibutuhkan sebuah teknik sinkronisasi suara yang bisa membuat model karakter animasi mampu menyelaraskan gerak bibir dengan kata atau kalimat yang dibunyikan secara otomatis. Proses sinkronisasi gerak bibir pada penelitian ini menggunakan *shape matching* dengan algoritma *Hidden Markov Model* (HMM). Dari permasalahan tersebut maka permasalahan dapat rumuskan adalah bagaimana mendesain bentuk mulut yang sesuai dengan Shape Priston Blair, bagaimana mensegmentasi suara dan teks

dan bagaimana mendapatkan kecocokan bentuk mulut dengan suara menggunakan algoritma *Hidden Markov Model* (HMM). Algoritma HMM menggunakan prinsip-prinsip rantai markov yaitu ucapan masukan akan dibandingkan (dihitung probabilitasnya) dengan parameter ucapan yang sudah dimodelkan. Nilai koefisien HMM dari bentuk bibir yang aktif dalam sinkronisasi gerak bibir untuk kata Selamat Datang adalah  $\alpha$  adalah sebesar 0.8 untuk Se,  $\beta$  sebesar 0.193. L sebesar 0.502, dan WQ sebesar 0.193 untuk phonem La,  $\gamma$  sebesar 0.4, L sebesar 0.026, dan MBP sebesar 0.8 untuk phoneme Mat, E sebesar 0.8 dan  $\delta$  sebesar 0.193 untuk phoneme Da, dan Rest sebesar 0.8 untuk phoneme Tang.

*Kata kunci* : Lips\_Sync, KarakterAnimasi, HMM

## 1. Pendahuluan

Ketika berbicara manusia menggerakkan mulut untuk mengucapkan sebuah kata. Gerakan bibir menggambarkan suatu *viseme* (*visual phoneme*). *Visual phoneme* atau visual fonem adalah bentuk bibir ketika mengucapkan suatu fonem. Bentuk bibir tersebut dapat digolongkan berdasarkan model bibir yang mirip, meskipun suara yang ditimbulkan berbeda seperti konsonan 'p', 'b', 'm'. Konsonan konsonan tersebut terdapat pada satu golongan yang sama. Gerakan bibir juga dipengaruhi oleh bentuk koartikulasi dalam suatu kata dan kalimat (Gultom,2012). Koartikulasi muncul dari pertukaran bentuk bibir ketika mengucapkan fonem yang berurutan. Pergerakan bibir dari sebuah model karakter animasi pada waktu berbicara merupakan salah satu komponen yang sangat penting pada produksi film animasi. Seorang animator dituntut harus kreatif mencari sendiri padanan pola gerak bibir yang sesuai untuk setiap teks/kata yang diucapkan, agar suara yang dikeluarkan model karakter animasi terlihat lebih realistis dan lebih hidup. Dibutuhkan proses cukup panjang dan jumlah *keyframe* yang banyak untuk setiap bentuk mulut pada waktu pengisian suara. Banyak teknologi yang

memungkinkan seorang animator untuk mendesain model karakter animasi berkualitas tinggi, akan tetapi untuk membuat model karakter animasi bisa berbicara dengan kesesuaian gerakan bibir dan ucapan sangatlah rumit (Chen, 2012). Sehingga sering dijumpai gerakan mulut dari model karakter animasi ketika berbicara relatif lebih cepat dari suara yang diucapkan.

Kebanyakan proses editing suara banyak menggunakan software *LipSync* dan hanya memperhatikan posisi mulut terbuka dan tertutup saja tanpa memperdulikan bunyi atau *phoneme* yang diucapkan. Sangat dibutuhkan sebuah teknik sinkronisasi suara yang bertujuan bisa membuat model karakter animasi mampu menyelaraskan gerak bibir dengan kata atau kalimat yang dibunyikan secara otomatis. Proses sinkronisasi gerak bibir pada penelitian ini menggunakan *shape matching* dengan algoritma *Hidden Markov Model* (HMM). HMM merupakan salah satu algoritma yang sering digunakan dalam sistem pengenalan ucapan manusia, di mana *statenya* tidak dapat diamati secara langsung (tersembunyi), tetapi hanya dapat diobservasi melalui suatu himpunan pengamatan lain. Algoritma HMM menggunakan prinsip-prinsip rantai markov yaitu ucapan masukan akan dibandingkan (dihitung probabilitasnya) dengan parameter ucapan yang sudah dimodelkan. Nilai probabilitas akan menunjukkan dikenali sebagai apa ucapan masukan tersebut.

## 2. State Of The Art

Beberapa teknik *Lips-Cync* yang sudah pernah dilakukan oleh peneliti antara lain adalah *Automated Up\_Sync: Background and Techniques* (Lewis, 1991) menggunakan metode *Linier Prediction Approach*. *Real-Time Continuous Phoneme Recognition System Using Class-Dependent Tied-Mixture HMM With HBT Structure for Speech-Driven Lip-Sync*, (Park, 2008) menjelaskan *real time lip-sync* menggunakan *lip shape* avatar disinkronisasikan dengan sinyal suara. *Animating Lip-Sync Character whth Dominated Anime Models* (Chen, 2012) meneliti tentang konfigurasi dan

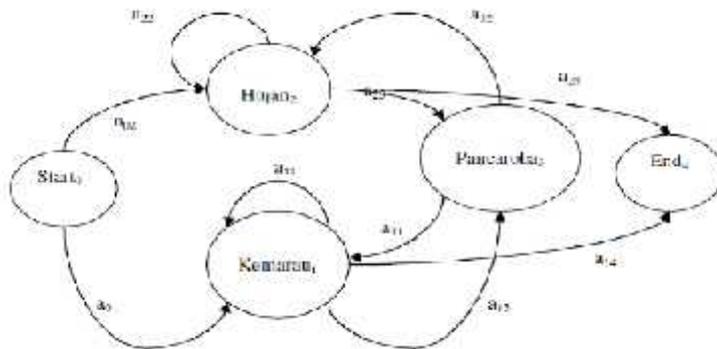
prinsip *lip sync* untuk *game*. Penelitian ini menjelaskan sinkronisasi pergerakan wajah dengan audio dari suara alami yang dihasilkan atau merubah *text – speech engine*. Untuk setiap *phoneme* satu atau banyak *Dominated Anime Models (DAMS)* pertama sekali harus dikelompokkan dengan sinyal kontrol. Rancang Bangun Aplikasi Pengenal Penutur Menggunakan Metode *Hidden Markov Model (Hmm)*, (Gultom, 2012), pengenal penutur (*speaker recognition*) adalah proses secara otomatis mengenali siapa yang bicara berdasarkan informasi individu yang terdapat di dalam gelombang wicara. Sistem pengenal penutur dibentuk melalui dua proses utama yaitu pelatihan (*training*) dan pengenalan (*recognition*). Penelitian yang dilakukan menggunakan metode *Mel Frequency Cepstrum Coefficients (MFCC)* untuk ekstraksi fitur dan metode *Hidden Markov Model (HMM)* untuk mengenali penutur. Hasil penelitiannya menunjukkan bahwa pengujian secara *real time* menggunakan mikrofon tingkat akurasi 30%. Sedangkan pengujian dari file rekaman 100%. Tingkat akurasi sangat bergantung pada kemampuan clustering dan klasifikasi.

## 3. Kajian Kepustakaan

### 3.1 Markov Model

Markov model merupakan bagian dari stokastik dengan ruang *state* (keadaan) diskrit, serta mempunyai sifat dimana *state* saat ini hanya tergantung pada *state* sebelumnya. Pada rantai Markov setiap *state* dapat diamati secara langsung, contohnya seperti keadaan cuaca. Keadaan cuaca besok dapat di prediksi melalui keadaan cuaca hari ini. Model ini merupakan bagian dari *finite state* atau *finite automation*. *Finite automation* adalah kumpulan *state* yang transisi statenya dilakukan berdasarkan masukan observasi. Pada Markov Chain setiap busur antar *state* berisi probabilitas yang mengidentifikasi kemungkinan jalur tersebut akan diambil. Jumlah probabilitas semua busur yang keluar dari sebuah simpul adalah satu. Gambar 1

memperlihatkan contoh Markov Chain yang menggambarkan kondisi cuaca.



Gambar 1. Markov Chain untuk Cuaca

Pada gambar 1  $a_{ij}$  adalah probabilitas transisi dari state  $i$  ke state  $j$ . Misalkan dari simpul start '0' keluar dua kemungkinan,  $a_{02}$  dan  $a_{01}$ . Maka jumlah probabilitas  $a_{01} + a_{02}$  adalah satu. Hal ini juga berlaku bagi simpul-simpul yang lain. Markov chain bermanfaat untuk menghitung probabilitas suatu kejadian teramati yang secara umum di rumuskan sebagai berikut:

$$P(\sigma_1) = P(\sigma_t \mid \sigma_{t-1}, \sigma_{t-2}, \sigma_{t-3}, \dots)$$

(1)

$\sigma_1$  adalah kondisi saat ini, dan  $\sigma_t$  adalah kondisi pada waktu tertentu yang berhubungan dengan  $\sigma_1$ . Sedangkan  $\sigma_{t-1}$  adalah kondisi sebelum  $\sigma_1$ . Diasumsikan sebelah kanan persamaan bersifat invariant, yaitu dihipotesakan dalam keseluruhan sistem, transisi diantara keadaan tertentu tetap sama dalam hubungan probabilitasnya.

Berdasarkan asumsi tersebut dapat dibuat suatu set keadaan probabilistik aij diantara dua keadaan  $S_i$  dan  $S_j$ .

$$A_{ij} = p(\alpha_t = S_i \mid \alpha_{t-1} = S_j), \quad 1 \leq i, j \leq N$$

.....(2)

Karena  $i$  dan  $j$  dapat sama, maka berlaku batasan berikut:

$$A \geq 0, \quad \text{dan} \quad \sum a_{ij} = 1$$

.....(3)

### 3.2 Hidden Markov Model

HMM merupakan model stokastik dimana suatu sistem dimodelkan diasumsikan sebagai markov proses dengan kondisi yang tidak terobservasi. Setiap kondisi memiliki distribusi kemungkinan di setiap output yang mungkin. Oleh karena itu, kemungkinan dari transisi antar kondisi menjadi satu-satunya parameter teramati. Dalam HMM, keadaan tidak terlihat secara langsung. Tetapi output yang bergantung terhadap keadaan tersebut terlihat.

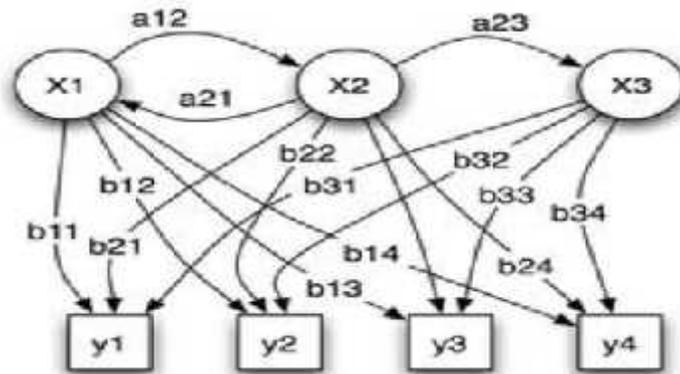
Setiap kondisi memiliki distribusi kemungkinan di setiap output yang mungkin. Oleh karena itu urutan langkah yang dibuat HMM memberikan suatu informasi tentang urutan dari keadaan. Sifat *hidden* (tersembunyi) menunjukkan kepada kondisi langkah yang dilewati model, bukan kepada parameter dari model tersebut.

HMM adalah variasi dari *finite state machine* yang memiliki kondisi tersembunyi  $Q$ , suatu nilai output  $O$  (observasi), kemungkinan transisi  $A$ , kemungkinan output  $B$ , sebuah kondisi awal  $\Pi$ . Kondisi saat ini tidak terobservasi, tetapi setiap keadaan menghasilkan output kemungkinan  $B$  biasanya,  $Q$  dan  $O$  dimengerti, jadi HMM disebut *triple*  $(A, B, \Pi)$ .

1. Himpunan observasi state  $O = O_1, O_2, \dots, O_N$

2. Himpuna Hidden State :  $Q = q_1, q_2, \dots, q_N$
3. Probabilitas transisi:  $A = a_{01}, a_{02}, \dots, a_{n1}, \dots, a_{mn}$ ;  $a_{ij}$  adalah probabilitas untuk pindah dari *state*  $i$  ke *state*  $j$ .
4. Probabilitas emisi atau obsevation likelihood:  $b = b_i(O_t)$ , merupakan probabilitas observasi  $O_t$  dibangkitkan oleh state  $i$ .
5. State awal dan akhir :  $q_0, q_{end}$ , yang tidak terkait dengan observasi.

Representasi parameter Hidden Markov Model di tunjukkan seperti dalam Gambar 2.



Gambar 2 Representasi Parameter HMM

Dimana;

$x$  = kondisi

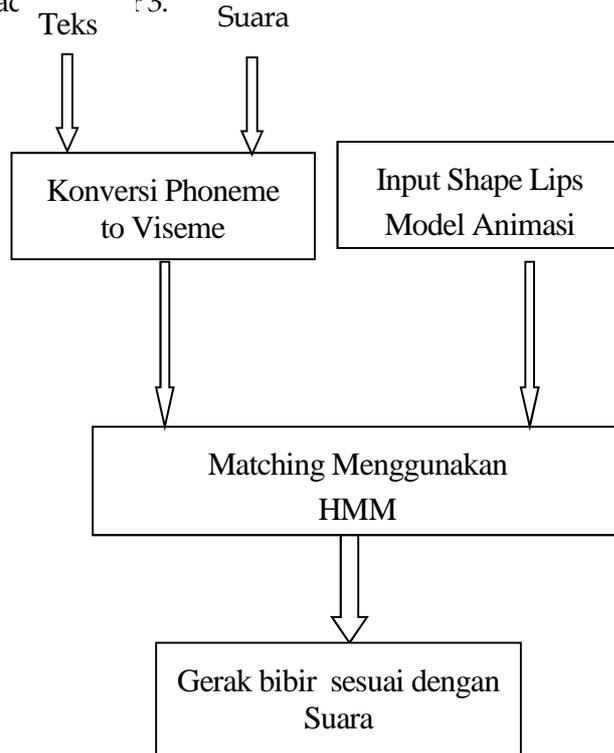
$y$  = observasi yang mungkin

$a$  = kemungkinan keadaan transisi

$b$  = kemungkina output

## 4. Metode Penelitian

Adapun metode penelitian *Lip-Sync* karakter animasi ini seperti ditunjukkan pada Gambar 3.

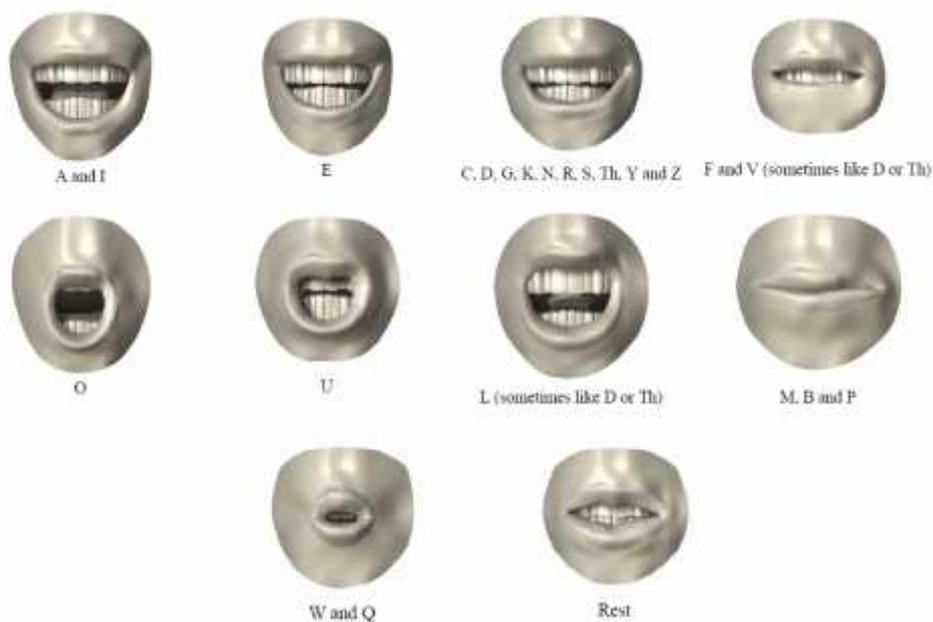


Gambar 3. Diagram Alir Penelitian

Tahapan penelitian *Lip-Sync* dilakukan seperti ditunjukkan pada Gambar 3 dimana proses pertama sekali yang dilakukan adalah proses Akuisisi Data. Data berupa sinyal suara diperoleh dengan cara merekam ucapan melalui mikrofon yang dihubungkan dengan komputer. Kemudian suara dan teks dikonversikan dari *phoneme* ke bentuk *viseme* (pola bibir) dengan menggunakan aplikasi Papagayo

untuk dibatasi (*splite*) tiap suku katanya sesuai dengan bentuk *shape* Preston Blair. Kalimat yang di rekam adalah kata “Selamat datang”.

Pada tahapan kedua adalah medesain model *shape* mulut dari karakter animasi yang digunakan dalam penelitian menggunakan Aplikasi Blender. *Shape* mulut yang didesain sesuai dengan *shape* Preston Bleir sebanyak sepuluh bentuk *shape* mulut dengan nilai koefisien mulai dari 0 sampai 1. Untuk posisi mulut tertutup nilai koefisien sebesar 0 dan posisi mulut terbuka maksimal adalah 1. Adapun bentuk shape seperti pada Gambar 4.



Gambar 4. *Shape Phoneme* Preston Bleir

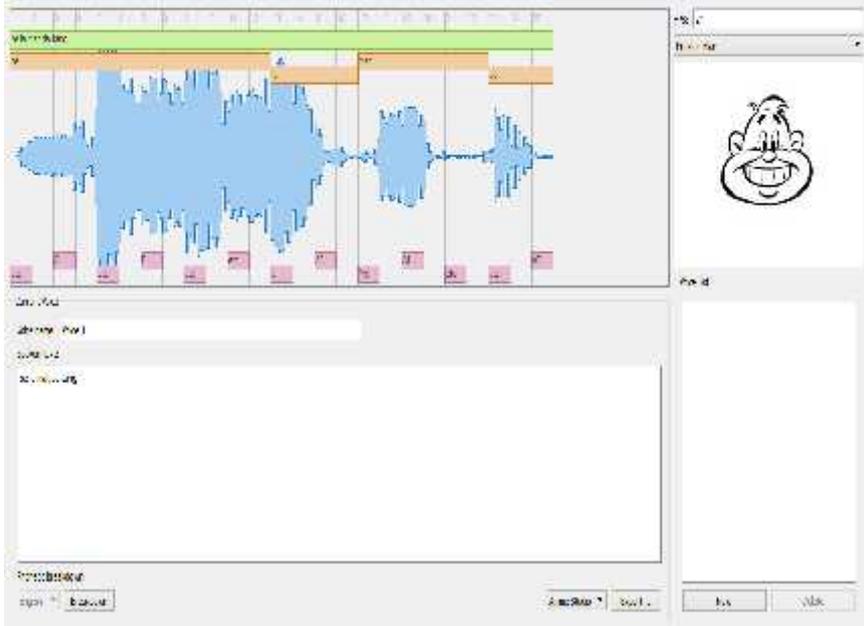
Untuk selanjutnya dilakukan proses pencocokan gerak mulut dari model karakter animasi yang sudah dimodelkan dengan suara yang sudah di klasifikasikan menggunakan algoritma Hidden Markove Model. Diagram trellis dapat digunakan untuk

memvisualisasikan kemungkinan dalam perhitungan Hidden Markove Model ditunjukkan pada tabel 1.

	E	O	U	MBP	F	V	L	MBP	NO	EST
0	0	0	0	0	0	0	0	0	0	0
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
10	10	10	10	10	10	10	10	10	10	10

## 5. Hasil Eksperimen dan Pembahasan

Hasil segmentasi rekaman suara dengan teks " Selamat Datang" dimana huruf ' Se' di klasifikasi kedalam bentuk mulut "ect, O, ect, E, ect, ect", untuk teks ' La' di klasifikasi kedalam bentuk mulut L dan AI, untuk teks 'mat' diklasifikasikan dalam bentuk mulut "MBP, AI dan ect ", teks 'da' di klasifikasikan dalam bentuk mulut " ect dan AI" dan untuk teks ' tang' di klasifikasikan dalam bentuk mulut " ect, AI dan ect", seperti ditunjukkan pada Gambar 5 dan pada Tabel 2.



Gambar 5. Hasil segmentasi suara “ Selamat Datang”

Tabel 2. Bentuk *shape* yang aktif untuk teks ‘Selamat Datang’ Pada Aplikasi Papagayo. Pada tahap proses learning dimana bentuk wajah dan mulut yang sudah di desain menggunakan Software Blender sebanyak sepuluh bentuk sesuai dengan bentuk mulut *Priston Blair* yang ada di Papagayo.

	SE	LA	MAT	DA	TANG
AI		✓		✓	✓
O	✓				
E	✓		✓		

Tabel 2. Bentuk shape yang aktif untuk teks 'Selamat Datang Pada Aplikasi Papagayo	U					
	ECT	✓		✓	✓	✓
	L		✓			
	WQ					
	MBP			✓		
	FV					
	REST					

mat Datang Pada Aplikasi Papagayo

Untuk setiap bentuk memiliki range nilai dari 0 sampai 1. Nilai 0 untuk posisi mulut tertutup dan nilai 1 untuk kondisi mulut terbuka maksimum. Tampilan kondisi awal dari aplikasi Lip\_Sync yang sudah terkoneksi dengan Aplikasi Papagayo ditunjukkan pada Gambar 6.



Koefisien HMM untuk Phoneme Priston Blaire yang aktif kondisi ucapan SE yaitu hanya dua buah bentuk mulut yang aktif yaitu bentuk mulut WQ dengan koefesien sebesar 0.8 dan bentuk mulut REST sebesar 1 di tunjukkan pada Tabel 4.

Tabel 4. Phoneme Priston Blaire yang aktif kondisi ucapan SE

Phoneme Priston Blaire yang aktif kondisi ucapan SE									
AI	O	E	U	ETC	L	WQ	MBP	FV	REST
0	0	0	0	0	0	0.8	0	0	1.0

Koefisien HMM untuk Phoneme Priston Blaire yang aktif kondisi ucapan La yaitu bentuk mulut AI sebesar 0.8, ect sebesar 0.193, l sebesar 0.502 dan WQ sebesar 0.193 di tunjukkan pada Tabel 5.

Tabel 5. Phoneme Priston Bleire yang aktif kondisi ucapan LA

Phoneme Poison Blaire yang aktif kondisi ucapan LA									
AI	O	E	U	ETC	L	WQ	MBP	FV	REST
0.8	0	0	0	0.193	0.502	0.193	0	0	0

Hasil klasifikasi dari Papagayo untuk phoneme 'La' di klasifikasikan kedalam kombinasi bentuk mulut 'L dan AI'.

Koefisien HMM untuk Phoneme Priston Blaire yang aktif kondisi ucapan Mat yaitu bentuk mulut 'ect' sebesar 0.4, L sebesar 0.026 dan MBP sebesar 0.8 di tunjukkan pada tabel 6.

Tabel 6. Phoneme Poison Blaire yang aktif kondisi ucapan MAT

Phoneme Poison Blaire yang aktif kondisi ucapan MAT									
AI	O	E	U	ETC	L	WQ	MBP	FV	REST
0	0	0	0	0.4	0.026	0	0.8	0	0

Koefisien HMM untuk Phoneme Priston Blaire yang aktif kondisi ucapan 'Da' yaitu bentuk mulut 'ect' sebesar 0.193, E sebesar 0.8 di tunjukkan pada Tabel 7.

Tabel 7. Phoneme Priston Bleire yang aktif kondisi ucapan DA

Phoneme Priston Bleire yang aktif kondisi ucapan DA									
AI	O	E	U	ETC	L	WQ	MBP	FV	REST
0	0	0.8	0	0.193	0	0	0	0	0

Hasil klasifikasi dari Papagayo untuk phoneme 'Da' di klasifikasikan kedalam kombinasi bentuk mulut ect dan AI.

Koefisien HMM untuk Phoneme Priston Blaire yang aktif kondisi ucapan 'Tang' yaitu bentuk mulut 'Rest' sebesar 0.8, di tunjukkan pada Tabel 8.

Tabel 8 Phoneme Priston Bleire yang aktif kondisi ucapan TANG

Phoneme Priston Bleire yang aktif kondisi ucapan TANG									
AI	O	E	U	ETC	L	WQ	MBP	FV	REST
0	0	0	0	0	0	0	0	0	0,8

Secara keseluruhan nilai koefisien HMM untuk teks "Selamat Datang" ditunjukkan pada Tabel 9.

Tabel 9 Nilai koefisien HMM untuk phoneme Selamat Datang

	SE	LA	MAT	DA	TANG
AI	0	0.8	0	0	0
O	0	0	0	0	0
E	0	0	0	0.8	0
U	0	0	0	0	0
ECT	0.8	0.193	0.4	0.193	0
L	0	0.502	0.026	0	0
WQ	0	0.193	0	0	0
MBP	0	0	0.8	0	0
FV	0	0	0	0	0

REST	0	0	0	0	0.8
------	---	---	---	---	-----

Dari Tabel 9 dapat dilihat nilai koefisien HMM yang aktif untuk ucapan Selamat datang adalah kombinasi dari beberapa shape mulut yang mirip. Ada beberapa shape yang aktif tidak sesuai dengan shape di papagayo seperti pada pengucapan teks Da dan Tang. Hal ini bisa disebabkan salah satunya adalah bentuk mulut di papagayo mengikuti phoneme Bahasa Inggris yang mempunyai perbedaan pengucapan phoneme dengan bahasa Indonesia. Dari lima phoneme dalam dalam teks Selamat Datang hanya dua yang berbeda dengan aplikasi Papagayo yaitu pada phoneme Da dan Tang.

## 6. Kesimpulan

Dari penelitian yang telah dilakukan dapat disimpulkan:

Nilai koefisien HMM dari bentuk bibir yang aktif dalam sinkronisasi gerak bibir untuk kata Selamat Datang adalah ect adalah sebesar 0.8 untuk Se, ect sebesar 0.193. L sebesar 0.502, dan WQ sebesar 0.193 untuk phonem La, ect sebesar 0.4, L sebesar 0.026, dan MBP sebesar 0.8 untuk phoneme Mat, E sebesar 0.8 dan ect sebesar 0.193 untuk phoneme Da, dan Rest sebesar 0.8 untuk phoneme Tang., dan dari penelitian yang dilakukan kombinasi bentuk mulut yang aktif hasil yang di learning dari lima phoneme yang dilatih hanya dua bentuk phoneme yang bebeda dengan bentuk mulut dari Papagayo.

### Referensi.

- Chen, Y.-M., 2012. Animating Lip-Sync Character Whith Dominate Anime Models. *IEEE*.
- Gultom, M., 2012. Rancang Bangun Aplikasi Pengenal Penutur Menggunakan Metode Hidden Markov Model (Hmm), s.l.: Teknik Informatika STIMIK GI MDP.

Park, J., 2008. Real-Time Continuous Phoneme Recognition System Using Class-Dependent Tied-Mixture HMM With HBT Structure for Speech-Driven Lip-Sync. *IEEE*, Volume 10.

Firdaus, dkk, 2008. Hidden Markov Model, Jurnal Matematika FMIPA Universitas Padjadjaran.

<https://cartoonsolutions.com/store/catalog/Lip-Sync-Mouth-Animation-sp-79.html>, Mouth Movements Instructional.