

## BAB II

### TINJAUAN LITERATUR

#### 2.1 Tinjauan Pustaka

Dalam pengembangan animasi karakter yang dapat berbicara letak realistiknya ada pada proses sinkronasi gerakan bibir (lipsync) karena proses tersebut mempunyai kompleksitas dan waktu pengerjaan yang tergolong lama. Proses sinkronasi gerakan bibir pada dialog atau lipsync merupakan aspek penting dalam dunia animasi. Adapun beberapa penelitian yang membahas tentang pentingnya teknik auto lipsync untuk meningkatkan efisiensi. Beberapa penelitian seperti Sari dan Ramadhan (2023) telah membahas mengenai teknik *auto lipsync* pada karakter 3D yang menekankan bahwa pembentukan viseme dari berbagai bentuk fonem mempunyai kesulitan tersendiri bagi animator, lalu hal itu diperkuat dengan adanya penelitian yang dilakukan oleh Kristama et al (2022) tentang bagaimana teknik auto lipsync dapat memangkas waktu proses produksi animasi pada lingkungan studio profesional. Berlandaskan problem tersebut, maka bab ini akan meninjau secara mendalam terkait landasaan teori yang mendasari pengembangan sistem lipsync otomatis pada animasi karakter.

Jika mengacu pada kedua penelitian diatas, meskipun berhasil membuktikan manfaat efisiensi dari auto lipsync adapun kekurangan dari fundamental sistem diatas. Penelitian dari Sari dan Ramadhan (2023) menyebutkan secara eksplisit bahwa mereka menggunakan set viseme standart yang mempunyai basis Bahasa Inggris. Bahkan mereka mengakui bahwa harus melakukan penelitian lebih lanjut untuk membentuk viseme dalam bahasa lain seperti Bahasa Indonesia agar karakter lebih natural. . Kesenjangan fonetik ini yang juga dibahas oleh Kristama et al (2022), yang memberikan contoh perbedaan terhadap gerakan bibir seperti huruf 'A' yang dibaca 'ei' dalam bahasa inggris namun dalam Bahasa Indonesia

dibaca ‘a’. Model fonem dan viseme inilah yang menjadi akar permasalahan, yang menyebabkan hasil lipsync otomatis untuk dialog Bahasa Indonesia seringkali terasa tidak akurat.

Kebutuhan diatas telah diuraikan sebelumnya dan terjawab secara langsung oleh penelitian fundamental yang dilakukan oleh Setyati et al (2015). Dalam jurnal berjudul “Phoneme-Viseme mapping for Indonesian

Language based on Blend Shape Animation”, dimana mereka telah mengusulkan sebuah standar untuk viseme Bahasa Indonesia melalui pendekatan linguistik, penelitian itu berhasil mengidentifikasi 49 fonem dan mengelompokkannya menjadi 12 kelas viseme yang secara visual mewakili pengucapan Bahasa Indonesia yang diharapkan dapat menjadi acuan untuk pengembangan selanjutnya. Hasil dari penelitian ini menyediakan sebuah “kamus” visual yang penting namun masih menyisakan kekurangan dalam hal implementasi praktisnya kedalam produksi animasi modern secara real-time

Berdasarkan beberapa penelitian diatas, terdapat kesimpulan meskipun kebutuhan akan lipsync yang efisien telah diidentifikasi (Sari dan Ramadhan, (2023); Kristama et al., 2022), implementasinya masih terjadi kendala oleh ketergantungan pada model fonem bahasa inggris. Penelitian fundamental oleh Setyati et al. (2015) telah berhasil memetakan fonem ke viseme untuk Bahasa Indonesia. Namun, hingga saat ini masih terjadi kesenjangan antara model teoretis tersebut dengan alur kerjayang praktis dan dapat diakses oleh animator. Oleh karena itu, penelitian ini diharapkan akan setidaknya mengisi celah dan menguji sebuah alur kerja spesifik dalam pemanfaatan perangkat lunak seperti iClone 8 dan Unity untuk pengaplikasian dan penyesuaian parameter viseme Bahasa Indonesia.

Oleh karena itu, penelitian ini ada untuk mengisi celah tersebut dengan memanfaatkan alur kerja dari iClone 8 dan Unity. Fokus dari penelitian ini bukan hanya untuk menghasilkan animasi secara otomatis, melainkan juga untuk melakukan penyesuaian parameter viseme Bahasa

Indonesia melalui teknik *fine-tuning* manual. Dengan integrasi dari hasil pemetaan Setyati et al (2015) dalam teknologi *real-time engine* saat ini dapat diharapkan terciptanya standart baru dalam produksi karakter virtual yang lebih natural dan memiliki akurasi artikulasi yang tepat. Hal tersebut akan memperkuat kualitas interaksi antara manusia dan komputer dalam dunia teknologi yang semakin canggih.

## 2.2 Kerangka Teori

### 2.2.1 Konsep Fonem dan Viseme

#### a. Fonem

Fonem dapat diartikan sebagai unit terkecil dari suara dalam sebuah bahasa yang berfungsi untuk menandakan perbedaan makna. Dalam Bahasa Indonesia, terdapat beberapa contoh sederhana dari fonem seperti /p/ dan /b/ pada kata 'pari' dan 'bari' yang menghasilkan makna yang berbeda. Sistem fonem pada Bahasa Indonesia cukup konsisten terutama pada fonem vokalnya yang terdiri dari lima vokal murni: /a/, /i/, /u/, /e/, /o/.

#### b. Viseme

Viseme adalah bentuk visual dari fonem yang berupa bentuk mulut ketika diucapkan. Seringkali, beberapa fonem menghasilkan bentuk bibir yang sangat mirip sehingga dapat dikelompokkan ke dalam satu kelas viseme yang sama. Sebagai contoh, fonem bilabial seperti /p/, /b/, /m/ yang menghasilkan visual hampir identik karena dimulai dengan bibir yang tertutup. Penelitian fundamental oleh Setyati et al. (2015) berhasil memetakan 49 fonem Bahasa Indonesia ke dalam 12 kelas viseme yang menjadi acuan dalam penelitian ini.

Tipe Fonem	Kelas Fonem	Jumlah fonem	Fonem Set	Kelas Viseme
Consonant	1	3	/b/, /m/, /p/	V1
	2	4	/f/, /v/, /w/, /ph/	V2
	3	8	/d/, /dh/, /dl/, /dz/, /l/, /n/, /t/, /th/	V3
	4	1	/r/	V4
	5	12	/c/, /j/, /s/, /z/, /ts/, /ps/, /ks/, /sh/, /sy/, /x/, /y/, /ny/	V5
Vowels	6	8	/g/, /gh/, /h/, /k/, /kh/, /q/, /ng/, /ngg/	V6
	7	1	/a/	V7
	8	3	/i/, /I/, /oi/	V8
	9	4	/ə/, /e/, /ε/, /ai/	V9
	10	3	/o/, /ɔ/, /au/	V10
	11	2	/u/, /u/	V11
-	-	-	Silent	V12
	Total	49		

**Tabel 1** Tabel Kategori Fonem

### 2.2.2 iClone 8 dan Fitur *Text-to-Speech*

iClone 8 adalah perangkat lunak utama dalam alur kerja ini yang memiliki fungsi sebagai studio animasi dimana tempat proses lipsync dibuat. Salah satu fiturnya yang digunakan dalam penelitian ini adalah kemampuan *Text-to-Speech* internal yang ada didalamnya.

Metodologi ini memanfaatkan kemampuan iClone untuk secara langsung menerima inputan teks yang bisa disesuaikan. Setelah memasukkan teks, maka iClone akan melakukan dua proses yaitu:

1. Melakukan sinkronisasi dari teks yang dimasukkan yang selanjutnya akan menghasilkan sebuah fonem dan viseme secara standart
2. Menghasilkan animasi lipsync awal yang dimana sistem ini dihasilkan untuk secara otomatis membuat timeline animasi viseme pada fitur AccuLIPSRe

Timeline yang dihasilkan secara otomatis ini yang akan menjadi baseline dalam penelitian ini. timeline editor pada iClone 8 akan menyediakan fungsi penuh untuk animator melakukan penyesuaian manual. Pada tahap ini merupakan inti dari penelitian ini, yaitu untuk mengoreksi dan menyesuaikan keyframe viseme yang dihasilkan secara otomatis agar akurat dengan fonem Bahasa Indonesia.

### **2.2.3 Unity**

Unity merupakan game engine yang berfungsi sebagai platform implementasi akhir, setelah animasi lipsync disesuaikan di iClone 8, maka seluruh aset diimport ke Unity. Dalam Unity animasi tersebut diintegrasikan dengan audio yang dikontrol oleh skrip C# untuk ditampilkan secara real-time.