Ultimart: Visual Communication Journal
Vol 18 No. 1, June 2025
ISSN: 1979 - 0716, e-ISSN 2615-8124
Faculty of Art and Design
Universitas Multimedia Nusantara

ANALYSIS OF THE IMPLEMENTATION OF THE VISEME-BASED METHOD IN FICUSIA ANIMATION

Selly Artaty Zega[1]
Adric Filbert[2]
Gerson Julyfer Parulian Tambun[3]
Muhammad Fauzi Ardhi[4]

Received May. 18, 2025; Revised June. 13, 2025; Accepted June. 24, 2025.

Abstract: *Based on interviews with the Ficusia animation team and professionals from Infinite Frameworks Studios, it was found that the lip and jaw movements in Ficusia were too fast, making them appear less realistic. This study aims to improve the quality of lip-sync animation in Ficusia using the viseme-based method, which allows for more natural lip and jaw movements. The viseme-based method was chosen because it is more suitable for 3D animation, producing smoother lip movements that are better integrated with facial expressions compared to the phoneme-based method, which tends to be rigid. A qualitative case study approach was conducted on the first episode of Ficusia, which contains scenes with predominantly human mouth shapes. Primary data were obtained through interviews with experienced animators and supervisors, while secondary data were gathered from literature. The results show that the viseme-based method can enhance the quality of lip-sync, create more natural animation, and improve the integration between lip movements and character expressions. This method is considered effective by professional animators and is expected to serve as a reference for producing realistic lip-sync in 3D animation.*

Keywords: *3d animation; blender; lip-synch; viseme-based*

Introduction

Lip-sync, or lip synchronization, aligns lip movements with pre-recorded audio and is essential in animation for creating lifelike characters and natural dialogue (Hoon & Shaharuddin, 2019; Shukurov, 2024). Poor lip-sync is easily noticeable and diminishes animation quality (Hoon & Shaharuddin, 2019). Successful shows like The Forces of Evil, My Little Pony, and Mr. Bean demonstrate how effective

[1]*Selly Artaty Zega is a lecturer at Department of Informatics Engineering, Animation Study Program, Politeknik Negeri Batam.* *e-mail : selly@polibatam.ac.id*

[2]*Adric Filbert is a student at Department of Informatics Engineering, Animation Study Program, Politeknik Negeri Batam.* *e-mail: adric.filbert@students.polibatam.ac.id*

[3]*Gerson Julyfer Parulian Tambun is a lab assistant at Department of Informatics Engineering, Animation Study Program, Politeknik Negeri Batam.* *e-mail: gerson@polibatam.ac.id*

[4]*Muhammad Fauzi Ardhi is Consultant 3D Layout Artist at Anima Vitae Point Sdn Bhd* *e-mail: ardhi.fauzi@gmail.com*

lip-sync supports storytelling and character (Aneja & Li, 2019).

Ficusia — an anti-drug animation by 69 Politeknik Negeri Batam students and collaborators—serves as a local example of lip-sync application (Alfi & Prasetiya, 2023; Hidayah & Zega, 2023). It tells the story of a tree obsessed with Cuta seeds to attract Flowna. Despite local animation advances, Ficusia struggles to meet international lip-sync standards.

Ficusia was selected as a case study due to its realistic style in facial expressions, the availability of production team insights, and its role as a representative example of emerging 3D animation practices in Indonesia. Compared to other local animations, Ficusia is the latest Indonesian animation that openly documents its process and showcases clear lip-sync challenges, making it suitable for in-depth qualitative analysis. The lip-sync challenges are found during interviews with Ficusia animators and professionals from Infinite Frameworks Studios. It was revealed that while mouth shapes were accurate, lip and jaw movements were too fast, resulting in unnatural animation. Neither team initially identified their lip-sync technique, but this study found two relevant approaches in literature: phoneme-based and viseme-based (Shukurov, 2024; Nakada et al., 2024).

The lip-sync with phoneme-based method, commonly used in 2D animation, matches mouth shapes to phonemes but can cause stiff or exaggerated motions in 3D (Osipa, 2010). The viseme-based method, preferred for 3D, groups similar sounds and blends movements of lips, corners, and jaw for a more natural look (Osipa, 2010; Hoon & Shaharuddin, 2019; Muhammad, 2023).

This qualitative case study analyzes viseme-based lip-sync in Ficusia episode one, selected for its clear dialogue scenes. Data were collected through interviews with experienced animators and supervisors, literature reviews, expert consultations, and shot analysis using Blender. This research also only applied to human-like facial structures as their facial structures demand more precise and realistic lip-sync. While several characters that meet the requirements are analyzed in this research too, due to page limitation, this study will only show the picture of Uci the little girl for comparison as she has the clearest close up shots.

This study aims to contribute to animation literature and offers practical guidance for local animators aiming to produce lip-sync animations that meet international standards.

Lip-sync

Lip-sync, or lip synchronization, matches mouth movements to audio and is key to making animated characters appear natural (Hoon & Shaharuddin, 2019; Shukurov, 2024). It's often done by linking mouth shapes to phonemes (Chen et al., 2010). Though no single method guarantees realism (Osipa, 2010), several criteria from research and interviews help evaluate it.

1. There is ease in And ease out, timing Good, and felt squash and stretch (William, 2001; Wahkid, personal communication, 2024)

2. Different mouth shape or asymmetry (Osipa, 2010; Danar, personal communication, 2024)

3. Character expressions are legible (William, 2001; Danar, personal communication, 2024)

4. Integration of other facial ele-

ments (William, 2001; Osipa, 2010).

Viseme-based Method

The viseme-based method links mouth shapes with facial expressions and character, making lip-sync more natural than strict phoneme matching (Hoon & Shaharuddin, 2019). It groups similar sounds (e.g., M, B, P) into one viseme and blends poses by focusing on mouth corners and chin (Osipa, 2010). This suits realistic 3D animation, as people often speak with less precision (Chen et al., 2010; Osipa, 2010). Below are combined rules from Osipa and Williams for applying this method.

B/M/P: Closed mouth with a tip that matches the letter before and after it. This sound must be made, with the lips turning back for emphasis. The width of the mouth can widen or narrow according to context

EE (expanded): The tip of the

mouth widens from the previous shape. The width of the mouth can change, adapting to the shape of the mouth before and after.

F/V: The lower lip rolls back, the upper lip rises slightly to reveal the upper teeth. This form should always be clearly visible in the dialogue.

OR: Lips curl forward, corner of mouth narrows. U is more oval and smaller than O. The width of the mouth must be maximum to read clearly, except in certain sentences. This form is not affected by expression.

SH/CH: The tip of the mouth widens from its previous shape, revealing teeth with a proportion of 70% upper teeth and 30% lower teeth, with a closed chin. The width of the mouth depends on the context.

Th: The tip of the mouth adapts to the previous shape, with the chin lower except after M, and not closed completely.

Mouth shapes that are not mentioned can be adjusted to the mouth shape rules above because one visemes can represent several words (Osipa, 2010).

1. Define an expression.

Character expressions can be determined based on storyboard or the sound when the character speaks (Osipa, 2010). The characteristics of expressions according to Osipa are based on Building stylized brow shapes :

Happy: Focus on the upward movement at the corners of the lips and the small wrinkles around the eyes.

Sad: The position of the lips is lowered, the eyebrows are curved upwards, and the lower eyelids are slightly raised.

Angry: The eyebrows are drawn together and lowered, the eyes are narrowed, and the lips may be pressed.

Scared: Eyes wide open, eyebrows raised, and lips tensely parted.

Surprised: Eyes very open, eyebrows raised high, and mouth wide open.

Disgusted: The nose is wrinkled, the upper lip is raised, and asymmetrical expressions are often used.

2. Adapt viseme with phoneme and overall expression animation

To adjust viseme with phonemes, determination is required in dialogue. The following Fig.1, is an illustration of how to determine viseme for character based dialogue waveform from audio.

Selly Artaty Zega[1]
Adric Filbert[2]
Gerson Julyfer Parulian Tambun[3]
Muhammad Fauzi Ardhi[4]

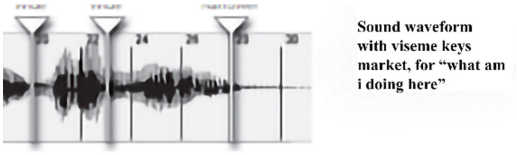

Figure 1. Determine viseme in dialogue
(Source: Stop Staring : Stop Staring : Facial modeling and animation done right)

3. Simplifies movement of the tip of the mouth and chin

According to Osipa, mouth shapes are simplified by adjusting the corners (narrow or wide) and using squash and stretching on the chin. Not all phonemes are equally important (Osipa, 2010; Williams, 2001). Animators should prioritize key shapes and blend similar sounds like A and E. Smoothing can be done in the viewport or graph editor (Osipa, 2010; Zega, 2021).

4. Working on teeth and tongue

The character's teeth must sync with the chin, with only the lower teeth moving in line with it (William, 2001). During speech, characters typically show upper teeth, lower teeth, or a 70:30 ratio (William, 2001). Tongue movements should be quick, usually without ease-in or ease-out (William, 2001; Osipa, 2010).

5. Create asymmetry

Asymmetry in expressions or mouths can be created using uneven mouth corners, misaligned lips, and differently raised eyebrows (William, 2001; Osipa, 2010). Asymmetry is applied last to enhance expression and is sometimes minimized if unnecessary (Osipa, 2010).

Methodology

This study uses a qualitative case study to analyze the impact of the viseme-based method on Ficusia's lip-sync animation, with conclusions based on expert feedback. The viseme-based method then applied in episode 1 of the Ficusia series.

The research process includes the following steps:

1. Identification of primary and secondary data
2. Determination of time and location
3. Identification of necessary shots
4. Determination of production stages
5. Implementation of the production stages
6. Data analysis

This qualitative case study focuses on the third, fourth, and sixth stages to address the research questions. The unit of analysis is each shot from the Ficusia film, described and critiqued based on interviews. Problematic shots are improved using the viseme-based method, then evaluated by experts. The study aims to enhance Ficusia's lip-sync quality according to research and interview criteria.

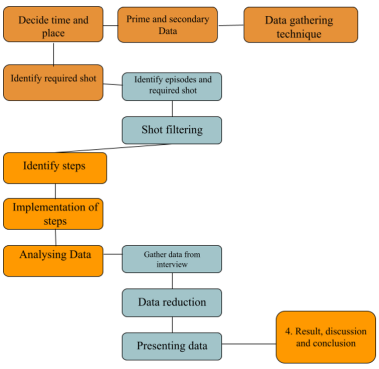

Figure 2. Animated interactive wall
(Source: Personal research documentation)

Data Collection Techniques

Data collection will be based on 3 things:

1. Observation

Observations will be carried out at Batam State Polytechnic to analyze differences before and after implementation viseme-based method, and at Infinite Studios Batam to determine the criteria for shots and episodes that will be used for implementation viseme-based method.

2. Interview

Interviews will be conducted with sources from Infinite Framework Studios who have at least more than 5 years of experience to obtain valid information.

3. Recording

Recording will be carried out to document work, collect data from interviews, and analyze before and after implementation viseme-based method.

Determination of Primary and Secondary Data

Primary data will come from interviews with Infinite Frameworks Studios' professional animators: Wahkid Joko Sayekti and Danar Donianto (lead animators, more than 5 years experience), as well as Denis James Deegan and Phil Mitchell (supervisors, 20 years of experience). Secondary data is taken from literature such as Animation Survival Kit and Stop Staring.

Determination of Time and Place

The interview was conducted at Infinite Frameworks Studios according to the resource person's schedule, while the production process was carried out in the Galang room, Polibatam Techno Building. The research takes place from October 20 to November 28, 2024.

Identify the Shot that is required

1. SC01_SH07_ANM
2. SC01_SH08_ANM
3. SC01_SH12_ANM
4. SC04_SH16_ANM

Determination of Manufacturing Stages Lip-Sync

Determination of stages will be based on the Osipa and William's rule:

1. Define an expression.
2. Adapt the viseme with the phoneme and the overall expression animation.
3. Simplifies the movement of the tip of the mouth and chin.
4. Working on teeth and tongue.
5. Create asymmetry.

Implementing Stages

The implementation stages will be carried out in Blender 2.9. The documentation and animation files are placed inside Gdrive and Youtube.

1. Define an expression

Expressions will be determined from the character's voice based on observation results. As for the expression for each shot, that is:

- SC01_SH07_ANM = Happy
- SC01_SH12_ANM = Happy
- SC01_SH08_ANM = Happy to sad
- SC04_SH16_ANM = Happy

2. Adapt the viseme with the phoneme and the overall expression animation

An adaptation of viseme with audio can be taken from the waveform in the audio. The shot of SC01_SH07_ANM below.

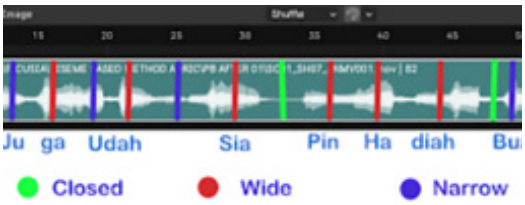

Figure 3. Results of determination analysis viseme shot SC01_SH07_ANM.
(Source: doc. Adric Filbert)

As can be seen in Fig. 3, it is shown that the phoneme will widen, narrow, or close according to the waveform of the sound. Based on this, the animation is refined using viseme principles from William and Osipa's guidelines.

3. Simplifies the movement of the tip of the mouth and chin

The animation lip-sync was then smoothed at the corners and chin to create a smooth animation. Simplifying movement is created with the corner mouth rig and chin as can be seen in Fig. 4 below.

Figure 4. Image of the mouth corner rig (Left) and chin (Right)
(Source: doc. Adric Filbert)

Fig. 5 showcase the graph editor for the jaw in rotate x.

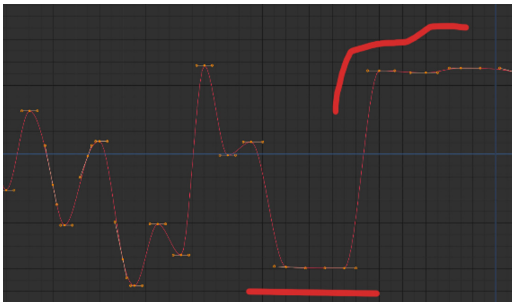

Figure 5. Example of non-smooth rotation of the chin shot SC01_SH07_ANM.
(Source: doc. Adric Filbert)

In Fig. 5, the Graph Editor revealed uneven jaw animation curves. These were refined into smoother, curved motions to enhance the animation's fluidity (Zega, 2021). In Fig. 6, are the results after the adjustments.

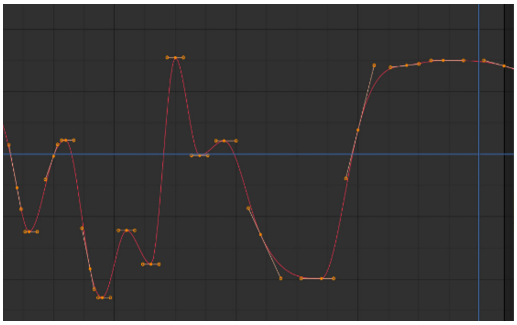

Figure 6. Example of subtle rotation of the chin shot SC01_SH07_ANM
(Source: doc. Adric Filbert)

4. Working on teeth and tongue

Figure 7. Rig Uci's teeth
(Source: doc. Adric Filbert)

As can be seen in Fig. 7 & Fig. 8, the work on teeth and tongue is carried out using rig upper and lower teeth in character. Work will use grab to raise or lower gears according to the 70:30 ratio (William, 2001).

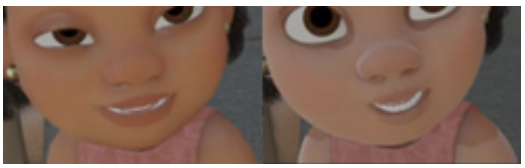

Figure 8. Pictures of teeth before and after
(Source: doc. Adric Filbert)

5. Create asymmetry

Asymmetry adjustments on Uci's mouth and eyebrows were made to keep a subtle, realistic look. Based on Gerson's interview, Ficusia's style is realistic, not cartoonish. As Osipa states, asymmetry should be present but not exaggerated. Fig. 9, as can be seen below are the results of this work.

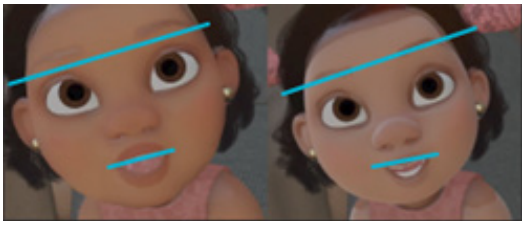

Figure 9. Asymmetry before (left) and after (right).
(Source: doc. Adric Filbert)

At frame 42 ("Pin" in "Siapin"), the original shows parallel eyebrows and mouth. The revision adds subtle asymmetry in eyebrow and mouth angles, with lips and teeth adjusted to align with the chin.

Analyzing Data

This study uses thematic analysis to identify patterns and themes from interview data (Ramdhani et al., 2024). It aims to assess the realism of viseme-based lip-sync and find areas for improvement.

Data Collection

Data was collected through interviews with sources from Infinite Frameworks Studios. According to the interviews, the viseme-based method has increased the quality of the lip-sync. However, there are still several things to do in order to improve the animation which are ease in, ease out, good timing, squash and stretch, mouth shape or asymmetry, readable expression and facial integration.

Reducing Data

After improving the animation according to the revision. Research is continued with another animation review to check the effectiveness of viseme-based methods. The reduction of data is done by determining the themes based on the Osipa and William's lip-sync realism criteria and reevaluating the overall sentences of each interviewee whenever it matches each of the themes or not (Nur & Saihu, 2024). Data reduction then served in a table with shot name, theme, and result. Results are shown for each theme by having a plus or minus in each opinion of the interviewee, which total will be 4 symbols

or either plus or minus, then determined whenever it's good or bad by overall plus or minus in each column. Table 1 can be seen below.

Table 1. Quality Improvement Second Attempt

Data Presentation 2				
Shot	Criteria			
	ease in, ease out, good timing, squash and stretch	mouth shape or asymmetry	Readable expression	Facial integration
SC01_SH07_NOTE	Good (++++)	Good (++++)	Good (++++)	Good(+++)
SC01_SH08_ANM	Good (++++)	Good (++++)	Good (++++)	Good (++++)
SC01_SH12_NOTE	Good (++++)	Good (++++)	Good (++++)	Good (++-+)
SC04_SH16_ANM	Good (++++)	Good (++++)	Good (++++)	Good (++++)

Result

The initial use of the viseme-based method improved timing and reduced hard stops in most shots. However, some issues still remain in the first try. The shot SC01_SH08_ANM and SC01_SH12_ANM still had slight jaw popping that disrupted motion, and SC01_SH07_ANM needed better asymmetry and more natural mouth rest positions. In general, facial expressions were unclear, and eye and eyebrow movements were either underused or not linked to the lip-sync. The revised version fixed these issues with more detail. Jaw movements became smoother with squash and stretch, and viseme-phoneme alignment improved, especially during vowel changes. Facial expressions were more dynamic with small tweaks to eyebrow angles and mouth corners. SC04_SH16_ANM stood out for its strong integration of lips, jaw, and facial features, clearly showing better animation quality.

Discussion

The second attempt has significantly enhanced animation lip-sync quality by emphasizing full facial integration. Coordinated counter movements between the nose and lower face added realism, while mouth corner adjustments made the character's expressions create more positive and emotionally engaging. This approach brought the animation closer to professional standards, as also emphasized in Osipa and Richard William's literature. The revised shots have met all key criteria for high-quality lip-sync, including smoothness, mouth shape, and facial integration. These results confirm that the viseme-based method when combined with expression focused refinements effectively improves lip-sync animation quality. Such improvement can be seen in Fig. 10.

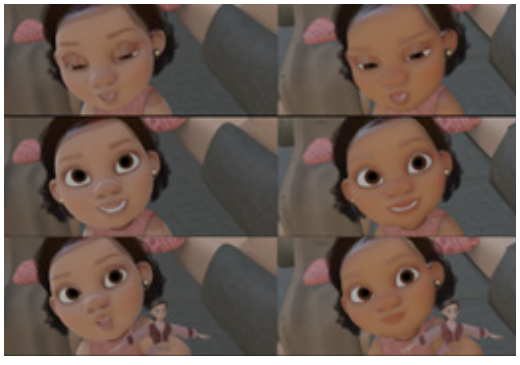

Figure 10. Top for integration (Left before, right after), bottom for expression.
(Source: doc. Adric Filbert)

Conclusion

Interviews have shown that the viseme-based method has succeeded in

improving lip-sync quality. As we can see from the result section, the facial integration, smoothness and timing of the mouth in SC01_SH08 until SC04_SH016_ANM has improved considerably and was approved by interviewee. Below are the overall ways to create realistic lip-sync animation quality with viseme-based method:

1. Define an expression

Character expressions are determined by the storyboard or the character's spoken sound, as done in this study. This research focuses on happy and sad expressions, chosen based on Osipa's observations and supporting literature:

Happy: Focus on the upward movement at the corners of the lips and the small wrinkles around the eyes.

Sad : The position of the lips is lowered, the eyebrows are curved upwards, and the lower eyelids are slightly raised.

2. Adapt viseme with phoneme and overall expression animation

Viseme adjustments based on the audio waveform follow Osipa's guidance: large waves correspond to wide mouth corners, while small waves indicate narrow corners. For example, the letter O starts with narrower corners than the letter A. Expressions follow Osipa's rule that the nose area moves opposite to the lower face.

3. Simplifies movement of the tip of the mouth and chin

The mouth and chin shapes can be simplified by following the Osipa's rules, narrow and wide mouth corners and squash and stretch for the chin based on the second stage's identification. This step focuses on achieving smooth ease-in/out, timing, and squash and stretch, while simplifying mouth movements as described by Osipa.

4. Working on Teeth and tongue

The character's teeth must sync with the chin, with only the bottom teeth moving in line with the chin (William, 2001). Characters typically show upper teeth, lower teeth, or a 70:30 ratio while speaking (William, 2001). The tongue moves quickly and usually doesn't require ease-in or ease-out (William, 2001; Osipa, 2010).

5. Create asymmetry

Asymmetry in expressions can be created using mouth corners, misaligned lips, and uneven eyebrows (William, 2001; Osipa, 2010). It is applied at the end to clarify expressions and may be minimized (Osipa, 2010). Interviews added two terms: "popping" (excessive spacing) and "hard stop" (abrupt transitions without ease-in/out), emphasizing the importance of easing and spacing. These terms can aid future animation quality reviews.

References

Alfi, N. F., & Prasetya, H. Y. (2023). Semiotic analysis of John Fiske's message on drug use in the 3D animation film "Ficusia". DECA: Journal of Visual Communication and Multimedia Design, 12(3), 75–85. https://jurnal.polibatam.ac.id/index.php/DECA/article/view/5271

Aneja, D., & Li, W. (2019). Real-time lip-sync for live 2D animation. arXiv preprint, arXiv:1910.08685. https://arxiv.org/pdf/1910.08685

Blaine, J. (2020). The complete guide to Blender graphics: Computer modeling and animation (5th ed.). CRC Press. https://www.taylorfrancis.com/books/

mono/10.1201/9780429196522/complete-guide-blender-graphics-john-blain

Chen. (2010). Animating lip-sync characters (1st ed.). https://www.academia.edu/110410717/Animating_Lip_Sync_Characters

Heiderich, T. (n.d.). Cinematography techniques: The different types of shots in film [PDF]. Ontario Media Arts. Retrieved November 24, 2024, from https://oma.on.ca/en/contestpages/resources/free-report-cinematography.pdf

Hidayah, A. (2023). Application of animation principles in the animated series film "Ficusia" episode 1. Journal of Electrical and Computer Engineering (DECA), 12(2), 85–92. https://jurnal.polibatam.ac.id/index.php/DECA/article/view/6036/2174

Hoon, L. N., & Shaharuddin, S. S. (2019). Accuracy performance and potentiality of real-time avatar lip-sync animation in different languages. International Journal of Recent Technology and Engineering (IJRTE), 7(6S2), 57–66. https://www.researchgate.net/publication/343442180

Muhammad, Y. (2023). Unlocking the power of Azure audio-to-viseme conversion: Explained [PDF]. ResearchGate. Retrieved November 26, 2024, from https://www.researchgate.net/profile/Rakan-Alwabel-2/publication/385381042

Nakada, K., Matsuda, S., Tanaka, Y., Sasaki, H., & Ueda, T. (2023). Lip-Sync ML: Machine learning-based framework to generate lip-sync animations in FINAL FANTASY VII REBIRTH. ResearchGate. http://www.jp.square-enix.com/tech/library/pdf/LipSync-ML_SIGGRAPH_Talks_2024_abst.pdf

Nelli, N., & Aliyyah, N. R. (2024). Implementation of learning media in the independent curriculum in junior high schools: Thematic analysis study. Karimah Tauhid, 3(8), 8541–8552. https://ojs.unida.ac.id/karimahtauhid/article/view/14412

Nur, M. A., & Saihu, M. (2024). Data processing. Scientica: Scientific Journal of Science and Technology, 2(11), 163–175. https://jurnal.kolibi.org/index.php/scientica/article/view/2764

Osip, J. (2010). Stop staring: Facial modeling and animation done right (3rd ed.). Wiley.

Shukurov, S. (2023). Audio to viseme and viseme to lip sync using Azure. ResearchGate. https://www.researchgate.net/publication/380895958

Williams, R. (2001). The animator's survival kit (Expanded ed.). Faber & Faber.

Zega, S. A., Soon, S. H., Firmanda, A., & Lumombo, L. (2020). Automatic in-between frames of snappy animation using linear interpolation. In Proceedings of the 3rd International Conference on Applied Engineering (ICAE 2020) (pp. 12–19). SciTePress. https://www.scitepress.org/Papers/2020/103510/pdf/index.html