

**“CAT STRANGE BEHAVIOUR”: PERCEPTUAL REALISM OF 2D
ANIMATED LIP-SYNC**



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS LAPORAN

JUDUL: “CAT STRANGE BEHAVIOUR”: PERCEPTUAL REALISM OF 2D ANIMATED LIP-SYNC

SESI PENGAJIAN: [2020 / 2021]

Saya: NUR HAZIRAH BT ALIMAN

mengaku membenarkan tesis Projek Sarjana Muda ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. * Sila tandakan (✓)

SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan)

TIDAK TERHAD

(TANDATANGAN PELAJAR)

Alamat tetap: 114, JALAN CEKAL 16,
TAMAN SRI LAMBAK, 86000,
KLUANG, JOHOR

Tarikh: 7 September 2021

(TANDATANGAN PENYELIA)

(DR. MOHD ADILI BIN NORASIKIN)
Nama Penyelia

Tarikh: 12 September 2021

CATATAN: * Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak

“CAT STRANGE BEHAVIOUR”: PERCEPTUAL REALISM OF 2D ANIMATED
LIP-SYNC

NUR HAZIRAH BT ALIMAN



This report is submitted in partial fulfillment of the requirements for the
Bachelor of Computer Science (Interactive Media) with Honours.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

DECLARATION

I hereby declare that this project report entitled
**“CAT STRANGE BEHAVIOUR”: PERCEPTUAL REALISM OF 2D ANIMATED
LIP-SYNC**

is written by me and is my own effort and that no part has been plagiarized
without citations.

STUDENT :  _____ Date : 7 September 2021
(NUR HAZIRAH BT ALIMAN)



I hereby declare that I have read this project report and found
this project report is sufficient in term of the scope and quality for the award of
Bachelor of Computer Science (Interactive Media) with Honours.

SUPERVISOR :  _____ Date : 12 September 2021
(DR. MOHD ADILI BIN NORASIKIN)

DEDICATION

Dedicated to my beloved family and supportive friends who always give me strength and spirit throughout my journey.



ACKNOWLEDGEMENTS

I would like to offer my heartfelt gratitude to a number of people and organizations for their unwavering support throughout my studies. Firstly, I wish to thank my parents for giving me refuge and love support for my entire life and my friends for always being there during ups and downs.

Next, I would like to thank my supervisor, Dr. Mohd Adili Norasikin, for his excitement, patience, insightful remarks, valuable information, practical guidance, and never-ending ideas, all of which have greatly aided me in completing my project. I could not have imagined having a better supervisor in my study.

Last but not least, I also wish to express my sincere thanks to lecturers of Faculty of Information and Communication Technology for sharing some useful information on this project and thanks to Universiti Teknikal Malaysia Melaka for having me here until my final year. This project would not have been able to accomplish without all of the help and supervision. Thank you for all of your encouragement.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ABSTRACT

“Cat Strange Behaviour” is a 2D animation that focuses on the perceptual realism of animated lip synchronization. Lip-sync (short for lip synchronization) is a term that refers to the process of matching lip movements with sound or spoken vocals. Moreover, real-time lip-sync animation is a method that will bring a virtual computer-generated character to life by synchronizing precise lip movement and sound in real-time. However, the development of realistic lip movement is difficult because the lip shape and position must synchronize with the spoken sounds. Consequently, viseme is used as a technique to the basic animation parameters in assessing visual similarities between various phonemes that included voice processing, speech recognition and computer face animation based on human speech mainly focus on the English language. The accuracy of lip motion capture of human with character performance in real-time and speech recognition is analyzed. As a result, perceptual realism is crucial for lip-sync animation for applying viseme-based human lip shapes in the English language to map the mouth and sound in real-time animation character. It is also provided as a guide for lip-sync animation, demonstrating how to employ simple synchronization tactics to increase the accuracy and create a more realistic visual impression, as well as how to include advanced features into lip synchronization application.

ABSTRAK

"Cat Strange Behavior" adalah animasi 2D yang fokus pada realisme persepsi animasi penyegerakan bibir. Penyegerakan bibir adalah istilah yang merujuk kepada proses memadankan pergerakan bibir dengan suara atau vokal lisan. Selain itu, animasi penyegerakan bibir masa nyata adalah kaedah yang akan menghidupkan watak yang dihasilkan oleh komputer maya dengan menyegerakan pergerakan bibir dan suara yang tepat dalam masa kehidupan sebenar. Namun, perkembangan pergerakan bibir yang realistik sukar dilakukan kerana bentuk dan kedudukan bibir mesti diselaraskan dengan bunyi yang diucapkan. Oleh itu, viseme digunakan sebagai teknik untuk parameter animasi asas dalam menilai persamaan visual antara pelbagai fonem yang merangkumi pemrosesan suara, pengecaman pertuturan dan animasi wajah komputer berdasarkan pertuturan manusia terutama dalam bahasa Inggeris. Ketepatan tangkapan gerakan mulut manusia dengan prestasi watak dalam masa nyata dan pengecaman pertuturan dianalisis. Oleh itu, realisme persepsi sangat penting untuk animasi penyegerakan bibir bagi menerapkan bentuk bibir manusia berasaskan viseme dalam bahasa Inggeris untuk pemetaan mulut dan suara dalam watak animasi masa nyata. Hal ini juga disediakan sebagai panduan untuk animasi penyegerakan bibir, menunjukkan bagaimana menggunakan taktik penyegerakan sederhana untuk meningkatkan ketepatan dan membuat kesan visual yang lebih realistik, serta bagaimana memasukkan ciri-ciri canggih ke dalam aplikasi penyegerakan bibir.

TABLE OF CONTENTS

	PAGE
DEDICATION	III
ACKNOWLEDGEMENTS	IV
ABSTRACT	V
ABSTRAK	VI
TABLE OF CONTENTS	VII
LIST OF TABLES	XI
LIST OF FIGURES	XII
LIST OF ABBREVIATIONS	XIV
LIST OF ATTACHMENTS	XV
CHAPTER 1: INTRODUCTION	1
1.1 Introduction.....	1
1.2 Problem Statement.....	2
1.3 Objectives.....	2
1.4 Scope.....	2
1.5 Project Significance.....	2
1.6 Conclusion.....	3
CHAPTER 2: LITERATURE REVIEW AND PROJECT METHODOLOGY	4

2.1 Introduction	4
2.2 Domain	4
2.3 Literature Review	5
2.3.1 Real-time Lip-sync Animation Concept	5
2.3.2 Lip Shapes and Positions in Humans	5
2.3.3 Speech in Animation	7
2.3.4 Speech, Phoneme and Viseme	7
2.4 Project Methodology	12
2.5 Project Requirements	13
2.5.1 Software Requirement	14
2.5.2 Hardware Requirement	14
2.6 Conclusion	14
CHAPTER 3: ANALYSIS	15
3.1 Introduction	15
3.2 Requirement Analysis	15
3.2.1 Project Requirement	15
3.2.2 Software Requirements	16
3.2.3 Hardware Requirements	17
3.3 Project Schedule and Milestones	18
3.3.1 Project Schedule	18
3.3.2 Milestones	18
3.4 Conclusion	19
CHAPTER 4: DESIGN	20
4.1 Introduction	20

4.2 Preliminary Design	20
4.2.1 Storyboard Design	20
4.2.2 Character Design	21
4.2.3 Lip-sync Design	22
4.2.4 Script	23
4.3 Conclusion	24
CHAPTER 5: IMPLEMENTATION	25
5.1 Introduction	25
5.2 Media Creation	25
5.2.1 Text Production	25
5.2.2 Graphic Production	27
5.2.3 Audio Production	27
5.2.4 Animation Production	28
5.3 Media integration	31
5.4 Product Configuration Management	32
5.4.1 Configuration Environment Setup	32
5.5 Implementation Status	34
5.6 Conclusion	35
CHAPTER 6: TESTING	36
6.1 Introduction	36
6.2 Test Plan	36
6.2.1 Test Prototype	36
6.2.2 Test Environment	37
6.2.3 Test Schedule	37

6.3 Test Strategy	37
6.4 Test Implementation	39
6.4.1 Test Description	39
6.4.2 Test Data	39
6.5 Test Results and Analysis	40
6.6 Analysis Testing	42
6.7 Conclusion	44
CHAPTER 7: CONCLUSION	45
7.1 Observation on Project Issues and Futures	45
7.1.1 Project Issues	45
7.1.2 Project Futures	46
7.2 Propositions for Improvement	46
7.3 Project Contribution	47
7.4 Conclusion	47
REFERENCES	48
APPENDIX A	49
APPENDIX B	54

LIST OF TABLES

	PAGE
Table 3.1 : Milestone of Project Development	18
Table 5.1 : Text Style Used in 2D Animation: Cat Strange Behaviour	26
Table 5.2 : Configuration Setup for Adobe After Effect	32
Table 5.3 : Configuration Setup for Adobe Premiere Pro	33
Table 5.4 : Implementation Status	34
Table 6.1 : Test Schedule	37
Table 6.2 : The Whole and Outline of Human and Character Lip Shape	38
Table 6.3 : Comparison Result between Animated and Real Lip-sync in Different Dimension (Whole Lip Shape)	40
Table 6.4 : Comparison Result between Animated and Real Lip-sync in Different Dimension (Outline of Lip Shape)	41

LIST OF FIGURES

	PAGE
Figure 2.1 : A Symbolic View of the Virtual Presenter by Zoric (2005).....	5
Figure 2.2 : Illustration of Human Lip Shapes by Martin J. Ball & N. Muller (2005).....	6
Figure 2.3 : The Chart of 44 English Sounds by Dawn Reithaug (2002).....	8
Figure 2.4 : The Chart of 44 English Sounds by Dawn Reithaug (2002).....	9
Figure 2.5 : The Chart of 44 English Sounds by Dawn Reithaug (2002).....	10
Figure 2.6 : The International Phonetic Alphabet (IPA) Phonemic Chart, Cambridge.....	11
Figure 2.7 : Chloe's Viseme Set by Deepali Aneja and Wilmot Li (2019)	12
Figure 3.1 : Gantt chart of Project Development	18
Figure 4.1 : Storyboard.....	21
Figure 4.2 : Character Goldennie.....	21
Figure 4.3 : Lip Shape of Viseme.....	22
Figure 4.4 : Script.....	24
Figure 5.1 : Text Production Process.....	26
Figure 5.2 : Graphic Production Process.....	27
Figure 5.3 : Audio Production Process.....	28
Figure 5.4 : Movements and Effect Edited in Adobe After Effects.....	29
Figure 5.5 : Parts of The Character Edited in Adobe Photoshop.....	29
Figure 5.6 : Animating and Computing Lip-sync in Adobe Character Animator.....	30
Figure 5.7 : Combining Process in Adobe Premiere Pro.....	30
Figure 5.8 : Animation Production Process.....	31
Figure 5.9 : The Media Integration Process.....	31
Figure 5.10 : Composition Setting in Adobe After Effect.....	33
Figure 5.11 : Sequence Setting in Adobe Premiere Pro.....	34

Figure 6.1 : Image Comparison Online Tool (https://www.textcompare.org/image/)	39
Figure 6.2 : Average Accuracy of 2D Animated Lip-sync	43



LIST OF ABBREVIATIONS

FYP	-	Final Year Project
2D	-	Two-Dimensional
Hz	-	Hertz
fps	-	frame per Second
IPA	-	International Phonetic Alphabet



اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF ATTACHMENTS

		PAGE
Appendix A	Collection Data of Accuracy of 2D Animated Lip-sync	49



CHAPTER 1: INTRODUCTION

1.1 Introduction

In animation, lip-sync is frequently used in the post-production phase of animated films. Yet, drawings do not have the ability to speak, the language must first be captured and analyzed before key-framing and animating them to speak. According to Huang and Chen (1998), real-time lip-sync animation is a methodology for generating mouth movement from acoustic speech data that is driven by the human voice. It also considers a human-to-computer interaction interface. The lip-sync technique is still used in animated films and television broadcasts and it is an important part of the animation process today. This is a method of performing a virtual computer produced character's speaking in real-time by synchronizing an accurate lip movement and speech signal.

A study of two researches at Adobe Research and the University of Washington, Deepali Aneja and Wilmot Li (2019) said that to enable live animation, a system must be able to capture a human actor's performance and map it to matching animation events in real time. In Adobe Character Animator, it translates a performer's facial expressions to a cartoon figure using face tracking. While these features allow performers to express themselves through the animation, speech is the most important aspect of practically any live animation performance. Lip-sync that is convincing allows the character to embody the live performance, but lip-sync that is not convincing removes the illusion of the characters being live participants. Therefore, in this project, lip movements and patterns must be perfectly coordinated with the audio in order to provide a genuine outcome.

1.2 Problem Statement

Creating a precise lip-sync animation, on the other hand, would be considerably more complex, particularly in terms of keyframe value. In fact, mapping the lip motions and sounds to be matched is very difficult. It's a time-consuming approach. To be detailed, the procedure is carried out manually, with frame-by-frame adjustments that typically require many passes of fine-tuning to fit the sound. As a result, a real-time approach is required to overcome the challenges of standard lip-sync techniques and to achieve realism in lip-sync animation. It is essential to match the lip shape to the sound to make the character animation genuine in this project.

1.3 Objectives

The objectives of this project are as below:

1. To study lip synchronization for 2D animation.
2. To produce 2D animation that incorporates lip-sync.
3. To compare the accuracy of animated with the real lip-sync.

1.4 Scope

The target audience for this project is anyone who wants to watch an educational animation while experiencing the reality of the animation's lip synchronization. This animation project is fully provided with subtitles for the accessibility of all kind of audience.

1.5 Project Significance

This paper will provide an automated digital speech model of viseme classification mapping to match the key phoneme sounds for English. This project also delivers some knowledge about lip synchronization for the audience to learn and feels its realism.

1.6 Conclusion

In conclusion, Chapter 1 covers the project topic “Cat Strange Behaviour”: Perceptual Realism of 2D Animated Lip Sync with the objective and goal of the project. To guarantee that the project runs smoothly within the time-frame, the project will follow the activities schedule of the Gantt chart and milestones given in the proposal. Literature review and project methodology will be undertaken in the next chapter.



CHAPTER 2: LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

This chapter explains the literature review and project methodology on related topic about the realism of 2D animated lip-sync with multiple perspective such as lip shape, speech and etc. In the literature review, previous studies and researches from publishing materials such as case studies, technical documents, and an online library play a significant role. For the project methodology, it will cover the process and the direction of the development and also the requirements such as software and hardware in order to develop the application.

2.2 Domain

The domain of the project in perceptual realism of 2D animated lip sync is animation. Figures are altered to appear as moving images in the animation process. Masson (2007) claimed that on the computer, 2D animation figures are made or edited using 2D bitmap images and 2D vector graphics. The two-dimensional image is a distinct artefact with added semantic meaning, not just a representation of a real-world object. With today's rapid increase in computational power, animation's attributes provide an extra layer of visually convincing realism. Hence, commercial tools for real-time performance-based 2D animation have now made it possible for 2D characters to appear on live broadcasts and streaming platforms.

2.3 Literature Review

2.3.1 Real-time Lip-sync Animation Concept

Hoon L. et al. (2014) studied the automatic lip sync or digital speech are terms used to describe real-time lip-sync animation. Automatic lip-sync or digital speech is also often known as computer speech recognition, which is a type of automated digital speech system that understands the computer's voice and performs any required activity. In the context of live performances or recordings, the speech drives the virtual character in real-time. The speaker is speaking while lip-syncing, interacting with the audience, and displaying animation on the screen. Figure 2.1 shows a symbolic representation of real-time lip-sync animation.

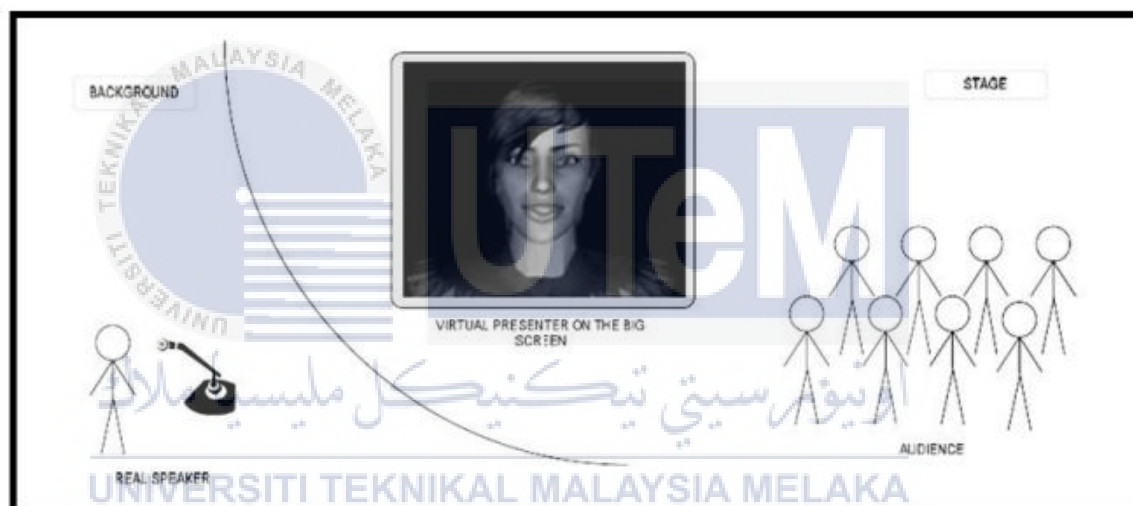


Figure 2.1: A Symbolic View of the Virtual Presenter by Zoric (2005)

In addition, real-time animation will also introduce the audience to the dynamics and changes of new media in live performances such as theatre, television, education, and live presentation. Thus, to produce an output perfectly in this animation project, the simulation of the avatar mouth's movement must always be synchronized with actual time and speech sound.

2.3.2 Lip Shapes and Positions in Humans

The phonemes are thought to be functions of the mouth's position. The variation in mouth position or lip shape will be compensated for by different speech intonations in vowels and consonants. L. Pardew (2012) stated that the movement of

the mouth as it generates the sounds of the dialogue is depicted in human speech. When we speak, the shape of our mouths change, and each sound has a particular appearance. S. Roberts (2012) suggested that the importance of mouth forms during speech is because they are very visible to the audience. Furthermore, according to M. J. Ball & J. Rahilly (1999), human lip shape is typically described as either rounded or unrounded, with rounded, neutral, and spread being a somewhat more extensive examination. The following is how they explained phonetics in English. With /ʌ/ having a neutral lip shape and /ɔ/ having a rounded lip shape, /i/ is interpreted as a spread lip, which is also known as an unrounded lip. Likewise, according to Peter Roach (2010), human lips can take on a variety of forms and positions, but there are only three possibilities. When speaking, the option is using a rounded lip, a wide lip, or a neutral lip. Figure 2.2 shows illustrations of human lip shapes.



Figure 2.2: Illustration of Human Lip Shapes by Martin J. Ball & N. Muller (2005)

Viewers benefit from the virtual human's lip shape since it provides visual information about speech sounds. It is classified using visual clues and viseme analysis, in which each spoken sound must be assigned a shape as part of the viseme method. Therefore, this project will follow the works in creating an accurate automatic lip sync animation performance.

2.3.3 Speech in Animation

In animation, a character's speech is prompted now no longer simply by voice input however additionally by visible cues which include lip motion and facial expressions. As stated by Zoric (2005), eyes, eyebrows, cheeks, mouth, nose, ears, chin, and jaw are the primary feature points of facial components that will influence facial movement. It is needed to decide human speech behavior so that it will map natural speech to real-time lip shape animation. Furthermore, Hofer et al. (2008) stated that the movement of the mouth throughout the speech is a crucial aspect of facial animation. A character that can accurately imitate lip sync will provide a sense of natural and genuine emotions. Besides, determining the motion of the mouth and tongue during speech is essential in the way to produce real-time speech animation. Rathinavelu et al. (2006) studied that the articulatory modelling of the lip alone is insufficient to achieve realistic speech animation. Both tongue and jaw additionally need to take into account. According to Tang, Liew, and Yan (2004), simulating a speaking animation requires controlling the proper muscle tissues from diverse phonemes. Muscles are used to construct greater practical mouth shapes, that are then used to extract lip parameters. According to studies, lip pattern, facial emotions, jaw, chin, tongue, facial muscle, and speech signal are all factors that influence human speaking behavior. These elements are necessary for the phase of visual mapping in which the speech is processed and classified into viseme categories throughout this animation project. The generated visemes are utilized to animate the face of the character in this animation.

2.3.4 Speech, Phoneme and Viseme

Speech is a vocalized mode of human communication in which articulate sounds are used to represent thoughts and feelings. The phoneme is determined by the position of the mouth, lip pattern, and tongue, which defines the intonation qualities of speech. Eventually, it is determined independently of the intonation features of speech. In a specific language, a phoneme is the perceptually unique unit of sound that distinguishes one word from another. L. Pardew (2012) claimed that there are 44 distinctive speech sounds in English. Vowels and consonants are the two major types of consonants. When making a vowel sound, the air flow is unobstructed. On the other hand, a consonant sound is one in which the air flow is cut off partially