Physically-based Animation of 'Sticky Lips'

Matthew Ian Leach
March 2020

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Supervisor: Dr Steve Maddock
Department of Computer Science
The University of Sheffield

Abstract

Producing a realistic animation of the face is challenging due to the familiarity people have with facial expressions and movements. In recent years there has been increased activity in the use of physically-based models to create realistic animations of soft-tissue structures, as well as interest in modelling more subtle effects occurring in the mouth. This thesis presents a physically-based model of the mouth. In particular, the model recreates the effect of saliva on the movement of the lips, a largely unexplored topic.

The research is composed of four novel components. The first component is a physically-based model of the mouth featuring a new stickiness model, recreating the effect of the saliva on the movements of the mouth. The model is supported by a novel moisture model which controls the stickiness level over time. The stickiness model itself provides more realistic behaviour than the few other current models and reproduces complex effects which can be seen in real mouths. The second component is a perceptual evaluation of the realism of mouth animations which incorporate stickiness. The evaluation concludes that the inclusion of the stickiness model results in an improvement in perceived realism of animations of the mouth. The third component is a new analysis process for capturing information about mouth movements from video. This analysis process is used to evaluate the developed model by comparing it against videos of real mouths. The analysis demonstrates that the stickiness model provides an improvement in accuracy of animation compared to models that do not incorporate stickiness. The fourth component is a corpus of mouth videos in which utterances and actions are recorded at varying levels of lip stickiness to produce high frame rate close up mouth videos which show stickiness effects in a variety of participants. This corpus is used in the objective evaluation.

Acknowledgements

I would like to begin by thanking the Engineering and Physical Sciences Research Council (EPSRC) and Department of Computer Science for funding my work. I am grateful to Dr Zeike Taylor for his advice regarding NiftySim and the use of the TLED finite element method for soft tissue modelling. The advice given by my panel members, Professor Jon Barker and Dr Paul Watton has also been invaluable. A big thank you to my friends at GBC and the Latin and Ballroom society who have kept me laughing, smiling, and probably at times a little too distracted, as well as to Billie for many a coffee-fuelled reassuring chat. Finally, I would like to extend my deepest gratitude to my supervisor, Dr Steve Maddock for his patience, hard work and continuous support, my family for their faith and advice, and my partner Lydia, for telling me I need to sit down and get on with it. These people were instrumental in the completion of my thesis, and I really wouldn't have made it had any of them not been there for me.

Contents

List of Figures

List of Tables

List of Abbreviations

1D 1-dimensional

3D 3-dimensional

AAM/ASM Active Apearance Models/Active Shape Models

ALE Arbitrary Langrangian-Eulerian

CDT Constrained Delaunay Tetrahedralization

CG Computer Graphics, Computer-generated

CGI Computer-generated Imagery

CT Computed Tomography

FACS Facial Action Coding System

FEM Finite Element Method

GPU Graphics Processing Unit

HSV Hue, Saturation, Value

MPM Material Point Method

MRI Magnetic Resonance Imaging

RGB Red, Green, Blue

RGBD Red, Green, Blue, Depth

TLED Total Lagrangian Explicit Dynamics

Chapter 1

Introduction

The mouth is the most expressive part of the human face due to the number of muscles involved and the mobility of the tissue (Expression and Series, 1969). As such, effective modelling and animation of the mouth is paramount to producing convincing and realistic simulations of human speech and expressions. Improvement of mouth modelling could provide benefits to many industries, most notably the entertainment industry for films and games, and the medical industry for visual speech applications and simulation of surgery.

The first major work on animation of the mouth was conducted by Frederic I. Parke in the 1970s. Parke produced the first three-dimensional facial animation, and later in 1982, a parameterized model (Parke, 1982). Since then, the field has developed considerably due to advances in both research and computational power. Today, many methods exist for facial animation, including parameterizations, blend shapes, motion capture, generative neural networks, mass-spring muscle simulations and finite element (FEM) method models (Noh and Neumann, 1998).

In recent years, motion capture techniques have become viable for producing animations. These can use markers or markerless techniques, but, in both cases, involve cameras tracking the movement of the body or face. Motion capture provides great advantages in terms of producing realistic movements very quickly, although has issues capturing fine details, particularly in the face where this is important to producing a convincing animation. Additionally, motion capture data is often used to train learning-based statistical models (Huang et al., 2011; Chai and Hodgins, 2007). Hybrid models have also been developed which use motion capture data to drive physically-based or blendshape-based models (Sifakis et al., 2005; Deng et al., 2006). Generative models use neural networks to generate new faces and animations. They also require training data, which may be in the form of video footage or motion capture data. Generative methods are very powerful, but are highly dependent on the training data. Physically-based models could be used to provide high quality, realistic training data in cases where it may be difficult to reliably capture particular effects or movements from real people.

Physics-based models vary in complexity, but generally model the muscles and skin and additionally sometimes the skull. The two most common approaches to physical modelling are *mass-spring* systems and the *finite element method*. The mass-spring method simulates physical systems by approximating them as a series of point masses connected by springs (Kahler et al., 2001; Zhang et al., 2001; San-Vicente et al., 2012). The method is fast and produces reasonable results, however, it is not accurate enough if realistic results are desired. In comparison, the finite element method (Zienkiewicz et al., 2005) produces considerably more realistic and accurate results, at the expense of computation time.

This thesis uses the finite element method to simulate the behaviour of the mouth. A novel saliva model is implemented to include the effect of saliva between the lips. Specifically, the saliva can act as a glue between the lips, causing them to stick together when the mouth is opened or the lips are drawn apart. This glue effect leads the lips to open gradually outwards

from the center rather than along their entire length at once. Varying levels of stickiness across the lips can lead to asymmetric openings or the formation of multiple mouth openings. The inclusion of the saliva model also helps to introduce subtle differences in the movement of the lips, which are lost in traditional animation techniques. There have been some previous attempts to recreate these stickiness-dependent behaviours in animation. Most attempts come from artists wishing to include the effect in their works, leading to plugins for 3D modelling software such as Maya (Kimon Matara, 2017). These, however, are added on to an already existing animation, meaning the effect of the stickiness is not transmitted to the tissues surrounding the lips, are not physically-based and are artist driven, leaving plenty of room for error. Recently, academic work has begun to tackle the problem using physically-based models. Barrielle's work uses a series of zero-length springs with a stochastic breaking model (Barrielle and Stoiber, 2018). However, this is not representative of the physical behaviour of the saliva itself.

1.1 Research Questions

This research aims to improve the quality of animation of the human mouth by replicating the physical behaviour of the real mouth. Specifically, the aim is to reproduce the behaviours caused by the presence of saliva between the lips. In order to achieve this, the following research questions are considered:

- In what ways does the saliva between the lips affect the movements of the mouth?
- Does the effect of the saliva change over time?
- What is the best approach for modelling the effect of saliva on mouth movements?
- Does the inclusion of sticky lip effects improve animations of the mouth?
- Does the speed of a mouth movement affect the preferred stickiness level?
- Does the inclusion of audio affect the preference for stickiness in an animation?

1.2 Contributions

The thesis makes the following contributions:

- The creation of a finite element model of the human mouth. The model is constructed from reduced integration linear 8 node hexahedra and solved using the total Lagrangian explicit dynamics formulation of the finite element method. The simulation uses novel modelling of the effect of saliva on the lips, producing sticky lip effects (Leach and Maddock, 2018). The simulation itself is based on a modified version of NiftySim (Johnsen et al., 2015). The modifications were developed as part of the thesis work. The software has been updated to run with the latest versions of CUDA and extended with custom material models to support the saliva elements. Additionally, a new constraint type was developed to support the different types of muscles and their activation patterns. The material model is supported by a novel moisture model, which drives changes in the level of stickiness over time. The moisture model is based on the physical equations describing evaporation. This allows for natural changes in behaviour over time as the lips dry, based on real environmental parameters.
- A perceptual evaluation of the impact of stickiness on mouth animations. The evaluation used a study asking participants to view pairs of mouth animations and select the one

they thought was more realistic. It was found that participants generally preferred animations which incorporated stickiness.

- A method for evaluating and comparing the movement of mouths, animated or real, using a set of tracked metrics. This method is used to evaluate the model created in this thesis. The evaluation process was initially tested using a variety of mouth openings. Three different mouths (two male and one female) opening were used, performing a range of movements. The method successfully detects distinct mouth behaviours such as 'zippering' as the mouth opens. These are recreated in the simulation, and behaviour profiles quantitatively demonstrate the improvements as a result of the inclusion of the sticky lip effect (Leach and Maddock, 2019).

- The creation of a corpus of videos of 15 people's mouths, speaking a set of words and phonemes with varying levels of lip moisture. Each video captures the lower half of the face, and features the participant performing a range of mouth movements. Prior to and at certain points throughout the recording, participants licked their lips, leading to varying levels of stickiness throughout the recordings. The aim is to publish this video corpus as part of a new paper on evaluating sticky lip effects.

1.3 Thesis Outline

Chapter 2 provides relevant background knowledge. It covers the human anatomy, particularly that of the mouth. It also reviews previous approaches to animating the mouth, focusing on physically-based models and attempts to model lip stickiness. It also covers evaluation techniques for testing mouth models.

Chapter 3 provides details of the simulation approach. An overview of the process used in the project is given. This includes information regarding the pipeline developed for producing a mouth model from photos of a user and a summary of the software used and simulation techniques and an outline of the evaluation method created. The remainder of the chapter covers the simulation of the mouth model using the total Lagrangian explicit dynamic (TLED) finite element method, as well as details of the novel saliva modelling used to recreate the sticky lip effects.

Chapter 4 shows the results of the model. It shows a range of effects that the model developed in chapter 3 is capable of producing.

Chapter 5 looks at the evaluation process created for comparing the realism of different mouth models. This incorporates the creation of a dataset of videos of mouths speaking a series of letters and words, with rests between each repetition to allow the saliva time to dry to impact the stickiness between the lips. This video corpus is analysed using a novel evaluation method and metrics to generate profiles of various mouth movements at varying levels of stickiness. The evaluation method is then applied to the results generated by this project, as well as current state of the art animations. Comparisons are drawn to demonstrate the effectiveness of the presented method.

Chapter 6 concludes the thesis, highlighting key results and the potential for follow-on work.

Chapter 2

Literature Review

2.1 Introduction

This chapter provides background knowledge providing context for the work presented in this thesis. Section 2.2 gives an overview of human anatomy, specifically covering the anatomy of the mouth. This includes the properties of soft and hard tissues found within the body, as well as the musculature and structure of the mouth. Additionally, the properties of saliva and its influence on the movement of the mouth are discussed here. Section 2.3 introduces methods of facial animation which are not physically-based. Section 2.4 introduces a range of physically-based modelling methods and explains challenges common to the approaches, along with their typical solutions. Section 2.5 builds on these sections to explore the state-of-the-art literature regarding physically-based mouth animation. Section 2.6 covers typical evaluation methods used for facial animation.

2.2 Anatomy

2.2.1 Anatomy of the Face and Mouth

The human mouth, technically known as the oral cavity is the beginning of the human alimentary canal. The primary roles of the mouth are to enable digestion via reception and chewing of food, to aid communication by producing different shapes to affect the sound produced by the larynx, and as an air inlet for breathing. The mouth can be thought of as a combination of the oral cavity and the tissues surrounding it. The oral cavity itself is the space in the mouth. This can be further divided into two regions by considering the region between the teeth and lips as one region, the vestibule, and the region behind the teeth as another, the oral cavity proper.

The oral cavity proper is bounded by the hard and soft palates, the cheeks, the teeth and the tongue. The hard palate is found towards the front of the mouth on the upper surface. It is a bone plate which divides the oral and nasal cavities. The side bounding the oral cavity proper is lined by oral mucosa. The soft palate is muscular in nature and moves to open or close pathways to the interior of the body. Together, the hard and soft palates form the roof of the mouth. The remaining sides of the oral cavity proper are the cheeks, and the teeth. The cheeks are oral mucosa on the inside of the mouth, skin on the outside, and the buccinator muscle runs through them. They contract via the buccinator muscle to keep food in the correct area for chewing, and are also used in some facial expressions.

The majority of the floor of the mouth is taken up by the tongue, connected to the lower oral mucosa via the frenulum of the tongue. The tongue is used for speech as well as transporting food to the back of the mouth to swallow.

The lips form the mouth opening. Humans have two lips, the upper lip, “labium superius

oris" and the lower lip, "labium inferius oris". The skin of the lips is thinner than most of the skin on the face, consisting of only three to five layers of cells, in contrast with sixteen for average skin on the face. The skin here also contains less pigment, meaning the blood beneath the skin shows through more strongly giving rise to the characteristic red colour of the lips. The skin on the outside of the lips is stratified squamous epithelium, while the portion of the lip inside the mouth is covered by a mucous membrane.

The main red area of the lips is called the vermillion zone, and the border between the vermillion zone and the remainder of the face is called the vermillion border. The orbicularis oris muscle runs through the centre of the lips forming a continuous circular band and is surrounded on either side by adipose tissue. This in turn is covered by the mucous membrane on one side and the dermis on the other.

Bones

The bones affecting the movement and structure of the mouth can be considered to be the facial bones. These are the fourteen bones forming the lower half of the skull (Figure 2.1). The remainder of the skull is called the neurocranium and surrounds the brain and upper half of the eyes. The bones making up the outer facial skeleton are as follows

- Mandible
- Maxilla
- Zygomatic Bones
- Nasal Bones

Of these, the first 3 form the area we consider the mouth. There are two zygomatic bones, one on either side of the face. The maxilla is formed of two bone segments which are fused, and the mandible is a single bone. Colloquially, the zygomatic bones are known as the cheekbones, and the mandible is known as the jaw. The mandible is hinged where it joins with the rest of the skull. The lower teeth are set in the mandible, and the upper teeth in the maxilla.

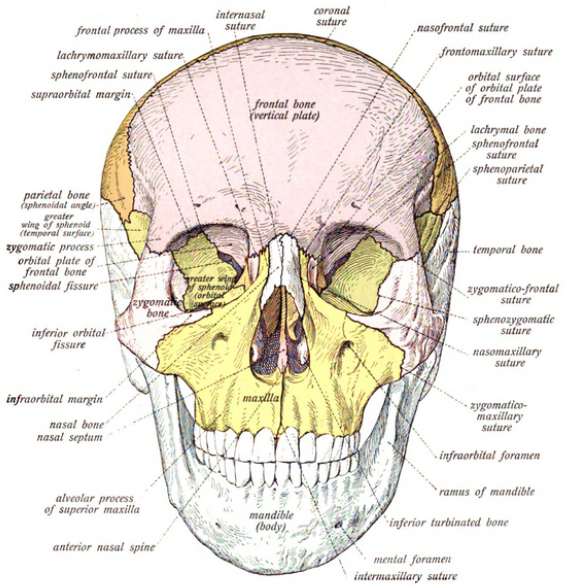


Figure 2.1: The bones of the skull. Image from Sobotta's Human Anatomy (1909). Public domain.

Muscle	Type	Action
Levator labii superioris alaeque nasi	Linear	Dilates nostrils and elevates upper lip
Levator labii superioris	Sheet	Elevates outside of upper lip
Zygomaticus major	Linear	Draw the corners of the mouth upwards at an angle
Zygomaticus minor	Linear	Elevates the upper lip
M depressor anguli oris	Linear	Depresses the corners of the mouth
M depressor labii inferioris	Sheet	Depression of the lower lip
Mentalis	Linear	Wrinkles and raises chin, pushes lower lip outwards
Risorius	Linear/Variable	Draws mouth outwards and slightly upwards
Levator anguli oris	Linear	Elevates mouth angle as in a genuine smile
Orbicularis oris	4 Quadrants of muscle fibres, roughly a sphincter	Puckers the lips
Buccinator	Quadrilateral	Draws cheeks against the teeth

Table 2.1: The muscles affecting the movement of the mouth and their types and actions

Muscle	Attachment	Insertion
Levator labii superioris alaeque nasi	Upper process of maxilla	Nostril and labii superioris
Levator labii superioris	Medial infra-orbital margin, beneath the eye	Outside edge of labii superioris
Zygomaticus major	Outside of zygomatic bone	Upper side of modiolus of mouth
Zygomaticus minor	Zygomatic bone	Skin of labii superioris
M depressor anguli oris	Tubercle of mandible	Lower side of modiolus of mouth
M depressor labii inferioris	Mandible, between the symphysis and the mental foramen	Integument of labii inferioris, orbicularis oris fibres
Mentalis	Anterior mandible	Chin
Risorius	Parotid fascia	Modiolus of mouth
Levator anguli oris	Maxilla	Modiolus of mouth
Orbicularis oris	Maxilla/mandible	Skin of the lips
Buccinator	Alveolar processes of the maxillary bone and mandible	Fibres of orbicularis oris

Table 2.2: The muscles affecting the movement of the mouth and their attachments and insertions

Muscles

The muscles in the face can be categorised into three groups: linear, sheet and sphincter (Kahler et al., 2001). Linear muscles are linear in shape and contract along the length of the muscle. The zygomaticus major is an example of a linear muscle. Sheet muscles are planes of muscle which contract isotropically along one axis. The occipitofrontalis is an example of a sheet muscle. Sphincter muscles are rings of muscle which contract towards their centre. The most relevant example of a sphincter muscle in this project is the orbicularis oris. The muscles influencing the mouth and their types and actions can be seen in table 2.1. Table 2.2 shows where the muscles are attached and inserted. Figure 2.2 shows the muscles of the face.

2.2.2 Saliva

In humans, saliva is a water-based fluid found in the mouth. It is produced by salivary glands which are located in various locations in the mouth (Figure 2.3). Saliva is made up of more than 99% water (Humphrey and Williamson, 2001), as well as sugars, electrolytes, mucus, enzymes, white blood cells and epithelial cells (Schwarz, 1987; Levine, 1993). The function of saliva in the mouth is to lubricate the upper portion of the digestive tract, facilitating swallowing of food. Additionally, the saliva prevents the oral mucosa from becoming dried out.

Humans commonly lick their lips - most notably when eating. This leads to a transfer of saliva from the tongue to the surface of the lips. Once on the lips, the saliva will begin to dry out. Indeed, the saliva inside the mouth is also continually drying and must be replenished. If insufficient saliva is produced, a condition known as *dry mouth* can present (Fox, 1997). This is often caused by dehydration, although can also be induced by certain medicines, breathing overly through the mouth, or anxiety, among other things. In these cases, it is common for the saliva transferred to the surface of the lips to dry a considerable amount.

As the saliva dries, it leaves behind a sticky residue. This is formed of the non-water components. The other components form viscoelastic fluids and intermolecular interactions can result in the formation of a cohesive gel. The study of flow of materials is known as

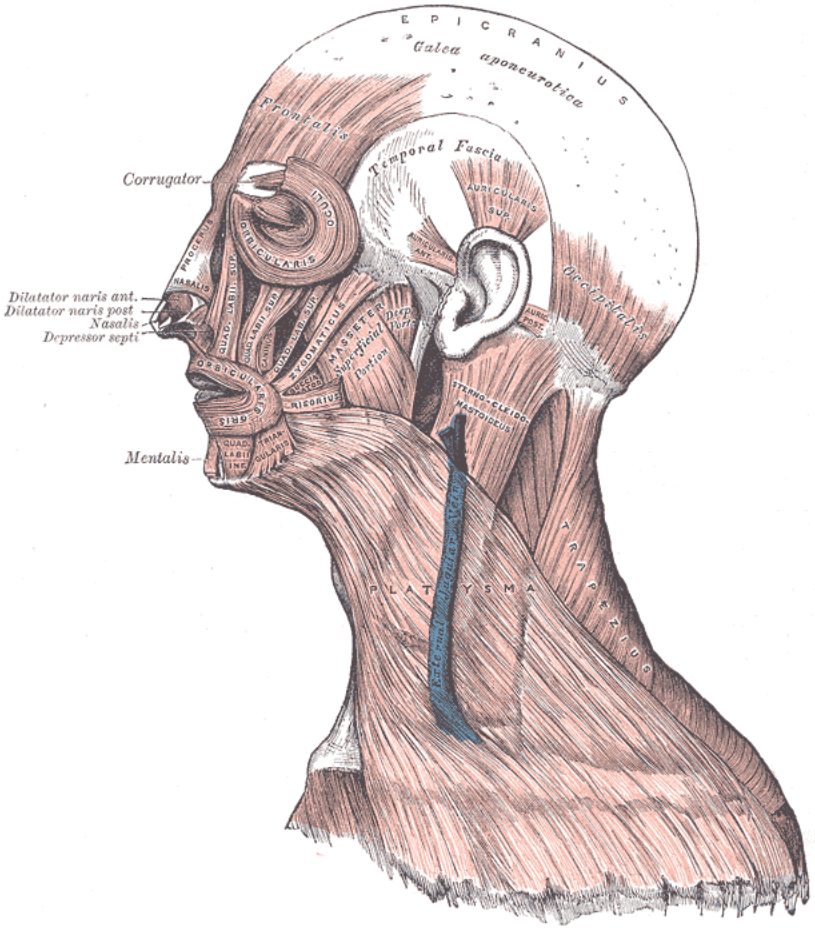


Figure 2.2: The muscles of the face. Image from Gray's Anatomy (1918). Public domain.

rheology. The drying of the saliva can greatly affect its rheological properties, and in turn affect the interactions between surfaces it is found on.

When the water content is high, the saliva acts as an effective lubricant between the lips, with an effective coefficient of friction of $\mu \approx 0.02$ (Bongaerts et al., 2007). In this case, the lips easily slide over each other and separate easily if the mouth is opened. The drier the saliva becomes, the more viscous it becomes as the relative water content decreases, and the density of the dissolved sugars etc. increases. The increased density leads to more interactions, resulting in a thicker, stickier fluid. At this stage, frictional shearing forces between the lips are increased, and the cohesive properties of the saliva can begin to 'glue' the lips together. As the saliva continues to dry, the stickiness and viscosity increase up to a certain point. Beyond this point, the water content decreases so much that the gel begins to solidify. Once completely dry, there is no stickiness between the lips, and the lips move over each other more easily, as the gluing effect is no longer in play. Figure 2.4 shows a graph representing this behaviour. The shape of the graph is assumed to fit the aforementioned properties in the absence of available data on the true relationship. With no moisture, the lips are completely dry and there is no stickiness. As moisture increases, the stickiness level increases as the lips become tackier until there is sufficient moisture to start thinning the saliva. As the moisture level increases further, the stickiness tapers off until there is only the surface tension due to the liquid remaining, which is negligible. The saliva can cause asymmetric mouth openings if it is stickier on one side (Figure 2.5), as well as causing the corners of the mouth to stick (Figure 2.6).

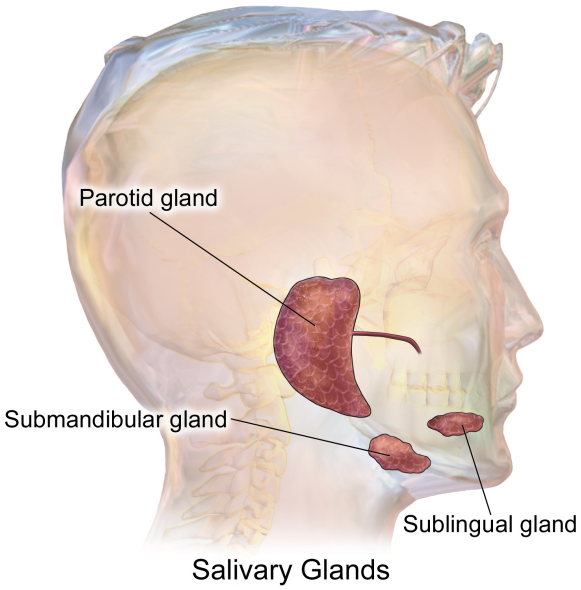


Figure 2.3: The location of the salivary glands in the human mouth. Image reproduced under CC BY 3.0 License, courtesy of Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014".

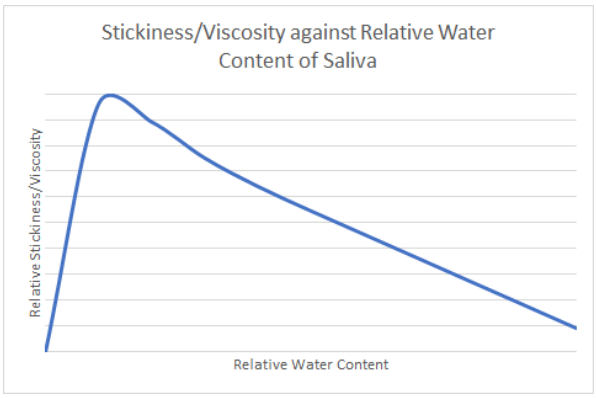


Figure 2.4: Representative graph showing how water content could affect stickiness. Zero water content leads to no stickiness as the saliva is in a solid phase. With a small amount of water present, the saliva forms a cohesive gel. At higher values, the saliva behaves as a lubricant.

Figure 2.5: An asymmetric mouth opening.

Figure 2.6: Saliva is visible forming a thin film in the corners of the mouth. Here it is not yet dry enough to exhibit gluing behaviour.

2.2.3 Capabilities of the Mouth

The primary movement of the mouth is its opening and closing. This movement is most strongly driven by the movement of the mandible via the contraction and relaxation of the masseter muscle. This movement is used within many common human actions, such as chewing food, speaking, and producing various facial expressions. The jaw is also capable of moving forwards and backwards through activation and relaxation of the lateral pterygoid muscle.

The second largest scale area of movement in the mouth region is the lips. The lips contain the orbicularis oris muscle; this is a sphincter muscle. The lips are also affected by other muscles throughout the face, and are capable of forming a range of different shapes. The shape of the mouth can also be affected by the teeth and the tongue. The tongue is capable of applying pressure to most of the mouth.

Another factor which can influence the movement of the mouth and lips is the energy an utterance is spoken with. Differences in energy could be categorised into two types: differences in the energy of the speech and differences in the energy between different movements. The first category is dependent upon the volume and force with which a particular sound is produced. The second category is dependent upon the type of movement, for example, bilabial plosives feature a sudden release of energy as air is released from the mouth cavity, whereas speaking a vowel sound might produce almost no movement in the lips. Figure 2.7 shows a range of shapes which the mouth can produce.

Figure 2.7: A range of shapes the mouth is capable of producing. Courtesy of James Edge (Edge, 2004)

2.3 Non-physically-based Facial Animation

Animation and modelling of the human face and mouth is a difficult yet rewarding task, with applications ranging from films and games to medical simulations. Simulations of face and mouth movement can be grouped into one of two fundamental categories: non-physically-based or physically-based. Whilst the focus of the thesis is a physically-based approach, it is

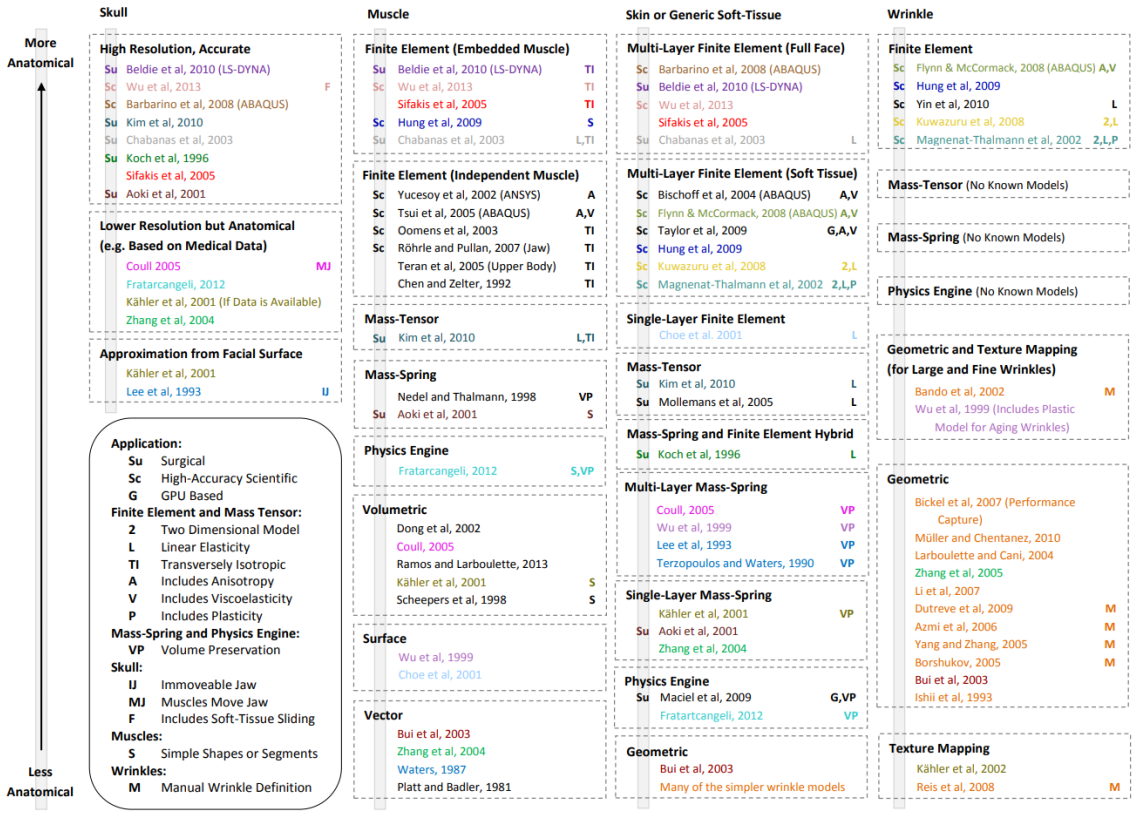


Figure 2.8: A categorisation of soft tissue modelling work. Courtesy of Mark Warburton (Warburton and Maddock, 2015).

worth preceding this with an overview of other approaches to facial animation, since some of this work is relevant to the discussion of mouth animation in section 2.5. As such, this section gives an overview of common non-physically-based approaches, namely surface deformation models, blend shapes, motion capture and learning-based models.

Figure 2.8 shows the various approaches to modelling soft tissues, who they were taken by and when. Approaches are grouped by similarity, and are ordered according to how closely they mimic the actual underlying anatomic structure. Physically-based models are towards the top, and non-physically-based approaches such as blendshape models towards the bottom.

2.3.1 Surface Deformation Models

Surface deformation models simulate the face as a single continuous surface, which is manipulated directly. There is no attempt to model the underlying musculature. The first work of this kind was carried out by Parke in 1974 using a model face. Manipulation of the face was restricted to basic operations: translation, rotation, scaling and interpolation between positions (Parke, 1974). The motion of the lips was produced by vertex position interpolation between a smiling expression and a neutral expression, scaling, and translation towards or away from the teeth. The lower lip moved with rotation of the jaw, and the corners of the mouth could also be translated (Parke, 1974).

In 1986, Lewis and Parke produced another model used for automatic synchronisation of mouth movement with speech, which produced more accurate lip movement using key framing (Lewis and Parke, 1986). The main disadvantage to this method is that extreme positions must be pre-programmed and it is difficult to generate alternate motions. For example, generating an expression such as a smile could be very difficult without a smiling key frame present.

Due to the variability of the mouth and lips, accurately reproducing arbitrary motions of the mouth through key framing is challenging at best.

In 1990, Lance attempted to improve the realism of facial animation by mapping a 2D animation onto a 3D model, which he described as a "3D mask" (Williams, 1990). The model looked considerably more realistic than other work of the time as photos of real people were used for the texturing, however the animation itself was not hugely effective. The texture was morphed based on motion capture point data, resulting in very unnatural warping of the textures. The model itself was also static, and the approach didn't allow for opening of the mouth and eyes (Williams, 1990). This technique has fallen out of favour in comparison with direct animation of the model itself.

2.3.2 Blend shapes

A significant number of animations of the face are currently produced using blendshapes (Lewis et al., 2014). A blendshape animation is constructed using a set a rest frame of the face in a default position, and a series of other key expressions. These represent the set of expressions which best maximally represent the individual motions of the face. Each is recorded as a set of differences from the rest pose, and these can be combined in linear combinations to produce different expressions. Although blendshapes are typically created manually by an artist, they can also be generated from motion capture data (Liu et al., 2008).

To drive a blendshape model with motion capture data, an inverse solver finds the blendshape weights which optimally reproduce a captured movement by minimising the difference between the movement of the blendshape animation and motion capture data. More recently there has been work to automatically retarget animations (Ribera et al., 2017). Any inaccuracies caused by the blending process or troubles with the reconstruction of expressions on the new model must then be hand corrected by the artist. Motion capture can be used to drive a blendshape animation. Motion capture driven blendshape models can be augmented with classification techniques to improve prediction of optimal blendshape weights (Ravikumar et al., 2016).

Blendshape animations can often include inaccuracies as a result of the linear interpolation used, as well as being limited by which expressions can be constructed from the available blendtargets. Corrective are specifically constructed blendshapes which can be introduced to try to address key problems with an animation. Stickiness between the lips is often currently managed using correctives in industry.

2.3.3 Motion Capture

Motion Capture is a technique using some form of camera and software to track and record movements of a live subject. The motion can then be replayed as is, or transferred to another rig to animate another model. It has become a ubiquitous technique in the production of film animations, whereby human actors are used to ensure lifelike motion of CG characters. The most common approach to this is using a full body suit with attached markers. Cameras located around the actor accurately capture the motion of the markers. A mapping is created between the markers and the model of the CG character, and the data subsequently is used to drive the character's motion. Other major applications for the technology in its various forms are surveillance, control interfaces and analysis of motion such as gait analysis. In recent years with increased processing power and research, markerless motion capture has also become more practical.

Two methods have become common for animating the face with motion/face capture. The first involves tracking an actor's face to build a set of static expressions. The expressions are transferred to a facial rig, on which the expressions are then combined or interpolated between in some way to produce an animation. The second method involves tracking the surface of

the actor's face to create an animation directly. The first system offers advantages in that it is easier for an artist to modify the subsequent animation, however the animations are generally not as convincing as only facial expressions which can be expressed as a combination of the static poses can be created. It is possible for animations using the second method to be edited, although only by simply chopping and recombining pieces of the original animation, or frame by frame editing from an artist.

Motion Capture techniques loosely developed from a process called rotoscoping used as early the 1930s for animation. A live action film was projected onto a glass pane, and traced over by an artist to produce realistic animated actions. This was used in Disney's 1938 film, "Snow White and the Seven Dwarves". With the introduction of computers, primitive motion capture techniques began to develop, initially used for biomechanical investigation.

Gollum in "The Lord of the Rings" was animated using a system much closer to modern motion capture techniques. Gollum was also intended to be animated using facial motion capture. The process involved the actor temporarily wearing a mask with holes drilled through. A make up pencil was used to mark dots on the actors face through the holes in the mask, and these acted as tracking points for the system. The mask ensured the same points were in place each day. The points were mapped to particular muscle groups/movements in the CGI model and used to recreate expressions. In the end there was not enough time to get the system functioning fully, and Gollum's face was key-frame animated instead, demonstrating the young age of the technology (Serkis, Andy, 2016).

The entertainment industry has also made use of facial motion capture in the gaming industry. L.A Noire provided notable development in the industry. The developers used a rig of cameras surrounding the actor seated in the centre to capture a volume around the actor (Rockstar Games, 2016). The cameras are placed at varied angles too to ensure no information is missed. The resulting 3D videos are generally convincing, but must be compressed for use in game. This compression loses a lot of the fine detail that is key to making movements look realistic, particularly when it comes to the mouth.

The data generated by these motion capture techniques generally falls into one of two categories. The first of these is skeleton-based. The most commonly used example of a format of this type is the Biovision BVH format. It consists of 2 parts. The first part, a header, defines the hierarchy of the components of the skeleton and the initial pose. The second part is a data section, describing the movement of the hierarchy components through time.

The second category of motion capture formats are found more often in facial motion capture than full body motion capture, and are based on recording data about an entire surface or mesh. This data type is generally produced by markerless systems. The scene is recorded with multiple cameras at different angles. A stereo solver is used to produce a mesh capturing the movement of the surface. Although this approach captures a lot more detail than marker-based solutions, fine movements can still be lost as the mesh resolution may be limited and there is significant room for error in the reconstruction process.

Motion capture approaches track the movement of the face overall very well, however, they also suffer from difficulties attempting to reproduce fine details. Marker-based solutions only track certain points on the face and use these to drive a mesh. This means that there isn't any direct information about what is going on between the markers. A higher resolution of markers placed on the subject goes someway to solve this, but there is a limitation to the number that can be used. Markerless motion tracking doesn't suffer from this, however it is limited by other things such as the resolution of the cameras used, occlusion of particular areas (the inside of the lips often cannot be seen), or the framerate of the recording. The framerate is particularly important when aiming to capture dynamic effects. The lip edges are often also difficult to distinguish in video, especially when in motion, leading to artists correcting the animations produced by motion capture. A typical method for improving the

capture of the behaviour of the lips is to use very small lip markers, either point based, or a grid of differently coloured segments. This aims to increase the capture fidelity in the mouth region.

Recent motion capture methods are capable of producing effective results from monocular video. Xu et al. (Xu et al., 2018) use a neural network to perform 2D and 3D pose estimations, from which a 3D skeletal animation of a human in motion is reconstructed. This method does not however provide accurate animation of the face. Zollhöfer et al. offer an excellent summary of the state of facial animation using monocular video (Zollhöfer et al., 2018).

2.3.4 Learning-based methods

Learning-based models are trained on data to construct a statistical model. Learning-based methods are often used to drive blendshape models, such as in Xu et al.'s work (Xu et al., 2013) and Joshi et al.'s work (Joshi et al., 2006). Learning-based methods can produce effective results in real time (Weise et al., 2011). They can also be extended with corrective shapes to improve animation quality (Li et al., 2013). Typically learning-based methods are applied to 2D problems, such as generation of video. There are a lot of additional challenges involved in reproducing 3D animations using learning-based methods, such as correctly identifying depth information in the source footage. The increased availability of RGB-D cameras has led to some developments in recreation of 3D scenes (Malleson et al., 2019). Other approaches include training a 3D geometric model to respond to input (Basu et al., 1998; Xu et al., 2013). Recent developments also allow the editing of facial animations through the use of learning methods (Berson et al., 2019). Data driven methods have also been used for skinning (Liu et al., 2020).

Statistical or stochastic models are very effective at capturing general behaviours. They can, however, struggle to reproduce the finer details of a process, or indeed those which happen infrequently. In addition, a stochastic model typically requires training on large amounts of data. Such a dataset of significant size does not currently exist for lip stickiness. Given that the interest is in these fine details which stochastic methods may miss, and that the data to train a model isn't readily available, statistical modelling is inappropriate for this task.

Another class of learning-based methods is generative modelling using deep learning neural networks. Ranjan et al. (Ranjan et al., 2018)used a convolutional neural network trained on a set of face meshes. The model effectively reproduces the expression-space with a lower reconstruction error than another state-of-the-art face model. It uses a non-linear representation allowing representation of extreme deformations and non-linear expressions. Generative adversarial nets (Goodfellow et al., 2014) form the basis of many modern photo-realistic generative models. These networks use both a generator and classifier, with the generator aiming to produce an output which the classifier maximally estimates to be real. Tewari et al. (Tewari et al., 2020) give an excellent summary of works using such methods. Within the domain of face modelling, this approach has notably been used by Li et al. (Li et al., 2020) to high fidelity digitized faces with physically-based attributes.

One potential issue with using learning based methods is that they require large amounts of training data to produce accurate behaviour. The capture process can take a considerable amount of time and effort, and the equipment required to capture the data can be expensive. Physically-based models could potentially be used to produce large quantities of physically-plausible training data. This is particularly advantageous for effects which may be hard to capture using traditional techniques - for example, the lips forming multiple mouth openings due to a local region of high stickiness happens relatively infrequently in regular speech, but could easily be recreated using a physically-based model.

2.4 Physically-based Facial Animation

Physical modelling techniques aim to create animations using modelling of physical behaviours. Physical modelling typically involves the production of a 3D mesh representation of the domain and a physics-based mathematical description of how the system should behave. The system is then solved for at each time-step until the desired animation length has been reached. The 3D meshes are usually constructed of individual elements, which do not necessarily have to match the dimensionality of the model, for instance, a cube can be constructed out of one dimensional line elements. Although a cube and other 3 dimensional objects can be represented using line elements, a volumetric element is usually superior as solution fields are defined fully across the elements.

Common examples of mesh-based physical models are the mass-spring method, the finite element method and energy-minimisation methods, e.g. (Ichim et al., 2017). Mesh-free methods also exist which typically use particles to represent portions of the continuum rather than elements. Examples of mesh-free methods are smoothed particle hydrodynamics and the finite point method. Although all the aforementioned methods are discussed in the following sections, the emphasis is on the finite element method and hybrid models as these best match the solution approach in this thesis.

Although physically-based methods are very effective at producing plausible, realistic behaviours, they are not without their limitations. Many of these limitations relate to the art-directability of the model. Blendshapes present a very intuitive method for editing animations with direct blending between expressions. In contrast, physically-based models are driven by physical parameters, such as muscle activation values. A further limitation is that the computational time required to produce an animation means that making even minor adjustments requires waiting time for the simulation to run again.

Modern hardware goes someway to alleviate these problems, as parallel simulation techniques leveraging GPUs provide much faster simulation times. The performance issue can also be tackled by producing a much faster approximation of the model through the use of reduced-order modelling or using the model to train a neural network. This can then be used by the artist to get a good impression of how a parameter change will affect an animation very quickly, and then the physical model can be used for the final render to provide the most accurate behaviour.

2.4.1 Mass-Spring Models

Mass-spring models aim to simulate a problem by approximating the problem as a series of point masses connected by springs. The points can be organised into layers and physical properties can be mimicked by adjusting the properties of the springs within a layer. A basic mass-spring model of the face includes 3 layers of points. The innermost layer represents the bone, the middle layer represents the interface between muscle and skin, and the outer layer represents the epidermis. The two gaps between the layers represent muscle and the dermis itself (Keeve et al., 1998). Each spring can have a different stiffness to model different physical properties. Typically, Hookean, or classical springs are used. These have a linear force-extension relationship. For a classical spring, the force applied to each node is given by equation 2.1:

$$\mathbf{F} = k\mathbf{x} \tag{2.1}$$

where $\mathbf{F}$ is the vector of the resultant force, k is the spring constant, corresponding to the stiffness, and $\mathbf{x}$ is the extension of the spring from its resting length. Where the nodes are connected, a set of simultaneous equations can be constructed by summation of the forces from each spring on a given node. These can be formulated into a matrix equation and solved via common techniques such as Gaussian elimination or matrix inversion.

The first use of the mass-spring method for facial animation was by Platt and Badler (Platt and Badler, 1981), using vector muscles and a spring-based model. Although this is described as a tension net, it can be considered a mass-spring model in which each point had uniform masses. Waters used a mass-spring model incorporating bi-phasic springs to model the non-linear stress-strain response of the skin (Waters and Terzopoulos, 1992). This model was also used to recreate animations from video (Terzopoulos and Waters, 1993). Waters also used CT scan data to accurately model the depth of the skin using a mass spring model controlled by muscles (Waters, 1992). Teschner et. al used a similar mass-spring model in order to directly compute updated resting tissue positions as a result of surgical operations affecting the skull (Teschner et al., 2000). Additional springs can be introduced to model more effects. Kähler et al introduced additional springs which pulled the skin outwards to model volume preservation (Kähler et al., 2001). Sera et. al (Sera et al., 1996) modelled the epidermis, dermal-fatty layer, fascia surface, muscle layer and skull. They produced a tool for generating mouth shapes and expressions through the use of a graphical user interface controlling forces applied to different muscles. Their work particularly focused on generating correct mouth positions for speech.

Waters, Lee and Terzopoulos had developed a dynamic model of facial tissue controlled by muscles with an estimated skull (Lee et al., 1995). The model was formed from point masses connected by springs, arranged in layers representing different tissues in the face. They modelled the behaviour of the different layers through modification of the equations controlling the motions of the springs. Force was applied to the points representing the skin by a weighted combination of the force from the muscle and volume preservation forces applied to mimic the incompressibility of human skin (Lee et al., 1995). They modelled 28 muscles in the face, and also included separate eye and tooth models. The teeth however were not interactive and simply scaled to fit behind the mouth of the facial model. To animate the mouth, they modelled the orbicularis oris, the muscle which circles the mouth, as a piecewise linear combination of muscle fibres.

The mass spring method is very computationally efficient and stable, but suffers due to its simplicity. Even with its already impressive speed, it can be further accelerated at the expense of precision (Liu et al., 2013). As it is only constructed of springs, it is difficult to model the physical behaviour of 3D volumes. These behaviours must be recreated using heuristics or the use of additional springs. As such, the mass-spring method is not well suited to modelling realistic soft tissues (Keeve et al., 1998). The mass-spring method is still employed for modelling of other parts of the body such as the hair (Selle et al., 2008). Applying extra constraints to a mass-spring model can also allow the modelling of non-elastic materials such as cloth (Provot et al., 1995).

2.4.2 Muscle Modelling

In 1987, Waters developed a parameterized model of the face with the intention of improving on the flexibility of facial models. He achieved this by implementing a muscle model, and parameterizing the muscles rather than aspects of the facial expression directly (Waters, 1987). His model is not actually physically-based, but represents an early attempt to model the muscles and is a forerunner to more modern approaches. He validates his model against the Facial Action Coding System (FACS) developed by Ekman and Friesden (Ekman and Friesen, 1978) which uses 50 key independent facial actions to produce a multitude of expressions.

A significant early attempt at modelling muscles came from Waters and Frisbie (1995). They modelled 10 muscles throughout the lower face. Using a two dimensional model, with a mass spring framework for the actual muscles. Each muscle was mapped to points on the skin surface which it affected. Observing that all the large scale deformation comes from the muscle, and the skin more or less follows its movement, they simply had the skin move in the same way as the muscle.

In their 1996 work, Sera et. al (Sera et al., 1996) simulated the musculature of the face by defining a series of linear muscle fibres. They modelled 10 muscles in the face using 34 linear fibres. The individual muscle fibres were not physically represented in the model, although there was a volumetric layer representing the muscle layer in the face. Instead of the individual muscles being modelled, they were defined as a series of properties. These properties included: mesh nodes the fibre was attached and inserted to, width, stiffness and whether the muscle was linear or piecewise linear. The force exerted by the muscles could be varied to produce different mouth shapes.

Kähler, Haber and Seidel further developed the piecewise linear muscle model originally developed by Lee. They introduced a volumetric representation for the muscles by assigning each linear segment of the muscle an ellipsoidal volume (Kahler et al., 2001). Scheepers also used this idea, using volume preserving ellipsoids to represent the muscle bellies. The muscle bellies were attached and inserted via tendons which retained their length. Some muscles were formed of multiple bellies distributed over a surface generated from spline curves, allowing for muscles of more complex shapes (Scheepers et al., 1997)

2.4.3 The Finite Element Method

Before discussing the facial animation work using finite elements, this section gives an overview of the different aspects of the finite element approach.

The finite element method is a method for simulating physical situations described by partial differential equations in single or multiple dimensions. The technique works by dividing a problem into a large number of small simple pieces, known as elements, which are simulated individually and aggregated to produce a model of the whole problem (Zienkiewicz et al., 2000). Elements are connected at their edges by nodes, which can be thought of as points along the edges where elements join and share the same value for a given property. The behaviour of each element is described by a differential equation. This equation is solved using a numerical approximation method for each element.

The method in general always includes certain steps. To begin with a mesh representing the problem domain is created. This is the collection of elements and nodes, and can be a conforming mesh, where nodes in the mesh match those in the problem domain, or a non-conforming one where they do not. Once a mesh is created, elements are assigned material properties describing how they behave physically. Boundary conditions are then applied which enforce constraints on the simulation, either of the value being solved for directly, whether displacement, temperature or something else, or on the derivative of those values such as the velocity or heat flux. The final stage involves computation of the solution coefficients. For a static simulation, these are computed in a single step and represent the equilibrium position of the object. A quasi-static simulation solves for the equilibrium position at several timesteps, but does not include inertial effects. A dynamic simulation is also discretised in time, and the coefficients are solved for each timestep.

Discretization/Meshing

The first step in applying the finite element method to a problem domain is the discretization or meshing stage. This is essentially defining the elements over the problem domain, and involves constructing a mesh similar to those found in 3D-modelling over the whole domain. The resolution of the simulation can be increased by increasing the resolution of the mesh, resulting in smaller elements. This means that important areas can be studied in higher detail without wasting processing time on unimportant areas. As an example, in a simulation of a boat moving through water, the water close to the hull of the boat and the hull itself could be modelled at a high resolution to provide more accurate behaviour, while the water some distance from the boat could be meshed at a lower resolution as its behaviour is not as

important. The mesh has nodes defined along its edges, which are the points connecting the elements in the FEM simulation.

For multiphase simulations such as the boat in the water, or a simulation consisting of skin, bone and muscle, the mesh is divided into regions defining which elements/nodes belong to which materials. Each material can have different physical properties so the method handles multiple phases well. Physical properties are generally available in literature, though work also exists to tune them based on scan data (Kadleček and Kavan, 2019) The elements used generally have simple shapes. For 3D problems, the tetrahedron is commonly used as an element shape.

Boundary Conditions

A finite element simulation is concerned with two types of boundary conditions: Dirichlet and Neumann. In the context of a simulation evaluating forces and displacements, a Dirichlet boundary is one where the property has a known value, while a Neumann boundary is one where the derivative has a known value. In the finite element method, the problem is formulated into a large series of simultaneous equations, expressed in a matrix form. The boundary conditions are applied by setting specific entries in the matrix and vector to enforce values at particular nodes.

Computation of Solution Coefficients

For a simple static simulation, the problem statement has now been reformulated into its approximated matrix form, it is possible to compute the approximated solution. The problem is currently stated in the form:

$$\mathbf{Ku} = \mathbf{f} \tag{2.2}$$

Given the vector $\mathbf{f}$ and the matrix $\mathbf{K}$, the aim is to find the vector $\mathbf{u}$. There are two approaches to this. The first involves computing the inverse of the matrix $\mathbf{K}$, $\mathbf{K}^{-1}$ and left multiplying both sides of equation (2.2) by it to give:

$$\mathbf{K}^{-1}\mathbf{Ku} = \mathbf{K}^{-1}\mathbf{f} \tag{2.3}$$

which gives:

$$\mathbf{u} = \mathbf{K}^{-1}\mathbf{f} \tag{2.4}$$

This can be solved directly to give the coefficients in $\mathbf{u}$, giving the approximate solution to the problem. The issue with this method is that inversion of the matrix becomes increasingly costly with the matrix's rank with $O(n^3)$ complexity for a square matrix. This quickly becomes prohibitive for more complex simulations. The alternative to this is to use an iterative method to solve for $\mathbf{u}$ such as the generalised minimal residual method (GMRES) (Saad and Schultz, 1986).

Mesh distortion

As the mesh becomes deformed from its original configuration, elements lose their shape causing the errors in the simulation grow. This is particularly a problem for large deformations where the elements change shape considerably, such as those that will take place in the lips/cheeks. One approach to solving this problem is adaptive meshing, in which the mesh is periodically regenerated (Lo, 2002). The improved mesh leads to a higher stable time step which can counteract some of the cost of re-performing the meshing process.

2.4.4 Finite Element Method Formulations

This section provides an overview of the key ways in which finite element method simulations can vary. This includes different formulations and element shapes. One of the key difficulties

of using the finite element method effectively is selecting the correct way to approach it, as particular methods can be impractical for certain problems.

Eulerian vs Lagrangian

To demonstrate the difference between these two formulation, two types of points are considered throughout the problem space. Material points describe the object being simulated, and would be located as an example across the surface of the face. Nodes on the other hand, are the points at which the elements of the discretisation meet. The primary difference between the Eulerian and Lagrangian approaches is that in the Eulerian approach, the nodes are fixed in space and the material points move through them, while in the Lagrangian formulation, the nodes move with the material points. As a result of the mesh nodes being fixed in the Eulerian formulation, there is zero mesh deformation. However, because the mesh nodes do not coincide with the material points, the interfaces between different parts of the simulation may be poorly modelled. The Eulerian approach is often used for simulations of fluids as there may be less interfaces.

In the Lagrangian formulation, the mesh nodes are fixed to the material points. This makes dealing with boundaries or interfaces trivial, as mesh nodes are simply defined on the boundary and remain fixed there as the model deforms. This does, however, introduce deformation of the mesh if the material undergoes large deformation. Greater deformation contributes to a greater error in the approximation of the problem. In general the Lagrangian approach is applied to solid objects

The arbitrary Lagrangian-Eulerian formulation uses a mixture of the Lagrangian and Eulerian formulation. The solution mesh is fixed neither to spatial points or material points and can instead move in arbitrary ways. This allows the element shapes to be optimized. An additional benefit is that parts of the mesh can remain fixed on boundaries, allowing proper handling of boundaries between multiple materials. The ALE formulation is commonly used in the modelling of hyperelalstoplastic materials (Rodríguez-Ferran et al., 2002). The ALE formulation can reduce to either the Lagrangian or Eulerian formulations if desired.

Total vs Updated

In a time dependent finite element simulation, each iteration must be computed in relation to a reference configuration. The reference configuration for both total and updated formulations, is the initial configuration. In a total formulation, all iterations are computed relative to the initial configuration. In an updated formulation an iterative process is used where each iteration uses the configuration of the previous iteration as its reference configuration.

The main advantage of the total implementation is that a lot of the information can be precomputed as the reference frame is constant. However, the total formulation also requires a much smaller timestep to be stable, so more steps must be computed to cover the same period of time. The updated formulation however is stable with larger timesteps, but requires more computation per frame as the reference configuration changes with every iteration.

Simulation in Time

The simplest type of analysis done with the finite element method is static analysis. This simply computes the equilibrium state for a system. For many types of engineering problem this is sufficient; buildings and other structures are often simply under load from their own weight. For animation purposes however, importance is placed on how the system changes over time, known as dynamic analysis. To achieve this, the system must be further discretised in time. In its most basic form, the equation of motion for a system in time is a second order differential equation:

$$M\ddot{x} + C\dot{x} + Kx = f \tag{2.5}$$

where M is the mass matrix, C is the damping matrix, x is the displacement and f is the vector of resultant forces.

A set of initial conditions are required for the simulation, including the positions and velocities in the case of x representing position. The system can now be discretised at each point in time separated by a period of time called the timestep. To advance the state of the system from one timestep to the next, an integration scheme must be used. This can be implicit or explicit. Explicit schemes are generally simpler to implement and work by computing the next state as a function of only the current state. Implicit schemes involve finding an explicit equation describing the relation between the two states and solving this. Often implicit schemes have no analytic solution and an iterative root finding method must be used to solve the equation. Implicit schemes can however be helpful when the equation to be solved is mathematically very stiff, requiring a very small timestep, and hence large number of iterations, for an explicit scheme to approximate it accurately.

The time integration itself can be performed using a variety of methods, with the most common being the Euler method and the fourth order Runge-Kutta method. The euler method is used commonly as it is the easiest to implement and highly performant. It also however has a high upper bound for error and often requires a very small timestep to produce a stable solution. The fourth order Runge-Kutta method provides a much more stable approximation by iteratively evaluating the function and gradient at multiple points within the time interval between the two states. These are combined in a weighted fashion to produce the final result. As a fourth order method, the overall error is $O(h^4)$.

Another commonly used method is the central difference approach. Whereas forward and backward schemes such as the aforementioned Euler scheme use values from only one point in time (assuming integration in time), the central difference scheme takes values from between two points in time. This necessitates having two initial values present for the computation, though this is simple to implement as it requires only that two time steps worth of displacements are stored.

Element Shapes

As previously stated, a finite element mesh is constructed from a series of connected individual elements. Selection of the shape of each element can have a critical effect on the results produced by the simulation. The elements can vary in dimension. It is important to note that the dimension of the element is not directly related to the number of dimensions in the simulation, as a 3D model could be constructed from 1D elements.

The simplest element is the 1D spring element. This represents a truss in engineering terms and is commonly used for static analysis of structures such as bridges. The elements are connected at their nodes ensuring a continuous basis. The simplest 2D element is the triangle element. This allows for interpolation over an area, unlike the spring or truss element. The linear tetrahedron is the simplest 3D element, and allows interpolation throughout a volume. This is one of the most common element types used for finite element simulations.

Hexahedral elements have 6 faces and 8 nodes, as opposed to the 4 faces and 4 nodes of tetrahedral elements. This allows the element to model bending effects across an element, whereas tetrahedra may require a much finer mesh to achieve the same simulation. The hexahedral mesh however requires more computation due to the increased number of nodes. Hexahedral elements can also often produce a better quality mesh than tetrahedra, as their shape lends itself naturally to building up more complex shapes. This in turn means that a hexahedral mesh may be much more stable, in which case fewer elements may be required, giving a trade off in performance.

Aside from producing a finer mesh (known as h-refinement) to improve the accuracy of the simulation, it is also possible to use higher order functions to approximate the physical behaviour across an element. If a particular property is not linear across an element, the

use of a higher order shape function will allow a better approximation of the property over the element, reducing the error. This technique is known as p refinement. The primary disadvantage is that with the additional degrees of freedom present in the element, more computation is required. Quadratic elements are particularly useful in bending problems. Considerable bending takes place in the face, such as when the lips are pushed outwards.

Total Lagrangian Explicit Dynamics Formulation

The Total Lagrangian Explicit Dynamics (TLED) formulation is useful for simulations involving large nonlinear deformations. The steps involved in computing a solution differ slightly from the simple static method described earlier. The mesh and boundary conditions are created in the same way, however subsequent steps differ.

The algorithm as described by Miller (Miller et al., 2007) involves 3 key stages. In the first stage, precomputation is carried out. The TLED method allows many of the necessary calculations required for each element to be computed in advance as they are invariant. Initially, the mesh and boundary conditions are loaded, and for each element the spatial derivatives of the shape functions, determinant of the jacobian and the linear constant strain-displacement matrices are computed. The diagoalized, or lumped, constant mass matrix is also computed at this stage. The second stage involves initializing the model. Here initial nodal displacements and force loads are applied.

The third stage is advancement of the model through time. This step is repeated until the desired simulation time is reached. At the beginning of each time step, the displacements of each node from the previous time step are taken and use this to compute the deformation gradient tensor, which describes how a shape has deformed from its original configuration. Next, the full strain-displacement matrix is computed, by multiplication of the initial strain-displacement matrix with the deformation gradient. The second Piola-Kirchoff stress vector is computed for each integration point. This is a vector representation of the second Piola-Kirchoff stress tensor, which has little physical interpretation on its own, but loosely describes the state of stress in the reference configuration, which is the configuration at time 0 in the total formulation.

From these tensors, nodal reaction forces are computed by taking the integral of the strain-displacement tensor multiplied by the second Piola-Kirchoff tensor, over the volume of the element. As these steps are per element, they can be computed in parallel and this fact combined with the precomputation steps give the TLED algorithm its performance benefits. Finally, the forces are integrated to compute the displacements for the time step.

2.4.5 Finite Element Approaches to Facial Animation

Koch et. al (Koch et al., 1996) describes a more physically accurate approach to modelling the face: use of the finite element method. Their reasoning for using this method was that previous approaches had been tailored towards real time animation and thus didn't attempt to accurately model the behaviour of the different tissue layers (Koch et al., 1996). Developing their work with the intention of using it for predicting the results of surgery, the real time factor was much less important than correct behaviour of tissue.

The first advance their work made was using models reconstructed from CT scan data for the skull and facial surface. This was critical for their work as the simulation must be patient specific to correctly predict the surgery results. The second advancement was the use of the finite element method to simulate the skin surface. Forces on each node are computed by a mass-spring model simulating the muscles and bone. They concluded that their work had been a success and research of a 3D volume-based FEM solution using tetrahedral element would be worth pursuing (Koch et al., 1996).

Essa et. al (Essa et al., 1996) used a isoparametric shell-based finite element mesh to produce a system which recreated facial animations extracted from video via estimation of

FACS parameters. Choe et. al (Choe et al., 2001) made use of a finite element model of the face controlled by 19 muscles and a jaw rotation value to allow artists to edit animations produced by motion capture. Specifically they explored the relationship between surface deformation of the skin in video and muscle activations. The muscle activations were estimated from video and could be edited to produce animations using the model.

Chabanas et. al (Chabanas and Payan, 2000) describes a 3D finite element model targeting plastic and maxillo-facial surgery. Rather than simply model the skin using the finite element method, they also modelled the muscles. They note that the muscles have generally the expected impact on the animation of the face. The correct regions of the face move in more or less the right way, however they note a problem with the upper lip retracting too far when the orbicularis oris is activated. They suggest this is due to the lack of simulation of the teeth (Chabanas and Payan, 2000).

Continuing with the 3D finite element model, Gladilin et. al (Gladilin et al., 2004) produced a finite element model of the face including muscles which considered the muscle tension acting along the fibre tangents within the muscle. As the muscle fibres are too fine details to resolve in a CT scan they employed a heuristic approach, producing a field representing the direction of the force throughout the muscle. They also considered the way that the muscles attach to the bones. Rather than simply stopping, the muscle splays out across the bone. This was modelled using a conic extension at the end of the muscle. They used their model to simulate single muscle actions and FACS coded expressions, producing a more realistic result than previous conventional techniques (Gladilin et al., 2004). Gérard et. al used a finite element model to estimate the physical properties of the tongue and cheek by modelling said tissues and adjusting constitutive parameters to recreate the results recorded in a physical indentation experiment (Gérard et al., 2005). Bickel et. al also used a finite element to simulate indentation of soft tissue. They used coloured markers to validate their model (Bickel et al., 2009).

Sifakis et. al (Sifakis et al., 2006) aimed to use a finite element model to simulate speech. The work offered a dramatic improvement over Sera's 1996 work, generating considerably more detailed and realistic expressions at 7 minutes per frame. The mouth was capable of interacting with external objects, allowing for simulation of speaking with a lollipop in the mouth as an example. As part of their future work considerations, they suggest modelling the air pressure within the oral cavity (Sifakis et al., 2006). Barbarino et. al (Barbarino et al., 2008) produced a finite element model of the face reconstructed from MRI scan data. Their model included modelling of many of the anatomical features of the face including bone structures, muscles and connective tissues but noted that better simulation of muscle contraction behaviour was required to improve the model.

Nazari et. al (Nazari et al., 2013) explored using muscle elements to better recreate muscular behaviour in finite element models of the face and the application of the muscle elements to speech reproduction. Stavness et. al (Stavness et al., 2014) used a finite element model to produce subject specific orofacial models. They particularly looked at coupling the hard and soft tissues in the face and demonstrated that the impact of this coupling is significant in facial modelling.

Cong et. al (Cong et al., 2015) used the finite element method to animate a range of characters. Their work looked at morphing a base finite element mesh to fit other characters based on landmark features and profile curves. Rohan et. al (Rohan et al., 2017) used the finite element method to produce a model of the human tongue. Using their model they demonstrated that a mixed-element approach in which multiple types of finite elements are used to construct a mesh can provide a more performant alternative to fully hexahedral meshes.

2.4.6 Mesh-Free methods

In contrast with the mass spring and finite element methods, mesh-free methods operate with no mesh. Instead, they typically use balls to represent portions of the continuum (Belytschko et al., 1996). Unlike mesh-based methods such as the finite element method which can struggle with advanced material behaviours due to the mesh structure being imposed on the solution or drastically changing geometry, mesh-free methods experience no such issues.

Mesh-free methods initially developed with smoothed particle hydrodynamics (Lucy, 1977; Monaghan, 1988) which was used for the analysis of problems in astronomy. One of the key limitations of smoothed particle hydrodynamics is the difficulty in treatment of boundary conditions (Shadloo et al., 2016). Although traditionally used in fluid simulations, smoothed particle hydrodynamics has been extended for use in solids (Allahdadi et al., 1993). Nonetheless, traditional SPH formulations suffer from a lack of consistency. Belytschko et al. demonstrated that SPH could not be expected to converge for non-linear continuum problems (Belytschko et al., 1998). Various methods have been applied to tackle this problem, such as the reproducing kernel particle method (Liu et al., 1995), stress point integration (Belytschko et al., 2000) and the smoothed particle applied mechanics method (Hoover and Hoover, 2001). Other common meshless methods are the element-free galerkin method (Belytschko et al., 1994), and the finite point method (Oñate et al., 1996).

2.4.7 Hybrid methods

Hybrid methods attempt to combine the best of both mesh-based and mesh-free methods. Typically these methods involve particles which store continuum information and properties, along with a background mesh which is fixed in space. Properties are mapped to the mesh to perform a finite element time step and then back to the particles. The most significant hybrid method is the material point method (Sulsky et al., 1994). The material point is particularly well suited to modelling granular problems, most notably in recent years, Disney used the material point method to simulate snow in the film Frozen (Stomakhin et al., 2013). This has renewed interest in the method, particularly in the modelling of viscoelastic foams (Yue et al., 2015; Ram et al., 2015). The material point method has found further applications in the modelling of anisotropic cloths (Jiang et al., 2017). There has also been work modelling soft tissue with the material point method, though typically this is for modelling the failure of the tissue, e.g. (Ionescu et al., 2006, 2005). Bao et al. combined blendshapes with physical modelling, using blendshapes to control the muscle shapes which then drive a physically-based deformation (Bao et al., 2019; Cong et al., 2016). Another hybrid approach is applying physically-based secodary motion to blendshape models (Kozlov et al., 2017; You et al., 2019). Other hybrid models include the generalized interpolation material point method (Bardenhagen and Kober, 2004). Hybrid models utilising motion capture also exist, such as Barrielle and Stoiber's work generating real-time physically-based animations from live motion capture (Barrielle and Stoiber, 2019).

2.5 Animation of the Human Mouth

Producing an accurate representation of the movement of the lips is something that has not had much research in its own right. Most work aims to model the whole face. In 1969 Madsen suggested (Madsen, 1969) as a minimum lip animation must produce:

- an oval mouth when speaking u, o and w.
- open lips for the vowels a, e and i.
- closed lips for b, m and p.

- tucking of the lower lip behind the upper front teeth for f and v.

Guiard (Guiard-Marigny et al., 1996) developed a model of lips. The contours of the lips were described by algebraic equations. The model was developed by finding continuous functions capable of represesnting 22 reference lip shapes. The output of the model was controlled by 5 parameters. The model was adaptable to different speakers via modification of the coefficients in the equations, and collision was handled by developing a volumetric implicit surface-based model.

This idea is still in use in recent work on modelling the lips. Kuratate (Kuratate et al., 2010) takes a similar approach to build speaker-specific lip models. Using 3D surface data and MRI scans, spline parameters describing the surface of the lips are adjusted to minimize error between the nodes of the model and the scan data.

Sera et. al (Sera et al., 1996) used a mass-spring model of the face in which forces were computed from linearly interpolated key frames to produce animations. They found their model struggled to correctly simulate situations where the lips are pushed forward, such as when producing an “n” sound (Sera et al., 1996). On the hardware of the time this proved too slow for real time animation, requiring 1 minute 15 seconds per frame. They conclude their work stating that modelling the back of the lips and the tongue is essential to create more realistic animations.

Badin et. al (Badin et al., 2002) developed a technique for controlling animation of the mouth based on using MRI and photogrammetry to determine lip contours, and a subsequent parameterisation including lip protrusiion and aperture. This work also used the same approach to model the tongue. Badin’s further work (Badin and Serrurier, 2006) expanded on this, fitting a generic model of the tongue to the contours for a specific speaker. He also found a relationship between the parameters obtained using linear component analysis and specific muscle actions in the face. Birkholz (Birkholz and Kröger, 2006) used a similar model, a parameterised vocal tract based on MRI scans.

Koch and Chabanas (Chabanas et al., 2003) produced a finite element model of the face with the intention of predicting the effects of maxillofacial surgery. They constructed a generic face mesh which consisted of two layers, the outer representing the dermis, and the inner layer representing the hypodermis. They only modelled muscles directly affecting the movement of the lips, treating them as part of the outer layer as they are relatively superficial. They opted to represent muscles that perform similar actions on the lips by grouping them into a single muscle structure in the mesh (Chabanas et al., 2003), presenting one possible source of error. They also chose to assume a linear stress response for the facial tissue, which they suggest makes their model incorrect for relative strains above 15% (Chabanas et al., 2003). While less of an issue for their work, production of lip animation will require strains of this level. They simulated their model using a quasi-static analysis.

Gerard et. al (Gérard et al., 2006) produced a finite element model of the tongue. Their mesh was designed with the arrangement of muscles in the tongue in mind, and the muscles themselves were implemented as force generators acting directly on nodes of the mesh.

Nazari (Nazari, 2011) concluded that a dynamic simulation is better than a quasi-static simulation for generating correct lip motion. Nazari’s work focused on the stress stiffening effect in which muscles become stiffer under stress. He demonstrated that this effect significantly affects the shaping of lips, particularly under contraction of the orbicularis oris for lip protrusions. Running acoustic simulations he showed that this could affect the sound produced in speech, giving insight into why a poor lip simulation can look so unnatural in speech animations. His model included muscles which were specified independent of the mesh, as ”cable” elements which apply force to nearby related elements. With this model he demonstrated the ability of muscles not in direct control of the lips to influence their movement, noting that ”stiffening of the cheeks due to the activation of the buccinator induces a limitation of the amplitude of the upper lip protrusion” (Nazari, 2011).

Edwards et. al (Edwards et al., 2016) developed a model which generates a mouth animation from an audio file. The system is called JALI, a nod to its mapping of jaw and lip usages which underpin the method. The animation is generated in a format using FACS which is easy for artists to further adjust if necessary.

2.5.1 Hybrid Models

Hybrid models are discussed separately as they typically combine multiple techniques or methods, usually those given in the previous section in order to combine benefits from two different methods, such as the physical realism of a physically-based simulation with the ease of editing of blendshapes.

Secondary Motion

Secondary motion describes physical effects due to the primary motion of an object or character. For example, flesh jiggling with the movement of a limb. This is applicable to facial animation in cases such as the cheeks relaxing after a smile, or the jaw, cheeks and lips wobbling with a rapid shake of the head. In a traditional blendshape-based approach to creating an animation, whether artist or motion capture driven, these effects are not accounted for. These are difficult to animate directly, as they can often involve complex interactions with other objects in the environment, or be caused by movement in multiple degrees of freedom.

Adding these effects to a blendshape-based animation using some sort of physical model can improve the quality of animation, and allow limited environmental interaction. Lasseter (Lasseter, 1987) pioneered the use of secondary motion in computer graphics, following its use in conventional animation. Opalach et. al (Opalach and Maddock, 1994) expanded on this, applying 'Disney' effects to objects using implicit surfaces.

A possible example is adding a physically-based skin layer over the top of the model. Another possibility is to directly replace the mesh with a physically-based model. Constraints can be applied to the model, ensuring its motion matches the original animation. By replacing blendshape targets with a mass-spring model, Ma et. al (Ma et al., 2012) produced a system capable of dealing with self- and external collisions. Ichim et. al (Ichim et al., 2016) produced a similar system, through the use of a data driven volumetric blendshapes technique.

Blendforces

Blend forces are a recent alternative to blendshape-based systems (Barrielle et al., 2016; Barrielle, 2017), and represent the a strong foray into the practicality and effectiveness of hybrid systems. Based on the idea of blendshapes, their work focuses on encoding forces which drive a physical model of the face, rather than simply capturing the displacements of it. They suggest that this allows for simulation of more advanced phsyical interactions including better lip response due to collision. They implement a physical model based on minimising nonlinear energy potentials due to elastic forces.

They state that their work is limited by some of the same problems as blendshape-based systems, notably that an expression cannot be constructed if it cannot be expressed as a combination of the available blendshapes/blendforces. They also state that their system is limited in its ability to produce more complex dynamics of skin such as wrinkling, although it does offer an improvement over blendshape-based systems. This work demonstrates the practicality of using a hybrid system to improve facial animations. To close, they state that they hope the work encourages further research towards use of physical modelling to improve this area of animation (Barrielle et al., 2016).

2.5.2 Sticky Lips

The 'sticky lip problem' describes the sticking together of the lips as a result of saliva and moisture between them as they are separated. Recreation of this behaviour in computer generated animations could improve the accuracy and realism of the animations. Literature describes the sticky lip effect as "difficult to attain with existing real-time facial animation techniques" (Olszewski et al., 2016). Olszewski et al. attempted to solve this problem using a motion capture method. Their approach generates effective results, but is not easy to edit. They use a convolutional neural network to train parameters which control a digital avatar. The Digital Emily (Alexander et al., 2009) project also highlights the problem. They state that the sticky lip effect has been recreated in various facial rigs for film characters such as Gollum in The Lord of the Rings, showing demand for an effective solution to the problem. More recent work in a similar vein by Helzle et. al (Helzle and Goetz, 2018), used a curve blending approach to recreate the sticky-lip effect in their Digital Albert Einstein project.

Academic literature to date presents limited work on addressing the sticky lip problem. Gascón et al. (Gascón et al., 2010) demonstrate a set of sticky lips as part of a more general work on adhesion, however they make no attempt to produce a realistic simulation of the effect. The other key work including sticky lips is that of Barrielle and Stoiber (Barrielle and Stoiber, 2018). They use spring-based constraints to enhance a projective dynamics model to include the sticky lip effect. Their model goes some way to reproducing sticky lip behaviours, however the stochastic nature of their solution does not produce physically realistic behaviour. Chapter 3.5 of this thesis will show new work in this area, presenting a deterministic stickiness model which produces more realistic behaviours.

Garrido et. al (Garrido et al., 2016) used a statistical regression model to improve recreation of the lips from monocular video. They produced a high quality lip database of videos which was used to train the model. The recorded actors had temporary tattoos applied to the lips in order to allow strong tracking of the lip movement. The model uses a vector valued radial basis function based network. They state that their training set includes examples of sticky-lips and that their work improves lip shape recreation, although they do not give examples of sticky-lip behaviour being reproduced.

Dinev et. al (Dinev et al., 2018) used a sketch-based interface to train a corrective model for improving surface-based lip animations as produced by motion capture or from video footage. The method produces effective results but still relies upon an artist suggesting corrections. The method is not physically accurate and it is not clear what degree of control the artists have over produced effects.

2.6 Evaluation of Mouth Models

Evaluation approaches to mouth animations can be grouped into two categories: perceptual (also known as subjective) and objective. Objective evaluations provide a numerical measure of the accuracy of a model. Perceptual, or subjective evaluations test whether people perceive an animation as realistic or not (e.g. Walker et al. (1994); Hong et al. (2002)). Perceptual evaluation methods differ depending on the goal of the animation being evaluated and the method used to create the animation (Geiger et al., 2003).

The main motivations for creation of mouth animations are creation of an artistic or aesthetically pleasing animation, creation of a physically accurate animation, or creation of an animation which aids in visual speech recognition. The artistic and visual-speech animations are not necessarily realistic. The artistic animations may be exaggerated in some way in order to evoke a particular emotional response in a viewer, and visual-speech animations may be modified or enhanced in some way to improve the intelligibility of the speech. As an example, Alghamdi (Alghamdi et al., 2015) experimented with artificially colouring the lips to enhance intelligibility of speech. Similar ideas could be applied to animations.

Brand (Brand, 1999) conducted a perceptual evaluation of his work, in which "naive viewers" were asked to distinguish between ground truth and synthesized animations, selecting the one they believed to be more natural. He found that three observers preferred the ground-truth animations, three preferred the synthesized animations and one selected from both sets. He concluded from this that although the synthesised animations could be distinguished (6 of the 7 viewers chose consistently), they were equally plausible. Pandzic (Pandzic et al., 1999) also conducted a perceptual evaluation, comparing various aspects of different types of animation. He was able to draw statistically significant conclusions from the evaluation, supporting the efficacy of perceptual studies. This type of test is particularly effective for animations which produce synthesized videos. There is less evidence of its use in works producing an animated model, due to it being more difficult to produce direct comparisons between real and synthesized footage without it being obvious which is which due to the limitations of the rendering of the model.

Perceptual evaluations are, by their nature, subjective. As such the results can be affected by things such as the uncanny valley (Seyama and Nagayama, 2007) in which animations which are very close to realistic are deemed unnerving, or the McGurk effect, which describes how a viewer's perception of a spoken word can be affected by the visual cues they are receiving (McGurk and MacDonald, 1976). For example, upon hearing audio of 'bat' being spoken, whilst watching a video of 'vet' being spoken, viewers may actually hear the word 'vet'. This effect has also been used in the perceptual evaluation of talking heads (Cosker et al., 2005).

Various works on mouth animation have used objective tests which give a precise measure of the accuracy of a synthesised animation. These objective tests typically evaluate the animation against ground-truth data. The main difficulty involved in conducting an objective evaluation is obtaining accurate ground-truth data. For some types of animation this is more practical. If the source material is also an animated mouth, it is likely that vertex positions can be directly extracted. This is useful for models which attempt to recreate other animations, or provide a more efficient method of generating an animation. To obtain ground-truth data from video footage is more difficult. Cao (Cao et al., 2013) used an RGBD kinect camera to record real faces. 2D face landmarks were manually labeled and used to compare their animations against. Other works have also used marker-based tracking. Sera et. al (Sera et al., 1996) used markers placed around the mouth to record correct positions of the areas of the mouth when forming particular phonemes. Their model was then evaluated by comparing the position of the markers with the position of virtual markers on their model for each facial expression. Other examples of ground truth distance measures being used do exist, using a variety of modelling techniques (Cosker et al., 2004; Barrielle et al., 2016). Although ground truth distance measurements do give an objective description of how close an animation is to its real counterpart, subtle effects which could have a significant impact on the perceived realism of an animation may not be highlighted by this method.

Aside from ground truth based methods, active shape and active appearance models (ASM/AAM) (Hill et al., 1996; Edwards et al., 1998b,a; Farnell et al., 2019) can be used to track movements from video footage of real mouths. Such techniques use a statistical model trained on image data and attempt to fit the model to the newly presented frames of the video. One example of software using this approach is Faceware Analyzer (Faceware, 2019). The reliance of active models on matching a predefined shape to the footage means that they can struggle to accurately map to unusual behaviours which may not often occur in the training data. They additionally offer limited precision as they rely on matching only a set of points rather than capturing full surface information.

2.7 Summary

This chapter has covered background anatomical knowledge, highlighting the capabilities of the mouth and lips, reviewed common techniques for modelling physical behaviours and examined state of the art techniques for animating the mouth. A review has been given of works which include discussion of or modelling of the stickiness between the lips. Throughout the mouth animation works, relatively few mention the stickiness effects, and those that do typically only mention it incidentally or as a desirable feature to recreate. This thesis will expand on this area, presenting a physically-based stickiness model which simulates the dependence of the stickiness level upon the moisture content of the saliva present on the lips. There are various components which make up an animation and determine its realism. Most notable is the distinction between rendering and modelling/animation. Rendering deals with lighting and texturing, whilst model creation and animation deals with the geometry and movement of the model. A further consideration is how easy it is for an artist to direct or modify an animation. This thesis specifically focuses on the model creation and animation aspects, with some concern given to how the method can be included in an artist's workflow, however, it is recognised that improvements to rendering will also be required in order to produce a truly realistic animation.

The second section of the literature review has discussed evaluation methods for mouth animations. As shown, the most common approach is using a perceptual study and many works do not include an evaluation. An objective measure of animation accuracy is desirable as it provides more information about the effectiveness of a modelling technique than is typically given in published works. Additionally, due to the uncanny valley effect (Seyama and Nagayama, 2007), it is possible that people may perceive a more physically accurate animation as less realistic, further demonstrating the need for objective evaluations. This thesis presents a new computer-vision-based method for evaluating mouth animations which can also be applied to videos of real mouths to draw direct comparisons of a range of metrics. This does not however negate the importance of perceptual evaluations in the area, as humans will be the ultimate judge of any animation produced. As such, both types of evaluation are used in this thesis. The objective evaluation is conducted using the newly presented method along with a novel data set containing videos of mouths producing actions at varying levels of lip stickiness.

Chapter 3

Modelling and Simulation

Section 3.1 describes the aims of the simulation and specifies which parts of the face and which saliva effects will be modelled. The proposed simulation method uses the total Lagrangian explicit dynamics formulation of the finite element method to simulate the mouth. The first novel contribution in this chapter is the extension of the finite element method with a novel approach to handling stickiness between the lips based on the physical behaviours of glues. An earlier version of the stickiness model was published (Leach and Maddock, 2018). The stickiness level is controlled by a novel moisture model which models evaporation of the water content from the saliva. Compared to existing works, the presented stickiness model produces better deterministic behaviour, would be easier for artists to control and automatically handles changing levels of stickiness over time.

3.1 Aims

This section outlines the aims of the simulation. It will highlight the features that the model should include and effects it should reproduce. Additionally, those details which are not included in the model will be discussed. A rationale is presented to explain why features were or were not selected. Specifically this will cover which features of the face are modelled, which scale of biological processes are modelled and which effects and movements the model aims to reproduce. The mouth and saliva must be understood together before choosing a simulation method. As such the mouth features and saliva effects which will be reproduced are discussed ahead of any discussion of a specific method of simulation.

3.1.1 Physical Features Modelled

The key focus of the research is the movement of the mouth and lips. As such the lower half of the face is the area that must be simulated. The upper half of the face has little influence on the movement of the mouth. In this work, the upper half of the face is defined as that which is above the bridge of the nose. Of the muscles which are above the bridge of the nose, only the *levator labii superioris alaeque nasi* is connected directly to lip movement. This muscle passes from just under the nose to the corner of the eye. As such, the action of this muscle on the lip can be replicated without including the full length of the muscle.

The nose itself is not included in the simulation. Although some facial muscles pass by the nose, it does not have any direct effect on mouth movement itself. It can impact the movement of the mouth as it has limited vertical movement. This effect is replicated by imposing a boundary condition fixing the very top of the upper lip in place where it would be joining the nasal area. This is only applied to the central area of the lip, as the skin further towards the sides of the face is more mobile.

The teeth and tongue can both have an effect on the movement of the mouth, although

in typical speech the effect is limited. The tongue in particular has no effect unless pressed against part of the outside of the mouth. The tongue is also a very complex muscular structure which is notoriously difficult to model (Stavness et al., 2012). The teeth do have an impact in some movements, however it is not believed that they will affect the impact of stickiness. As a result of this and the additional complexity involved in modelling them, the teeth are omitted from the model.

The skin is modelled as a single substance with uniform material properties. In reality, the skin is made up of multiple layers with varying material properties. Work has shown that the modelling of the different layers does increase the accuracy of the model, particularly in the case of wrinkle formation (Warburton and Maddock, 2015). For the purposes of research it was not considered necessary to model to this level as the focus is on making a comparative study between a model with and without stickiness.

Muscles are modelled as virtual muscles. The key muscles modelled are the orbicularis oris and the zygomaticus major, the levator labii superioris and the buccinator. The actions of the other muscles in the lower half of the face have less influence, and it is common practice to approximate their behaviour using the other muscles.

The soft tissue in the face is capable of sliding over the bones. At the extremes of movement the connection between the flesh and bone then limits the movements of the soft tissue. As such, for a lot of movmements the bones do not significantly affect movement of the tissue, particularly in the mouth where the lips are very free and are only really limited by the frenulum. The region of tissue where the top of the upper lip meets the base of the nose does not have significant freedom of movement, so this area will be fixed in place. The action of the jaw is modelled by applying consistent force to tissues which would contain the jaw.

Physical processeses are considered at a scale in the order of millimetres. Thermal effects, fluid dynamics and chemical interactions are not considered as these are beyond the scope of the project and do not significantly affect the movements of the mouth.

3.1.2 Saliva

In modelling the saliva, the aim is to reproduce the effect it has on the movement of the lips. The behaviour of the molecules within the saliva is not simulated. The inclusion of modelling the saliva using fluid dynamics is a potential area for further work. Combined with contact mechanics and a tongue model, the movement of the saliva could be automated, whilst also generating a more precise stickiness distribution across the lips. This would however require a multi-scale simulation and is outside the scope of this thesis. Additionally, the fluid behaviour of saliva is complex, as it displays non-newtonian behaviours such as shear thinning.

The spatial volume occupied by the saliva also has little effect on the mouth movements, as the saliva forms a thin film over the lip which is negligible compared to the scale of movements taking place and the size of the lips. The saliva film has a large effect on the contact mechanics between the lips, particularly on friction forces between the lips in lateral movements of the jaw. In this thesis, the focus is on recreating movements which do not include such lateral motions. Rather, reproduction of typical speech motions and common actions is attempted, targeting effects caused by separation of the lips rather than those produced by the lips being in contact.

The primary effect of saliva the work aims to reproduce is the ‘zippering‘ effect between the lips as they are drawn apart. As the saliva dries on the lips, the saccharides and other substances are left behind when the liquid content evaporates. These substances are cohesive and have a gluing effect, forming a solid layer attaching the lips together. When the mouth is opened, this pulls the two lips together until it breaks somewhere, typically in the center of the mouth where the displacement of the lips is greatest. Throughout the opening of the mouth the saliva layer then continues to break, creating an effect which visually appears very similar to a zip being undone. Throughout the thesis, this effect is termed *zippering*.

Depending on the amount of saliva between the lips, dryness of the saliva and mouth, speed of mouth opening and movement of the lips, the zippering can vary considerably. If the saliva has a high moisture content, then it may exert very limited force on the lips. As a result it does not pull on the lips much, whilst in cases where the saliva has had more time to dry, the pulling effect can be quite marked. If the mouth is only partially opened, the zippering effect may not reach all the way to the edges of the mouth. This can lead to stuck regions in the corners where the two lips meet.

The mouth openings do not always form symmetrically. Due to variations in the level of stickiness across the lips, parts may be more or less sticky. As an example if the lips are licked from one side to the other, more saliva may be deposited on the initial side, which could lead to a smooth variation in stickiness. This could lead to asymmetry even in cases where the muscle forces drawing the lips apart are evenly distributed. Asymmetric openings can also form as a result of asymmetry in the face or muscle activation.

The zippering is not always symmetric. Extreme variations in stickiness levels across the lips can even result in the formation of multiple mouth openings. This is a fairly rare effect, as it seems to require very high stickiness in a localised area of the lip.

3.2 Selection of a Method - The Finite Element Method

Certain physically-based methods are better suited to particular types of problems. For example, techniques such as smoothed particle hydrodynamics are typically used for simulating liquids. For the simulation of solids, commonly used approaches are the mass-spring method, the finite difference method, the finite element method and the material point method.

The mass-spring method is limited in its ability to reproduce realistic complex physical behaviours. With all types of forces being handled via springs, the parameters must be chosen carefully to match a real materials behaviour. This is not easy to do, as the springs do not have behaviour which directly corresponds to material properties of volumetric materials. As any mass-spring model is constructed of one dimensional springs, even fairly trivial material behaviours must be modelled in awkward ways, for example for a block of material to resist rotations, diagonal springs must be added across the element, but this also affects the way normal stretches take place.

The finite difference method is simple to implement, but is not used widely in practice for more complicated problem domains. It is limited by its solution points being arranged regularly, making it difficult to model curved or complicated geometries. In simple cases, the finite element method reduces to the finite difference method.

Both the material point method and finite element method are powerful techniques for solving problems involving complex partial differential equations governing a continuum. The material point method Nairn (2003) is an alternative simulation method to the finite element method which was examined due to two key advantages it holds: (i) as a meshless method, distortion due to large deformations is not an issue; (ii) it handles contact 'for free'. The method, in summary, tracks material information on particles and in each frame transfers information to a regular background grid, computes a finite element step on the grid and then transfers the updated properties back to the particles. After each computation step, the grid is reset. To determine whether the method was a good fit for the problem, a 2-dimensional test program was produced (Figure 3.1).

In the end it was determined that although the method does handle contact, it does not naturally handle separation of specific boundaries, a very desirable property for handling the separation of the lips. In addition, the contact is handled as a result of particles being mapped to the same grid element - the result of this is that there is always some gap present - given the fine scale of the saliva in relation to the rest of the model, this makes the method inappropriate. There are solutions to this, although implementation would be as much work

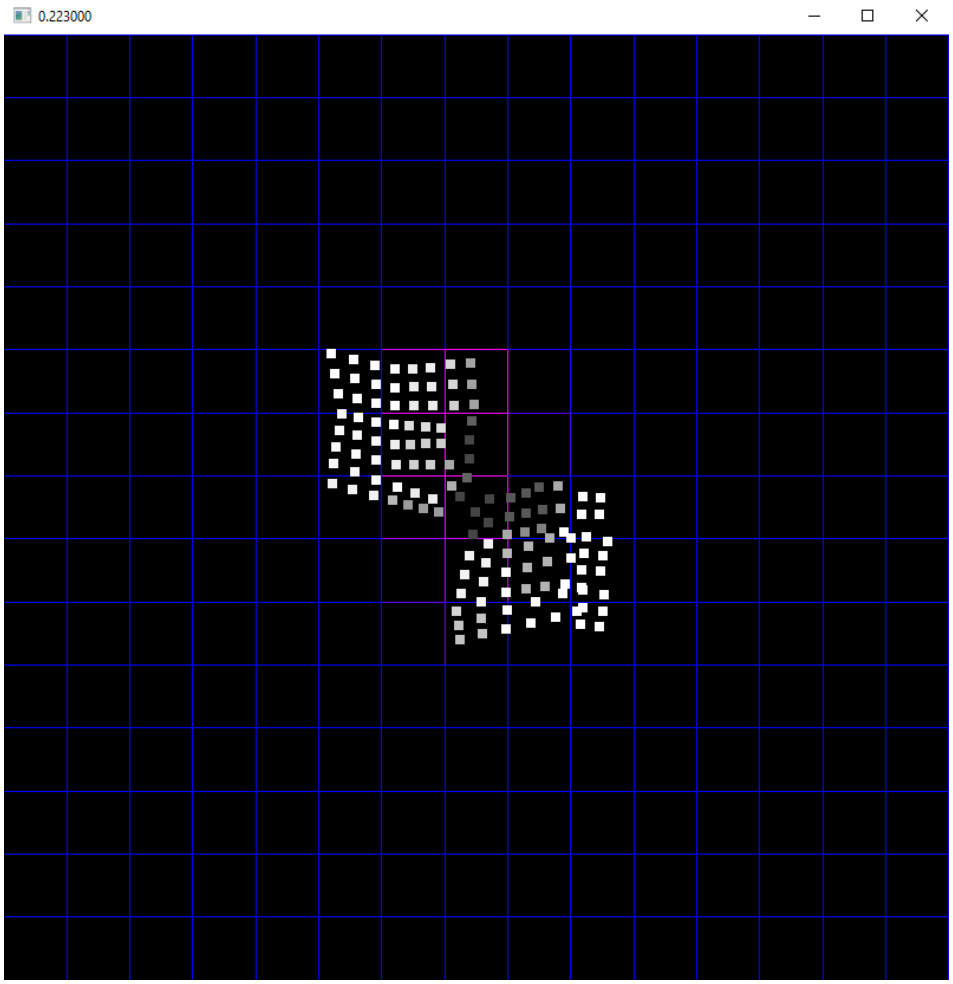


Figure 3.1: A 2D simulation using the material point method. The white points initially formed two separate squares. In the frame shown, the two squares of material are colliding and deforming. The darkness of the white points represents the pressure at each point. The darker points in the center of the figure show the forces and deformations taking place as a result of the collision. The blue grid is the background mesh which the material properties are mapped to for the finite element step.

as handling contact within the finite element method anyway.

In contrast to the material point method, the finite element method is a mesh-based method, making specification of regions of material and boundaries very simple. Although large deformations can cause issues for the finite element method, there are specific formulations which are capable of handling these non-linear behaviours. Additionally, the finite element method has been used extensively in research, including the research of soft-tissue mechanics. As such there is a lot of supporting research available.

3.2.1 Selecting an FEM formulation - The Total Lagrangian Explicit Dynamics Formulation

Within the finite element method there are many different formulations, depending on how particular aspects of the method are handled. A more detailed description of these differences is given in section 2.4.4 within the literature review. The TLED formulation is chosen mainly for the following benefits:

- The total form avoids significant recomputations in each timestep.
- The Lagrangian form is very convenient for modelling solid masses, as the material coordinates match exactly with the vertices of the mesh which describes the object/system.
- With the large, dynamic deformations involved, a small timestep is required to produce an accurate animation. Explicit timestepping lends itself better to small timesteps, as implicit solves can take much longer.
- As interest lies in recreating realistic movement, not just a realistic equilibrium/end position, a dynamic simulation is the only appropriate option.
- The element forces can be solved in parallel, meaning the solution can be computed using the GPU. This provides a dramatic speedup in computation time.
- The method supports the modelling of geometric and material non-linearities.

The details of the formulation are presented in more detail in the following sections.

3.2.2 NiftySim - A TLED Framework

A finite element solver was required to perform the simulations. The two main industry-standard finite element packages are ABAQUS [1] and ANSYS [2]. ABAQUS has various packages for performing different types of simulations, with ABAQUS/explicit available for performing dynamic analyses. ANSYS also features an explicit dynamics method for dynamic analysis. However, neither of these tools are open source so are not easily extensible, so alternatives were explored. NiftySim is an open-source finite element solver designed to support high performance medical simulations of soft-tissue. As NiftySim was designed to be applicable to a range of applications, it is easily extensible, providing base classes which can be extended for more specific applications. For example, the boundary force class can be extended to a muscle force class modelling a particular profile of force application. A validation was carried out in which virtual blocks of materials were deformed under the same loading conditions in each of ABAQUS, ANSYS and NiftySim. The results were consistent across all three.

The package also includes a range of constitutive models. This extensibility meant it was possible to integrate the saliva elements and muscle behaviour into the simulation. In addition

[1] https://www.3ds.com/products-services/simulia/products/abaqus/
[2] https://www.ansys.com/

to these features, NiftySim also has a publication record demonstrating its suitability for research. NiftySim uses the TLED algorithm which is highly parallel, and NiftySim implements this algorithm making use of the GPU, a device with parallel architecture. This dramatically increases the performance of the solver, allowing considerably more complex simulations to be solved in practical time frames.

NiftySim supports multiple constitutive models. These determine the relation between stresses and strains in the material. Constitutive models can be expressed as a tensor, C, relating the stress to the strain:

$$\sigma = C\varepsilon \tag{3.1}$$

A constitutive model can be described as linear or non-linear depending on whether the tensor is dependent on the strain. For a non-linear material model, the above expression becomes:

$$\sigma = C(\varepsilon)\varepsilon \tag{3.2}$$

NiftySim natively includes 7 constitutive models, 1 linear and 6 non-linear. The linear model is the linear elastic model and the non-linear models are neo-Hookean hyperelastic, Arruda-Boyce hyperelastic, polynomial hyperelastic, transversely isotropic hyperelastic, neo-Hookean visco-hyperelastic and transversely isotropic visco-hyperelastic. Skin exhibits a non-linear response and so the linear model is not appropriate here. Of the remaining models, the Neo-Hookean model is chosen to model the soft tissue in the simulation.

Various element types are supported within NiftySim. The choice of element is one of the most important aspects of creating an accurate finite element simulation. Elements can also be categorised as linear or non-linear, according to the shape functions which describe how properties vary across them. The other major variation between element types is the shape of the element itself. Elements can vary in number of dimensions, ranging from 1D springs to 3D bricks. It is important to note that the dimensionality of the element does not have to match the dimensionality of the domain - a 3D shape can be constructed out of differently oriented 1D springs. Typically there is a trade off between complexity and accuracy.

NiftySim supports the following volumetric element shapes: linear tetrahedra, nodally averaged pressure tetrahedra and reduced integration 8 node hexahedra. For this application, reduced integration 8 node linear hexahedra were selected. For a more detailed analysis of the different element shapes, refer to section 3.2.3.

NiftySim uses an XML format to describe the simulation models. The format consists of four main sections. The first describes the geometry of the problem domain. The first part of this involves specification of all nodes in the problem domain, along with the number of degrees of freedom per node. In this case, nodes are specified in 3-dimensional space, with 3 degrees of freedom per node to represent displacement along each of the three coordinate axes. The second part of the geometry specification is defining the elements. Elements are described by giving an element shape and a list of indices of the nodes which construct the said element. The order nodes are specified in dictates their position with an element.

The second section of the XML file specifies which elements are governed by which constitutive models. In order to facilitate this, NiftySim introduces the concept of an element set. An element set is a collection of elements which share common material properties. An element set is specified by supplying the list of indices of elements which belong to the set, a material type which references one of the constitutive material models supported. To fully define the material properties for the element set, any parameters of the material model selected and a density are specified.

The third section of the XML file details any loads and boundary constraints which are present in the model. The default constraint models included are *Fix* and *Force* constraints. Fix constraints specify Dirichlet boundary conditions, i.e. points at which displacements are

fixed. Force constraints specify Neumann boundary conditions, i.e. external forces. Both types of constraint require specification of the nodes to which they apply, and a degree of freedom in which they are active.

The final section of the file gives the system parameters - these are global settings which apply across all elements, as well as instructions regarding how long the simulation should run for. The key parameters used are the time step, total time and damping coefficient. The damping coefficient is used for Rayleigh damping in the simulation. The total time specifies the amount of in-simulation time that should be computed, i.e. the length of the final animation, not the amount of time the computer should spend running the simulation. An example NiftySim XML model file is shown below.

```
<?xml version="1.0" encoding="utf-8"?>
<Model>
<Nodes DOF="3" NumNodes="64">

</Nodes>
<Elements NumEls="1" Type="H8">
0 1 2 3 4 5 6 7
</Elements>
<ElementSet Size="1">
    <Material Type="NH">
        <ElasticParams NumParams="2">
            800 7000.0
        </ElasticParams>
        <Density>1050</Density>
    </Material>
0
</ElementSet>
<SystemParams>
    <HGKappa>0.08</HGKappa>
    <TimeStep>5e-5</TimeStep>
    <TotalTime>1.5</TotalTime>
    <DampingCoeff>90</DampingCoeff>
    <DoDeformableCollision>0</DoDeformableCollision>
</SystemParams>
<Constraint NumNodes="2" DOF="1" Type="Disp" LoadShape="RAMP">
    <Nodes>
    0
    1
    </Nodes>
    <Magnitudes Type="UNIFORM">-0.015</Magnitudes>
</Constraint>
<Constraint NumNodes="8" DOF="1" Type="Fix" LoadShape="RAMP">
    <Nodes>
    3
    4
    5
    6
    </Nodes>
    <Magnitudes Type="UNIFORM">0.015</Magnitudes>
</Constraint>
<Output Freq="100">
```

```
42 <Variable>U</Variable>
43 </Output>
44 </Model>
```

3.2.3 Evaluation of Element Types in NiftySim

Having established NiftySim is a suitable and valid tool for performing the simulation component of the work, it was important to select an element shape for the model. This can have a significant effect on results and performance. Three types of elements are available in NiftySim natively:

- Linear 4 node tetrahedra (T4)
- Averaged nodal pressure 4 node tetrahedra (T4ANP)
- Linear hexahedra (H8)

This evaluation is limited in scope to only these elements. The most obvious type of element missing from this list, is anything of a higher order, such as the quadratic 10 node tetrahedron. Quadratic elements have been ruled out early on as unsuitable for this project. The key advantage quadratic elements offer is being able to directly represent curved surfaces at an intra-element scale. This offers a slight increase in accuracy for displacement across an element, although most of the benefit here is in so called 'derivative' quantities.

If an element is linear, it means it is described by linear shape functions. These represent a linear mapping of primary properties across an element. In this case, this is the displacement. 'Derivative' quantities such as strain, can only be expressed across the element by the derivatives of the shape functions, so for linear elements, strain etc. are constant across an element. This is unsuitable for many engineering applications, however as in this application interest lies in only the displacements, this is not so relevant here, with quadratic elements offering only a slight improvement in accuracy. Further to this, as the model is rendered using flat polygons, non-linear displacement information is of little use with regards to rendering.

Performance

The effect the type of element has on performance is an important consideration. Performance will be measured against 4 criteria:

- Scaling with number of elements
- CPU vs GPU
- Critical timestep (What does this vary with, mass scaling, force size)
- Number of elements required for a model

Scaling with number of elements - A simple model was run using the same number of elements across each type. As can be seen in figures 3.2 and 3.3, the total time taken to complete a simulation scales approximately linearly with the number of elements used, for T4, T4ANP and H8 elements.

CPU vs GPU As expected, the GPU handles larger numbers of elements much more effectively than the CPU. It is clear from figure 3.4 that the GPU can handle very large numbers of elements with speed. Figure 3.4 shows that for all 3 of the element types, the GPU is roughly an order of magnitude faster than the CPU for large numbers of elements.

Figure 3.5 shows a breakdown of where compute time is spent on the GPU. This information was generated using the NVIDIA Visual Profiler, nvprof. If further performance is

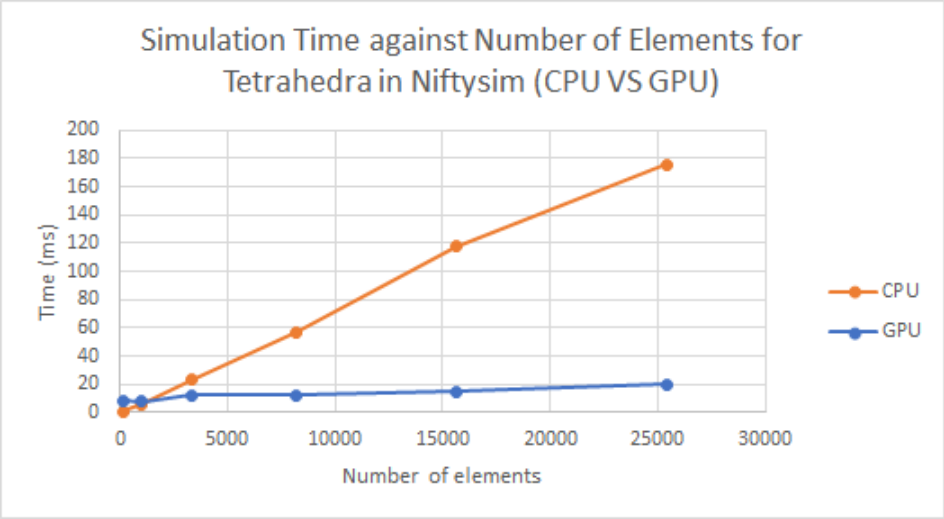


Figure 3.2: Simulation time against number of elements for 4 node linear tetrahedra.

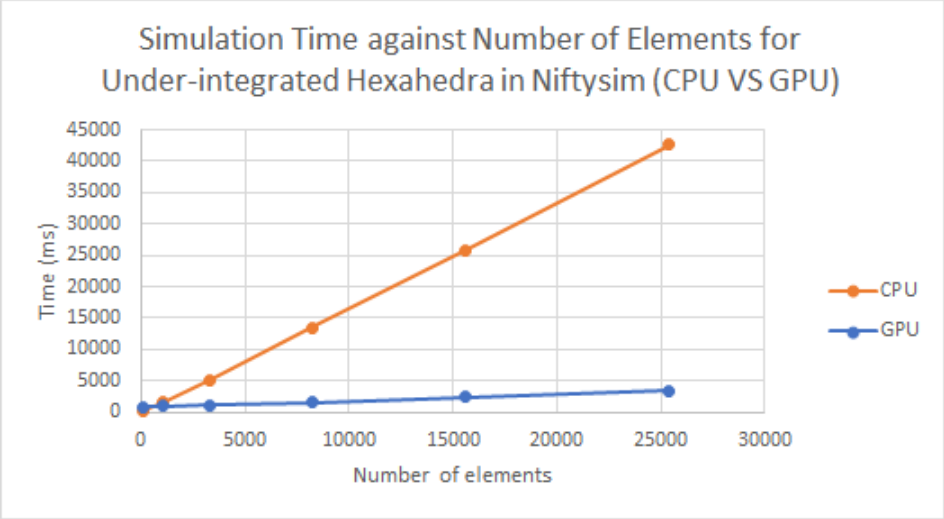


Figure 3.3: Simulation time against number of elements for 8 node linear reduced integration hexahedra.

required, 3 of the 4 memcopies taking place could be grouped to remove some of the overhead and reach a higher throughput. Potentially some of the memcopies could be removed entirely, as they are per frame boundary conditions being transferred. If they are unchanged there is no need to copy them again, in addition if the values are being interpolated the computation could be carried out on the GPU. There may also be scope for optimising the kernels, providing a range of options for further increasing the performance of the GPU code if necessary.

Critical Timestep The critical timestep was evaluated by constructing a simple beam model, fixed at one end with a vertical load at the other. The timestep was increased gradually until the model no longer converged. Overall, the critical timestep for hexahedral elements was found to be roughly an order of magnitude greater than the tetrahedral ones.

This has various implications. With a larger critical timestep, the same total simulation time can be computed in less timesteps. Alternatively, it is possible to generate a more accurate solution in the same number of timesteps. The absolute values for the critical timestep vary with the size of elements and forces involved.

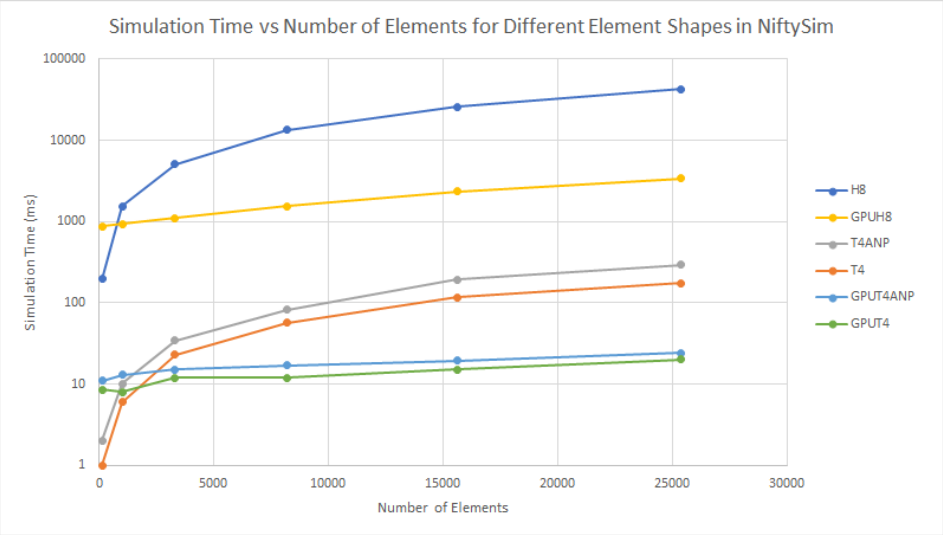


Figure 3.4: Simulation time against number of elements for all elements on the CPU and GPU.

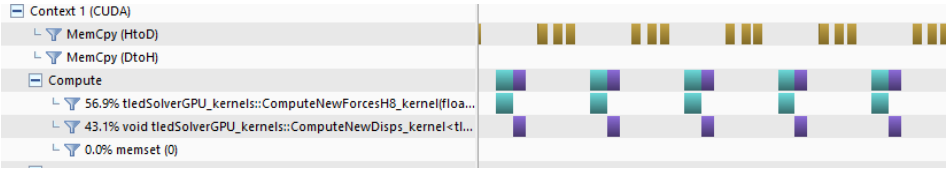


Figure 3.5: Performance analysis of time spent in the TLED algorithm for 5 frames of animation.

Ease of Meshing

Creating meshes is an important part of the project as multiple meshes will be required to run experiments, and the final face mesh will be required for the end result. In terms of meshing, T4 and T4ANP elements are identical. Both are represented by tetrahedra in the mesh. As such, the comparison here is between tetrahedral and hexahedral elements.

In general, creating a hexahedral mesh is complicated due to tesselation constraints. Whereas tetrahedra can easily be connected to fill any space or shape, hexahedra are much more heavily constrained given their increased number of faces and vertices.

Creating a hexahedral mesh further becomes complicated with increasing geometry complexity. Often experts divide models into simpler parts which are meshed individually, before either being combined, or simulated individually with common boundary conditions enforced between them. Tetrahedral meshes do not suffer from this problem, and there are techniques available for creating good quality tetrahedral meshes automatically (Si and TetGen, 2006), even for geometry containing holes and oddly shaped regions. The most common algorithm for automatic tetrahedral meshing is constrained delaunay tetrahedralization (CDT). This is employed by tools such as TetGen (Si, 2015).

By examining the curvature, these algorithms can automatically increase mesh density in more complex regions. Parameters can also be supplied constraining features such as the maximum edge length and minimum allowed internal angle in an element, ensuring well formed elements.

One solution available to this problem for hexahedral meshes is to use a non-conforming mesh. This entails creating a regular, voxelised mesh which is an approximation of the actual geometry. The simulation is performed on this mesh, and the geometry deformed accordingly. This approach was used by Warburton (Warburton and Maddock, 2015) when simulating

wrinkles in the forehead using the finite element technique. However, this approach can impose properties of the mesh geometry on the animated model.

Behaviour under bending

Perhaps the most important factor for which element type is appropriate for this project is the behaviour under bending. The lips undergo highly non-linear deformations, involving significant bending components. To evaluate bending behaviour, a block was constructed. Half the block is fixed, and the other half is loaded vertically.

The block was constructed using a python script to generate a regular grid of elements. One important consideration for hexahedral (H8) elements is hourglassing. As the elements use reduced integration rather than full integration, certain modes of deformation do not generate any internal stress and therefore resistive force. As such, these are called zero-energy modes as it requires no energy for the element to deform this way. This leads to an artificial introduction of extra energy into the system, and can easily cause the system to destabilise.

To counteract this, additional shape functions are used in conjunction with a parameter setting a limit on the proportion of strain energy due to zero energy modes allowed of the total strain energy of the element.

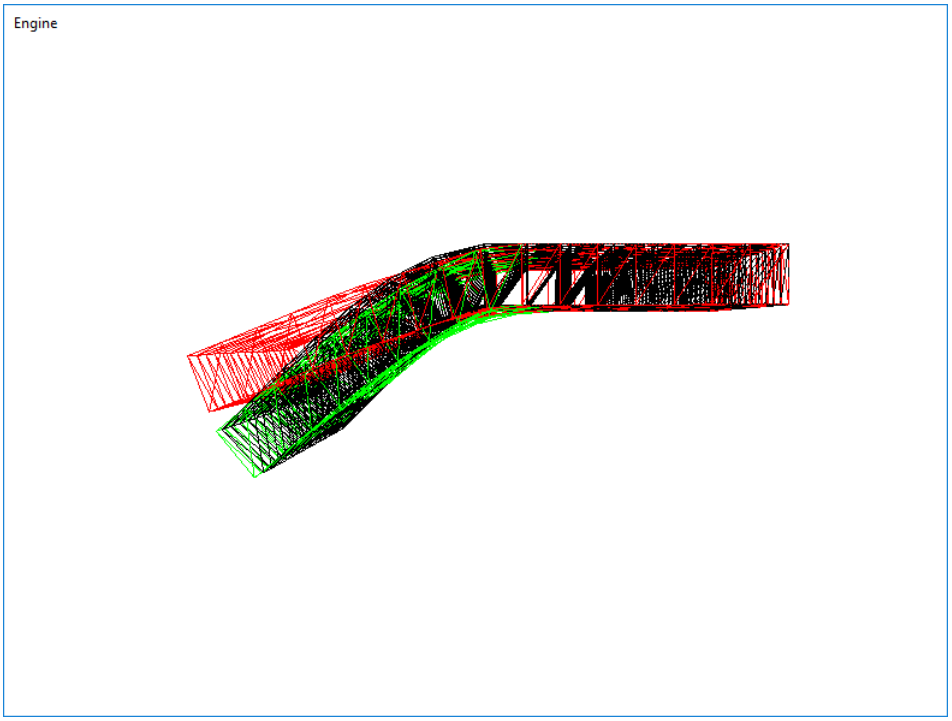


Figure 3.6: A comparison of elements bending with hourglass correction.

As described in various other literature, linear tetrahedrons are overly stiff (Joldes et al., 2009). Aside from the first experiment in which very little hourglassing was allowed, the linear tetrahedra were always the least capable of bending. The T4ANP elements were better at bending than the linear tetrahedra. With the adjustment of the hourglassing parameter, the hexahedral elements proved to be most capable of bending, bending further than the T4 or T4ANP elements could, and proved to be more stable as well. Figure 3.6 shows blocks constructed of the three different element shapes subject to the same boundary conditions. The red T4 elements bend considerably less than the black T4ANP and green H8 elements, showing their overly stiff behaviour. The T4ANP and H8 elements with hourglassing demonstrate similar bending behaviours, although can still be overly stiff for large bending problems. Adjustment of the hourglassing parameter allows modelling of more realistic behaviours for

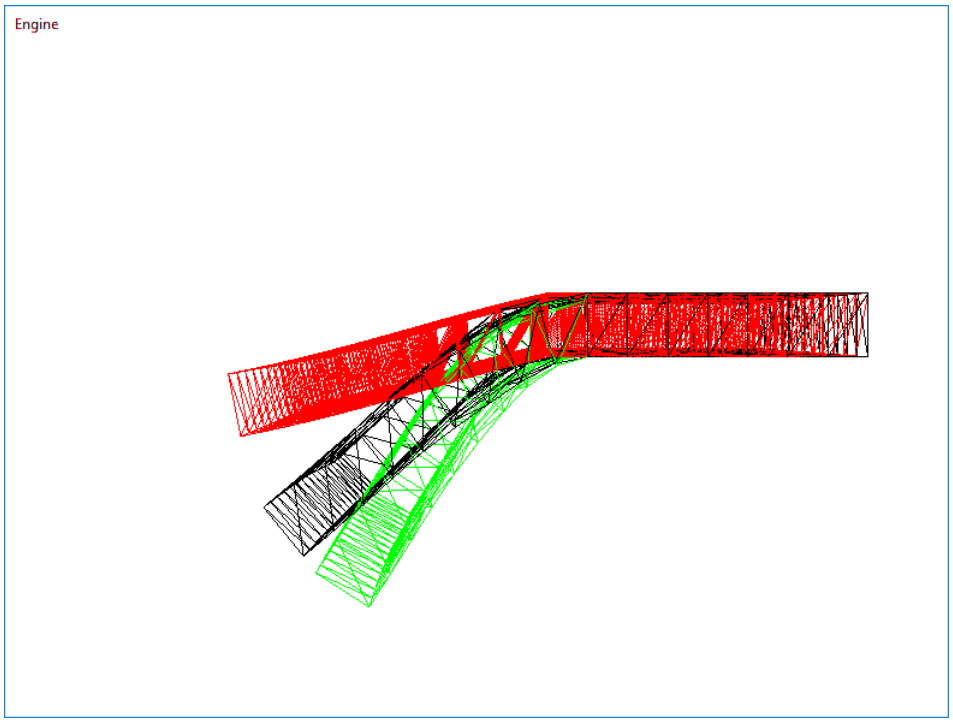


Figure 3.7: A comparison of different element types bending with no hourglass correction.

large bending. Figure 3.7 shows the three element types under the same boundary conditions, but with the hourglassing parameter adjusted to allow the H8 elements to bend more.

Conclusions

To briefly summarise the findings of this work, T4 elements are easy to mesh and very highly performant, but display very poor behaviour when modelling bending problems, and can require a very small timestep or become unstable. T4ANP elements have the same benefits as T4 elements, but are more capable of dealing with bending problems. H8 elements are more difficult to mesh, but produce better results, particularly where bending is involved. Although hourglassing is required to handle the reduced integration scheme, it has been shown in other work that this often actually better represents bending problems than fully integrated linear hexahedra.

Both types of tetrahedral element display higher like-for-like performance. This is, however, counteracted by the hexahedral mesh requiring fewer elements to model the same geometry, and the increased stability of H8 elements meaning a much larger timestep can be used (approximately an order of magnitude.) As such, H8 elements have been determined as the ideal element choice for this project. Despite the increased difficulty of meshing complex geometry, it is expected not many complex geometries will be required. The key one, is of course, the full mouth model. Other experiments can generally be carried out on vastly simplified models, such as cubes, planes of elements or other easily analytically defined shapes. These are simple to generate programmatically.

3.3 Construction of the Model

This section describes the process used to create the finite element mesh used in the simulation. An overview of the construction process can be seen in figure 3.8. Mesh generation approaches fall into one of two categories: automatically or manually generated. Automatically generated meshes are typically tetrahedral, as generating a good quality tetrahedral mesh is generally

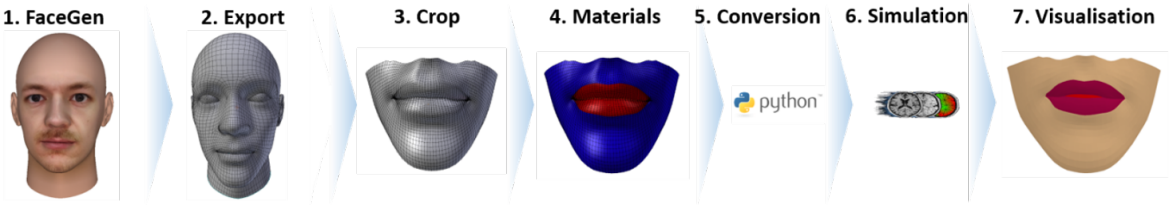


Figure 3.8: The modelling pipeline used to create finite element mesh used in this work.

easier than creating a good quality hexahedral one. Common tetrahedral mesh generation methods include the advancing front technique (Peraire et al., 1988; Zhang et al., 2009; Boyd and Müller, 2006) and constrained Delaunay tetrahedralisation (Si and Gärtner, 2005). Most automatic hexahedral meshing methods are either similar in concept to the advancing front technique (Stephenson et al., 1992; Tautges et al., 1996) or morph a base mesh to fit a particular geometry (Schneiders, 1996; Shepherd and Johnson, 2007; Cook, 1982).

In the interest of testing and evaluation, it was useful to have a mesh based on the author's face. Obtaining scans in order to produce an automatic mesh from the data proved too expensive. so a manually meshed mesh is generated from a surface model of the author's face. The software which generates the surface generates well structured quadrilaterals which can be converted to hexahedral elements.

3.3.1 Facegen Modeller

Facegen Modeller (Inversions, 2008) is a piece of commercial software capable of generating models of human heads. Faces are described by over 150 parameters, and these can be adjusted or randomized to create different faces. Facegen Modeller also includes a feature allowing the generation of a face based on a set of photographs of a human. The photographs are a head-on and side view of the person whose face will be modelled. In this work, the author was used as the model.

Once the photos have been entered, the software attempts to create a face which best matches those in the photographs (Figure 3.9). The result is a 3D head model. By default the model is in a static pose, however, the created model is also mapped to the commonly used *FACS* system. Blendtargets for each of the actions can also be exported. These are used for generating blendshape-based animations throughout the work.

The 3D models can be exported from Facegen Modeller in a variety of formats. In this case, they were exported to the Wavefront OBJ format as it is very simple to work with, in that it is both widely supported by 3D modelling software, and is easy to read and manipulate for a human or script. At this stage, the model is only a surface model, i.e. it has no thickness.

3.3.2 Blender

Cropping

The model is set up in Facegen Modeller with the mouth very slightly open. This ensures there is a clear separation between the lips when it is exported. This model is then imported into Blender, a free 3D modelling package.

The lower part of the face is not affected considerably by movements in the upper half of the face, but including the upper half of the face would significantly increase the complexity of the project for little benefit to the animations of the mouth. Inside blender, the full face model is cropped to the mouth region. Additionally, the teeth, tongue and mouth cavity generated by Facegen Modeller are removed as they are not used in this simulation.

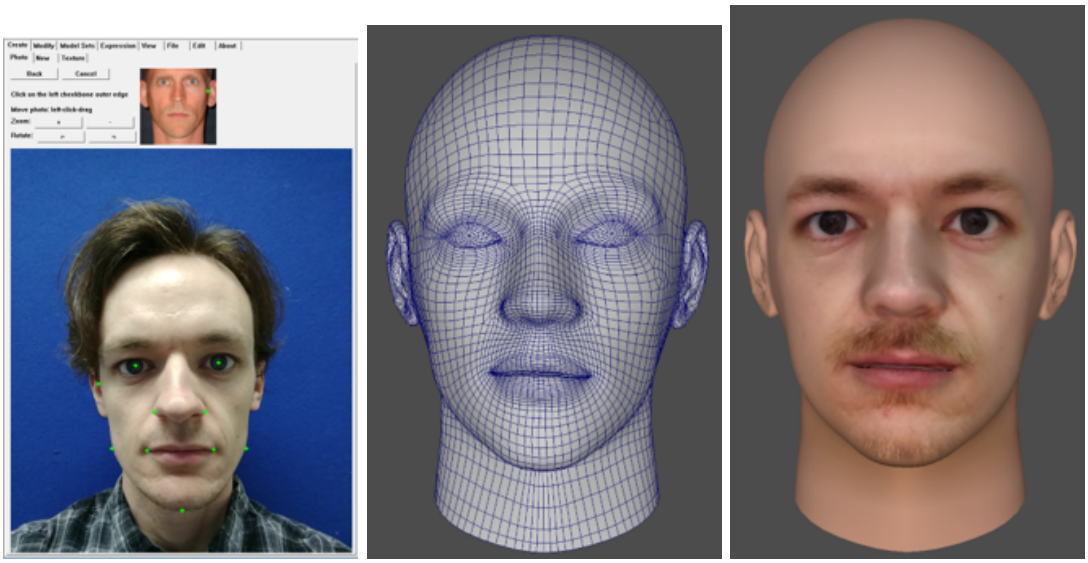


Figure 3.9: The 3D head model generated from the author's photographs as seen in Facegen Modeller.

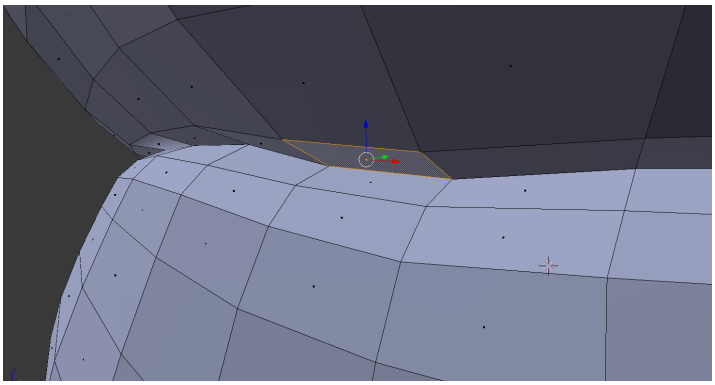


Figure 3.10: A saliva polygon generated between the lips. These are created along the full length of the lips and will form the basis for the saliva elements in the model.

Creation of Saliva Elements

Additional elements are created to represent the saliva between the lips. The mesh exported from Facegen is well structured, such that the upper and lower lips are constructed of the same number of polygons along their lengths. Polygons which will represent the saliva layer are created by joining vertices from the upper and lower lips to create new faces (Figure 3.10). These are very thin vertically, as the saliva layer is very thin and visually the saliva layer should not be noticeable directly, only its effect.

Material Region Definitions

In order to identify different regions of the model, different portions of the model are assigned different materials in Blender. When exported, the vertices/faces are grouped by their material ID as defined in Blender. In this model, three materials/groups are used: the skin, the lips and the saliva. In each case the respective faces are selected and assigned a material ID. The regions are visible in Figure 3.11. Blue represents normal soft tissue, red represents the lip region and green represents the saliva layer.

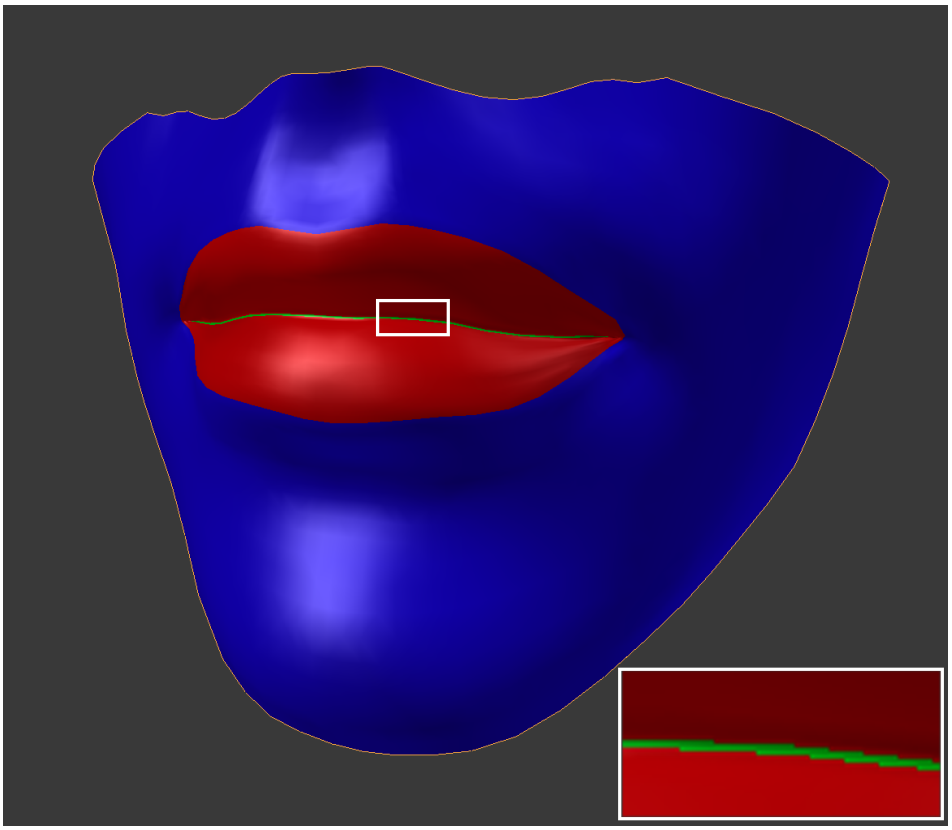


Figure 3.11: The material regions of the model. Blue represents normal soft tissue, red represents the lip region and green represents the saliva layer.

3.3.3 FEM Mesh Generation

In order to generate a valid finite element mesh from the Blender based .obj model, a set of python scripts were developed.

Mesh Extrusion

As discussed in section 2.4 of the literature review, volumetric meshes provide an advantage for this type of simulation. Specifically, a mesh generated of hexahedral elements is desired. In order to do this, the quad faces of the .obj mesh are extruded along the z axis to give the model depth. The extrusion of quad faces ensures all elements created are brick elements. A python script is used to extrude the vertices and form the brick elements. It then rewrites the vertex data and elements into the NiftySim XML format.

3.3.4 Material Property Definitions

The NiftySim XML format describing the model supports multiple materials. A Neo-Hookean material model was selected to model the soft tissue. The saliva is modelled with a modified Neo-Hookean material, dubbed *Neo-Hookean Breaking*.

For a Neo-Hookean material, two parameters are given to influence its behaviour. The elements in the XML description of the model are grouped according to the materials specified for them in the Blender modelling stage. The parameters for the soft tissue are:

$$\lambda = 800$$
$$\mu = 7000$$
$$\rho = 1050kg/m^3$$

For the saliva, the parameters are:

$$\lambda = 7700$$
$$\mu = 191000$$
$$\rho = 1050kg/m^3$$

The parameters are estimated from available data. Using precise values is not possible as the skin is being modelled as a uniform material, when in reality it is constructed of layers of differing mechanical properties. The bulk and shear modulus are chosen such that the poisson ratio is fairly close to 0.5 whilst maintaining the stability of the simulation. A poisson ratio of 0.5 represents an incompressible material. The skin is close to incompressible, however, as the poisson ratio in a finite element simulation approaches 0.5, the stability of the simulation decreases. The bulk and shear modulus for the saliva elements are chosen such that the saliva elements are significantly stiffer than the surrounding tissue, the reasoning for this is explained in section 3.4.10.

3.3.5 Bone and Muscle Definition

In this model, bones are represented using fixed displacement boundary conditions which fully prevent nodes from moving. Displacement boundary conditions are specified as constraints in the NiftySim XML model file. In order to apply them in an automatic way, objects were modelled in blender which could overlay the face model. These objects are then exported. Any nodes which lie inside the bone volume are set as fixed nodes. This is implemented using a python script which automatically checks which nodes lie inside the 'bone' volume and generates constraints for them. Although not many bones were implemented in this simulation as the soft tissue should be free to slide, this approach could be used in further developments of the model.

Muscles are implemented using virtual muscles. This means that the muscles are not modelled volumetrically, rather regions of the soft tissue are specified as being a part of the muscle and forces act upon them to model the muscle's behaviour. Muscles are specified in a similar manner to the bones. Shapes are modelled and nodes which lie inside the volume of the muscle model are given muscle constraints. A constraint is created for each muscle in the NiftySim XML model file. These are again generated by a python script which automatically detects the relevant nodes and generates the constraints.

In NiftySim, force applying constraints are controlled by functions which describe how the amplitude of the force changes over time. NiftySim natively includes a few different functions for this, such as the step function and ramp function. The functionality of these was limited and it is not easy to control the change in forces over time, as everything must be hard coded with time values specified between 0 and 1. To aid in creating animations, NiftySim was extended to support key-framed force activation patterns loaded from text files (Listing 3.1). These were used to control the muscle activations.

The format allows specification of different interpolation functions for each key-framed segment. This gives finer control over the activation pattern for the muscle. This also allows for easy programming of much more complicated activation patterns, such as that seen in

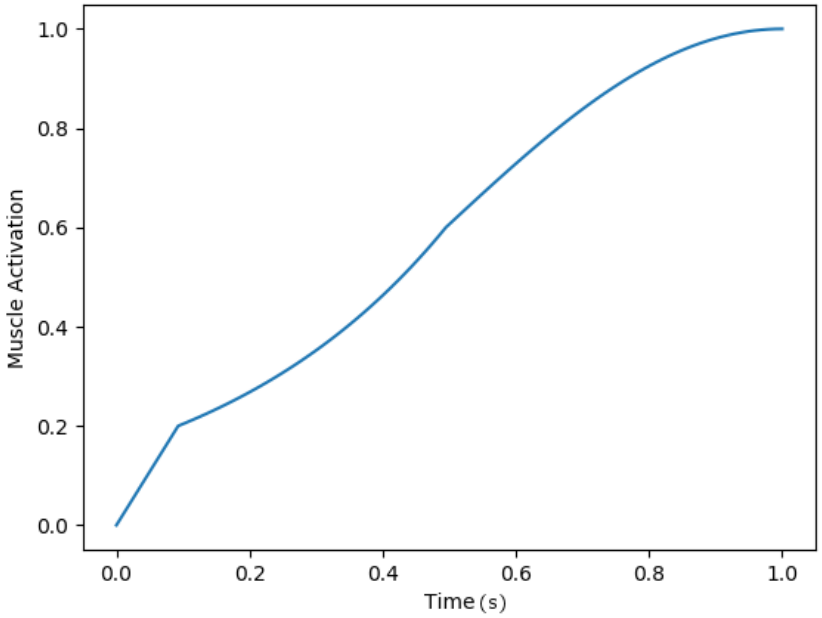


Figure 3.12: A potential muscle activation pattern implemented using the new MuscleForceConstraint class. This is an example curve which has been constructed to demonstrate the capabilities of the MuscleForceConstraint. It shows an initial linear response, followed by a geometric segment and finally, a sin wave segment, reaching maximum activation at 1 second.

Figure 3.12. The force function was implemented as an extension of the constraint class, in a new MuscleForceConstraint class.

```
0.0 0.0
linear
0.1 0.2
geometric
0.5 0.6
quarter_sin
1.0 1.0
```

Listing 3.1: Example code for controlling a muscle activation. The code generates the pattern shown in Figure 3.12

The directions the forces applied by muscles act in are specified in the constraint definitions in the XML file. For the majority of the muscles implemented, the forces act in a linear direction. In reality, muscles can only contract and exert an equal force at either end. The fixture of the muscle to the static bone at one end means that the tissue at the other end is pulled towards the fixed point. As the bone structures have not been implemented in the sense that the geometry has been recreated, this is replicated by only applying the resultant force on the tissue in the direction of the attachment to the bone.

The exception to this is for the orbicularis oris, which has no connection to bone and runs through the lips forming a circle. Once again, in reality the muscle contracts around its perimeter, however the effect of this is a contraction of the lips towards the center. The orbicularis oris is implemented in this way in the simulation, as a muscle whose nodes have a force applied towards the central point.

At this point, a model has been constructed of hexahedral elements which is based on

the author's face. The model consists of two different material types, soft tissue and saliva elements. The model contains muscles and constraints representing fixed points where the skin is unable to move significantly. The model is now ready for simulation.

3.4 Simulation

3.4.1 Overview

The finite element method is part of a family of methods used for solving partial differential equations. Systems described by partial differential equations are notoriously difficult to find analytical solutions to. The finite element method uses approximations to reformulate the problem as a set of problems which are simpler to solve. The problem domain is divided into discrete elements called *finite elements*. These elements are constructed using simple shapes, which the equation is easier to solve across. The solutions are then combined through knowledge of how the elements are connected, giving an approximation of the whole domain.

The method is used in many disciplines and can be applied to a range of problems, such as structural engineering, thermodynamics and magnetic or electric field modelling. Whatever the discipline, solving a problem using the method involves the same steps:

- Develop a mathematical formulation of the physical system
- Discretise the problem domain
- Construct the equations governing how an individual element should behave
- Construct the global set of equations by treating the set of element equations as a set of simultaneous equations
- Impose boundary conditions on the system
- Find the solutions to the global set of equations
- Use the element shape functions to extract the solution values at the desired points

Each of the following subsections will cover one of these steps. The maths presented in these subsections is common to all simulations using the TLED finite element method and is already implemented as part of NiftySim. However, using a custom solver was initially an option, so a prototype was developed which implemented the full TLED process.

3.4.2 Problem Statement - Mathematical Description

This section aims to develop the strong form of the equations describing the system which is to be modelled. The strong form is so named as it requires strong continuity of the solution variables and their derivatives. The strong form is typically a second order partial differential equation. To begin, consider the principle of conservation of momentum, which states that the change in momentum in a system is proportional to the resultant force on the system:

$$\overrightarrow{p}\,(t + \Delta t) = \Delta t\,\overrightarrow{F} \tag{3.3}$$

Dividing equation 3.3 by Δt and taking the limit $\Delta t \xrightarrow{0}$, gives the differential form of the equation:

$$\frac{d\overrightarrow{p}}{dt} = \overrightarrow{F} \tag{3.4}$$

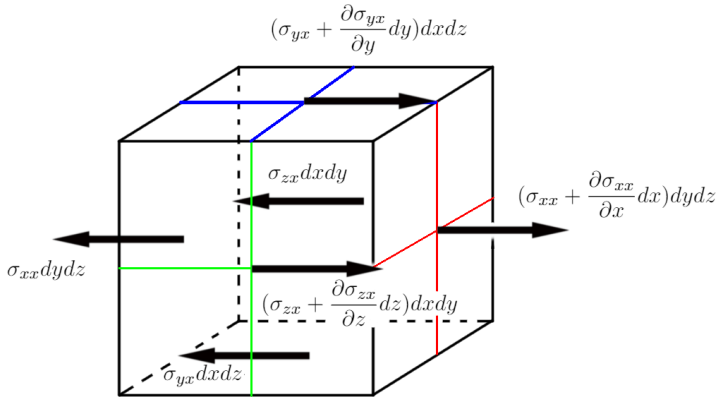


Figure 3.13: A stress body diagram for the forces on the x-axis of an infinitesimal element of a solid body.

To develop the equation further, consider an infinitesimal volume of a solid object (Figure 3.13).

The momentum of the volume element is given by its velocity multiplied by its mass (its volume multiplied by its density). The volume of the element is $dxdydz$ and its density ρ:

$$\overrightarrow{p} = \overrightarrow{v}\rho dxdydz \tag{3.5}$$

The mass of the element is constant in time, the above equation can be differentiated with respect to time to obtain an alternative form of the left hand side of equation 3.4:

$$\frac{d\overrightarrow{p}}{dt} = \frac{d\overrightarrow{v}}{dt}\rho dxdydz \tag{3.6}$$

To evaluate the right hand side of equation 3.4, consider that the total force, $\overrightarrow{F}$ can be expressed as a sum of surface forces which act on a single surface and body forces which act uniformly on the whole volume, $\overrightarrow{F} = \overrightarrow{F}_s + \overrightarrow{F}_b$. To express the surface forces, return to the force body diagram and resolve the forces along the x-axis:

$$\begin{aligned} F_s^x = &(\sigma_{xx} + \frac{\partial \sigma_{xx}}{\partial x} dx)dydz - \sigma_{xx}dydz + \\ &(\sigma_{yx} + \frac{\partial \sigma_{yx}}{\partial y} dy)dxdz - \sigma_{yx}dxdz + \\ &(\sigma_{zx} + \frac{\partial \sigma_{zx}}{\partial z} dz)dxdy - \sigma_{zx}dxdy \end{aligned} \tag{3.7}$$

Simplification of equation 3.7 gives:

$$F_s^x = \frac{\partial \sigma_{xx}}{\partial x}dxdydz + \frac{\partial \sigma_{yx}}{\partial y}dxdydz + \frac{\partial \sigma_{zx}}{\partial z}dxdydz \tag{3.8}$$

Repeating the process for the y and z axes, three equations are obtained describing the surface forces for each axis:

$$\begin{aligned} F_s^x &= \frac{\partial \sigma_{xx}}{\partial x}dxdydz + \frac{\partial \sigma_{yx}}{\partial y}dxdydz + \frac{\partial \sigma_{zx}}{\partial z}dxdydz \\ F_s^y &= \frac{\partial \sigma_{xy}}{\partial x}dxdydz + \frac{\partial \sigma_{yy}}{\partial y}dxdydz + \frac{\partial \sigma_{zy}}{\partial z}dxdydz \\ F_s^z &= \frac{\partial \sigma_{xz}}{\partial x}dxdydz + \frac{\partial \sigma_{yz}}{\partial y}dxdydz + \frac{\partial \sigma_{zz}}{\partial z}dxdydz \end{aligned} \tag{3.9}$$

Using the definition of the divergence operator, F_s can now be expressed as:

$$F_s = \nabla \cdot \boldsymbol{\sigma} \tag{3.10}$$

Substituting these into equation 3.4 an expression is obtained relating the accelerations, stresses and body forces:

$$\frac{d\overrightarrow{v}}{dt}\rho dxdydz = \nabla \cdot \boldsymbol{\sigma} + F_b \tag{3.11}$$

Restating the above values in terms of the displacement field, u, omitting the $dxdydz$ for convenience of notation and considering each field variable as a function of the Eulerian coordinate field x and time, t, the physical body's state can then be described by:

$$\rho\ddot{\boldsymbol{u}}(\boldsymbol{x},t) - \nabla \cdot \boldsymbol{\sigma}(\boldsymbol{u}(\boldsymbol{x},t)) = \boldsymbol{f}(\boldsymbol{x},t), \boldsymbol{x} \in \Omega, t \in [0,T] \tag{3.12}$$

where ρ is the material density, $\ddot{\boldsymbol{u}}$ is the acceleration field, $\boldsymbol{x}$ is a location in the problem space, t is the current time, $\boldsymbol{\sigma}$ is the Cauchy stress, $\boldsymbol{u}$ is the displacement field and $\boldsymbol{f}$ is the external body force field. Equation 3.12 is called the strong form as it exactly states the solutions which every point in the domain must satisfy. Note that at this point, the equations developed are valid for any domain, there has been no specification of any details of geometry at this stage

3.4.3 Developing the Weak Form

It is not practical to attempt to solve this equation (3.12) directly because of the complex geometry of the domain and the inclusion of partial derivatives. Additionally, it is difficult to incorporate boundary conditions to the strong form. To avoid these problems, it is possible to convert the equation into the weak form. The weak form reduces continuity requirements of the solution. Additionally the weak form is an integral formulation, which allows discretisation of the problem replacing the integrals with summations.

A differential equation can be converted into the weak form by multiplication of both sides by a function, $\omega(x)$ called the test function, and then integrating both sides. Now take $\omega(x) = \bar{u}$, where $\bar{u}$ is a kinematically admissible displacement field. For simplicity of notation, equation 3.12 is restated without expressing the parameters of the field variables:

$$\rho\ddot{\boldsymbol{u}} - \nabla \cdot \boldsymbol{\sigma} = \boldsymbol{f} \tag{3.13}$$

Rearranging gives:

$$\rho\ddot{\boldsymbol{u}} = \nabla \cdot \boldsymbol{\sigma} + \boldsymbol{f} \tag{3.14}$$

Multiplying both sides by the test function $\bar{u}$ and integrating gives:

$$\int \rho\ddot{\boldsymbol{u}} \cdot \bar{u} = \int \nabla \cdot (\boldsymbol{\sigma}) \cdot \bar{u} + \int \boldsymbol{f} \cdot \bar{u} \tag{3.15}$$

Using the identity $\nabla \cdot (Av) = \nabla \cdot (A^T) \cdot v + tr(A\nabla v)$ it can be stated that:

$$\nabla \cdot (\boldsymbol{\sigma}\bar{u}) = u \cdot \nabla \cdot (\boldsymbol{\sigma}^T) + tr(\boldsymbol{\sigma}\nabla\bar{u}) \tag{3.16}$$

However, through the relation $tr(A^T B) = A : B$ and the symmetry of the stress tensor, i.e, $\boldsymbol{\sigma}^T = \boldsymbol{\sigma}$, it is possible to further restate this as:

$$\nabla \cdot (\boldsymbol{\sigma}^T) \cdot u = \nabla \cdot (\boldsymbol{\sigma}\bar{u}) - \boldsymbol{\sigma} : \nabla\bar{u} \tag{3.17}$$

Using relation 3.17, equation 3.15 can be restated as:

$$\int \rho \ddot{\boldsymbol{u}} \cdot \bar{u} = \int \nabla \cdot (\boldsymbol{\sigma} \bar{u}) - \int \boldsymbol{\sigma} : \nabla \bar{u} + \int \boldsymbol{f} \cdot \bar{u} \tag{3.18}$$

Using the divergence theorem, $\nabla \cdot (\boldsymbol{\sigma} \bar{u})$ can be rewritten as a surface integral in terms of the traction vectors acting on the surface of the volume:

$$\int \nabla \cdot (\boldsymbol{\sigma} \bar{u}) = \int t \cdot u ds \tag{3.19}$$

Performing the substitution of equation 3.19 into 3.18 and rearranging the terms gives:

$$\int t \cdot u ds + \int \boldsymbol{f} \cdot \bar{u} = \int \rho \ddot{\boldsymbol{u}} \cdot \bar{u} + \int \boldsymbol{\sigma} : \nabla \bar{u} \tag{3.20}$$

This can be further decomposed by splitting the integral into one over s_u and one over $s_{\boldsymbol{\sigma}}$:

$$\int_{s_u} t \cdot \bar{u} ds + \int_{s_{\boldsymbol{\sigma}}} \bar{t} \cdot u ds + \int \boldsymbol{f} \cdot \bar{u} = \int \rho \ddot{\boldsymbol{u}} \cdot \bar{u} + \int \boldsymbol{\sigma} : \nabla \bar{u} \tag{3.21}$$

Now consider following the process for a second displacement field, u^*, such that $\partial u = u^* - u$ and $u^* = \bar{u}$ on s_u. Subtracting equation 3.21 from its equivalent containing u^* gives:

$$\int_{s_{\boldsymbol{\sigma}}} \bar{t} \cdot \partial u ds + \int \boldsymbol{f} \cdot \partial u = \int \rho \ddot{\boldsymbol{u}} \cdot \partial u + \int \boldsymbol{\sigma} : \nabla \partial u \tag{3.22}$$

Equation 3.22 is the weak form of equation 3.12. For physical problems such as tissue modelling, the weak form of the equation is also called the principle of virtual work.

3.4.4 The Principle of Virtual Work

The principle of virtual work is an expression of the equilibrium state of a deformable body. It states that if a body is in equilibrium, for any virtual displacements, the summation of the virtual forces on the body must be zero. For a body, the internal and external virtual work are considered. The internal virtual work is the work done by displacements and forces deforming the body, i.e. stresses. The external work is that done by external forces applied to the system.

3.4.5 Discretising the Domain

In deriving the weak form of the governing equations the continuity requirements on field variables have been relaxed and the problem has been restated in an integral sense. If the weak form is solved, an approximated solution is obtained - the solution is valid in the integral sense, but is not necessarily a classically correct solution to the strong form of the equations. Despite this approximation, it is still necessary to solve for a continuous domain, so this is tackled by approximating the continuous domain with a discrete representation.

The development of the weak form introduced the test function u, which was a continuous kinematically admissable displacement field. In the Galerkin method of the finite element method, approximation of the solution field variable is conducted using the same form of equation used for the test function. Therefore, the continuous displacement field is approximated using a set of connected finite elements. Elements are connected between nodes, and the aim is to find an approximate solution by finding the nodal displacement values and describing the variation between the nodes across an element using an interpolation function.

For the test function and solution approximation, piecewise linear basis functions are selected, which are then multiplied by coefficients describing the values at each node. The shape functions are chosen such that they have local support, that is to say they are only non-zero inside the elements they form a part of. Additionally, it is required that they have the value 1 at the node they are associated with, and vary to 0 at the edges of the elements

they form a part of. As such, take $u = \sum_i u_i \psi_i$, where u_i is the nodal coefficient and ψ_i is the shape function at node i. In other words, the solution is described as a weighted sum of the basis functions.

Substituting the approximation into the principle of virtual work and rewriting in matrix form gives the discretised version of the equation:

$$M\ddot{U} + D\dot{U} + R^{int}(U) = R^{ext} \tag{3.23}$$

where M is the diagonalised mass matrix, D is the diagonalised damping matrix and R^{int} is the vector of internal forces.

Note that the chosen shape functions have very convenient properties. Ensuring they are 0 along the boundary of connected elements means that each element can be treated individually. Furthermore, diagonalising the mass and damping matrices ensures that there is no dependency between elements, and it is possible solve the terms for each element individually and simply express the solution as a sum of the individual element solutions.

3.4.6 Element Equations

Construction of the equations governing a particular element requires several components:

- The shape functions for the element
- The derivatives of the shape functions with respect to the reference configuration
- The deformation gradient of the element, a measure of how the element has deformed
- The 2nd Piola-Kirchoff stress tensor which relates the stresses pulled back to the reference configuration

Shape functions

The form of the shape functions depend on the type of element chosen to discretise the problem space. In this thesis, Reduced Integration Linear Hexahedral elements are used. Linear elements have linear shape functions, that is to say, a linear change in properties between nodes is assumed. Consider a standard hexahedral element, consisting of 8 corner nodes. Assume that the element is in a regular reference configuration, that is to say an undeformed cube.

The shape functions are first considered in natural coordinates. Natural coordinates are material coordinates, and each side of the reference element is of length 2, along one of the axes. Consider 3 axes, ξ, η and ζ. The nodes are numbered 1 to 8 and have the following coordinates:

Node	ξ	η	ζ
1	1	-1	-1
2	1	1	-1
3	1	1	1
4	1	-1	1
5	-1	-1	-1
6	-1	1	-1
7	-1	1	1
8	-1	-1	1

Table 3.1: The natural coordinate values of each node

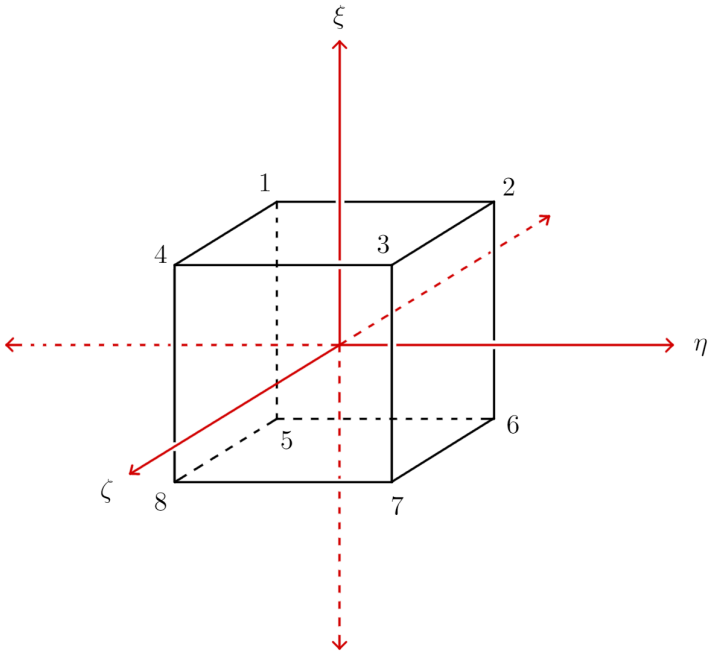


Figure 3.14: The reference hexahedral element and natural coordinate system

Each node will have associated one shape function, which should be valued 1 at the given node and 0 at all other nodes, varying linearly between them. These functions are lagrange polynomials, and they take the form:

$$N_i(\xi,\eta,\zeta) = \frac{(\xi-\xi_{conn})}{(\xi_i-\xi_{conn})}\cdot\frac{(\eta-\eta_{conn})}{(\eta_i-\eta_{conn})}\cdot\frac{(\zeta-\zeta_{conn})}{(\zeta_i-\zeta_{conn})} \tag{3.24}$$

where i is the index of the current node, and *conn* represents the index of the connected node along the relevant axis. For example, for $i = 1$, node 1 is connected to node 5 on the ξ axis, so ξ_{conn} will take the value of the ξ coordinate of node 5, i.e. $\xi_{conn} = -1$

Substituting the values of table 3.1 into equation 3.24 gives the eight shape functions for the linear hexahedral element in natural coordinates:

$$N_1 = \frac{(\xi+1)(\eta-1)(\zeta-1)}{8} \tag{3.25}$$

$$N_2 = \frac{(\xi+1)(\eta+1)(\zeta-1)}{8} \tag{3.26}$$

$$N_3 = \frac{(\xi+1)(\eta+1)(\zeta+1)}{8} \tag{3.27}$$

$$N_4 = \frac{(\xi+1)(\eta-1)(\zeta+1)}{8} \tag{3.28}$$

$$N_5 = \frac{(\xi-1)(\eta-1)(\zeta-1)}{8} \tag{3.29}$$

$$N_6 = \frac{(\xi-1)(\eta+1)(\zeta-1)}{8} \tag{3.30}$$

$$N_7 = \frac{(\xi-1)(\eta+1)(\zeta+1)}{8} \tag{3.31}$$

$$N_8 = \frac{(\xi-1)(\eta-1)(\zeta+1)}{8} \tag{3.32}$$

It is now possible to construct the matrix of shape functions:

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{N}_1 & \boldsymbol{N}_2 & \boldsymbol{N}_3 & \boldsymbol{N}_4 & \boldsymbol{N}_5 & \boldsymbol{N}_6 & \boldsymbol{N}_7 & \boldsymbol{N}_8 \end{bmatrix} \tag{3.33}$$

Where each of the submatrices N_i is:

$$\boldsymbol{N_i} = \begin{bmatrix} N_i & 0 & 0 \\ 0 & N_i & 0 \\ 0 & 0 & N_i \end{bmatrix}, i = 1, 2, ..., 8 \tag{3.34}$$

In order to use these shape functions in a useful way, they must also be expressed in terms of the typical cartesian coordinate system. The element that has been constructed here is called an isoparametric element. In isoparametric elements, it can be assumed that properties vary across an element in the same way as the geometry varies. This then means that the coordinate transformation can be given as:

$$x = \sum_{i=1}^{8} N_i x_i \tag{3.35}$$

$$y = \sum_{i=1}^{8} N_i y_i \tag{3.36}$$

$$z = \sum_{i=1}^{8} N_i z_i \tag{3.37}$$

Shape Function Derivatives

In addition to the shape functions themselves, the shape function derivatives are also required. These can be expressed in terms of the global coordinates through the use of the chain rule for partial differentiation:

$$\begin{bmatrix} \frac{\partial}{\partial \xi} \\ \frac{\partial}{\partial \eta} \\ \frac{\partial}{\partial \zeta} \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} = J \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} \tag{3.38}$$

and the inverse:

$$\begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} = J^{-1} \begin{bmatrix} \frac{\partial}{\partial \xi} \\ \frac{\partial}{\partial \eta} \\ \frac{\partial}{\partial \zeta} \end{bmatrix} \tag{3.39}$$

Deformation Gradient

The deformation gradient describes how an object has deformed from its reference configuration. For a deformed vector x and a reference vector X, the deformation gradient is given by:

$$F = \frac{\partial x_i}{\partial X_j} \begin{bmatrix} \frac{\partial x_1}{\partial X_1} & \frac{\partial x_1}{\partial X_2} & \frac{\partial x_1}{\partial X_3} \\ \frac{\partial x_2}{\partial X_1} & \frac{\partial x_2}{\partial X_2} & \frac{\partial x_2}{\partial X_3} \\ \frac{\partial x_3}{\partial X_1} & \frac{\partial x_3}{\partial X_2} & \frac{\partial x_3}{\partial X_3} \end{bmatrix} \tag{3.40}$$

However, note that a displacement u is given by $u = x - X$, and therefore $x = X + u$. Substituting this gives:

$$F = \frac{\partial (X_i + u_i)}{\partial X_j} = \frac{\partial X_i}{\partial X_j} + \frac{\partial u_i}{\partial X_j} = I + \frac{\partial u_i}{\partial X_j} \tag{3.41}$$

Note that as the deformation describes the transformation of a reference vector to a current spatial vector, the inverse of the deformation gradient can also be used to 'pull back' a spatial vector to its reference configuration. This is an important concept in the TLED method, in which the reference configuration is the initial configuration.

2nd Piola-Kirchoff Stress Tensor

Next the second Piola-Kirchoff stress tensor for an element is developed. The second Piola-Kirchoff (PK2) stress tensor is computed using a strain-energy function $\psi(C)$ which is a function of the deformation. For the Neo-Hookean material model used here, the deformation measure used is the right Cauchy-Green deformation, $C = F^T F$. The right Cauchy-Green deformation is split into isochoric (non-volumetric) and volumetric components. This decomposition also allows expression of the strain energy in terms of isochoric and volumetric components:

$$\psi(C) = \psi^{iso}(\bar{C}) + \psi^{vol}(J) \tag{3.42}$$

where J is the determinant of the deformation gradient and $\bar{C}$ is $J^{-2/3}C$, the modified right Cauchy-Green tensor. The PK2 stress tensor is then given by:

$$S = S^{iso} + S^{vol} = 2\frac{\partial \psi^{iso}}{\partial C} + 2\frac{\partial \psi^{vol}}{\partial C} \tag{3.43}$$

Strain energy functions are often expressed in terms of the the principle invariants of $\bar{C}$, $I_1 = tr\bar{C}$ and $I_2 = ([tr\bar{C}]^2 - tr\bar{C}^2)/2$. For a Neo-Hookean model, as used in this thesis, the strain energy function is then given by:

$$\psi^{NH} = \frac{\mu}{2}(I_1 - 3) + \frac{\kappa}{2}(J-1)^2 \tag{3.44}$$

where μ is the shear modulus and κ is the bulk modulus.

Nodal Forces

The internal forces of an element are those forces generated by the deformation of the object itself. These are typically restorative forces, attempting to return the body to its equilibrium state. The internal forces for an element are given by:

$$\boldsymbol{f}^e = \int_{0V^e} \partial_{\boldsymbol{X}} \boldsymbol{h} \boldsymbol{S} \boldsymbol{F}^T dV^e \tag{3.45}$$

where $\boldsymbol{f}^e$ is the internal forces for element e, V^e is the volume of element e, $\partial_{\boldsymbol{x}}\boldsymbol{h}$ are the shape function derivatives with respect to the reference coordinates, $\boldsymbol{X}$, $\boldsymbol{S}$ is the second Piola-Kirchoff stress tensor and $\boldsymbol{F}$ is the deformation gradient.

The integrals for the element are solved using reduced integration. More specifically, single point Gaussian Quadrature is used. This means that the integrals are evaluated using only the value at the centre of the element. This has two advantages. Firstly, displacement-based finite element solutions typically produce overly stiff elements. Using reduced integration will usually reduce the stiffness of elements, resulting in a better reproduction of real-life behaviour despite a loss of accuracy in computation. Second, reduced integration means that less computation must be performed to solve the integrals, leading to a faster simulation.

Gaussian quadrature is a numerical method which reduces an integral to a weighted sum of function evaluations. As an example:

$$\int_\Omega f(x)dV \approx \sum_i f(x_i)w_i \tag{3.46}$$

Where $f(x)$ is the function to be integrated over, i is the index of the integration point used, $f(x_i)$ is $f(x)$ evaluated at $x = x_i$, w_i is the weight for the relevant integration point.

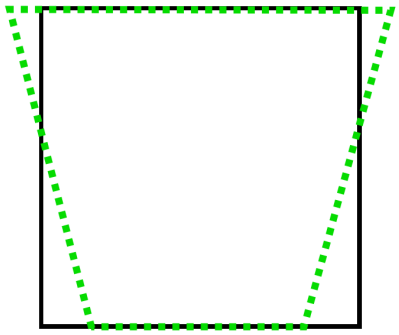


Figure 3.15: A representation of an hourglass deformation mode. The black line shows the undeformed element shape and the green dotted line the deformed shape. Note that the center of the element experiences 0 deformation, even though the element overall has deformed. Evaluating integrals using a single quadrature integration point at the center of the element will incorrectly report no stress.

The reduced integration 8-node hexahedra uses a single integration point located at the center of the element. For a single integration point, $i = 0$, $x_i = 0$, $w_i = 2$. In multiple dimensions, the weights are multiplied and so this weighting factor becomes 8.

Using the approximation from equation 3.46 in equation 3.45 gives:

$$\boldsymbol{f}^e \approx \sum_k \sum_j \sum_i \partial_{\boldsymbol{X}} \boldsymbol{h} \boldsymbol{S} \boldsymbol{F}^T w_i w_j w_k \tag{3.47}$$

and substituting the values for the weights and computing the sums:

$$\boldsymbol{f}^e \approx 8 \partial_{\boldsymbol{X}} \boldsymbol{h} \boldsymbol{S} \boldsymbol{F}^T \tag{3.48}$$

Hourglassing

The use of reduced integration via Gaussian quadrature benefits the simulation as it combats the volume locking typically caused by overly stiff fully-integrated elements, but it is not without its drawbacks. The Gaussian quadrature process involves sampling the value of the function being integrated at specific points. In the case of the 8-node linear hexahedron, 8 sampling points are required to accurately evaluate the integral. With the reduced integration element, only one sampling point at the center of the element is used. The function being evaluated when integrating to find the nodal forces is the evaluation of the stress tensor which describes the stresses within the material. The stress distribution is however not even across the entire element, so by sampling only at the center, information can be lost.

Certain deformations exist which do not deform the center of the element. As the central point is undeformed, when it is subsequently sampled for the integral, no forces are produced as the stress for the whole element is evaluated as 0. As such, these deformations cause no increase in the strain energy density of the element and no internal forces are generated to resist them. This violates the physical laws describing the system as deformation is created without any energy. These deformations are called *hourglass modes* and are physically impossible in the real world. Figure 3.15 shows an example hourglass deformation.

In order to counteract these zero-energy deformations, additional shape functions are introduced which describe the hourglass deformations. These are used in conjunction with the nodal displacements to evaluate hourglass nodal velocities. Based on these velocities, additional forces are applied based on the stiffness of the element and a user defined parameter which controls the maximum resistance to the hourglass modes. While additional computation is introduced in order to solve for the hourglass forces, it is considerably less intensive than fully integrating the element.

3.4.7 Global Equations

The forces generated by each element have been computed and may now be substituted into the global force vector which describes the forces on every degree of freedom in the system. Each local degree of freedom in the element has an associated global degree of freedom. The global forces are computed by summation of the forces of each node from each element to the relevant global degree of freedom. Each global degree of freedom may receive contributions from multiple elements, as multiple elements may contain the same global node. A single degree of freedom represents force along a particular axis at a node, so each node has 3 degrees of freedom. Force contributions from different elements at a node may reinforce each other or cancel each other out to give the resultant force at the node due to internal stresses.

The process of combining the local forces is termed *assembly* and can be described in an equation by the assembly operator:

$$R^{int} = \mathbf{A}_{e_i} f^{e_i} \tag{3.49}$$

3.4.8 Boundary Conditions

There are two forms of boundary conditions which can be imposed on the model. These are termed Dirichlet and Neumann boundary conditions, corresponding to imposing a displacement, or a force respectively. Dirichlet boundary conditions are described by:

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{u}^t_{constraint}, \boldsymbol{x} \in \Gamma_u \tag{3.50}$$

Neumann boundary conditions are described by:

$$\boldsymbol{f}(\boldsymbol{x}, t) = \boldsymbol{f}^t_{constraint}, \boldsymbol{x} \in \Gamma_f \tag{3.51}$$

These boundary conditions are imposed upon the global system matrices and vectors. Dirichlet boundary conditions are imposed by simply setting the displacement of a degree of freedom to 0. Neuman boundary conditions are applied by adding the force to the relevant degree of freedom. By adding it to the existing value rather than replacing the existing value, the internal forces due to deformation are preserved.

3.4.9 Advancing the Simulation in Time

The new displacements for the next time step can now be computed. The system is currently described by:

$$M\ddot{U} + D\dot{u} + R^{int}(U) = R^{ext} \tag{3.52}$$

A central difference scheme is used to express the accelerations and velocities in terms of the previous, present and next displacements. The central difference scheme is a finite difference method used to approximate derivatives. For a function $f(x)$, a finite difference equation is one of the form $f(x + a) - f(x - b)$. Dividing this by $a - b$ gives a quotient approximating the derivative between a and b. The central difference method evaluates the

derivative at a point x using $a = b$. Using this relation and setting $a = \delta t, b = \delta t$, it is possible to approximate the velocity and acceleration in terms of the displacements from the previous time step, current time step and next time step.

$$\dot{U}_n \approx \frac{1}{2\delta t}(U_{n+1} - U_{n-1}) \tag{3.53}$$

$$\ddot{U}_n \approx \frac{1}{\delta t^2}(U_{n+1} - 2U_n + U_{n-1}) \tag{3.54}$$

where δt is the timestep size. Substituting these relations into equation 3.52 and rearranging to solve for U_{n+1} gives:

$$U_{n+1} = A(R^{ext} - R^{int}) + BU_n + CUn - 1 \tag{3.55}$$

where A, B and C are the following matrix coefficients:

$$A_{ii} = 1 \Big/ \left(\frac{D_{ii}}{2\delta t} + \frac{M_{ii}}{\delta t^2}\right) \tag{3.56}$$

$$B_{ii} = \frac{2M_{ii}}{\delta t^2} \Big/ \left(\frac{D_{ii}}{2\delta t} + \frac{M_{ii}}{\delta t^2}\right) \tag{3.57}$$

$$C_{ii} = \left(\frac{D_{ii}}{2\delta t} - \frac{2M_{ii}}{\delta t^2}\right) \Big/ \left(\frac{D_{ii}}{2\delta t} + \frac{M_{ii}}{\delta t^2}\right) \tag{3.58}$$

These coefficients are not time dependent and as such can be precomputed. Recomputation is not necessary at each time step. The new displacements for this timestep have now been computed. The current and past displacements are updated, and the process is repeated until the desired simulation length has been reached.

3.4.10 Contact Handling

The contact handling approach used in this thesis draws inspiration from the Material Point Method. In the MPM, the continuum is described by particles but is mapped to a finite element background mesh at each step. Contact is handled implicitly here, as properties of nearby particles are mapped to nodes of the same background element, so the element behaves as though it is part of a single material. Given the physical properties of the elements, they cannot be inverted and so act as a contact 'buffer' between the two materials.

To handle contact in the simulation, elements between the lips that are about to collide are instantiated. Specifically, the saliva elements are used. The elements are set to be very stiff under compression relative to the surrounding tissue so that they form a consistent border between the two interfaces. Ensuring the elements only become highly stiff under compression prevents them from affecting the behaviour of the surrounding tissue. Although the saliva elements are used to support the stickiness behaviour, they can also act as normal elements so this does not necessarily mean that stickiness is always active when the lips are in contact. Compression and stickiness aside, they behave as normal soft tissue, to be consistent with the rest of the model.

Once a saliva element has broken, it must be re-instantiated when the lips come into contact, both to support further stickiness behaviour and in order to handle the contact between the lips correctly. Full details of the stickiness behaviour are given in section 3.5, but for now it is sufficient to know that once a saliva element breaks, it is 'deleted' by setting the force it exerts on its nodes to zero. By setting the force it exerts to zero rather than fully deleting the element from the mesh, its other properties can still be tracked. As such it is possible to track how much the element has stretched relative to its original height. As the lips approach each other and the saliva/contact elements return to a relative stretch of 1 or below, the element is toggled back on and begins exerting forces again. As it is

near to its original deformation state when the forces are reintroduced, the shock should be minimal. Nonetheless, in order to combat the sudden introduction of force and a (relatively) stiff element, the force is reintroduced gradually.

3.4.11 Simulation Process Summary

Pre-computation:

1. Load XML description of geometry and simulation parameters
2. For each element, compute the Jacobian, J, its determinant $det|J|$, the derivatives of the shape functions and the linear portion of the strain–displacement matrices.
3. Compute the lumped (diagonalized) mass matrix ${}^0M$

Initialisation:

1. Initialise the initial displacements and any strains/forces that are already present in the model.

Computing time steps:

1. Update past and present displacements.
2. For each element:
   (a) Compute the deformation gradient ${}^t_0F$.
   (b) Compute the full strain displacement matrix which accounts for the initial displacement effect.
   (c) Compute the second Piola-Kirchoff stress tensor at each integration point (in this case one, at the center of the element
   (d) Use gaussian quadrature to compute the nodal force values by solving the integral.
3. Obtain the net nodal reaction forces.
4. Compute the displacements for the next time step using the central difference formula
5. Apply new loads for the next timestep.

3.5 Developing a Stickiness Model

3.5.1 Targeted Effects

A number of cases were identified in which the stickiness of the lips produces effects or movements which would not take place without stickiness. The most prominent of these is the zippering effect the mouth often has when opened. The lips will typically first separate in the center and then open outwards towards the edges as the jaw is further lowered (Figure 3.16).

The second most noticeable effect is that sometimes multiple mouth openings appear when the stickiness across the lips is not uniform. This leads to a particular portion of the upper and lower lips being stuck together temporarily while the rest of the lips separate around the stuck point.

In addition to these two effects which are very obvious when seen, the stickiness between the lips also affects the dynamics of the lip movement. When less sticky, the lips separate with less force. In cases of higher stickiness, they must be stretched apart, leading to a sudden acceleration when the saliva bond breaks.

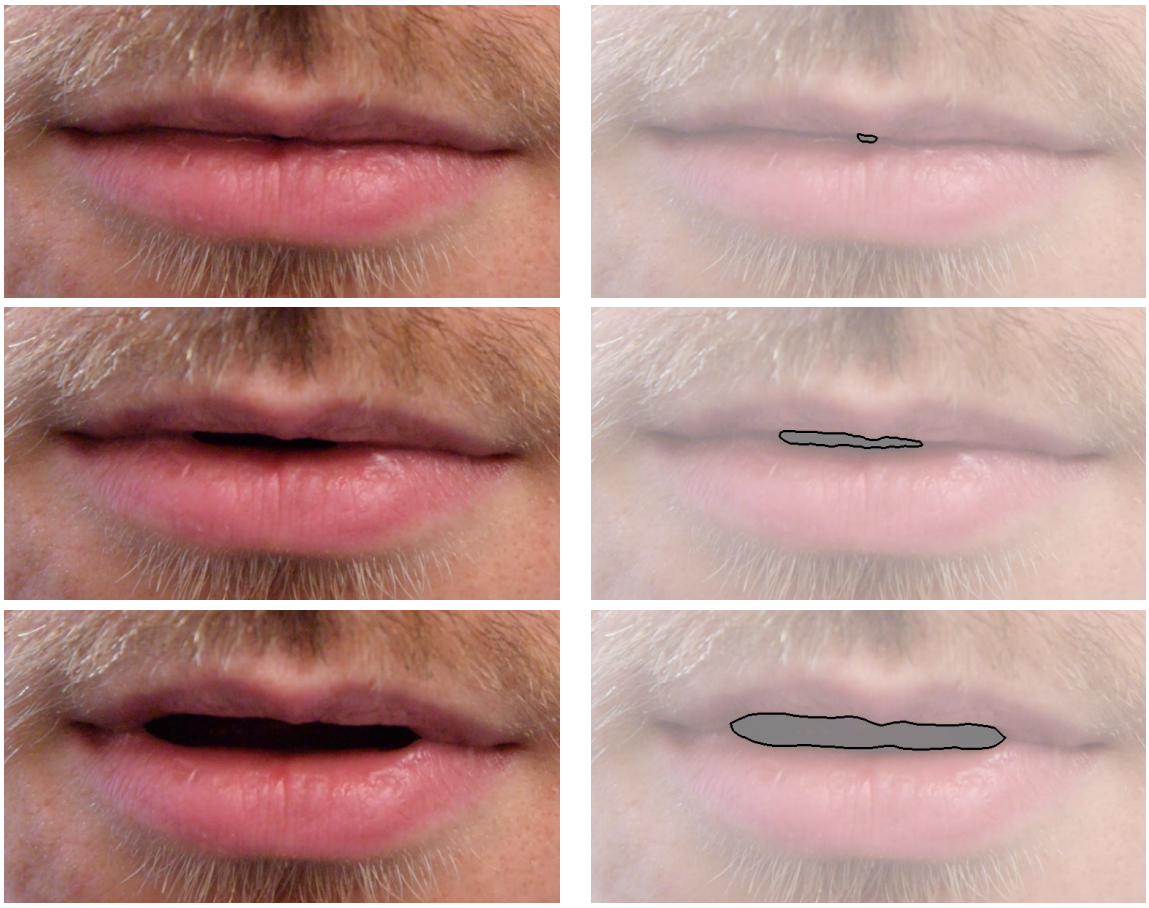


Figure 3.16: Zippering of the lips as the mouth is opened. The right hand column highlights the inner mouth contour to show its development.

3.5.2 Stickiness Behaviour

NiftySim was modified to support the stickiness behaviours. To model the stickiness caused by the saliva, the saliva is treated as a glue between the lips. There are three distinct states in which the saliva and lips can be considered to be in:

Completely Dry When completely dry, the saliva has no impact on the interaction between the lips.

Very wet When very wet, the saliva may act as a lubricant between the lips. This will affect any movement in which there is friction between the lips, such as if the jaw is moved from side to side while the mouth is closed. In this state, the saliva has little impact on the lips being drawn apart, as the liquid particles of the saliva have very little cohesion and as such apply only negligible forces to the lips.

Sticky Between the completely dry and very wet is the range where the saliva impacts the behaviour of the lips most. In this range, the saliva has dried sufficiently that the cohesive forces between the saliva molecules offer some resistance to being pulled apart. This in turn exerts additional forces on the lips.

This thesis aims to recreate those effects found in the *sticky* phase. As noted above, the forces involved depend on the cohesive forces between the saliva particles. The level of this force is represented with a stickiness level, Z. This stickiness level, or breaking point, is the amount a saliva element may stretch before breaking. The saliva elements are very

Figure 3.17: Multiple mouth openings produced as a result of uneven stickiness across the lips

thin and stiff, so the actual stretching of the element has minimal impact on the simulation, however, the stretch is proportional to the the force across the element. As such the stickiness level/breaking point is a representation of the force required to break the element.

To determine whether a saliva element should break, the current stretch of the element is compared with the breaking point or critical stretch, Z. In order to compute the current stretch of a given element, the deformation tensor is first computed. This is a second order tensor which maps a vector from its reference configuration to its current configuration. In the TLED formulation, the reference configuration is the initial configuration, i.e, the configuration at time 0:

$$F_{ij} = \delta_{ij} + \frac{\delta\ {}^{t}u_i}{\delta\ {}^{0}x_j} \tag{3.59}$$

where F is the deformation gradient at time t, relative to time 0, δ_{ij} is the Kronecker delta, ${}^{t}u$ gives the displacements at time t, and ${}^{0}x$ gives the initial positions. The global displacements are first transformed into a local coordinate system, so the initial orientation of the element in global space does not affect the deformation gradient.

From the deformation gradient, the right Cauchy-Green deformation tensor is computed. This tensor is rotation-invariant and physically gives the square of the change in lengths due to deformation:

$$C_{ij} = F_{ij}^{T} F_{ij} \tag{3.60}$$

where C is the right Cauchy-Green deformation tensor at time t, relative to time 0. The yy component of this tensor gives the square of the stretch in the local y direction of the undeformed finite element. This is the central component of the 3x3 tensor, C_{11}, which is used as the current stretch, Q:

$$Q = C_{11} \tag{3.61}$$

With the current stretch computed, the force that the saliva is exerting may now be evaluated. This is a modification to the normal force computation process for the element. The current stretch, Q, is compared against the critical stretch, Z. If the current stretch is below the critical stretch, the element behaves according to its normal behaviour, in this case, following a Neo-Hookean model. If the current stretch is greater than the critical stretch, the force across the element has exceeded its failure point and the saliva will break.

Typically, unless the saliva is very dry, it will undergo a ductile failure. This means that the force exerted by the saliva element gradually decreases, as opposed to a brittle failure in which the material breaks instantly and the force produced is instantly reduced to 0. In addition, each saliva element is representative of a continuous region of a saliva on the lip. In the 'zippering' effect in real mouths which has been identified in this work, the saliva breaks progressively along the length of the lip. This would also lead to a gradual decrease in force across a given element. The updated description of the force computation for an element is then:

$$ {}^{t}F = \begin{cases} {}^{t}f & \text{if } Q < Z \\ 0.99 \ {}^{t\text{-}1}F & \text{otherwise} \end{cases} \tag{3.62} $$

where ${}^{t}F$ is the modified internal force of an element at time t, ${}^{t}f$ is the unmodified internal force at time t (i.e. the direct result of the TLED force computation for the element) and ${}^{t-1}F$ is the modified force from the previous timestep (i.e. once the saliva element has broken, the force generated by the element at each timestep is 0.99 times that of the previous timestep). These additions are introduced in the form of a new material type in NiftySim. This gradual decay of force also avoids shocks in the system which can cause instabilities.

3.5.3 Dynamic Element Instantiation

To allow for further sticky behaviour after the first time the mouth has been opened and the saliva elements have broken, the saliva elements must be reset if the mouth is closed again. This process was briefly described in the discussion of contact mechanics 3.4.10 but is presented in more detail here.

The breaking of the saliva elements is operated by reducing the force that they exert on their nodes to 0. As such the elements still exist within the mesh, and their properties can be tracked. Specifically, the stretch measure used to determine when the element breaks, $Q = C11$, the current stretch given by the principle y component of the right Cauchy-Green deformation tensor is still tracked.

Q is the square of the stretch of the undeformed y axis of the element. This means that if an element has broken and Q becomes very close to 1, that the lips are about to contact. At this point the forces the element produces are restored. Only reinstating the element as the stretch is close to 1 ensures that the contact behaviour begins at the right time. Additionally, this helps to minimize the shock forces introduced as the element is reinstantiated, as in theory its deformation and therefore stresses are at a minimum.

Nonetheless, the introduction of the forces and the effective reintroduction of the mass of the saliva element to the system can produce shocks, so much like the force decay, the force is reintroduced gradually over a series of frames. For a timestep of $1e^{-5}$ seconds, a linear interpolation of the element forces over 1,000 frames was found to be effective. This corresponds to a total time of 0.01 seconds, insignificant in relation to the total time of the animation and imperceptible during an animation played at normal speed.

3.5.4 Implementation within NiftySim

Implementation of the stickiness behaviour in NiftySim required the creation of a new material model and its associated functions. As the material model is based on the Neo-Hookean material model, the new model was termed *Neo-Hookean Breaking*. Material models in NiftySim are defined by various components. Each has an associated abbreviation which is used to reference the material model in the XML file. A new abbreviation was created in the tledSolverGPU.cu file which controls reading of the XML model. In the same file, an additional case is added to the UpdateMaterialParams function.

The remaining modification was in the nodal force calculation kernel. The ComputeNewForcesH8_kernel was modified to include the additional NHB material type, which for the most part mimics the behaviour of the Neo-Hookean material. Additionally, the function was modified to take an additional parameter, int* g_broken which is an array containing an integer for each element. This integer is used to track the breaking state of the element and is initialised to 0 for each element, representing an element which is unbroken. The stress calculation proceeds in the same way, however the nodal force calculation at the end is modified.

If the element breaks, its associated value in the g_broken array is updated to 1. For an element with a value of 1, the force is computed as 0.999 multiplied by the previous force value. If a broken element's stretch returns to below 1.0001, then the element has its value in the g_broken array set to 2. This indicates that the element is being reinstantiated. For elements with a g_broken value of 2 or above, the force applied to their nodes is calculated as shown in equation 3.63

$$\hat{f} = f * 0.0001 * g \tag{3.63}$$

where $\hat{f}$ is the modified nodal force, f is the nodal force computed using the standard Neo-Hookean behaviour, and g is the value for the element in the g_broken array. Each time the force is computed, g is incremented by one, leading to a linear scaling of the force up to its proper value. Computing the force as a proportion of the standard force in this way still allows changes in the forces due to changes in the element shape as it is being reinstantiated. The scaling factors used here were determined empirically to provide good simulation stability for the selected timestep whilst minimising the time taken for the modified values to reach their targets. They would likely need adjusting for different timesteps as the timestep affects the stability of the simulation.

Algorithm 1: Modified Force Computation

```
foreach SalivaElement do
    if g_broken == 0 then
        finalForce = computeStandardForce();
    else if g_broken == 1 then
        finalForce = 0.999 * previousForce;
    else
        finalForce = computeStandardForce() * 0.0001 * g_broken;
        g_broken = g_broken + 1;
        if g_broken <= 10000 then
            g_broken = 0;
        end
    end
    previousForce = finalForce;
end
```

3.6 Automating Stickiness - A Moisture Model

Having developed a stickiness model and looked at some of the effects produced by it in more detail, it is clear that the stickiness level can have a significant impact on the movement of the mouth. In order to produce realistic animations over longer timer periods, it is desirable to also model how the stickiness level changes over time. This is primarily controlled by the moisture level on the lips.

3.6.1 Moisture Model Derivation

The moisture level on the lips is one of the key factors affecting the level of stickiness between the lips. For the purposes of this work, it is reasonable to consider two phases in which the moisture level of the lips changes significantly: the wetting phase and the drying phase.

Wetting

The wetting of the lips typically happens due to the lips being licked. The salivary glands inside the mouth cavity produce saliva, which is picked up by the tongue. This is transferred by physical contact to the surface of the lips. In relation to the time it takes for the lips to dry, the wetting process is very fast. It is possible for the tongue to only transfer saliva to parts of the lips, for example, a person may lick only one side of their lips.

The resulting level of moisture on the lips once they have been licked depends on a number of factors, namely: the level of moisture already on the lips, the amount of saliva on the tongue and the moisture content of the saliva.

Other mechanisms for moisture transfer to the lips include drinking, eating and weather effects. These other mechanisms are not directly modelled, but could be easily included if modelling of such a scene was desired.

In the presented model, it is possible to have a variable moisture level across the lips. This is achieved by giving each element its own moisture level. This is simply set initially - the model does not include a tongue, so there is no modelling of the transfer. However, it is possible to dynamically update the stickiness, which could be used to simulate the effect the tongue would have in a particular motion. The moisture distribution is typically quite uniform across the lips, so for most simulations a uniform initial moisture level is used.

Drying

To model the drying process, two parameters are considered: the moisture level, and the evaporation rate. To model the evaporation rate, the Herz-Knudsen equation is used as a starting point:

$$\frac{1}{A}\frac{dN}{dt} \equiv \varphi = \frac{\alpha(p_{liq} - p_{eq})}{\sqrt{2\pi m k_B T_{liq}}} \tag{3.64}$$

where A is the surface area of the fluid, N is the number of gas molecules, φ is the flux of gas molecules, α is a sticking coefficient for the gas molecules onto the surface, p_{liq} is the pressure of the liquid during evaporation, p_{eq} is the liquid pressure at which an equilibrium between evaporation and condensation is reached, m is the mass of a liquid particle, k_B is the Boltzmann constant and T_{liq} is the temperature of the liquid.

This states that the rate of evaporation for a given area is proportional to a constant and the pressure difference across the fluid-gas interface. It is inversely proportional to the square root of the mass of the liquid particles and the temperature of the liquid.

In the model it is assumed that the temperature of the liquid is constant. The saliva has come from inside the body, specifically the mouth, and as such will already be around body temperature. It is then transferred to the lips, at which point it may cool slightly depending on the atmospheric temperature. This cooling will take place very quickly and the temperature of the saliva will reach an equilibrium. The saliva film on the lips is only 0.07mm-0.1mm thick, so there is a very small volume of fluid for the temperature change to occur in. The time taken for the temperature to reach equilibrium is negligible in comparison with the time taken for the saliva to evaporate, so the temperature may be considered to be constant.

The pressure of a liquid or gas is given by

$$P = \frac{nRT}{V} \tag{3.65}$$

where P is the pressure, n is the number of molecules present, R is the ideal gas constant, T is the temperature of the gas and V is the volume taken up by the gas. As previously established, the temperature of the saliva can be considered constant. There are no volume constraints on the saliva, as such, the liquid pressure, pliq, can be considered to be a constant. The atmospheric pressure also deviates very little, so this is also treated as constant. As such, the difference in pressures is also a constant. Finally, the mass of the particles evaporating will not change over the course of the drying phase, and so this is also constant.

Each term in the right hand side of equation 3.64 has been demonstrated to be a constant value. As a result of this, the entire right side of the equation may be represented with a single constant. The physical interpretation of this is that there is a constant rate of evaporation of saliva from the lips.

The Herz-Knudsen equation describes the evaporation rate in stable conditions. In reality, and in the case of the mouth and lips, the evaporation rate is also affected by the velocity of the air moving over the surface. This means that the evaporation rate is affected by both breathing and speech. Humans may breathe either through the mouth or the nose. If breathing through the mouth, clearly there is airflow across the lips as air enters and exits the mouth cavity through the mouth. If breathing through the nose, there is still some airflow across the lips as air is expelled and drawn in through the nostrils. Breathing is a periodic process. Typically, a human breathes around 16 breaths per minute. This means an average breath takes a little under 4 seconds. This is relatively short compared with the time the evaporation happens over, and given its periodic nature can be treated as constant. This means that even with breathing taken into account, there is still an effectively constant evaporation rate for the saliva from the lips. This constant evaporation rate will be termed φ.

The moisture level for a given element at a given time now follows:

$$ {}^{t}M = \begin{cases} {}^{0}M - t\varphi & \text{if } {}^{0}M > t\varphi \\ 0 & \text{otherwise} \end{cases} \tag{3.66} $$

where ${}^{0}M$ is the initial moisture level for the element, t is the elapsed time since the lips were last wetted, φ is the evaporation rate and ${}^{t}M$ is the current moisture level. A moisture level of 0 represents a completely dry area of lip.

3.6.2 Controlling the Stickiness Level with Moisture

In order to construct a relation between the moisture level and critical stretch, a mapping function is needed. A linear mapping function is used. The form of this function does not have a confirmed physical basis and investigation of this relationship would be an interesting area of further work. The suggested mapping function, f is:

$$ Z = f({}^{t}M, M_c) = \begin{cases} 1 & \text{if } {}^{t}M > M_c \\ 1 + K\frac{{}^{t}M}{M_c} & \text{otherwise} \end{cases} \tag{3.67} $$

where K is a positive coefficient describing the maximum strength of the saliva's gluing effect. This gives a linear relation between the moisture level and stickiness level, with a maximum critical stretch given by K, and a minimum critical stretch of 1, representing 0 stickiness.

3.7 Visualisation

NiftySim had limited rendering options. As such, two custom rendering options have been implemented for visualisation of the results of simulations. The first is an internal renderer which is part of the program and displays output as soon as the simulation is complete. The

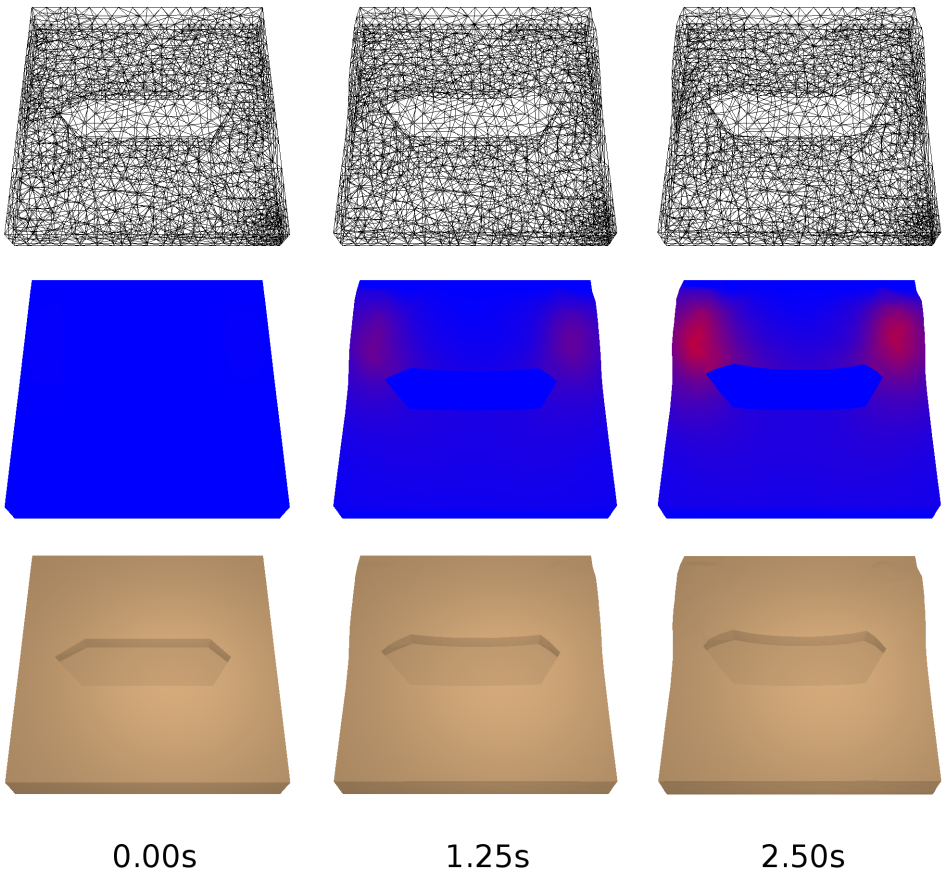


Figure 3.18: Different display modes of the internal renderer demonstrating an animation from an early version of the model.

second option provides increased graphical fidelity by exporting the results and rendering them in the 3D modelling package blender.

3.7.1 Internal Renderer

The internal renderer is implemented using OpenGL. As the simulation is computed, results are saved from NiftySim. The user may specify how often results should be saved out, allowing a frame-rate to be specified. NiftySim natively offers saving of displacement data, however information regarding the state of elements which can break is also required, so the NiftySim software was extended to support this.

The internal renderer provides various controls for adjusting the replay of animations. In order to facilitate viewing of dynamic effects which may happen very quickly, the replay rate can be adjusted, allowing replay speeds from 0.1x to 5x. Additionally the animation may be paused, and the frame may be advanced or rewound one frame at a time. Different display modes allow the visualisation of different properties. Figure 3.18 shows an early mouth model in wireframe mode, stress visualisation mode and normal rendering mode.

3.7.2 Export to Blender

In order to produce higher quality animations, the animation data may also be exported to the 3D modelling package, Blender. Blender contains an industrial quality rendering engine, Cycles. The animations produced are ray-traced using a GPU accelerated rendering engine

and support physically-based materials. This is important for skin rendering and allows modelling of advanced physical lighting phenomena such as sub-surface scattering.

3.7.3 Texturing

Typically the animations produced have been untextured as an untextured model provides clear definition between the lip region and surrounding skin. It is also easier to see the stickiness effects in these simplified renderings. With that said, it is possible to include texturing in animations produced by the model. The internal renderer supports texturing, and can use the textures produced by the FaceGen software used to generate the model initially. Textures can also be used in the Blender render path in conjunction with the physically-based materials.

3.8 Discussion

The face model used throughout the thesis is based on the author's face and is made up of hexahedral elements comprising a conforming mesh. The modelling approach can be used to semi-automatically produce user specific mouth models. Vector muscles and bone constraints are added to the NiftySim XML model using a script developed which allows specification of constraint regions by detecting mesh nodes which lie inside a 3D-modelled volume. The model is simulated using the TLED finite element method implemented in the NiftySim package. The TLED method handles large deformations as necessary in soft-tissue modelling. An evaluation within NiftySim of the most commonly used element types was conducted, providing a performance comparison between the types of element. Behaviour under bending was also investigated, demonstrating the volume locking issues present in linear tetrahedral elements.

With the primary aim being to recreate visually realistic mouth movement, not all biological structures and physical effects are modelled. Those which have limited to no effect on the movements of the mouth are either simplified or not implemented as they are not relevant. For example, simplified muscle structures are used in which the effect of multiple mouth muscles are combined if they act to produce similar movements. Further development of the model to include a more complex representation of the anatomy of the mouth could improve the animations, including features such as the wrinkling of the lips and interactions with the teeth.

The presented stickiness model uses a deterministic, physics-based material model which produces more realistic stickiness behaviour than other current models. The stickiness behaviour is controlled by an automated moisture model, allowing natural, dynamic changes of behaviour over time. The simulation incorporates a novel approach to handling the stickiness between the lips by physically modelling the properties of saliva. Saliva is modelled using finite elements using a Neo-Hookean material model with modified force computation. The use of the finite elements allows recreation of deterministic, physically-based non-linear behaviour, something not recreated in other works which rely on stochastic (e.g. (Barrielle and Stoiber, 2018)) or artist directed methods. Additionally, the saliva elements and modification of force computation allow trivial handling of contact mechanics between the lips. The inclusion of stickiness visually improves the realism of mouth animations generated with the finite element method.

To support the stickiness behaviour, a moisture model is proposed which replicates evaporation behaviour of the saliva from the lips. The model is derived from physically-based equations and empirical data from videos of mouths collected by the author. The moisture model acts for each element individually allowing variation in stickiness levels across the lips. This recreates asymmetric effects and multiple mouth openings. Moisture levels can be

set programmatically allowing updating of moisture levels to mimic the licking of the lips. Moisture evaporates at natural rates, providing a realistic change in stickiness over time.

The actual fluid dynamics are not modelled. This somewhat limits the ability of the model to automatically produce entirely correct behaviours based on interaction between the tissue structures of the mouth. For example, the tongue moving over the lips will not automatically apply a correct amount of saliva, although effects such as this can still be modelled by including a method for updating the moisture levels based on these interactions. This is an interesting area for further exploration, as the tongue and lips regularly come into contact during speech.

The simulation method is capable of producing animations with varied levels of stickiness on the lips. This could potentially also be useful in the study of medical conditions which affect the levels of saliva in the mouth, e.g. xerostomia. Further potential medical uses are prediction of mouth behaviour following a surgical procedure which affects saliva production or results in a change of structure in the patient's mouth region. The presented pipeline allows for user-specific models to be created semi-automatically, meaning patient-specific models could be produced.

The stickiness effect could also be used to model other scenarios involving soft tissue and other cohesive substances. This could have applications to medical research in the field of gynecology, or to the contraceptive industry. Aside from purely biological examples, it could be used to model the stickiness behaviour caused by glue on the skin. The moisture model presented in this work allows for the drying of the glue over time and the resulting changes in behaviour could be modelled. Although this research is primarily aimed at recreating realistic mouth behaviours, the stickiness model could also easily be used to create special effects, such as the mouth being glued shut or coated in something very sticky.

Chapter 4

Results

The results demonstrate that the inclusion of stickiness modelling enhances the capabilities of the finite element model, allowing recreation of natural lip movements and phenomena. Five sets of examples are presented, demonstrating various features of the model. The first set shows the effects which the model is capable of reproducing. The second set shows the effect of variation of stickiness. The third set shows how the moisture model can control the stickiness levels automatically. In the fourth section, examples are shown which use extreme amounts of stickiness which could be used for special effects. The final section shows frames of full animations. Although only stills from the animations are presented in this thesis, video animations can be viewed at `http://staffwww.dcs.shef.ac.uk/people/M.Leach/research.html`.

4.1 Mouth Model Capabilities

The first results section will demonstrate the capabilities of the model in forming particular expressions and show the effect of activating the different muscle groups which are modelled. The aim is to show how the model can function, then the remaining sections will highlight key features regarding the stickiness model and effects.

4.1.1 Activation of Muscle Groups

The jaw is responsible for the largest movements in the mouth as it is the main mechanism for opening and closing of the mouth - this action forms the basis for most speech and expressions. Although the jaw is capable of slight translations, its key movement is rotation around the temporomandibular joint resulting in a lowering of the chin. This action is modelled by applying displacement-based constraints to the chin region of the model. A typical maximal displacement for the chin is around 40-50mm. Figure 4.1 shows a partial opening of the mouth controlled by the jaw.

The second muscle group modelled primarily simulates the behaviour of the zygomaticus major. The group also includes the zygomaticus minor which is very closely aligned with the zygomaticus major. Collectively these muscles draw the corners of the mouth outwards and upwards. These muscles are used for smiling.

The oris orbicularis is the muscle which runs through the lips. It is unique in the face in that it is a sphincter and acts to contract towards a central point. It causes the puckering of the lips. Figure 4.2 shows the activation of the oris orbicularis in the model.

The risorius group represents the risorius and buccinator muscles. The group draws the corners of the mouth slightly outwards although its primary purpose is to stiffen the cheeks. However, it is also used in the creation of some facial expressions such as the smile. The levator labii group represents the levator labii superioris and serves to raise the upper lip.

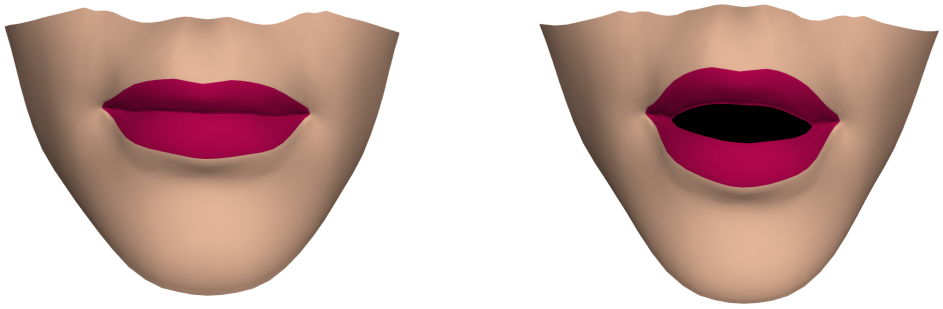


Figure 4.1: Activation of the jaw in the model

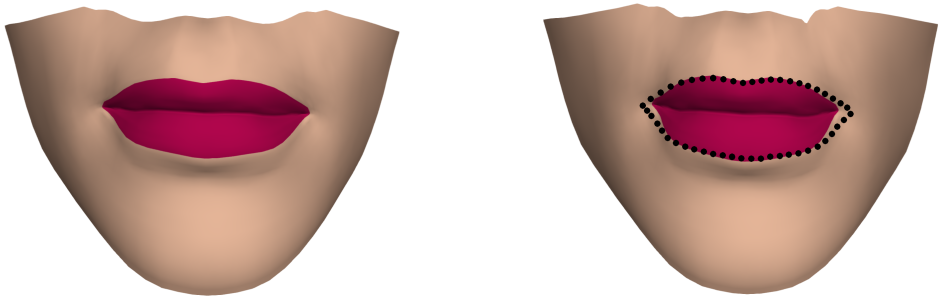


Figure 4.2: Activation of the oris orbicularis muscle in the model. The black dots show the contour of the resting mouth.

Figure 4.3 shows activation of the risorius group and figure 4.4 shows activation of the levator labii group.

Figures 4.5 and 4.6 demonstrate different mouth poses recreated in the model. Muscle activations for each of the different muscles groups are given with each of the examples.

4.2 The Impact of Stickiness

The first effect demonstrated is the change in the way the mouth opens. Rather than opening along its whole length at once as seen in the finite element model without stickiness (Figure 4.7), it can be seen that like the real mouth (Figure 4.8), the model including stickiness leaves the corners of the mouth closed as the central portion of the lips open (Figure 4.9)

This is a result of the zippering effect in which the lips gradually open outwards from the point at which they first separate. This can be better seen looking at the frames over an animation. Figure 4.10 shows a mouth opening over time, comparing a model with no

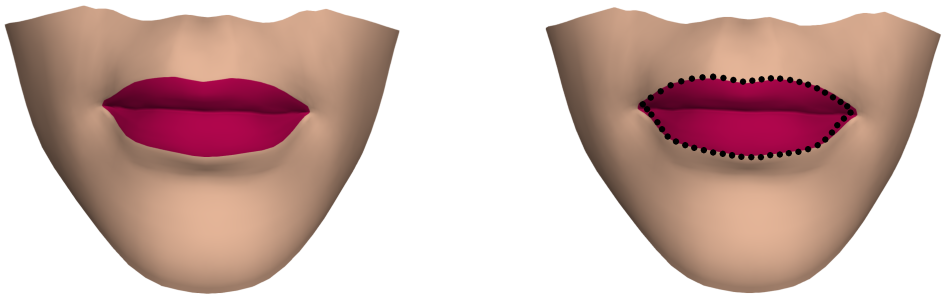


Figure 4.3: Activation of the risorius group in the model. The black dots show the contour of the resting mouth.

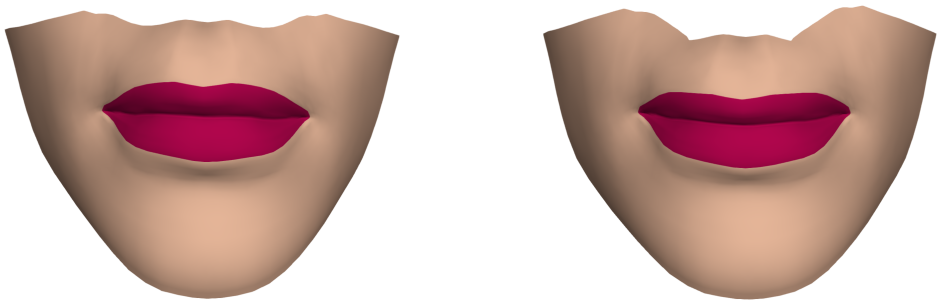


Figure 4.4: Activation of the Levator Labii group in the model

Figure 4.5: A frame from a video of a real mouth during speaking of the letter ‘m’ and the model reproducing the mouth shape.

Figure 4.6: A frame from a video of a real mouth during speaking of the ‘ee’ sound at the end of ‘mummy’ and the model reproducing the mouth shape.

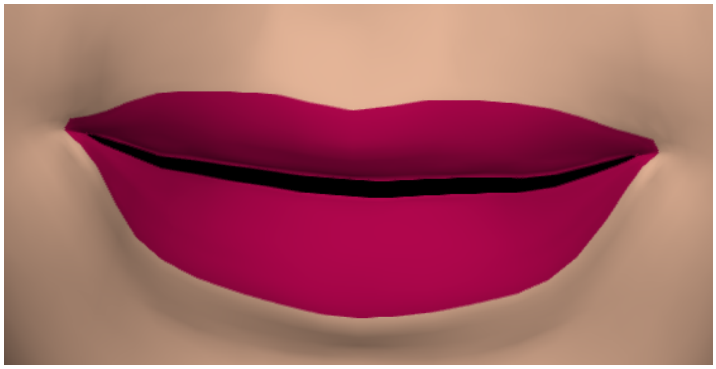


Figure 4.7: A frame from an animation produced using a finite element model with no stickiness.

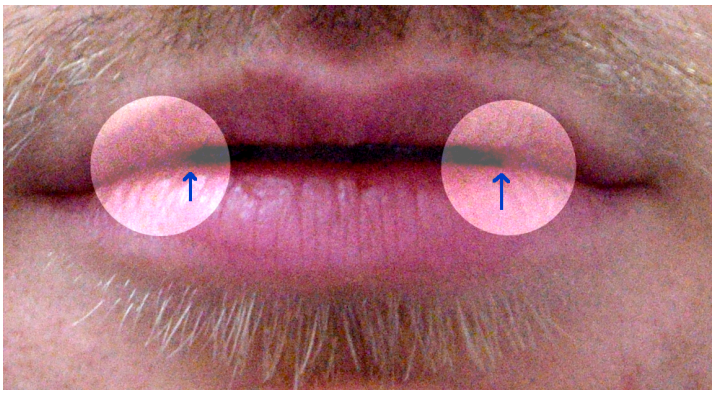


Figure 4.8: A real mouth during an opening action. Highlights show the stuck corners of the lips.

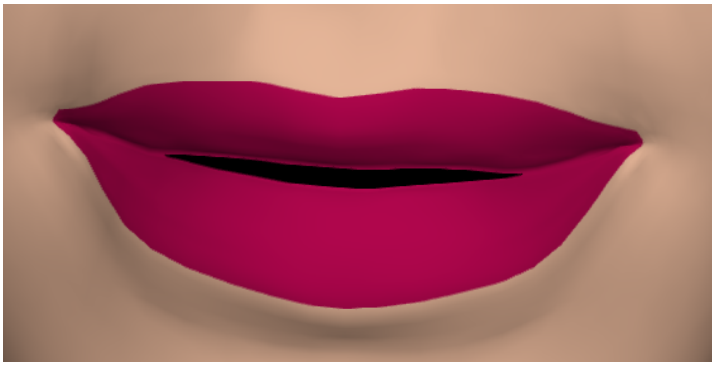


Figure 4.9: A frame from an animation produced using a finite element model with the presented stickiness model. The stuck corners of the lips are recreated.

stickiness, a real mouth and the model including stickiness. Visually it is clear that the overall movements are much closer to that of the real mouth when stickiness is included. Additionally, the final position reached in the animation is considerably improved.

Aside from recreating the zippering effect, the presented model also easily replicates asymmetric effects. Figure 4.11 shows an asymmetric mouth opening created by using non-uniform stickiness across the lips. Further imbalances in stickiness can cause multiple mouth openings to form. This behaviour is not simulated by other stickiness models, but is recreated effectively using the model presented in this thesis by setting the stickiness level for a central element to be higher than that of its neighbours (Figure 4.12).

Barrielle's (Barrielle and Stoiber, 2018) method is capable of reproducing some of these effects, however, the stochastic nature of his solution can cause large numbers of very small mouth openings to form in places which are not physically realistic, for example, near the corners of the mouth even when the mouth is mostly closed and opening from the centre. Figure 4.13 shows a real mouth and his method reproducing the real mouth's movement. The implausible openings can be seen in the first and second frames. In contrast, the method presented in this thesis generates deterministic and consistent behaviour, with mouth openings only forming in realistic places.

The stickiness behaviour is easily integrated with common rendering techniques such as texture mapping. Figure 4.14 shows a real mouth producing a shape and its texture-mapped counterpart produced by the presented model.

Figure 4.10: A comparison of an animation produced by a finite element model without stickiness (left column), a real mouth (middle column), and a finite element model with stickiness (right column). The stickiness provides a much more realistic movement of the lips throughout the animation. Time progresses vertically down the page with a 0.04 second interval between each row.

Figure 4.11: An asymmetric mouth opening created using non-uniform stickiness.

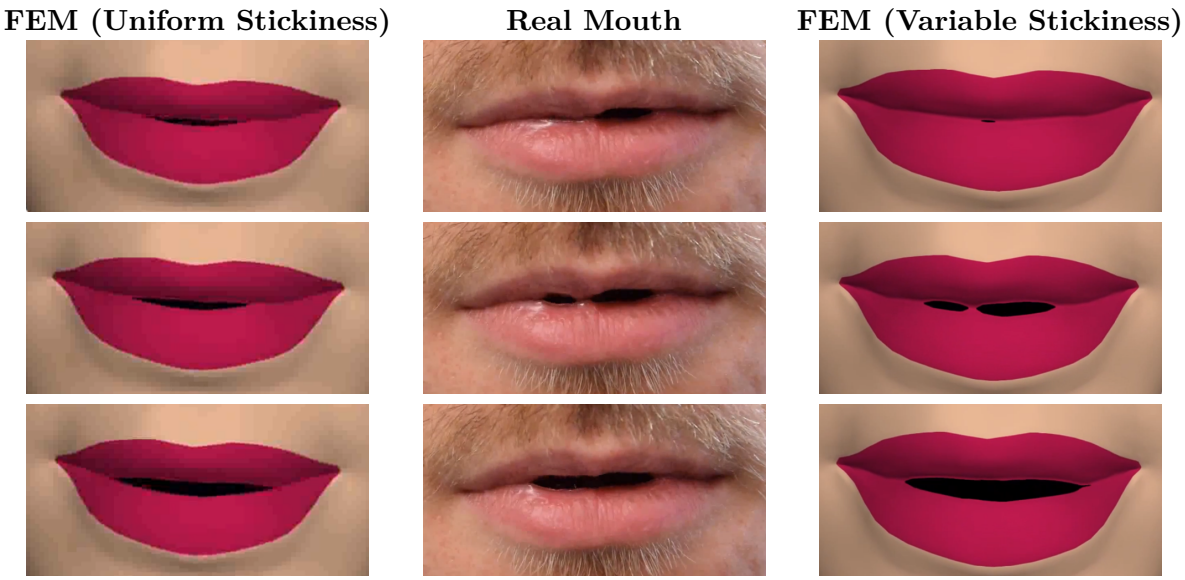


Figure 4.12: A comparison of an animation produced by a finite element model without stickiness (left column), a real mouth (middle column), and a finite element model with stickiness (right column), where a point near the centre of the mouth is stuck, producing multiple mouth openings. The stickiness provides a much more realistic movement of the lips throughout the animation. Time progresses vertically down the page with a 0.12 second interval between each row.

Figure 4.13: An animation produced by Barrielle's method (Right). Note that multiple tiny mouth openings are formed in the animation in areas which are completely closed in the real video (Left). The method presented in this thesis gives behaviours more consistent with real mouths. Additionally, the animation has the mouth opening centered whereas in the initial video it is to the right hand side of the mouth as viewed. This would be simple to recreate with the approach presented in this thesis through the use of specifying stickiness or moisture levels. Figure reproduced from (Barrielle and Stoiber, 2018) courtesy of Wiley, permission granted solely for use in this thesis.

Figure 4.14: Left: Real mouth. Right: Texture mapped mouth with physically-based sticky lips.

4.3 Varying Stickiness Levels

One of the key features of the presented model is the ability to adjust the level of stickiness. Figure 4.15 shows a comparison between three different levels of stickiness for the same animation. The animation is of the mouth speaking the letter 'm'. The three stickiness levels displayed, slight, normal and high, were identified based on analysis of collected mouth videos. The central column showing normal stickiness represents what has been determined to be a medium natural level of stickiness. The slight stickiness level is also plausible, but the high stickiness column represents an unnaturally high level of stickiness. Time progresses vertically down the page, with each row representing an increment of 4 frames of animation at 60 frames per second.

As the stickiness level increases, the mouth opening begins to form later in time. This is difficult to test in real mouths as there is not precise data on when a recorded participant is activating their muscles and cannot guarantee they move in exactly the same way each time. However, it would be expected that stronger cohesive forces between the lips delay the opening, as the saliva absorbs more energy before failure. It is also clear that the stickiness helps to produce a more rounded mouth opening and lip shape. This can be seen most obviously in the high stickiness animation, but is also the case for the normal stickiness animation relative to the slight stickiness one.

Comparing the slight and normal animations, we see that the mouth opening in the normal stickiness animation reaches a greater opening height than that of the slight animation by row 4. A possible explanation for this is that it is caused by inertial effects as a result of sudden acceleration as the stronger saliva bonds break. The stronger bond provides a greater impulse when the saliva breaks. Inertial effects such as this are difficult to recreate in blendshape models and may be missed by an artist, however, the physically-based model produces physically realistic inertial effects naturally. In the high stickiness model, the same effect is not observed. Here the strength of the sticky bonds is sufficient that the additional momentum caused by breaking of the previous saliva element is not enough to immediately cause the next one to break, and the breaking is instead controlled by the opening of the jaw. This leads to a slower zippering, smaller final mouth opening width and smaller mouth opening height.

Figures 4.16 and 4.17 show 16 frames of an animation of the mouth speaking the word 'puppy'. There are many similarities with figure 4.15. Mouth openings are still slower to form for higher levels of stickiness, and in general the mouth shapes produced are rounder. One interesting deviation is in rows 3-6. In the third row, each stickiness level presents quite different mouth shapes. By the fourth row, the normal and high stickiness levels have reached almost identical positions, while the slight stickiness animation opening has greater height. This trend is maintained through the next few frames. This is again potentially a result of inertial effects, with the stickiness in the two higher levels limiting the opening of the mouth.

Figure 4.15: A comparison of an animation speaking the letter 'm' produced by a finite element model with slight stickiness (left column), normal stickiness (middle column), and a finite element model with high stickiness (right column). Differing levels of stickiness lead to differences in the resultant animations. Time progresses vertically down the page with a 0.067 second interval between each row.

Figure 4.16: A comparison of an animation (first half) speaking the word 'puppy' produced by a finite element model with slight stickiness (left column), normal stickiness (middle column), and a finite element model with high stickiness (right column). Differing levels of stickiness lead to differences in the resultant animations. Time progresses vertically down the page with a 0.05 second interval between each row.

Figure 4.17: A comparison of an animation (second half) speaking the word 'puppy' produced by a finite element model with slight stickiness (left column), normal stickiness (middle column), and a finite element model with high stickiness (right column). Differing levels of stickiness lead to differences in the resultant animations. Time progresses vertically down the page with a 0.05 second interval between each row.

Figure 4.18: Frames from a mouth opening repeated three times with the saliva drying between each repetition.

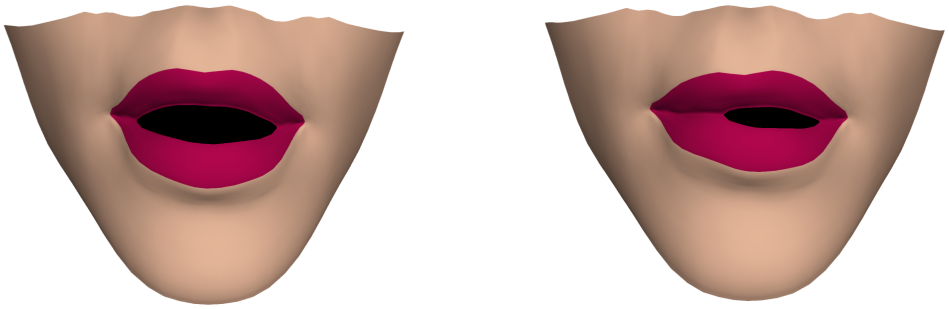


Figure 4.19: Frames from a mouth opening repeated three times with the saliva drying between each repetition.

4.4 Moisture and Evaporation Modelling

The moisture model in this work allows dynamic changes in stickiness behaviour. Figure 4.18 shows a dynamic but uniform change in stickiness level. A mouth opening is repeated 3 times, with the saliva drying between each repetition.

The evaporation rate can be varied across the lips. This can lead to asymmetry in the stickiness behaviour. Figure 4.19 demonstrates frames from an animation in which the evaporation rate is higher on one side of the mouth. Initially, the mouth opens symmetrically, however, as the animation progresses, the non-uniform evaporation rate produces non-uniform stickiness and the mouth opening is no longer symmetric.

4.5 Performance

Three different mesh resolutions were trialled (Table 4.1). From least to most complex, they contained 1784, 6982 and 27322 nodes and 852, 3414 and 13680 elements. The least complex mesh used the mesh topology at the resolution output by Facegen. The other meshes were created by subdividing the base mesh, and so increase in complexity by roughly a factor of four in each case. Even higher resolution meshes required considerably longer to simulate and so were not used in this evaluation, but could be used if a higher fidelity is desired. In a simulation in which the mouth is opened for one second with a timestep of 0.00001 seconds, the first two meshes took 22 and 28 seconds to solve for respectively. The highest resolution mesh required a smaller timestep to remain stable for the simulation due to the smaller elements and the relationship between the critical timestep and the size of the smallest element. As such, to compute the same 1 second animation, the higher resolution mesh took 354 seconds. The medium resolution mesh was found to provide a good balance of animation quality and performance. The lower resolution mesh did not feature sufficient resolution of saliva elements for the separations to appear smooth, and the higher resolution model was considerably slower.

The performance and stability of simulations can be improved through the use of a tech-

Model resolution	Nodes	Elements	Sim. Time (s)	Sim. Time with Mass Scaling (s)
Low	1784	852	22	2.2
Medium	6982	3414	28	2.8
High	27322	13680	354	35.4

Table 4.1: Statistics regarding the three model resolutions

nique called mass scaling. This method artificially increases the mass of elements to damp transient responses, leading to increased stability and the ability to raise the critical time step. The factor which the time step can be scaled by is proportional to the square root of the mass scaling factor. As an example, introducing a uniform mass scaling factor of 1000 allowed raising of the time step by a factor of 10, leading to overall simulation times of 2.2, 2.8 and 35.4 seconds. However, as it damps transient responses, dynamic effects are affected. Depending on the magnitude and location of the mass increase, the dynamic effects can be almost completely removed. Given the dynamic nature of this simulation, this is undesirable and mass scaling is not an appropriate technique for improving performance here. It is possible that mass scaling could be used in certain key areas to improve performance whilst not affecting the dynamics too much, however, as this work was not performance focused this was not explored.

Aside from the performance, the number of vertices and elements used also affects the quality of the model. A higher fidelity mesh will produce smoother results visually, but also a more physically accurate behaviour as increasing the finite element mesh decreases the error inherent in the approximation. In this case, the mesh has been uniformly subdivided. Subdivision of only the saliva elements would provide increased solution accuracy for the mouth whilst limiting the performance impact, however, it requires more advanced mesh manipulation. A further consideration with regards to quality is the frame rate of the final animation - more subtle or near-instantaneous effects may be lost in lower frame rate animations. It is important to note that this is independent of the simulation time step and that it is possible to produce animations with as high a frame rate as desired. For the effects demonstrated in this thesis, 30 frames per second is sufficient. If faster movements are produced, higher frame rates may be required to observe effects such as zippering.

4.6 Discussion

The presented model is capable of reproducing a range of mouth shapes and movements. Natural stickiness behaviours are produced including the recreation of the ‘zippering’ effect and asymmetric or multiple mouth openings. In comparison with other works in which the behaviour is stochastic and therefore not easily reproducible, these behaviours are produced in a deterministic manner. The stickiness effects evolve naturally over time due to the moisture model. The model includes dynamic effects as a result of the explicit simulation approach. Other animation techniques can be difficult for artists to control, either requiring significant work creating custom blendshapes or frame by frame modification. The method presented allows artists to easily control the behaviour of the stickiness model by adjusting the moisture level on the lips. Artist directed animations can often include physically-implausible behaviours as there is no physical simulation involved. In comparison, the presented model produces physically plausible effects.

Chapter 5

Evaluation

The typical approach to evaluating animations of the mouth is to use a perceptual evaluation, as reviewed in chapter 2, section 2.6. This approach is followed here to conduct an examination of the effect of stickiness on perceived realism of mouth animations. A user study is presented in section 5.1, testing the effects of the inclusion of different amounts of stickiness between the lips, giving new insights into the effects of lip stickiness in mouth animations. Perceptual evaluations are used as humans are the final judges of any animation produced, however, objective evaluations can provide different insights, potentially giving information about the way in which a model is more or less realistic. As such, the model is also evaluated using a second, objective method, developed and demonstrated in section 5.3. This objective evaluation is performed using a novel computer-vision-based mouth-opening detection algorithm, in combination with a new set of metrics which can be extracted from mouth videos. A new method for comparing the results of the metrics between different mouths speaking at different speeds is developed. Early work on this method was published (Leach and Maddock, 2019). The objective evaluation is performed using the developed method to analyse a new audio-visual corpus, captured as part of the thesis work, which features participants performing a variety of actions and utterances at different levels of lip stickiness. The corpus is described in section 5.2.

5.1 User Study

5.1.1 Motivation

Contrary to the objective evaluation which aimed to establish whether the model produced a more accurate mouth animation, the perceptual user study aims to find out whether it produces a perceptible difference, and if so, whether it is preferred or undesirable. Specifically the aim is to determine whether the inclusion of stickiness in an animation increases its realism. Additionally, it is useful to establish whether a particular level of stickiness is preferred/considered most realistic. It is also of interest to explore how the inclusion of stickiness affects different types of movement. Overall, the following questions should be answered through the perceptual study:

1. Does the inclusion of stickiness increase perceived realism in mouth movements?
2. Is there a difference in the impact of stickiness based on the speed of the mouth movement?
3. Does the inclusion of audio affect the perceived realism of animation using stickiness?
4. Are people able to differentiate between different levels of stickiness?

5.1.2 Trial Data

In order to answer the above questions, trial data was chosen to cover a range of actions. The data for the trials is a set of computer generated animations using the model developed in this thesis. In total, 6 actions were selected for the trials:

- Slow mouth opening
- Fast mouth opening
- Speaking 'P'
- Speaking 'M'
- Speaking 'Puppy'
- Speaking 'Mummy'

The slow and fast mouth openings are chosen as mouth opening is a simple, recognisable action which has a clear stickiness effect in reality. Having two variations of the movement at different speeds allows investigation of the impact of movement speed on preference of stickiness level. The fastest movements in speech are plosives (Canning and Rose, 1974) - 'P' and 'Puppy' are included to test these. 'M' and 'Mummy' are included as they are similar to 'P' and 'Puppy', but with a more gradual, softer opening instead of the plosive. 'P', 'M', 'Puppy' and 'Mummy' all support testing the impact of audio on the results. Audio was added to the relevant videos by manually aligning the audio from the source videos the animations were based on.

5.1.3 Methodology

Study Format

The study is conducted in an A/B format. Each participant completes a series of trials. In each trial, they are shown two videos side by side and asked to select the video they believe to be more realistic. Figure 5.1 illustrates the screen layout used. Participants may replay the videos at will and were allowed to view the videos as many times as they liked for a trial. The participants were not allowed to return to earlier questions once they had moved on to the next question. In total, there were 6 actions with 4 levels of stickiness. Additionally, the spoken actions were given with and without audio. Trials were not conducted in which the same movement with and without sound were paired, e.g. 'm' without sound against 'm' with sound.

As such, the 4 levels of stickiness give 6 possible pairings for each action, as pairs were only constructed of differing stickiness level animations. In total, this gives 12 trials for the unspoken actions, 24 trials for the speech actions without audio and 24 trials for the speech actions with audio. Every user completed all trials, so there were 60 trials per user overall. The study received ethical approval from The University of Sheffield. The details of the application and evidence of approval can be found in appendix B.

In order to test the study process, a pilot study of 3 users was conducted. At the end of the pilot study, the participants were asked for feedback on the process. Some minor adaptions were made based on the feedback, such as not auto-playing the videos, and increasing the size of the answer buttons.

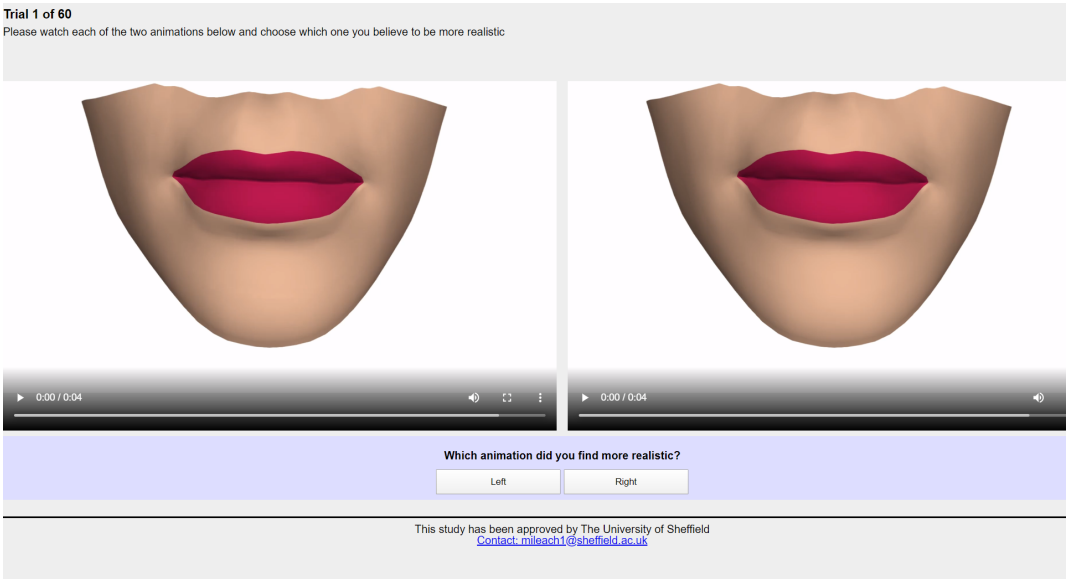


Figure 5.1: The user interface for the perceptual study.

Action	StickinessLevel	Audio
os	0	a
fs	4	(none)
p	7	
m	10	
pu		
mu		

Table 5.1: Sets of codes used to construct the video codes and file names.

Study Implementation

The web application was developed using php and a MySQL database. The application consists of three pages. On the start page, the user is presented with information regarding the study and instructions describing what will happen during the trial. Additionally, the user is assigned a unique ID which is tied to the web session. This ID is used to identify the user throughout the remainder of the process as no personal information is gathered.
The database used to store the results data consists of three tables:

- questions. This table contains three columns: (i) an ID identifying the question number; (ii) v1, a code identifying the first video of the question; (iii) v2, a code identifying the second video of the question. The codes take the form ActionStickinesslevelAudio, where Action, Stickinesslevel and Audio are chosen from the sets shown in table 5.1. The same codes were also used to name the video files.

- question_orders. This table contains three columns: (i) a uniquely identifying ID; (ii) session_id, which is the ID of the participant; (iii) question_order, which is a string representation of the order the questions will be presented to the participant in.

- answers. This table contains four columns: (i) a uniquely identifying ID; (ii) session_id, which is the ID of the participant; (iii) question, which references one of the questions from the questions table; (iv) answer, which is the code of the video the user selected as the preferred video.

Potential Problems

Order bias typically occurs in one of two forms: primacy bias (Van Erkel and Thijssen, 2016), in which users select the first option, or recency bias, in which users choose a later option because it is more fresh in their memory. Given there are only two options present in each trial, order bias is unlikely, nonetheless the display of the videos was randomized for each user. The order in which the trials were presented was also randomised to avoid any bias caused by the ordering of the videos.

5.1.4 Results

Twenty people participated in the study. The participants were students of The University of Sheffield. No background factors were identified which might influence the results of the study. The results of the study from the answers table were exported from the database in a CSV format. Full listings of the results are given in appendix A. A python script was written allowing the extraction of subsets of the data based on specific parameters, such as user ID, the presence of audio, or querying results based on the stickiness levels involved. The python script generates graphs of the requested data. Bar charts are used to demonstrate the frequencies at which a particular stickiness level was selected as the preferred level relative to the other options. For queries where only 2 options are compared, the script also automatically computes the confidence level in the result. These confidence levels are given alongside the graphs.

Confidence levels are given as p-values. This value represents the chance that the stated results have occurred due to random chance rather than as a result of a statistically significant outcome. For the trials used in this study which have only two options, the binomial distribution can be used to model a set of trials. In our null hypothesis, we assume that if the stickiness has no impact on which animation is preferred, participants will rate all levels of stickiness equally. As such, for each question we would expect each answer to receive 50% of the preference share. This can be modelled with a binomial distribution with the probability of each outcome being even, i.e. 0.5. This is alternatively called a coin-flip distribution, owing to the trials being akin to a coin flip.

The binomial distribution gives the probability of obtaining exactly r successes in a set of n trials as:

$$P(r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \tag{5.1}$$

The probability of obtaining at least r successes if the null hypothesis is true is given by the sum of the individual probabilities of obtaining r or greater successes:

$$P(\geqslant r) = \sum_{i=r}^{n} P(r_i) \tag{5.2}$$

Stickiness vs. No Stickiness

The first set of results examined are those looking at questions which included a video showing a model using no stickiness. Figure 5.2 shows the results of all questions in which no stickiness was compared with another stickiness level combined, for all users combined.

The graph demonstrates a clear preference for animations including stickiness, with a p-value of p=0.0000004. Despite this, it can be seen that animations with no stickiness were still selected regularly. In order to determine the reasons for this, subsets of the data are examined.

Figure 5.3 subplots (a), (b) and (c) show a breakdown of each of the different stickiness levels compared with no stickiness. In all cases the sticky option is preferred roughly equally often, with preference ratios of sticky to not sticky of 1.49, 1.40 and 1.65 for levels 4, 7 and 10,

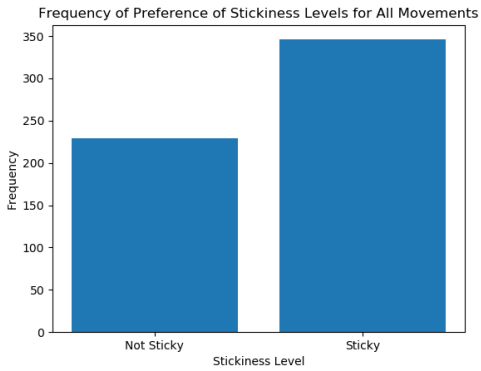


Figure 5.2: The combined results of all trials in which the user was asked to choose between an animation which included stickiness and one which did not.

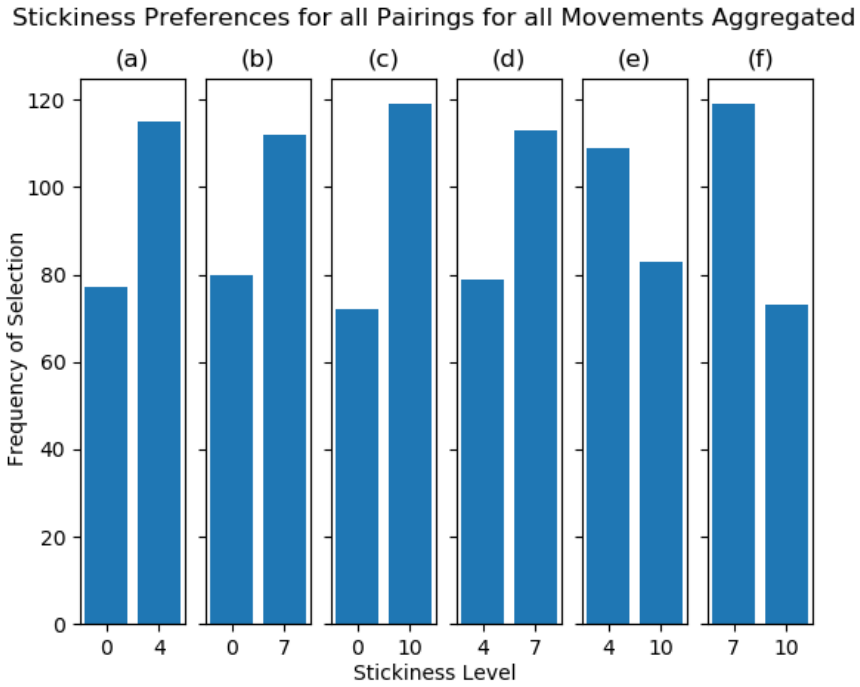


Figure 5.3: All pairings of individual stickiness levels for all speakers and actions

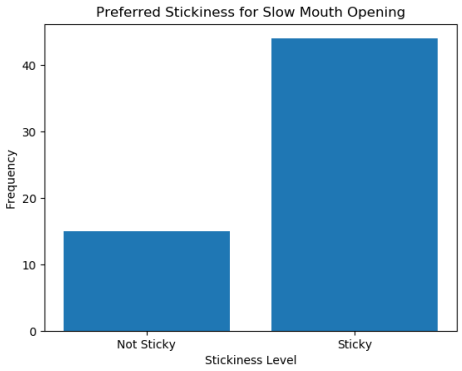


Figure 5.4: No stickiness vs all other stickiness levels for the slow mouth opening.

respectively. The chances of these results occurring if the null hypothesis is true are p=0.002, p=0.009 and p=0.0002.

Figure 5.3 subplots (e) and (f) support stickiness level 10 being too much and participants preferring levels 4 and 7. Subplot (d) shows a preference for level 7 over level 4, suggesting that overall the 'normal' stickiness level is preferred.

Slow Movements

One of the research questions for this study is whether the speed of movement affected the preference for stickiness. A discrepancy between slow and fast movements could also explain why no stickiness was preferred in some cases when data for all movements was aggregated. Figure 5.4 shows the data for all participants for all stickiness levels for the slow mouth opening. Looking at the slow movement in isolation gives a significant increase in the preference for sticky behaviour

Once again, looking at the individual stickiness levels, there is a clear preference for any of the sticky levels over no stickiness (Figure 5.5, subplots (a), (b), (c)). In spite of level 10 being the least preferable sticky level (Figure 5.5, subplots (d), (e), (f)) by some margin, for this slower movement it is still clearly preferred to no stickiness.

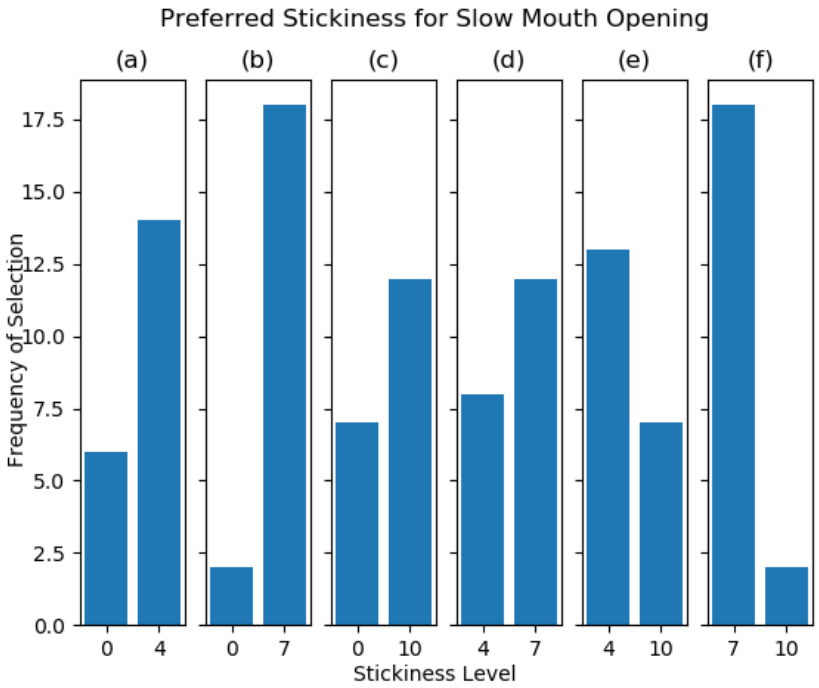


Figure 5.5: All pairings of individual stickiness levels for a slow mouth opening.

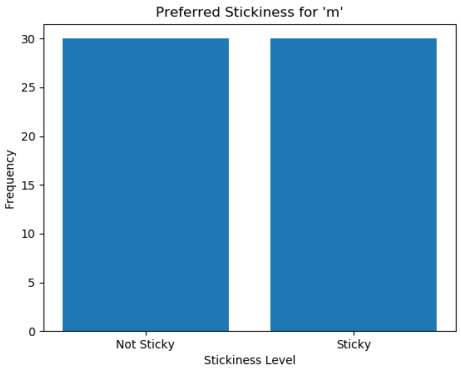


Figure 5.6: No stickiness vs all stickiness levels combined for the ‘m’ utterance.

The other slower movement in the set of trial actions is the speaking of the ‘m’ sound. Looking at the preference for no stickiness vs stickiness for this letter (Figure 5.6), it seems that for this action, there was no overall preference for stickiness vs no stickiness. This result seems at odds with the general trend in results and is perhaps explained by variance due to the relatively small sample size. A larger study would help to determine whether this result is statistically significant. The most significant result for ‘m’ was the preference for stickiness level 7 over level 10 (Figure 5.7, subplot (f)).

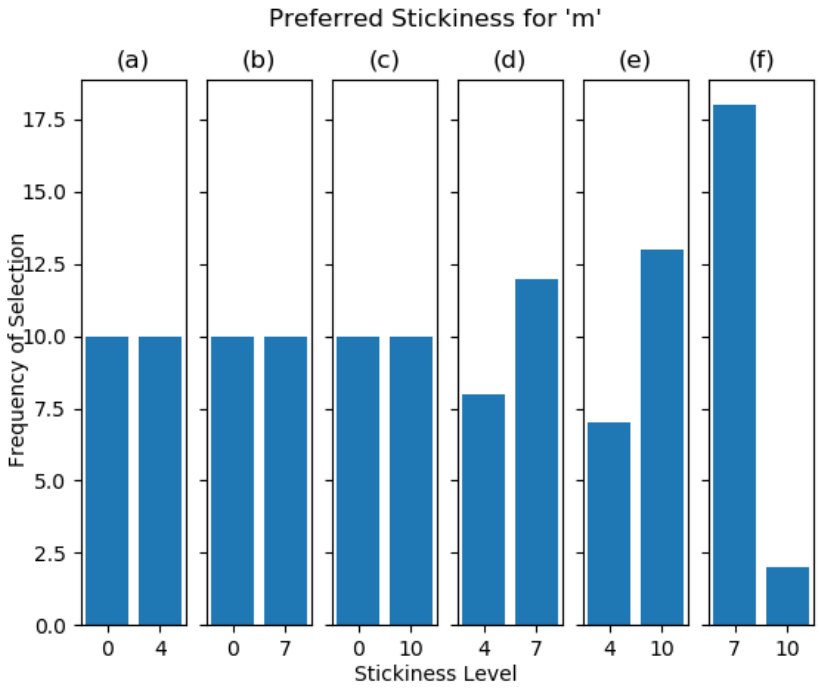


Figure 5.7: All pairings of individual stickiness levels for the 'm' utterance.

Fast Actions and Plosives

Examination of the faster actions is begun by looking at the fast mouth opening. This action is the same motion as the slow mouth opening and reaches the same equilibrium position, but the muscles are activated faster so the overall movement and mouth opening is faster. Figure 5.8 shows the aggregate non-sticky vs sticky preferences. Once again, there is a strong preference for sticky behaviour. Interestingly, the preference for stickiness in the fast mouth opening is stronger than for the slow mouth opening, with a ratio of 3.61, p=0.000002 rather than 2.93, p=0.00003. Looking at the individual level breakdown (Figure 5.9), it can be seen that in contrast with the slow mouth opening, the over the top stickiness level was preferred more strongly compared with no stickiness. This could be a result of it being more difficult to discern between stickiness levels in faster movements. For the fast mouth opening, higher stickiness levels were consistently rated as better (Figure 5.9, subplots (d), (e), (f)).

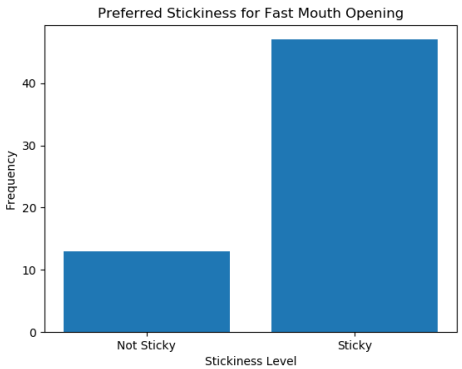


Figure 5.8: No stickiness vs all other stickiness levels combined for the fast mouth opening.

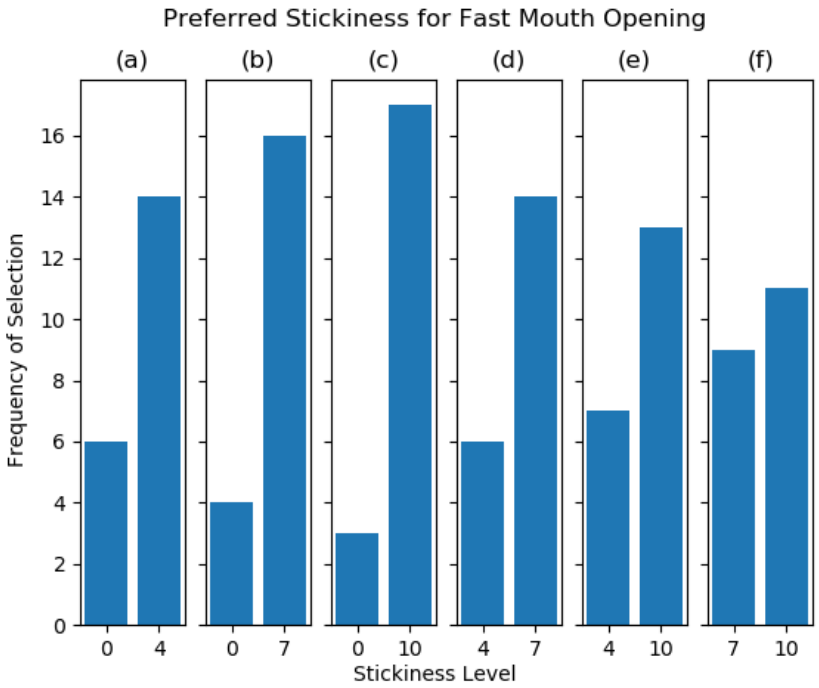


Figure 5.9: All pairings of individual stickiness levels for the fast mouth opening.

Next the 'p' action is examined. Once again there is a strong preference for the animations including the sticky behaviour (Figure 5.10). Looking at the individual breakdown it can be seen that lower stickiness levels were consistently preferred (Figure 5.11, subplots (d), (e), (f)).

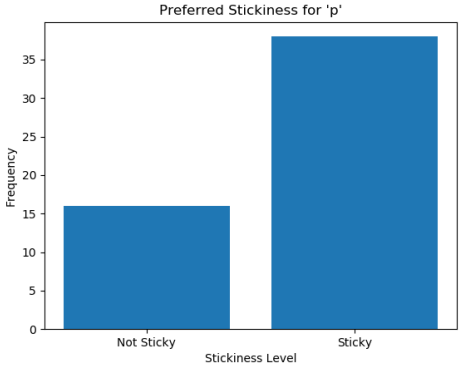


Figure 5.10: No stickiness vs all other stickiness levels combined for the 'p' utterance without audio.

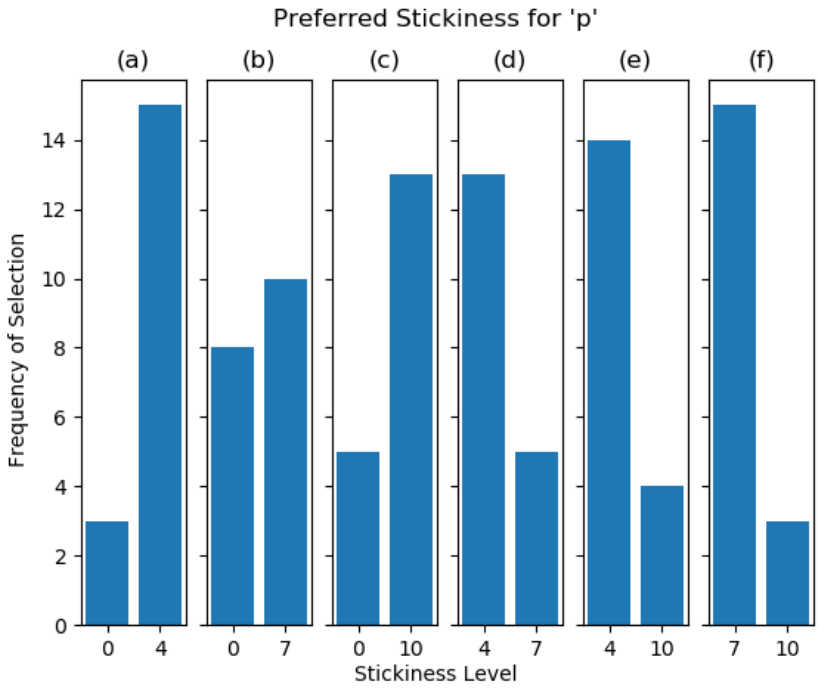


Figure 5.11: All pairings of individual stickiness levels for the ‘p’ utterance without audio.

The ‘puppy’ utterance tells a similar story, with a preference for the sticky animations (Figure 5.12). For this action, however, level 10 seemed to receive the strongest preference out of the no stickiness vs stickiness pairs (Figure 5.13, subplots (a), (b), (c)), although this was not reflected in the trials between sticky levels, in which there was little preference shown (Figure 5.13, subplots (d), (e), (f)). This could be a sign that it is difficult to distinguish between stickiness levels at higher speeds, but that the level 10 stickiness causes sufficient difference from the level 0 stickiness to be noticeable.

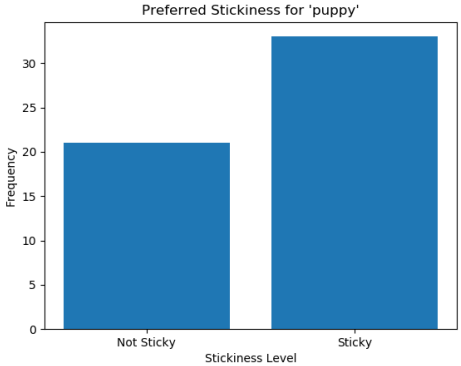


Figure 5.12: No stickiness vs all other stickiness levels combined for the word ‘puppy’ without audio.

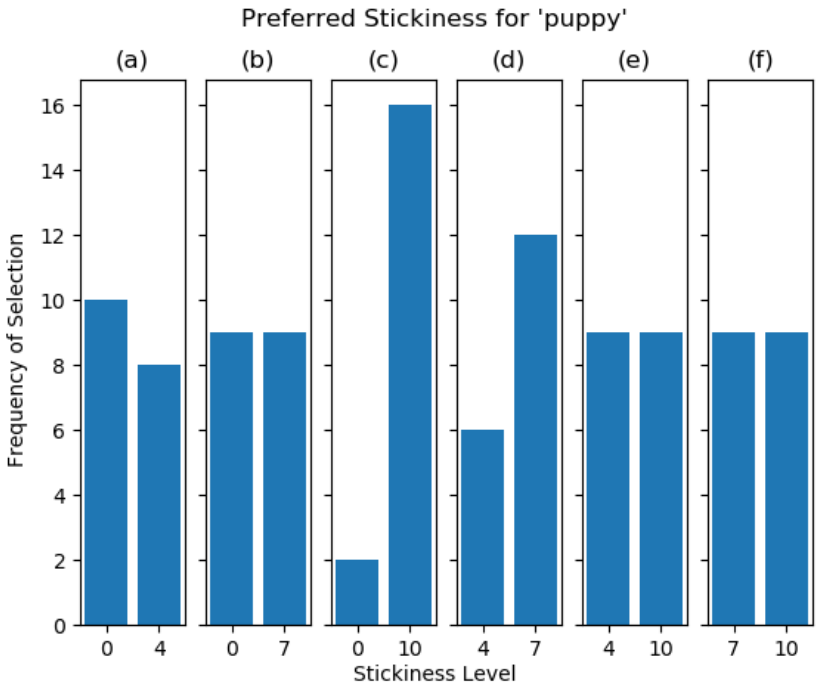


Figure 5.13: All pairings of individual stickiness levels for the word 'puppy' without audio.

Impact of Audio

The third research question to answer in this study is whether the inclusion of audio has an impact on whether stickiness is preferred or not. To facilitate this, trials were conducted for the 'm', 'p', 'mummy' and 'puppy' actions with and without audio. This analysis will begin by looking at the impact on each of the actions individually and then the aggregate effect across all actions.

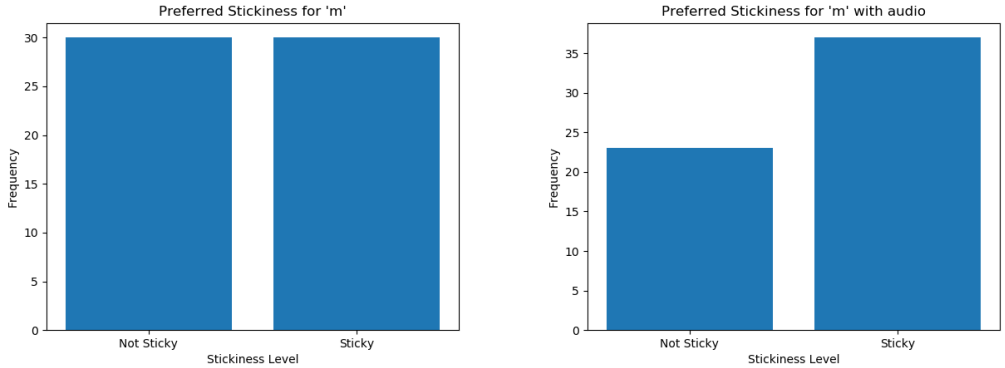


Figure 5.14: Aggregate stickiness preferences for the letter 'm' spoken without (left) and with audio (right).

Figure 5.14 shows that, overall, the inclusion of audio increased the preference for stickiness for the 'm' utterance. Looking at the breakdown of individual levels, it can be seen that there is a shift in preference from levels 0 and 4 to levels 7 and 10, with level 10 showing the greatest increase. This could indicate that the inclusion of the audio does affect the expected level of stickiness.

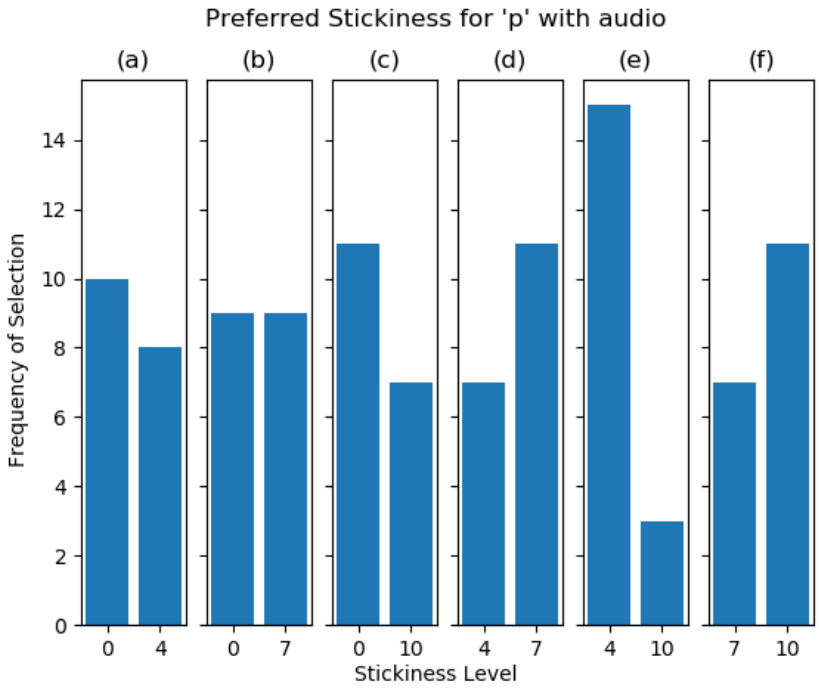


Figure 5.17: All pairings of individual stickiness levels for the ‘p’ utterance spoken with audio

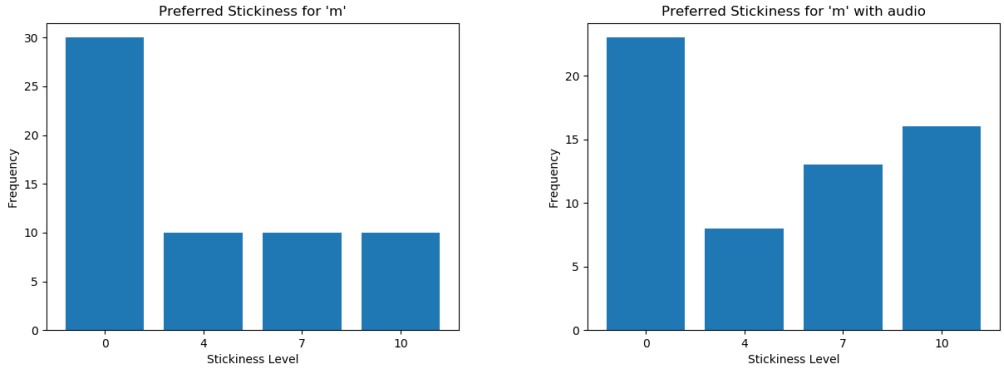


Figure 5.15: Individual stickiness preferences for the letter ‘m’ spoken without and with audio.

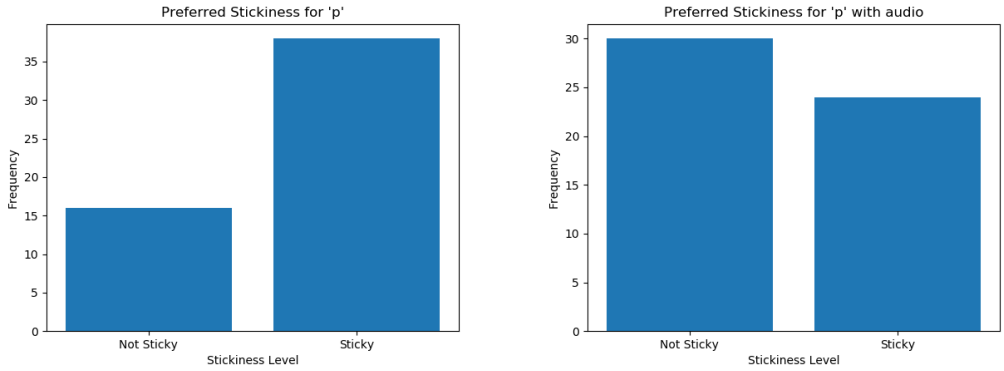


Figure 5.16: Grouped stickiness preferences for the letter ‘p’ spoken without and with audio.

Repeating the same analysis for the ‘p’ sound, there is a significant shift towards preference

of no-stickiness (Figure 5.16). This suggests that participants did not expect to see the sticky behaviour during the plosive movement. Such a dramatic shift suggests that perhaps the utterance was not correctly identified from video alone.

Figure 5.18 shows the overall change in preference for the word ‘mummy’. Again there is not much change, with a slight shift towards no stickiness being preferred when audio is included.

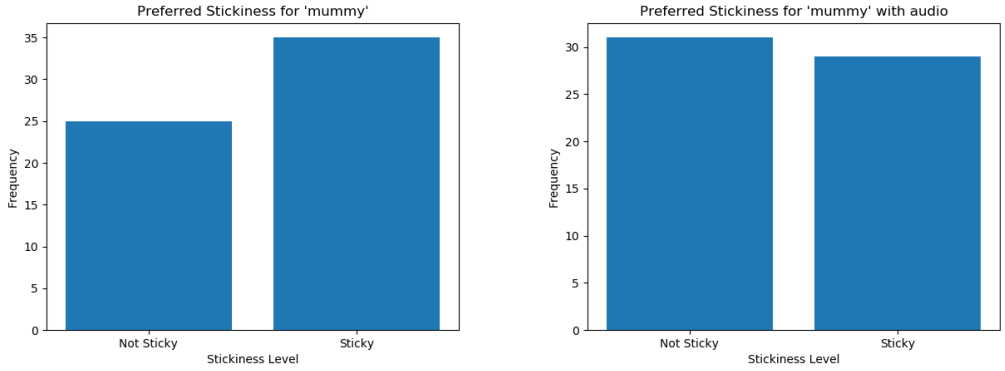


Figure 5.18: Comparison between the grouped stickiness preferences without (left) and with audio (right) for speaking of the word ‘mummy’

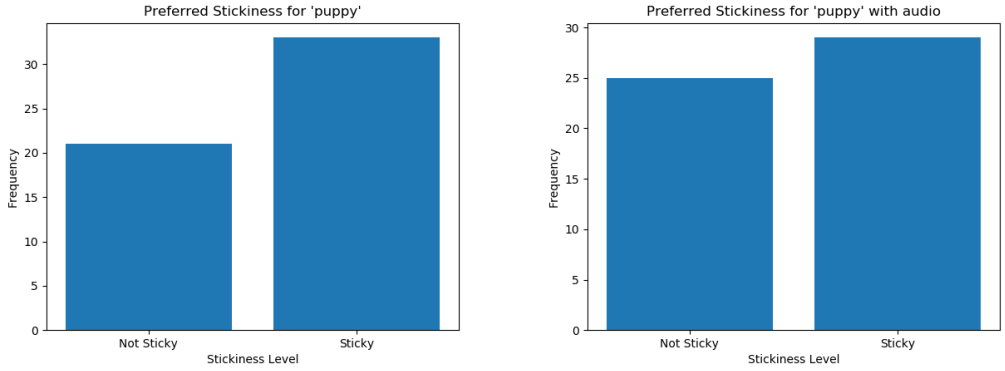


Figure 5.19: Comparison between the grouped stickiness preferences without and with audio for speaking of the word ‘puppy’

For the word ‘puppy’, Figure 5.19 shows that the overall preference once again did not change much with the inclusion of audio. However, examining the individual levels, it is clear that like the ‘p’ sound, there was a shift in preference towards the lower levels of stickiness (Figure 5.20). This supports the idea that perhaps users expect to see less sticky behaviour for plosives and that the fact that the animation represents a plosive movement could be difficult to identify from the video alone.

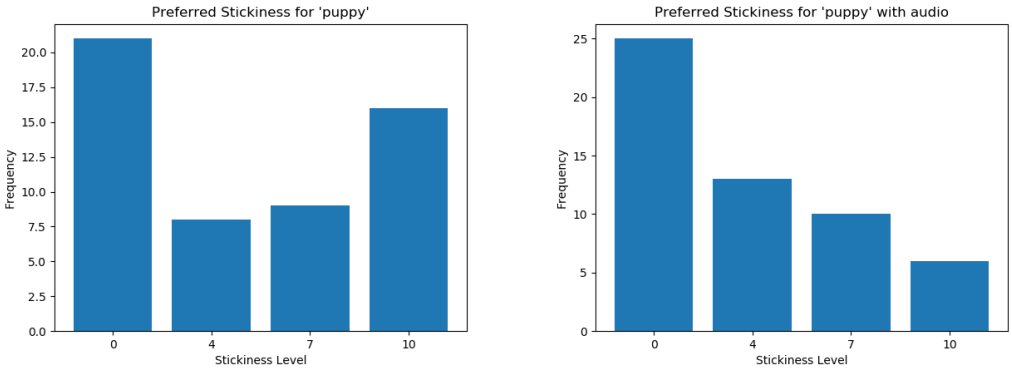


Figure 5.20: Comparison between the individual stickiness preferences without and with audio for speaking of the word 'puppy'

Inter-participant Variation

One of the most interesting results was the differences in results between participants. This could suggest that not everyone notices the effect of stickiness, but for those who are aware of it, it impacts realism significantly. Differences in human ability to recognise expressions are well documented (Palermo et al., 2013), so it does not seem unreasonable to hypothesize that different individuals may have different capabilities when it comes to picking up on nuances of the movement of the mouth.

A further explanation for this is the potential for different individual interpretations of the question which participants were asked, i.e, 'Which video is more realistic¿ The question was intentionally quite vague and did not mention lip stickiness to avoid any preconceived ideas of what a mouth movement 'should' look like from influencing the results. However, it does mean it is ambiguous in some ways, for example, in cases where the two stickiness levels presented are both physically plausible, how should the participant decide which is more realistic? Some might select the stickiness level they believe is more common, whereas others might say that although different in likelihood, both are equally plausible and treat the answers as equally realistic.

5.1.5 Insights and Conclusions

From the presented information, four main conclusions can be drawn:

- The inclusion of stickiness in general improves the perceived realism of the mouth model.
- Some individuals seem to have more awareness of the presence or lack of stickiness, whilst others cannot tell the difference or do not find it to be any more or less realistic.
- Higher stickiness levels were generally preferred on slower movements, whilst for faster movements, lower stickiness levels were preferred.
- The inclusion of audio can affect the preferred level of stickiness.

On average, there was a significant preference for sticky models over non-sticky models, with an overall preference ratio of 1.51, $p = 0.00004$. This result was not uniform across all actions or people, but in no case was the model with no-stickiness strongly preferred. At worst, the two models were considered equally realistic. From the outset it was understood that the effects being recreated are subtle, so it is reasonable to assume that it does not always go noticed, but in certain situations may be a deciding factor in whether an animation looks realistic or not.

The preference for stickiness level appears to depend on the speed of the movement involved. Slower movements consistently had clearer stickiness preferences with preference for higher levels. Preferences for faster movements were less distinct, but in general lower levels of stickiness were preferred. The inclusion of audio also seems to give rise to a lower preferred stickiness level. This could be explained by the inclusion of the audio suggesting a faster mouth movement than the participants visually recognised.

While the variety of actions and stickiness levels examined here has produced interesting results, it has also opened more questions. When looking in more detail at individual actions or participants' preferences, the data gives less statistical power than the aggregate conclusions. Further studies capturing more data, or examining each type of movement in more detail could produce more solid conclusions about the nature of the perception of mouth stickiness.

5.2 Creation of a Mouth Video Dataset

5.2.1 Motivation

In order to evaluate the effectiveness of physically-based modelling approaches for the mouth and the inclusion of stickiness in the model, it is desirable to perform an objective comparison against real mouths. In order to support this, a video corpus including actions recorded at different stickiness levels is required. This is developed in this section and then used for the objective evaluation in section 5.3. Most openly available libraries of videos of people speaking focus on the whole head. The data sets also often consist of relatively low resolution videos. This combined with the increased field of view means that it is harder to discern what is going on with the mouth at a finer scale. For this application where small scale effects are being studied, it is useful to have a close up view of the mouth and the tissue immediately surrounding it. A second motivation is that the videos in such datasets are typically recorded at a lower frame rate, often 24 or 30fps. In order to capture the dynamic effects, a higher framerate is desirable. Therefore, a new data set was required in order to perform an evaluation of the behaviours of the mouth when stickiness is involved.

5.2.2 Methodology

Equipment and Set Up

Filming was conducted using a Nikon D3300 DSLR camera. The lens was an AF-S Nikkor 18-105mm 3.5-5.6 DX Lens. Videos were recorded at a resolution of 1920x1080, with a framerate of 60 frames per second. Figure 5.21 shows the experimental setup. The participant was seated at a table, on which there was a tablet/laptop placed which played an instructional video. Behind the table, the camera was mounted on a tripod, with the height adjusted to match the participant's mouth. The participant's face was lit from either side in order to increase contrast between the skin and the mouth cavity. The lens of the camera was zoomed such that the lower half of the face was recorded.

What Data was Collected?

The aim of this data collection was to capture a variety of mouth movements to establish stickiness behaviour across a range of parameters. A key objective was to capture actions which involved different types of movements and speech sounds. Additionally, the collected data should include actions of varying complexity, ranging from the speaking of individual speech sounds to full sentences. To meet these criteria, categories of actions, letters, words and sentences were used. The letters and words categories are further divided according to the speech sounds which they contain, while sentences contain a mixture of movements and sounds. Each action or utterance is recorded at multiple levels of stickiness. The actions and utterances captured are given in table 5.2, where Grid 1 is "Bin blue with Q 7 please", Grid

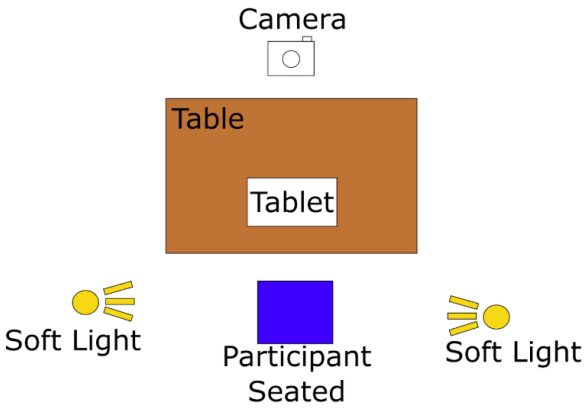


Figure 5.21: The layout of the room used in the data capture process.

	Actions	Letters	Words	Sentences
Gradual Opening	Mouth opening	m	mummy	Grid 1
Fricative		f, v	fish, shake	Grid 2
Plosive		p, b	puppy, baby	Grid 3
Lip rounding		o, q, u, w, y		

Table 5.2: The actions and utterances recorded in the video corpus

2 is"Place white by O 2 now" and Grid 3 is "Set green in U 4 soon". The sentences used are GRID sentences selected from The Audio-Visual Lombard Speech Corpus (Cooke et al., 2006). The sentences follow the form "command color preposition letter digit adverb" where each of the components are selected from the sets given in table 5.3.

These specific sentences were chosen to provide a good variety of mouth shapes, whilst including some of the letters and sounds used in the earlier sections on the basis that co-articulation could cause different behaviour in the lips and therefore in the effect of stickiness. Sentences from the GRID corpus were chosen as the corpus includes videos of speakers speaking every combination of the command, color and preposition.

Process Description

With the participant seated as described above, the recording is begun and the instructional video is started. The video instruction and participant actions are described below:

1. Participant is given initial information

   (a) Participant is informed that they will be asked to perform a number of actions and speak a number of utterances. Each action or utterance will be repeated

command	color	preposition	letter	digit	adverb
bin	blue	at	A-Z	1-9, zero	again
lay	green	by	excluding w		now
place	red	in			please
set	white	with			soon

Table 5.3: The sets of values from which the video codes are constructed.

   (b) Participant is informed that at times they will be asked to lick their lips.
   (c) Participant is asked to refrain from licking their lips between repetitions of actions/utterances
   (d) Participant is asked to try to breathe through their mouth between repetitions in order to aid the saliva drying.

2. Participant performs each of the following actions 5 times at a 10 second interval, licking their lips prior to each change of action:
   (a) Open mouth slowly
   (b) Open mouth quickly
   (c) Open left side of mouth
   (d) Open right side of mouth

3. Participant speaks each of the following letters phonetically 5 times at a 10 second interval, licking their lips prior to each change of letter:
   (a) m
   (b) b
   (c) p
   (d) f
   (e) v

4. Participant speaks the name of each of the following letters 5 times at a 10 second interval, licking their lips prior to each change of letter:
   (a) o
   (b) q
   (c) u
   (d) w
   (e) y

5. Participant speaks each of the following words 5 times at a 10 second interval, licking their lips prior to each change of word:
   (a) mummy
   (b) baby
   (c) puppy
   (d) fish
   (e) shake

6. Participant speaks each of the following sentences 5 times at a 10 second interval, licking their lips prior to each change of sentence:
   (a) Bin blue with Q 7 please
   (b) Place white by O 2 now
   (c) Set green in U 4 soon

The ten second interval was chosen based on an initial pilot test consisting of three users. This interval was found to be sufficient to see changes in behaviour without being so long that participants became uncomfortable.

Participants

Participants were sought who have a high colour contrast between their natural lip colour and the surrounding skin as the proposed evaluation technique depends upon computationally detecting differences in colour between the lips and their surroundings. This is more effective when there is a greater contrast in colour. This had to be the participants' natural lip colour, as lipstick etc. may have affected the stickiness between the lips. Participants were invited in person or by contact through e-mail/social media. In total there were 15 participants who took part. The data collection received ethical approval from the University of Sheffield. The application and approval letter can be found in appendix B.

5.2.3 Results

The results of the data collection are a set of 1550 videos, created from 15 participants. Each participant performed 4 actions and spoke 18 utterances, giving a total of 22 actions. For each action, 5 levels of stickiness were recorded, giving 110 clips per participant. Videos were recorded in single takes for each participant, but consequently segmented by action and stickiness level, with the footage of the participant waiting for the next instruction removed. Each segmented clip is 3 seconds in length. The total size of the data set is around 15GB, at roughly 10MB per clip. Figure 5.22 shows some sample frames from the data set.

5.2.4 Data Availability

Ethics permission has been received to make the data set available for research use through the University of Sheffield's ORDA system[1]. ORDA is the university's online research data repository, used for sharing research data. The corpus is not currently available through the system but the aim is that it will be released as part of a future publication which is being prepared. Of the 15 total participants, 15 agreed to have their data made available. Options will be provided to download the data as a whole, for a single participant, or for a single action/utterance. This is a novel data set in that it uses a higher frame rate and resolution, focuses on the lower half of the face from a fixed viewpoint and includes different levels of stickiness.

5.3 Objective Evaluation

5.3.1 Motivation

An objective evaluation provides a more definite answer to whether or not the inclusion of stickiness creates a more accurate mouth animation. However, a more accurate animation is not always preferred by users. For example, the uncanny valley describes an effect where something less accurate may be preferred to something more accurate, as the more accurate version seems too similar to a human whilst still being distinguishable. As such, a perceptual study may not necessarily indicate whether accuracy of animation has been improved or not. It is possible that the inclusion of lip stickiness could push the animation into the uncanny valley and actually be rated lower than an animation without. To this end, an objective evaluation gives us a clear yes or no answer as to whether the inclusion of stickiness produces a more accurate mouth animation. Additionally, through developing an evaluation process which is capable of drawing comparisons between animations and real videos, it is possible to extract more information about how real mouths move. Being able to analyse the real mouth in this way could lead to more insights around what correct behaviour of the mouth is.

[1] https://orda.shef.ac.uk/

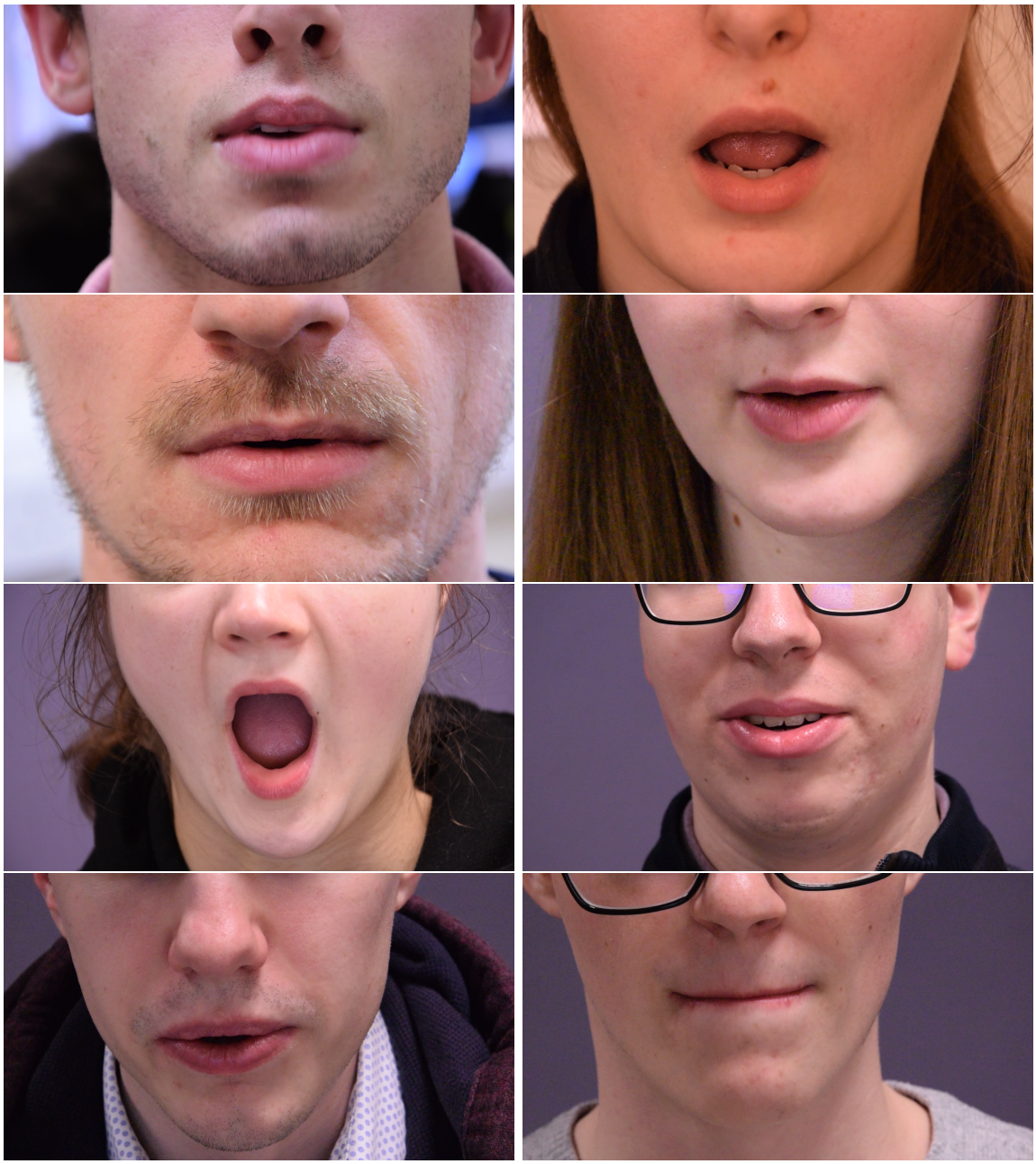


Figure 5.22: A set of example frames from the collected audio-visual mouth corpus.

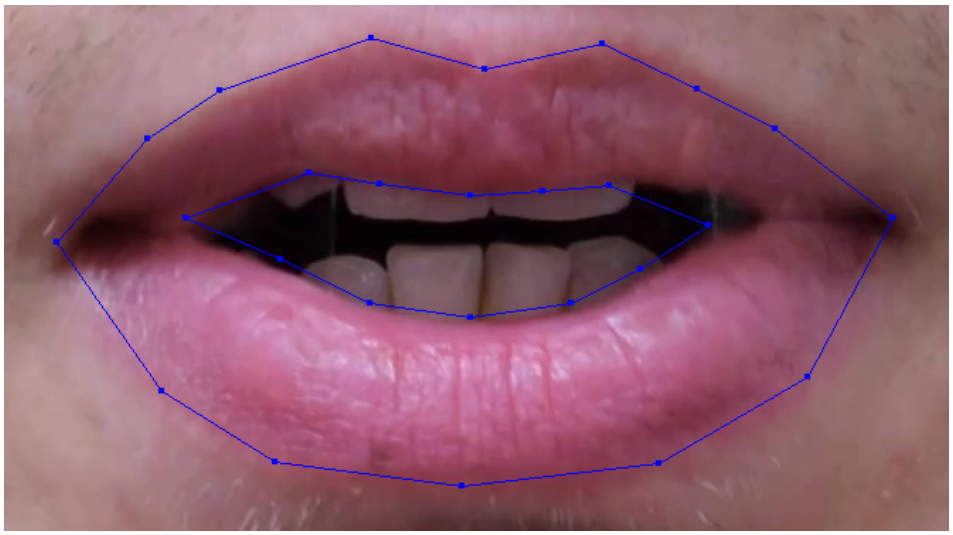


Figure 5.23: Inaccurate detection of the inner lip contour because of the limited sampling points in Faceware Analyzer.

5.3.2 Evaluation Process

Evaluation Process Overview

To produce a good objective evaluation, it is important to be able to examine and draw comparisons between real mouths and virtual mouths. In order to do this, a process has been developed which detects the inner lip contour, and then extracts various parameters from it. These parameters are graphed, and a method of comparing the graphs against an average for each type of movement is implemented. This is capable of handling different mouth widths and speeds. The comparison of a graph against the average allows matching of parts of particular movements, and comparisons can then be drawn between them based on rates of change and differences between extreme values of the parameters.

Lip Contour Detection

The human mouth has two lip contours. The outer lip contour is the divide between the vermillion zone and the normal skin. In this work, the inner lip contour is defined as the divide between the lips and empty space when the mouth is open. As this may depend on the viewing angle, videos recorded from a head-on perspective are used. Note that the given definition implies that there is no inner lip contour when the mouth is fully closed.

The most commonly used approaches for detecting the lip contours are active appearance models (Cootes et al., 1998), which match a statistically-based model of an object to an image. This approach is implemented in the software Faceware Analyzer[2], and this was initially trialled for the detection of the inner lip contour. Whilst the software was fairly effective at roughly matching the lip contour, it featured various inaccuracies.

The active appearance model is based around matching a specific set of points to an image. This inherently means that the number of points, and hence resolution/accuracy of the fit is limited. The inner lip contour consists of twelve linearly connected points in the Faceware Analyzer software. As such, there are inaccuracies in cases where the lips are particularly curved (Figure 5.23). A second issue was that the region detected would sometimes be incorrect, particularly in cases where the mouth opening was very small.

A final problem inherent with the active appearance model technique is that it is not able to detect more than one mouth opening (Figure 5.24). The model only features points to support a single mouth opening, and as such only one opening is detected. With the

[2]https://www.facewaretech.com/software/analyzer

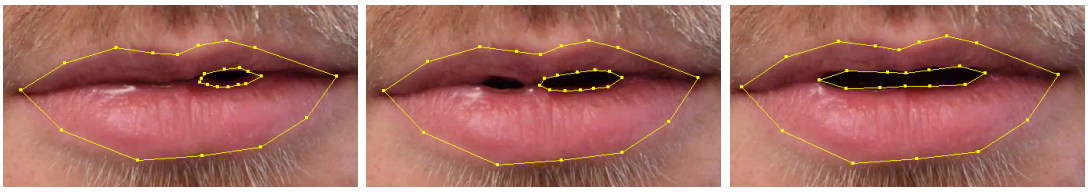


Figure 5.24: Left: Faceware Analyzer correctly detects the outer lip contour and the inner lip contour for a single mouth opening; Middle: Faceware Analyzer correctly detects the outer lip contour, but can only detect one of the two mouth openings present; Right: Once the mouth openings merge, Faceware Analyzer correctly detects the outer and inner lip contours.

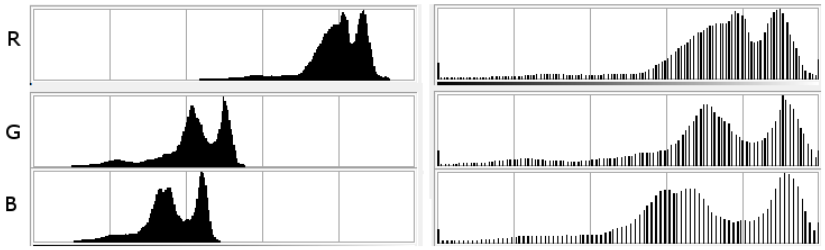


Figure 5.25: The histogram transformations involved in white balancing in visual form.

statistically-based method evaluating which of the two openings is 'more likely' to be the actual mouth opening, Faceware Analyzer's evaluation of where the inner contour is can also jump between the two mouth openings as they vary.

To combat these problems, a new approach was developed. The initial approach taken was to use colour-based thresholding to identify the lip region. Various thresholds were tested as suggested by literature (Hassanat and Jassim, 2010). It was found that while these were good at roughly identifying the lip area, they were often ineffective at correctly segmenting the lips towards the edges. In an attempt to improve this segmentation, thresholding in different colour spaces was considered. Previous work by Gritzman et al. (2015) used different colour spaces for lip segmentation, although the presented method first adjusts the colour of the images to improve contrast so these results are not directly applicable to this work and further investigation was required.

The use of different colour spaces improved the detection, but it was found that particular transforms/segmentation thresholds were better for different people's mouths. It was hypothesized that this was due to differences in the levels of skin pigmentation and lip pigmentation between people. To counteract this, a pre-processing stage was added. In this stage, the video frame is first auto-white balanced and then colour enhanced.

As the footage of the closed mouth will only contain the skin and lips, white balancing at this stage will help to balance out differences in skin tone. White balancing involves stretching each of the red, green and blue channels separately in order to remove any colour bias. The algorithm operates on the basis that the brightest points in an image should be pure white. As this assumption is not true for our image, in which the skin is likely to be the brightest thing, the skin is shifted to white. This means that the relative intensities of each channel have shifted, producing a hue shift. In a pure white, the pixel values for each of the R, G and B channels should be 255. As such, the aim is to produce a histogram of the intensities of each channel, find the maximum intensity, and scale the intensities up such that the maximum intensity becomes 255. By repeating this in each channel, the brightest points will be white.

Noise can impact the effectiveness of the algorithm. Sharp reflections or electronic interference may cause bright spots in a recorded image. To avoid including these in the process of finding the maximum intensity pixel in each channel, the histogram is clipped to cut out a proportion of the brightest pixels. Finally, an intensity scaling factor is computed by:

$$s = 255/i_{max} \tag{5.3}$$

where i_{max} is the maximum pixel intensity post clipping. Each pixel in the channel then has its intensity multiplied by this scaling factor to stretch the histogram (Figure 5.25).

With the image white balanced, the next step is to increase the distinction in colour between the lips and skin. To enhance the colours, a similar process is used. First, the image is transformed into the HSV colour space, where the hue represents the kind of colour, the saturation represents how intense the colour is and the value scales the brightness. The saturation channel is isolated and stretched, linearly scaling up the saturation of each pixel such that the maximum saturation is 255. The channels are then recombined and the image is converted back to RGB. Figure 5.26 shows how the white balancing and colour enhancement process creates consistent high contrast between the lips and surrounding skin even from source images with vastly different lighting and skin tones.

Figure 5.26: White balancing and colour enhancement applied to three different mouth images.

In order to improve segmentation, noise removal techniques are commonly employed. Various blurring filters were trialled, and a median blur was found to be most effective. The median blur, as its name implies, selects the median pixel value from a local domain. The key advantage of the median filter in this application is that unlike many other blurring techniques it is capable of preserving edges. This is a result of how the two filters handle areas of extreme brightness. A Gaussian filter essentially smears the bright point, whereas the median filter removes it completely. As the interest here is in finding the boundary, i.e, an edge, this technique is very well suited. Figure 5.27 demonstrates the effect of a median filter operating on a video frame.

From this point, two different methods are used to detect the lip region and correlated to increase confidence in the detection. One method aims to capitalise on the difference between the levels of red and green in the lip region, and the other uses a more typical hue-based thresholding. The introduction of two separate methods which are correlated is due to

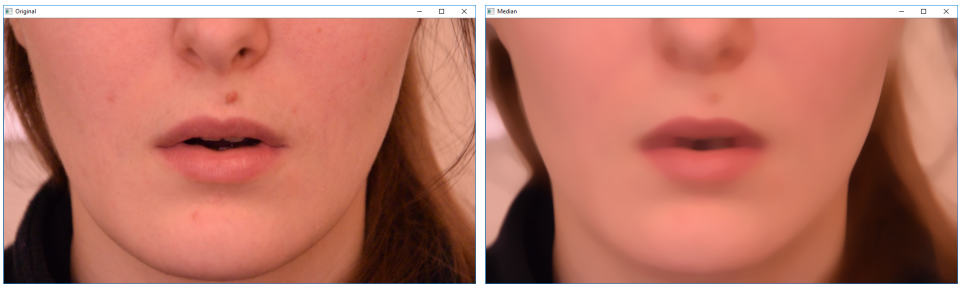


Figure 5.27: Median filtering of a frame.

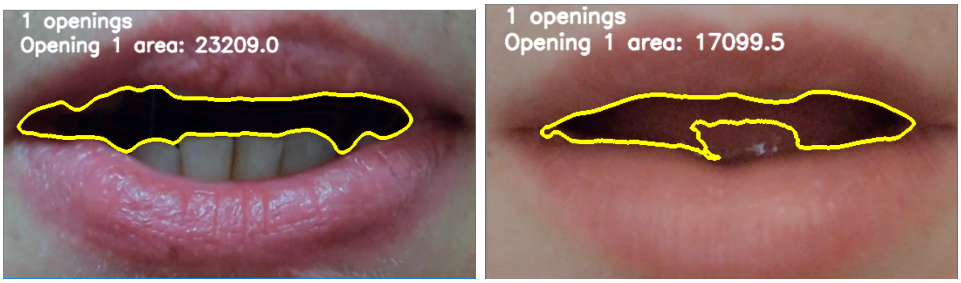


Figure 5.28: Problems with segmentation due to relatively bright teeth and reflections.

difficulties experienced with segmentation including the teeth and reflections of the tongue as part of the lips. Initial attempts using the algorithm (see (Leach and Maddock, 2019)) struggled with handling the teeth and reflections correctly due to their relative brightness (Figure 5.28).

An attempt was made to combat this by additionally detecting the teeth (as reported in (Leach and Maddock, 2019)). This was done by converting to the HSV colour space, in which the teeth typically show as very dark. The teeth region was segmented and dilated slightly (Figure 5.29), then this region was darkened in the original footage before the algorithm was applied. This had limited effectiveness. It would sometimes remove the teeth but was inconsistent. As such the algorithm was further developed to improve detection of the teeth.

Following the correlation of the two images, binary thresholding is performed. This aims to classify pixels as either belonging to or not belonging to a mouth opening. Finally, contours within the resulting image are detected to identify the edges of the mouth openings. Contour detection uses the findContours function from the OpenCV library. This function uses the method of Suzuki et. al (Suzuki et al., 1985).

The approach is very effective for synthetic mouths as there is no noise and edges are cleanly defined. Additionally, in this simpler rendering there are no reflections and the teeth are omitted. As such the process results in very accurate detection of mouth openings, even in

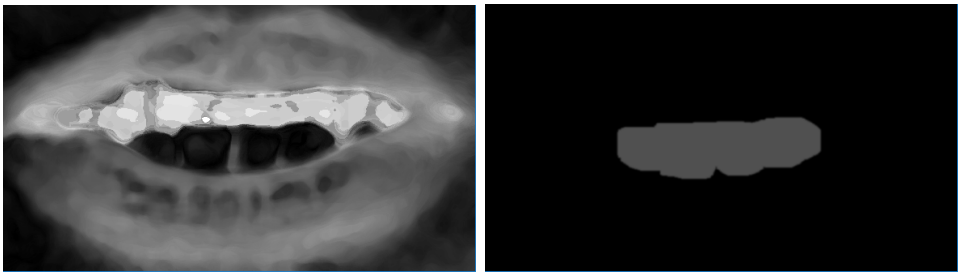


Figure 5.29: Left: The mouth viewed in the HSV colour space - note the dark teeth. Right: The segmented tooth region using the original algorithm.

Algorithm 2: Inner Lip Contour Detection

```
foreach Frame do
    Decompose into R, G, B channels;
    foreach Channel do
        Find intensity_max;
        s = 255/i_max;
        foreach pixel do
            i_pixel = i_pixel * s;
        end
    end
    Transform frame to HSV;
    Find S_max, the intensity of the brightest pixel in the saturation channel;
    s = 255/S_max;
    foreach pixel do
        S_pixel = S_pixel * s;
    end
    Transform frame to RGB;
    Blur frame with median filter;
    Transform to greyscale;
    Perform binary thresholding to segment image;
    Find contours using the method of Suzuki et al.;
end
```

the case of multiple mouth openings (Figure 5.30). For real mouths, the detection process is sometimes less accurate as a result of noise or sharp reflections. Figures 5.31 and 5.32 show the detection process operating on frames of two mouth videos and their synthetic counterparts.

Parameter Extraction

With the inner lip contour correctly detected, the next step of the process is to extract information about how it changes over time. In observing videos of mouths performing expressions and speaking, it was decided that certain parameters would be useful to extract to describe the motions of the lips. These parameters are:

- Number of mouth openings
- Opening width
- Opening height
- Opening area
- Ratio of opening width to height
- Rate of change of opening width
- Rate of change of opening height

The widths and heights are computed by computing the difference between the greatest and least pixel coordinates within the contour along the respective axes. The area is simply the number of pixels contained within a contour. Rates of change are computed by using a moving average over the previous 3 frames. This helps to smooth the data which can be erratic due to variation in lighting or video noise.

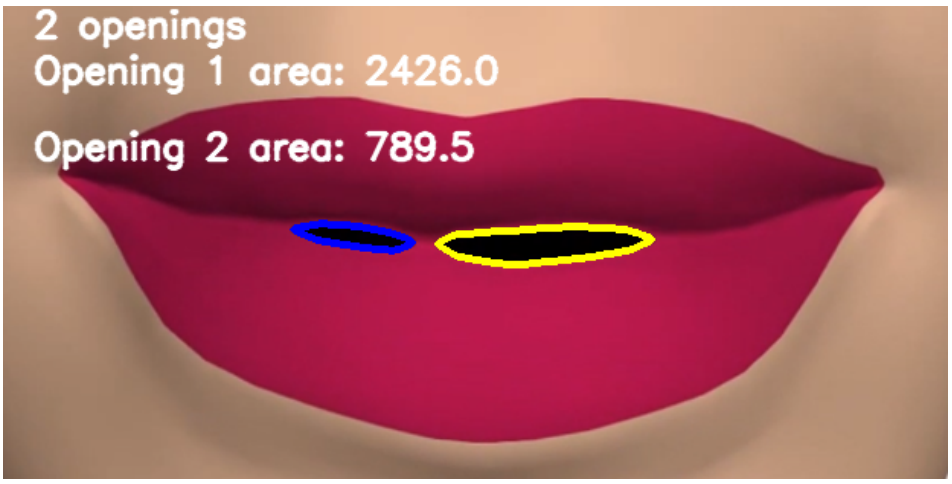


Figure 5.30: Multiple mouth openings accurately detected in a synthetic mouth animation.

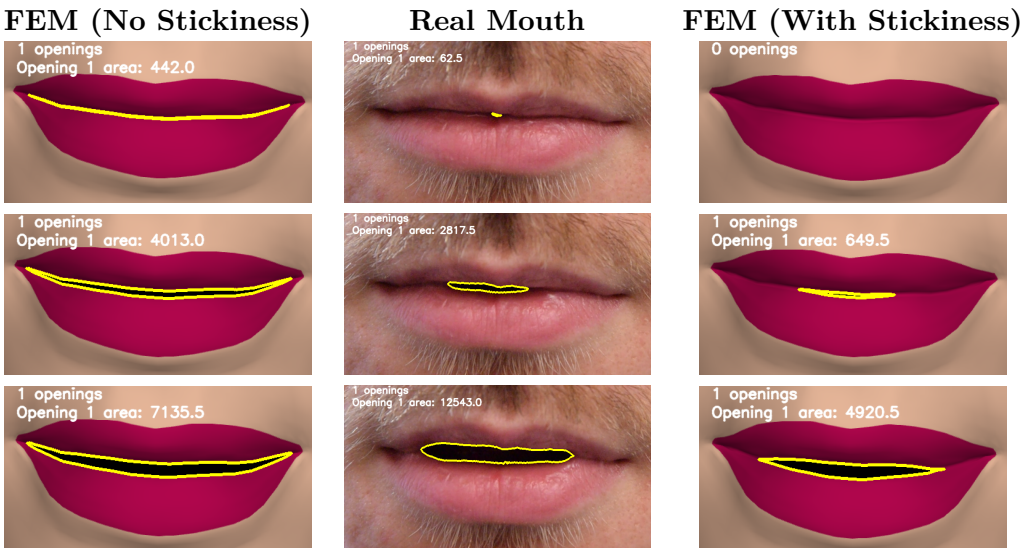


Figure 5.31: Analysis of an animation produced by a finite element model without stickiness (left column), a real mouth (middle column), and a finite element model with stickiness (right column). Time progresses vertically down the page with a 0.12 second interval between each row.

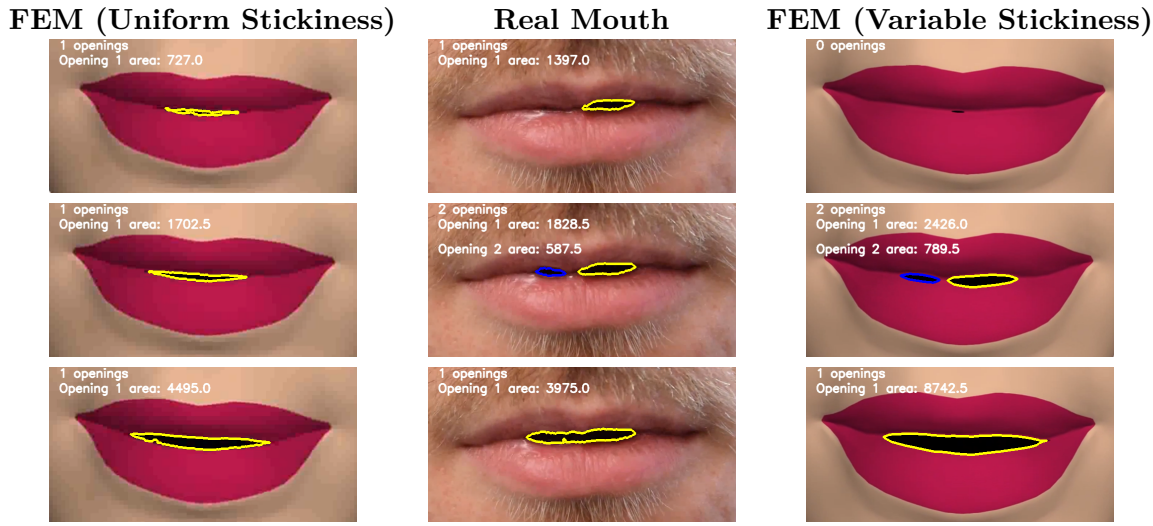


Figure 5.32: Analysis of an animation produced by a finite element model uniform stickiness (left column), a real mouth (middle column), and a finite element model with stickiness (right column). This example shows two mouth openings. Time progresses vertically down the page with a 0.12 second interval between each row.

As these parameters are determined using pixels, the distance of the speaker from the camera can affect them. For a single speaker this isn't a problem, but if comparing between speakers, results may be distorted. Additionally, to compare between users, it is more useful for these measurements to be relative rather than absolute as different speakers have mouths of different widths. To avoid this problem, the widths, heights and areas are normalized using the width of the mouth. It is assumed that the speaker remains facing the camera during the process.

Graph Generation

For the purposes of data analysis, it is enough to consider the data as vectors. That said, it is useful to be able to visualize the data in graph format. Graphs are generated using matplotlib (Hunter, 2007). Figure 5.33 shows an example graph generated from data collected in the analysis of a double mouth opening formed during the opening of the mouth.

Computation of Average Graphs

In order to build a picture of the profile of each movement analysed, the data for each movement is averaged across all participants. These average profiles serve as test patterns to match individual profiles against. In this way, specific regions of the test pattern can be labeled as corresponding to particular parts of the movement. It is then possible to identify these same regions in an individual video by matching the two profiles.

Time Invariant Signal Matching

In order to match an individual signal to an average, care must be taken to account for the differences in speed at which users may have performed an action or spoken an utterance. The dynamic time warping algorithm (Berndt and Clifford, 1994) was selected to handle this. Dynamic time warping has been used in speech recognition to cope with different speaker speeds, as well as for speaker recognition and signature recognition, as it also provides a measure of similarity between two signals.

The two signals can be considered as vectors of individual elements. Dynamic time warping functions by generating a discrete mapping between elements in one vector and those in the

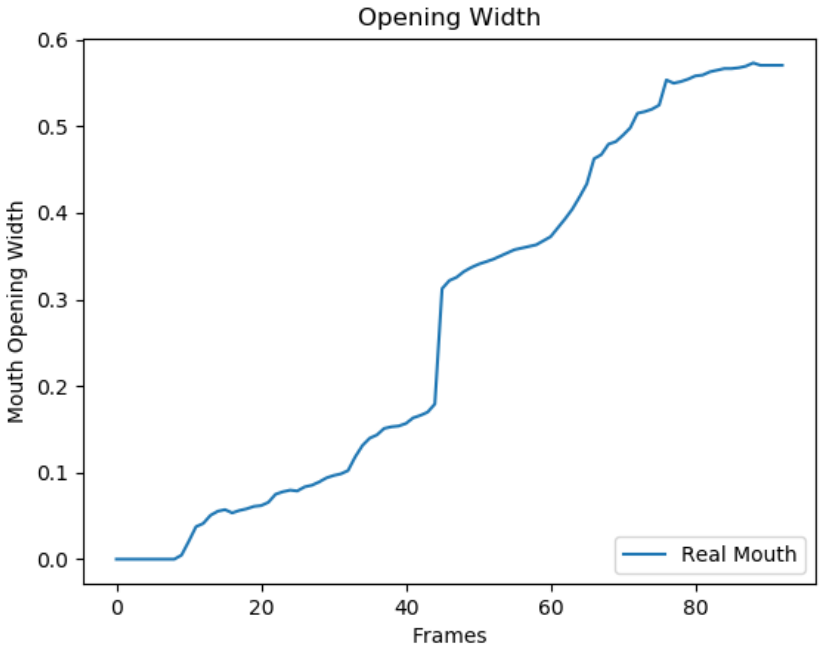


Figure 5.33: An example graph generated from a video including a double mouth opening. The data plotted is for the largest mouth opening present. The jump in mouth opening width around frame 40 is due to the double mouth opening merging into a single one.

other. This mapping is termed the *warping path* and may be used to temporally align the two signals.

Dynamic time warping generates an optimal match between two signals, given certain constraints:

- The first and last indices from each sequence must be matched, although each may have more than one match.
- Every item in either sequence must match at least one item in the other sequence.
- The matchings must be monotonically increasing. That is to say once the mapping advances to index i in the first sequence, no item from the second sequence may be paired with any element $x < i$ in the first sequence.

The algorithm in its naive form as implemented operates in $O(n^2)$ time complexity, however, as none of the sequences captured are very long this is not an issue and the mapping generation is quite performant, in the order of milliseconds per graph.

Hausdorff Distance

Once the mapping has been produced, the two signals may be aligned. To test the similarity of the two aligned signals, the Hausdorff distance measure (Hausdorff, 2005) is used. Conceptually, the Hausdorff distance finds the largest of all the distances between a point in one set and its closest point in the other set. This gives a strong similarity if every point in one set is close to a point in the other set. By computing the Hausdorff distance between a given signal and each of the average signals, it is possible to identify which action was performed to generate the signal.

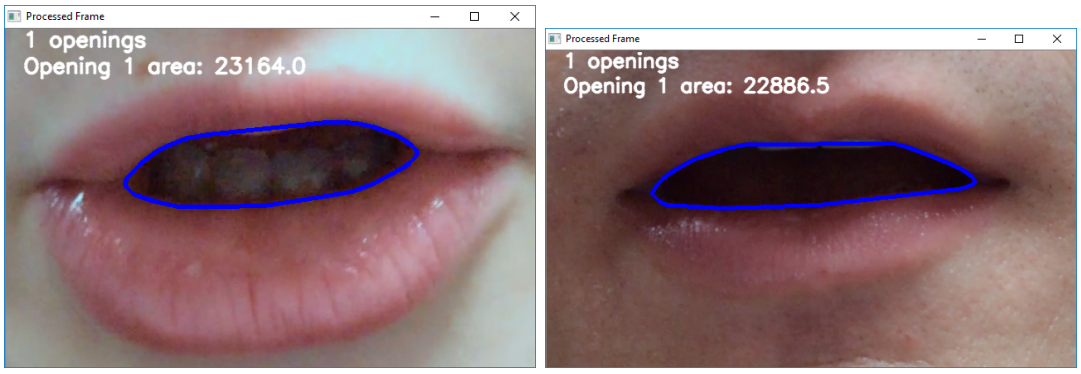


Figure 5.34: Examples of the analysis method detecting mouth openings.

Gradient Analysis

With knowledge of which action is being analysed and its warping path, a mapping can be drawn against the correct average signal and regions of interest in the signal can be labelled. This is useful as it is possible to extract information about how the lips are moving through a particular portion of an action, for example the initial mouth opening during speaking of the word *mummy*. By performing this process on multiple signals, direct comparisons between speakers or stickiness levels for a specific part of an action or utterance can be drawn.

5.3.3 Results

To demonstrate the evaluation process it is applied to real mouth videos collected in the data set. First, examples of the objective analysis process detecting a range of mouth shapes from videos of different participants are presented. Next, the average graphs are presented for a range of actions. Comparisons between the average graphs and some individual participant graphs are carried out to demonstrate how the process can be used to extract information about a movement. Following this, the process is applied to animations produced by the model using stickiness presented in this thesis. Next follows an examination of how the graphs change with the variation of stickiness level in the computer generated mouth videos. Finally, the results are related to those of the perceptual evaluation, examining correlations between objective improvements in stickiness behaviour and subjective preferences for particular stickiness levels in the animations.

Analysis of Real Mouths

This subsection demonstrates the analysis process applied to some real mouth videos. Figure 5.34 shows the analysis operating on some frames from the collected video corpus.

Graphs showing various parameters across different movements are shown in figures 5.35 and 5.36. Figure 5.35 demonstrates the graph profiles for three speakers, for three metrics speaking the letter ‘m’. The videos were chosen with three different levels of stickiness, with stickiness decreasing moving from left to right. The top row shows the opening area, the middle row shows the opening width and the bottom row shows the opening height. Visually, the profiles are similar for each of the three speakers, however, there is some inter-speaker variation. As an example, in the middle row which shows opening width, the second speaker’s profile has a plateau at its peak, suggesting that this speaker’s mouth reaches its fully opened width quickly for the utterance and then stays in a constant position. In contrast, the third speaker’s profile gradually rises a little after the initial opening. This could represent the sticky corners of the mouth opening even once gross jaw and lip movements have ceased. Figure 5.36 shows the same metrics generated for the same three speakers speaking the word ‘baby’.

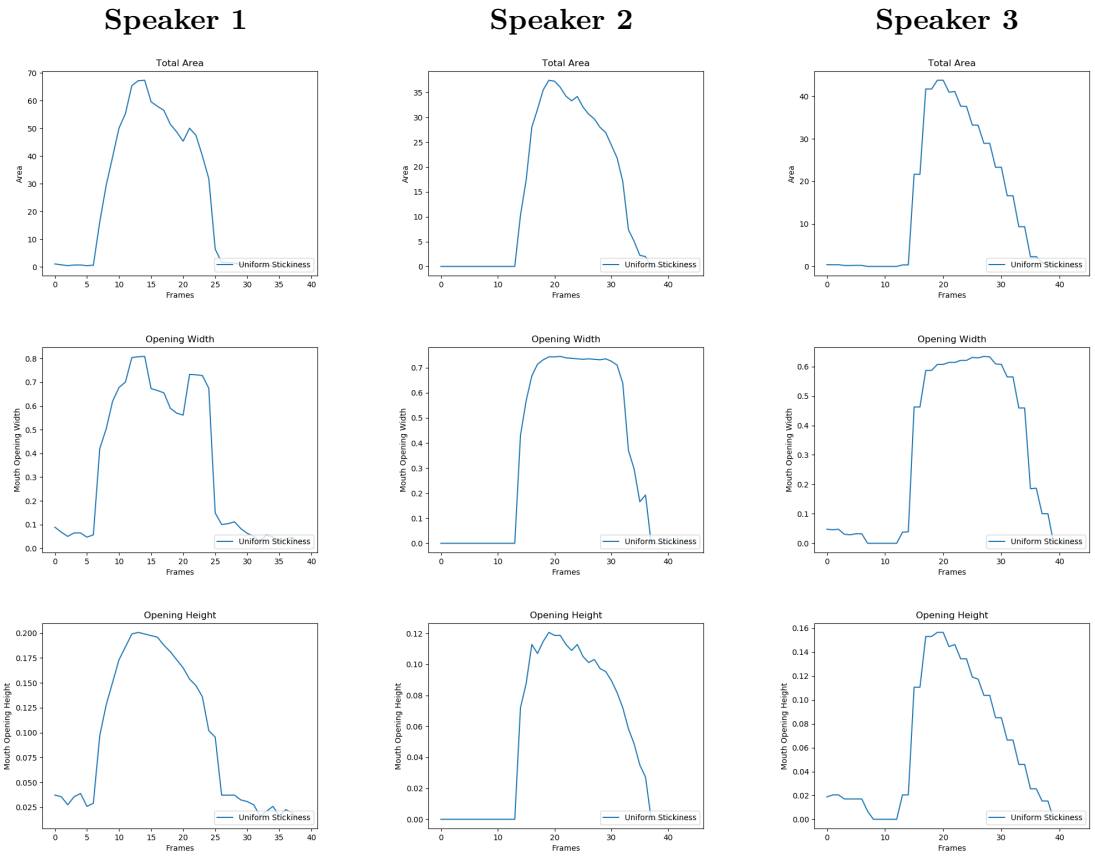


Figure 5.35: Graphs showing the opening profiles for 3 different speakers producing an 'm' sound. The top row shows the area, the middle row shows opening width and the bottom row shows the opening height.

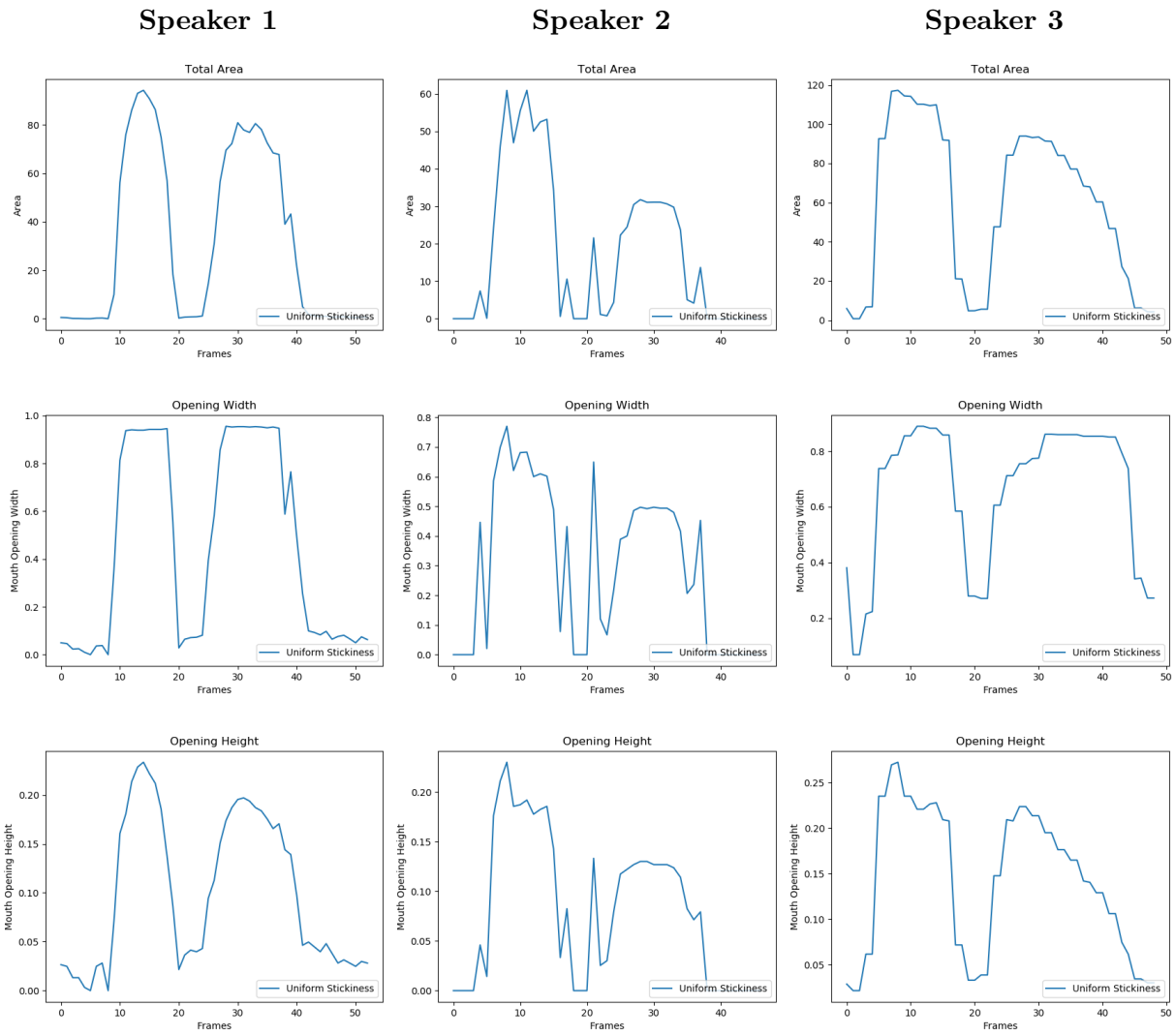


Figure 5.36: Graphs showing the opening profiles for 3 different speakers speaking the word 'baby'. The top row shows the area, the middle row shows opening width and the bottom row shows the opening height.

Comparison of Virtual Mouths and Real Mouths

Having shown how opening profiles are represented in the graphs and analysing some real mouth behaviours, the graphs are now used to evaluate the impact of the inclusion of stickiness on the model. Figures 5.37-5.39 demonstrate a range of metrics generated for a slow mouth opening action. In the videos these graphs are generated from, only a single mouth opening forms (figure 5.31). In the source videos, the mouth starts in an open position. It is then closed and reopened twice. Figure 5.37 shows that the profile for the model including stickiness better matches the opening width profile of the real mouth than that of a model without stickiness included. The same change can be seen in figures 5.38 and 5.39. The gradual increase in opening width seen in the sticky and real mouths shows the impact of the zippering effect as opening grows outwards from the center to the edges. The sudden rise in the opening width in the no stickiness graph demonstrates very clearly that the behaviour is incorrect, as the lips separate along their entire length at once. Using a traditional ground truth difference measure this would be difficult to spot. The displacement errors are very small in spite of the behaviour being categorically incorrect. The data shown in the graphs can also be compared using quantitative methods. Applying the L2 norm distance measure to the graphs, a distance of 30.4 is obtained between the no stickiness mouth and the real mouth, and a distance of 11.1 between the stickiness mouth and the real mouth. For the width profiles, the distance between the no stickiness mouth and the real mouth is 1.3 and the distance between the stickiness mouth and the real mouth is 0.5. This represents a clear quantitative improvement in the behaviour due to the inclusion of stickiness modelling.

In figures 5.40-5.42, graphs are shown for analysis of a video in which a particularly sticky section near the centre of the lips remains temporarily stuck as the lips are drawn apart. As such, there are temporarily two mouth openings until the central stuck region separates at around frame 45. From this moment onwards, there is only a single mouth opening. The graphs given show the properties of the largest mouth opening. As a result, when the saliva breaks, sharp jumps in the values of certain metrics can be seen.

In figure 5.40, the sharp jump in opening width at the 45 frame mark in the real and sticky simulations is absent from the simulation without stickiness, as with no stickiness only a single mouth opening forms. Aside from this distinct feature, the sticky simulation also matches the profile of the real mouth more closely than the non-sticky simulation for the first portion of the videos. Both the real and sticky profiles feature a fairly sharp initial rise in mouth opening (figure 5.41), which then taper off until frame 45 when the saliva breaks. At this point both the real and sticky profiles then again feature a sharp rise. The non-sticky simulation's opening width rises much more smoothly. In figure 5.42, there is a sudden step up in the area when the saliva breaks in both the sticky and real simulations.

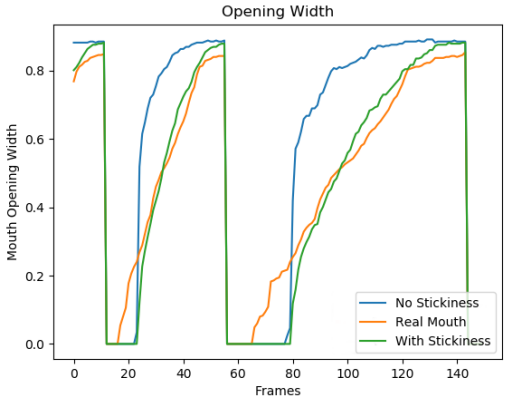


Figure 5.37: The model which includes stickiness matches the opening width profile of the real mouth much better than without stickiness.

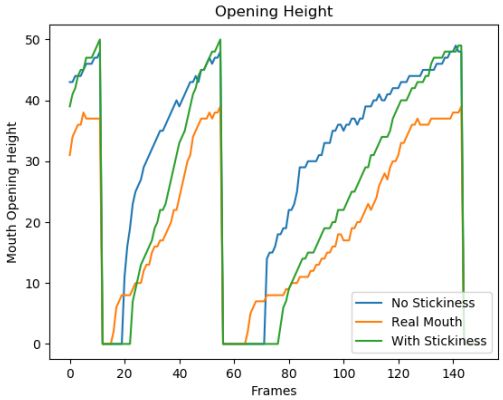


Figure 5.38: For a typical mouth opening action where only a single mouth opening forms, the inclusion of stickiness matches the opening height profile of the real mouth much better than without stickiness. The real mouth and sticky model both initially open slowly then more quickly as time progresses. The non-sticky model opens quickly initially then tapers off.

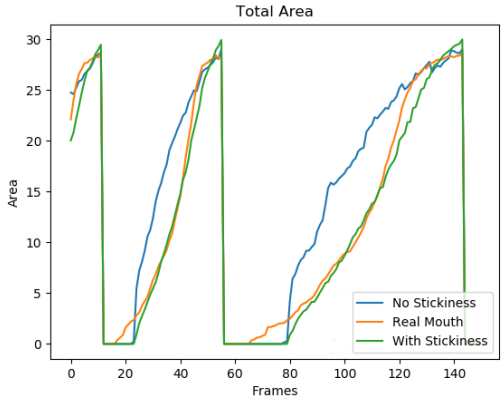


Figure 5.39: Without the inclusion of stickiness the area in the initial opening stage grows linearly, whereas with stickiness included the profile is curved, closely matching that of the real mouth.

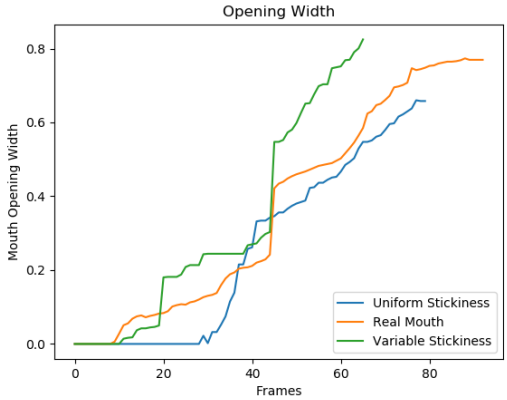


Figure 5.40: The sharp jump in opening width at the 45 frame mark in the real and sticky simulations is absent from the simulation without stickiness, as with no variable stickiness only a single mouth opening forms. Aside from this distinct feature, the sticky simulation also matches the profile of the real mouth more closely than the non-sticky simulation for the first portion of the videos.

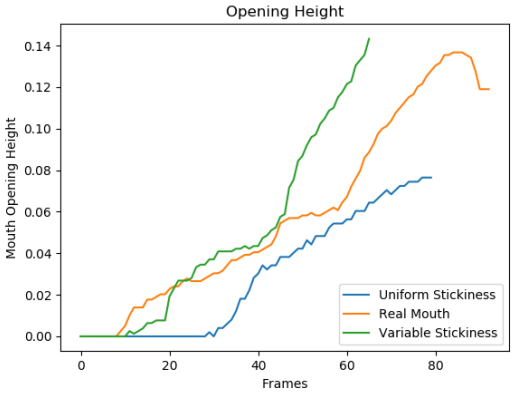


Figure 5.41: Both the real and sticky profiles feature a fairly sharp initial rise in mouth opening height, which then taper off until frame 45 when the saliva breaks. At this point both the real and sticky profiles then again feature a sharp rise. The non-sticky simulation's opening width rises much more smoothly.

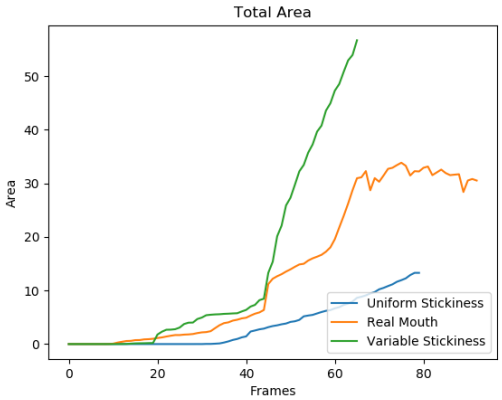
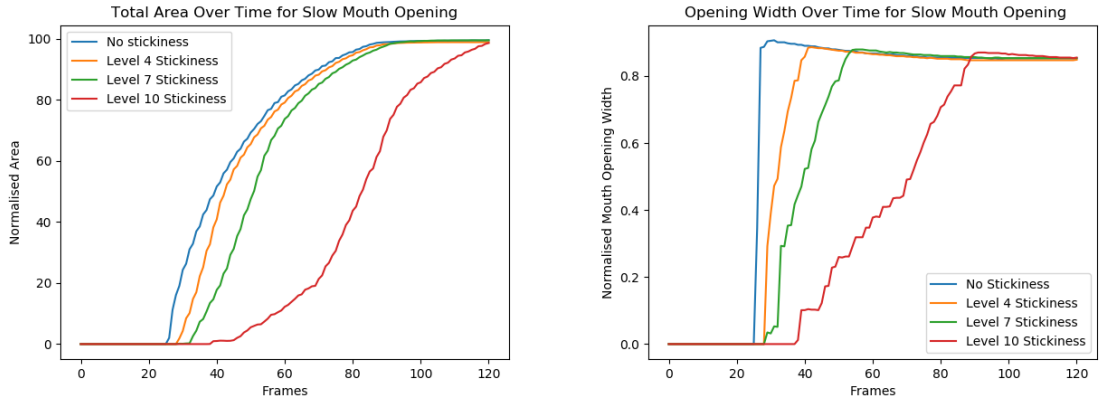
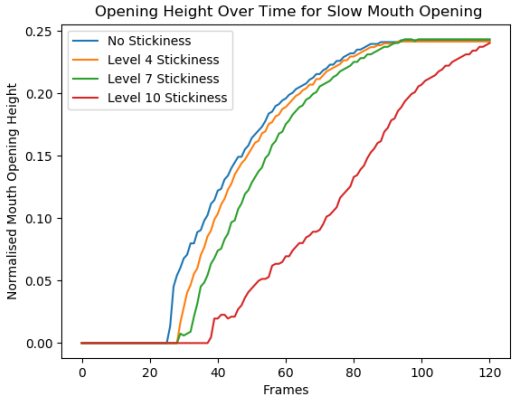


Figure 5.42: The area jumps when the saliva breaks in both the sticky and real simulations.



(a)

(b)



(c)

Figure 5.43: Graph profiles for the slow mouth opening. (a): Area, (b): Width, (c): Height

5.3.4 Analysis of the Effect of Stickiness Level in Virtual Mouths

Changing the level of stickiness also impacts the animations. The effects of these changes were measured by applying the analysis method to animations of the virtual mouth in which the same action was repeated with different stickiness levels. Four levels are used in the analysis: No stickiness (0), slight stickiness (4), normal stickiness (7) and high stickiness (10). These mirror the levels used in the perceptual study. Analysis is performed for each of the actions viewed in the perceptual study: slow mouth opening, fast mouth opening, 'm', 'p', 'mummy' and 'puppy'. For all actions and levels, uniform stickiness is used across the lips.

Figure 5.43 shows the profiles generated for the slow mouth opening. It can be seen that the first effect of the increased stickiness is to slightly delay the mouth opening. Logically this makes sense, as increased force is required to separate the lips. With the force being increased at a continuous rate, it will take longer before the lips begin to separate. This is demonstrated in the graphs by the point at which the begin to rise being further along the x-axis. From the figure 5.43 (a) and (b) graphs showing the change in area and height over time, we see that the high stickiness level produces a fundamentally different behaviour. For stickiness levels 0, 4 and 7, the profiles are quite similar, however, the changes in height and area for level 10 are significantly slower initially, requiring a ramp up period before rising significantly. This is in contrast with levels 0, 4 and 7 in which a very steep rise in area and height is seen almost immediately. For the opening width, graph (a), the profiles for each level are more similar, however, there are some significant differences. For stickiness level 0, there is an instantaneous rise to the maximum opening width, which gradually decreases.

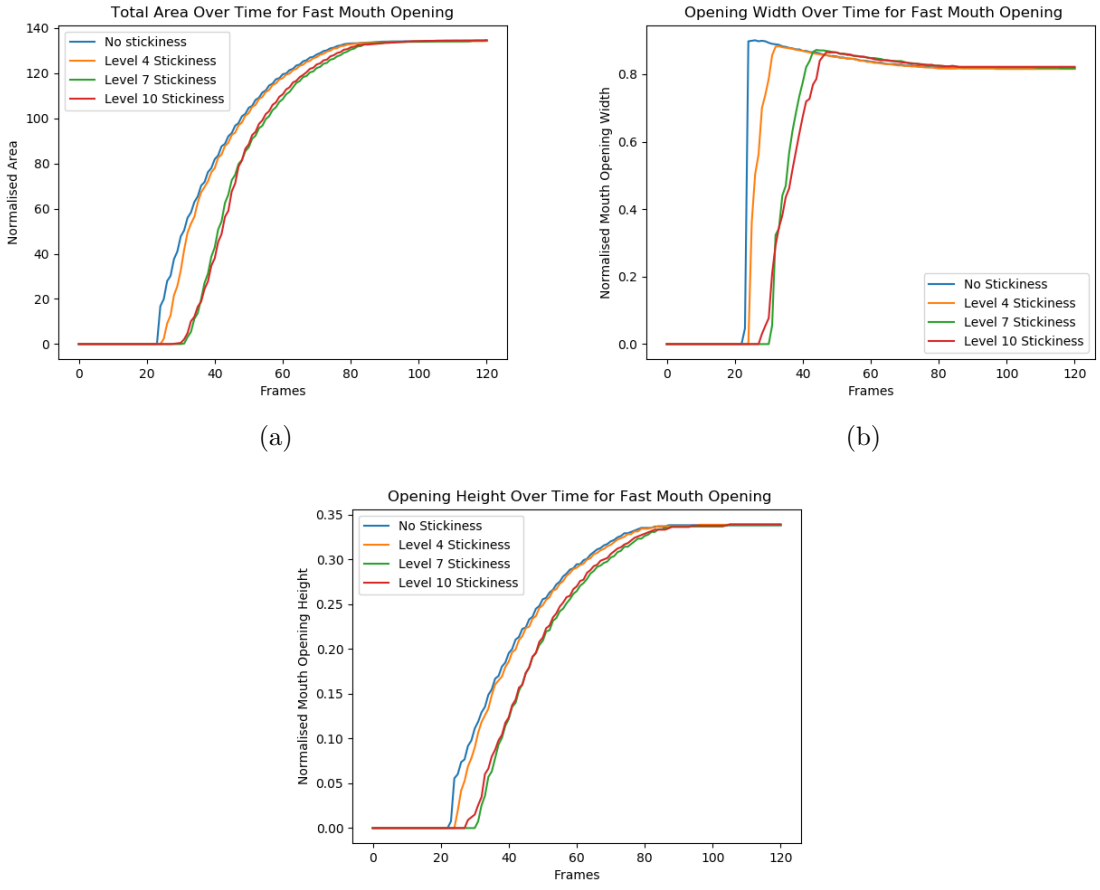
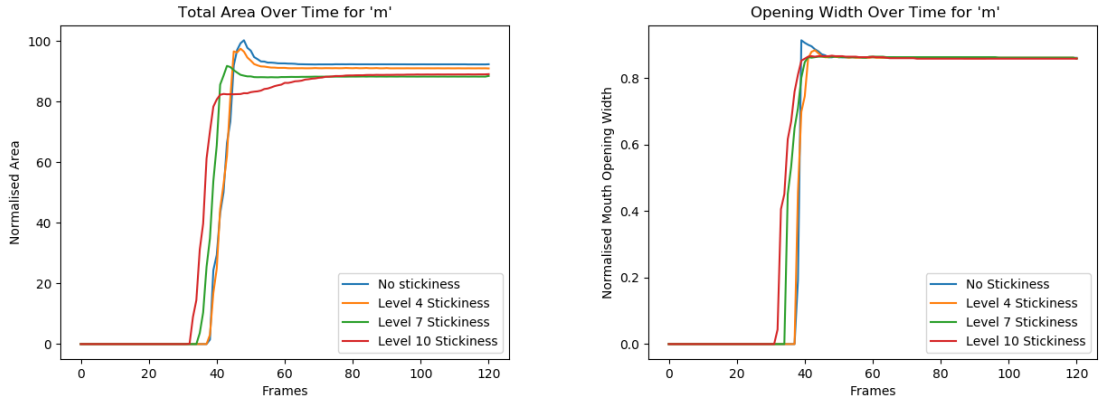


Figure 5.44: Graph profiles for the fast mouth opening. (a): Area, (b): Width, (c): Height

This decrease is a result of the mouth corners drawing inwards as the jaw opens further, leading the mouth opening to become more circular. The profiles of stickiness levels 4 and 7 are very similar to each other. Both feature a more gradual rise, with the rate decreasing over time until the maximum opening width is reached. We again see a gradual decrease in opening width following as a result of the rounding of the mouth opening, however it is not as marked a decrease. This is due to the corners of the mouth remaining stuck, meaning the mouth opening does not reach as great a width as that of the no stickiness model. The high stickiness level is significantly slower than the other two levels, showing a very gradual increase in opening width as the mouth slowly zippers open.

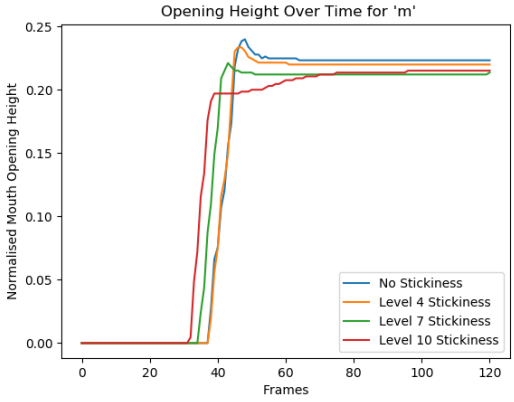
Figure 5.44 shows the opening profiles for the fast mouth opening action. The profiles are, as expected, very similar to those of the slow mouth opening. The shapes of the profiles seen are very similar, however the level 10 stickiness profile is significantly more similar to the level 4 and 7 profiles than for the slow mouth opening. In fact, the area and height profiles for level 7 and 10 are very similar. Objectively, there is much less difference between stickiness levels for the faster mouth opening than in the slow mouth opening. This fits with the preferences seen in the perceptual study, in which level 10 was strongly disliked for the slow mouth opening, but was roughly even with level 7 in the fast mouth opening, reflecting the much more similar behaviours.

Figure 5.45 shows the opening profiles for the ‘m’ utterance. Interestingly for the ‘m’ utterance, the higher stickiness levels’ mouth openings begin sooner than those of the lower levels. This is perhaps due to the way the distribution of force changes as a result of the



(a)

(b)



(c)

Figure 5.45: Graph profiles for the ‘m’ utterance. (a): Area, (b): Width, (c): Height

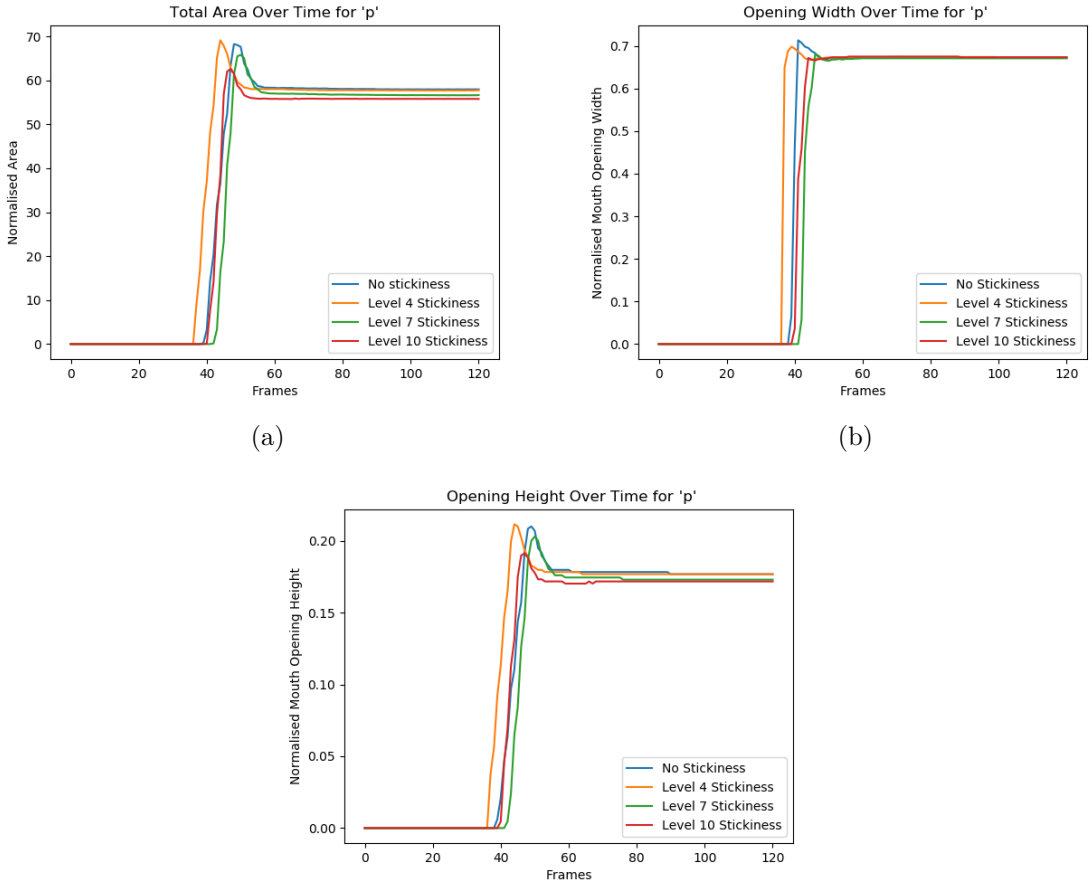


Figure 5.46: Graph profiles for the 'p' utterance. (a): Area, (b): Width, (c): Height

gluing effect between the lips. The width profiles for each level are very similar, with only a slight change in gradient, representing a fast zippering action whatever the stickiness level. More distinction is seen in the height and area graphs, which are very similar as the width is fairly constant. The lower levels see an initial spike which then levels off. This is likely due to the dynamic effects present. For the higher stickiness levels, the effect is damped by the saliva in the corners of the mouth absorbing some of the momentum. All stickiness levels tend towards a very similar equilibrium height and area. The level 10 stickiness is the only level which approaches the equilibrium position from below.

Figure 5.46 shows the opening profiles for the 'p' utterance. Generally, the graphs here are similar to those of the 'm' utterance shown in figure 5.45. This is expected as the actions involved in producing the two sounds are very similar, with the key difference being the plosive in 'p'. In the formation of the plosive, the lips are compressed together and suddenly released with greater force of separation caused by the internal air pressure. In this model, this is simulated by compressing the lips then applying an instantaneous larger force to separate the lips. The instantaneous force is sufficient to break the saliva bonding independent of the stickiness level, and so very similar profiles can be seen for all stickiness levels. The stickiness level does have a slight impact, as it can be seen that the peak height is slightly lower for the higher stickiness levels. As the lips are released, the force and movement is concentrated towards the center of the mouth, so the outer edges may still remain stuck and influence the final stages of the movement, leading to this decrease in peak height. It is worth noting the higher relative height of the peak maximum to the equilibrium height, showing that the

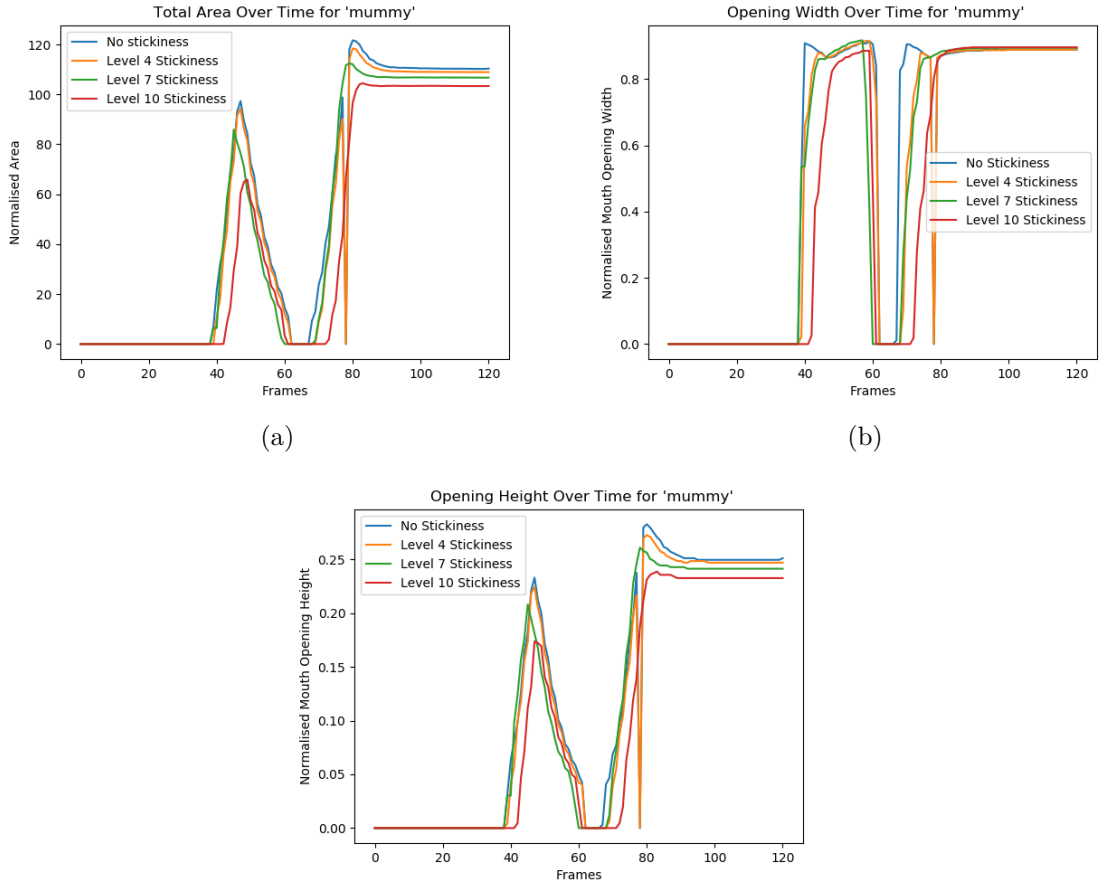


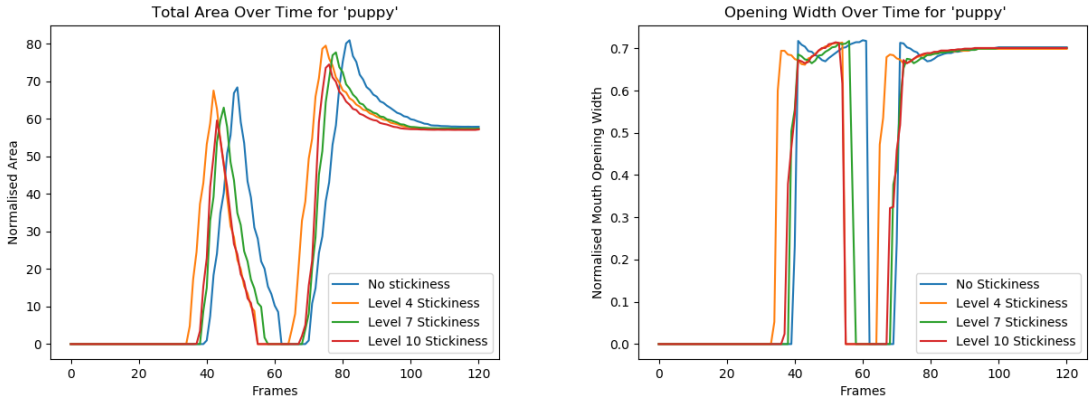
Figure 5.47: Graph profiles for the 'mummy' utterance. (a): Area, (b): Width, (c): Height

dynamic effects are stronger in this faster, more explosive movement.

Figures 5.47 and 5.48 show the opening profiles for 'mummy' and 'puppy' respectively. These both feature two mouth openings, as reflected in the graphs having two non-zero sections. The effect of the plosive can again be seen, with the stickiness level causing significant variation in the peak opening area and height of the first opening in 'mummy', but not in 'puppy'. This is again repeated in the second opening. The greater influence of dynamic effects is also represented in the greater relative peak height in the 'puppy' utterance's second mouth opening segment. The zippering effect caused by stickiness in the 'mummy' utterance is also demonstrated, with lesser gradients for higher stickiness levels.

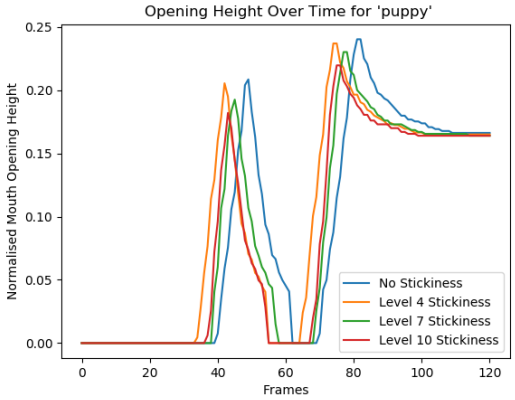
5.4 Discussion

A perceptual study was carried out in which participants were asked to rate which animation they thought was more realistic in pairs of animations. Six actions were demonstrated, four of which were given with and without audio. For each action, four different stickiness levels were compared. Animations including stickiness were preferred ($p=0.00004$). Some participants were unable to distinguish between animations or had no preference for any stickiness level, however, all those who had a preference preferred animations with stickiness overall. The preference for stickiness level was affected by the speed of the movement, with higher stickiness levels preferred for slower actions, and lower levels preferred for faster actions. Audio impacted the preferences, with the inclusion of audio resulting in a shift towards preference for lower



(a)

(b)



(c)

Figure 5.48: Graph profiles for the 'puppy' utterance. (a): Area, (b): Width, (c): Height

stickiness levels. This perhaps indicates that the audio suggests a faster movement than the participants visually recognised. This confirms the idea that the audio and visual component of perception of speech impact each other and could suggest that it is important to include audio in any perceptual evaluations of the mouth.

An objective evaluation of the mouth model produced was carried out. The purpose of this investigation was two-fold. Firstly, the investigation aimed to better determine how the mouth behaves and the effect that the stickiness of the lips has on said behaviours. Secondly, the evaluation aimed to determine whether the model successfully recreates physically plausible stickiness behaviour. To carry out the evaluation, a novel computer-vision-based analysis method was developed which does not require training or special treatment of the lips, such as application of markers or a highlighting lipstick.

The analysis method first homogenises the videos by white balancing and colour enhancing them. This improves consistency of skin and lip colour between videos, which may have different lighting conditions or different skin tones present. The video is then blurred using a median filter to remove highlights and noise from the frame. The image is transformed into two different colour spaces which both highlight the lips. Thresholding is performed on both images, segmenting the lips. At this stage, some features such as the teeth are present in one image but not the other. The results of the two transforms are correlated to isolate the lip region. Finally contour detection is performed to identify the lip contours. One interesting point is that although this contour is described as the inner lip contour, it is often actually part of the inner mouth forming it.

Following the contour identification, parameters are extracted giving information about how the shape of the mouth is changing over time. Data was collected from the videos in the mouth video corpus for different actions. The data were averaged to generate average curves for each action. In order to compare two new videos or animations, the analysis process is run on each one and dynamic time warping is used to temporally align both with the reference curve. With the two curves both aligned with the reference curve, gradient analysis can reveal information about differences in mouth behaviour between the two videos.

The evaluation process establishes a range of metrics which can be measured from video sources. These metrics provide information about the movement of the mouth over time, focusing on the mouth opening. Width, height, and mouth opening area are captured, normalized by the width of the mouth. This allows comparison of different mouths and videos with different resolutions or zoom levels. One limitation of the process is that it only works for video filmed directly facing the person. The evaluation process given here provides a method for testing all mouth models and producing a quantitative evaluation of their effectiveness. The evaluation process could also be applied to other mouth animation models. Further to this, the process can be used for videos of real mouths and could be useful to judge the results of medical procedures, for example in the case of palsy.

To support evaluation of the model, a new corpus of audio-visual data was collected. Films of mouths were recorded speaking a set of utterances and performing specific actions in which sticky lip effects are common. Speakers were asked to lick their lips, then repeat each action a number of times with a rest period between each repetition to allow the saliva to dry, affecting the stickiness level. The videos are recorded at 1920x1080 resolution, at 60 frames per second. This is believed to be the first data set capturing video footage with known variations of stickiness. The data is to be made available on the University of Sheffield's ORDA system. Although the data set provides improvements in resolution and frame rate, these could be improved further through the use of a high-speed camera. An even higher frame rate would allow better studying of dynamic effects and plosive movements, which can still be very fast at 60 frames per second. More control over the diet of participants in the run up to the study could also provide more consistent results between participants, as the composition of the saliva can impact its physical and rheological properties (de Almeida et al.,

2008). Another possibility for better controlling the stickiness level would be to use artificial coatings to modify the stickiness of the lips, such as lip stick or lip balm. Though these may change the behaviour of the lips, it is an interesting research area in its own right. Finally, more variables could be changed to test the effect of breathing through the nose or mouth and the air temperature/moisture on the rate of evaporation and changes in stickiness behaviour.

Although sentences were recorded in the data set, they were not used in the evaluation. Simpler isolated utterances and actions offer more consistency between speakers, although sentences could be used in future work. In the sentences in particular, coarticulation could affect how the sentence is spoken.

Comparing the results of the perceptual and objective analyses, both show that there is a significant difference in stickiness behaviour based on the speed of the movements. In the perceptual study, users had preferences for different stickiness levels for faster and slower movements. In the objective analysis, it was shown that the stickiness level had more impact in slower movements, with faster movements showing little difference between stickiness levels.

Chapter 6

Conclusions

This thesis has presented a physically-based virtual mouth model implementing stickiness behaviour, and validated the produced behaviours using both a perceptual study and an objective evaluation process. The research contributions are:

- The development of a finite element model of the mouth incorporating a novel stickiness model, controlled by a moisture model.
- A perceptual study giving new insights to the impact of stickiness on animations and the factors which affect it.
- The collection of a data set of high resolution, high frame rate mouth videos with varying stickiness. This data set is novel in that actions are recorded at varying stickiness levels.
- A novel objective evaluation method for tracking mouth movements from video.

The saliva was identified as causing stickiness between the lips, which leads to the effects of 'zippering' and formation of multiple mouth openings. The 'zippering' effect holds the lips together and leads to gradual opening of the mouth from the center rather than the lips separating along their entire lengths at once. Variable levels of stickiness can cause multiple or asymmetric mouth openings. It was also observed that the level of stickiness could vary and changed over time. Research into the behaviour of saliva identified that this is related to the moisture content of the saliva. The saliva could also act as a lubricant between the lips when present with a higher moisture content, leading to reduced shearing forces when the lips are moved transversely. The sticky behaviours, 'zippering' and formation of multiple/asymmetric mouth openings as well as their time dependence were chosen as the subject of this research.

The total Lagrangian explicit dynamics formulation of the finite element method was used and extended with a novel stickiness model to produce a simulation of the mouth which includes some of the saliva effects. The simulation is capable of reproducing the targeted effects of 'zippering' and formation of asymmetric or multiple mouth openings. The method is artist controllable through the dependence of the stickiness on a moisture level which can be easily controlled. This is also affected by a moisture model which accounts for evaporation of moisture from the saliva, leading to dynamically updated stickiness levels. In comparison with other state-of-the-art works, the model presented in this thesis produces more deterministic and realistically accurate behaviours as it is physically-based. With that said, the model is limited in its anatomical detail. The model could include more anatomical features such as volumetric bones and accurately shaped muscles. The tongue and teeth could also be included. Finally, the model could be integrated with another technique such as blendshapes to allow the use of the physically-based lip animation in more conventional workflows. This could work by matching boundary nodes of the physical simulation to nodes in the blendshape mesh, allowing insertion of the physically-based mouth region.

A perceptual study was conducted to evaluate whether the inclusion of stickiness improves the perceived realism of an animation, as well as investigate which stickiness levels are considered most realistic. The study concluded that animations including stickiness are perceived as more realistic with a p-value of 0.00004. Additionally, it was found that only some individuals seem to have a preference for sticky or non-sticky behaviours, though all of those who had a preference preferred the sticky model. The perceptual evaluation identified that there was a trend to prefer lower levels of stickiness for faster movements such as plosives and higher levels of stickiness for slower movements. It was also found that the inclusion of audio could affect the preferred level of stickiness. It has been demonstrated that visual cues can affect perception of audio (Alghamdi et al., 2015; McGurk and MacDonald, 1976). This research suggests that this may work both ways, and that audio cues could influence what people expect to see. The perceptual study could be improved by collecting data from a larger set of participants and widening the range of stickiness levels and actions.

A new corpus of mouth videos was collected in which the stickiness level was controlled. The videos are high resolution, high frame rate clips in which participants perform a set of actions and utterances at set intervals, allowing the saliva on their lips to dry between repetitions. At the start of each new action or utterance, participants licked their lips to ensure a higher initial moisture level. Ethical approval and consent from the participants has been received to make the corpus available on the University of Sheffield's ORDA research data database.

To further support the conclusions of the perceptual evaluation, an objective evaluation was conducted. This used a novel process to conduct analysis of the new corpus of mouth videos and animations produced by the model. This analysis also demonstrated that the inclusion of stickiness improves the realism of mouth animations. Further to this, the objective analysis demonstrates the differences in behaviours in the model between different stickiness levels. Specifically, it is shown that the stickiness level impacts the dynamic effects present, and demonstrates the importance of energies of movements, as the stickiness behaviours manifest differently for plosives. This result is also reflected in the perceptual evaluation, with participants expressing different preferences for faster and slower movements. This could be explained by the frame rate of the videos limiting how noticeable the effects are for fast movements at relatively slower frame rates. An analysis of which effects are visible in videos at different rates presents an opportunity for future work. The stickiness behaviour had the greatest impact on slower movements, particularly on the mouth opening width, with more severe zippering at higher stickiness levels. The metrics collected were effective at demonstrating differences between mouth movements, however, the analysis method sometimes struggled to distinguish between lips and tongue as these are very similarly coloured in real mouths, so this could present an area for further work. Additionally, further work could pursue improved recognition of the teeth and handling of reflections to better the consistency of results. Finally, the method could be integrated with other techniques to produce a hybrid recognition method, perhaps taking advantage of active appearance models to quickly identify the most relevant region or increase confidence in the result.

This work has focused on recreating natural stickiness between the lips. Despite this, the model could also be used to model artificial stickiness between the lips. This could include lips which have had a lip balm or lip stick applied. Application of such products can affect the stickiness level between the lips, as the lip balms and lip sticks are typically cohesive. This could be desirable for animations in films and games for characters with lip stick applied. This could also be of interest to the cosmetics industry. In the medical industry, the stickiness and saliva modelling could be developed to simulate results of operations or conditions which affect saliva production such as xerostomia.

Future work could improve the model by including mesh adaption to allow sliding of the lips laterally. The present method is effective for vertical mouth movements, but limits lateral

movement of the jaw. An even higher frame rate mouth video corpus could allow better analysis of dynamic effects and lead to further insights regarding the influence of stickiness on lip movement. Further statistical analysis could be performed on the collected video corpus to strengthen the conclusions drawn here, or identify further areas for research. The physically-based facial model could be used to aid generative modelling by generating training and test data for the networks. Additionally, a more fully featured face model could be developed using volumetric muscles, more detailed modelling of the skin and the inclusion of the teeth. The impact of different speeds of movements on the results in both the perceptual and objective studies suggest that further work recreating the correct energies of movements could also be important. To conclude, this thesis has successfully developed a model of the mouth which incorporates physically plausible stickiness behaviours and verified that the inclusion of stickiness produces more accurate animations of the mouth.

Bibliography

Oleg Alexander, Mike Rogers, William Lambeth, Matt Chiang, and Paul Debevec. The digital emily project: photoreal facial modeling and animation. In *ACM Siggraph 2009 courses*, page 12. ACM, 2009.

Najwa Alghamdi, Steve Maddock, Guy J Brown, and Jon Barker. Investigating the impact of artificial enhancement of lip visibility on the intelligibility of spectrally-distorted speech. In *AVSP*, pages 93–98, 2015.

Firooz A Allahdadi, Theodore C Carney, Jim R Hipp, Larry D Libersky, and Albert G Petschek. High strain lagrangian hydrodynamics: a three dimensional sph code for dynamic material response. Technical report, Phillips Lab Kirtland AFB NM, 1993.

Pierre Badin and Antoine Serrurier. Three-dimensional linear modeling of tongue: Articulatory data and models. In *7th International Seminar on Speech Production, ISSP7*, pages 395–402. UFMG, Belo Horizonte, Brazil, 2006.

Pierre Badin, Gerard Bailly, Lionel Reveret, Monica Baciu, Christoph Segebarth, and Christophe Savariaux. Three-dimensional linear articulatory modeling of tongue, lips and face, based on mri and video images. *Journal of Phonetics*, 30(3):533–553, 2002.

Michael Bao, Matthew Cong, Stéphane Grabli, and Ronald Fedkiw. High-quality face capture using anatomical muscles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10802–10811, 2019.

Giuseppe Barbarino, Mahmood Jabareen, Juergen Trzewik, and Edoardo Mazza. Physically based finite element model of the face. In *International Symposium on Biomedical Simulation*, pages 1–10. Springer, 2008.

Scott G Bardenhagen and Edward M Kober. The generalized interpolation material point method. *Computer Modeling in Engineering and Sciences*, 5(6):477–496, 2004.

V Barrielle and N Stoiber. Realtime performance-driven physical simulation for facial animation. In *Computer Graphics Forum*. Wiley Online Library, available online: 11 June 2018, 2018. doi: 10.1111/cgf.13450.

Vincent Barrielle. *Leveraging Blendshapes for Realtime Physics-Based Facial Animation*. Theses, Université Bretagne Loire, November 2017. URL `https://barrielle.xyz/`.

Vincent Barrielle and Nicolas Stoiber. Realtime performance-driven physical simulation for facial animation. In *Computer Graphics Forum*, volume 38, pages 151–166. Wiley Online Library, 2019.

Vincent Barrielle, Nicolas Stoiber, and Cédric Cagniart. Blendforces: A dynamic framework for facial animation. *Computer Graphics Forum*, 35(2):341–352, May 2016. ISSN 0167-7055.

Sumit Basu, Nuria Oliver, and Alex Pentland. 3d lip shapes from video: A combined physical–statistical model. *Speech Communication*, 26(1-2):131–148, 1998.

Ted Belytschko, Yun Yun Lu, and Lei Gu. Element-free galerkin methods. *International journal for numerical methods in engineering*, 37(2):229–256, 1994.

Ted Belytschko, Yury Krongauz, Daniel Organ, Mark Fleming, and Petr Krysl. Meshless methods: an overview and recent developments. *Computer methods in applied mechanics and engineering*, 139(1):3–47, 1996.

Ted Belytschko, Yury Krongauz, John Dolbow, and C Gerlach. On the completeness of meshfree particle methods. *International Journal for Numerical Methods in Engineering*, 43(5):785–819, 1998.

Ted Belytschko, Yong Guo, Wing Kam Liu, and Shao Ping Xiao. A unified stability analysis of meshless particle methods. *International Journal for Numerical Methods in Engineering*, 48(9):1359–1400, 2000.

Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

Eloïse Berson, Catherine Soladie, Vincent Barrielle, and Nicolas Stoiber. A robust interactive facial animation editing system. In *Motion, Interaction and Games*, pages 1–10. 2019.

Bernd Bickel, Moritz Bächer, Miguel A Otaduy, Wojciech Matusik, Hanspeter Pfister, and Markus Gross. Capture and modeling of non-linear heterogeneous soft tissue. *ACM Transactions on Graphics (TOG)*, 28(3):1–9, 2009.

Peter Birkholz and Bernd J Kröger. Vocal tract model adaptation using magnetic resonance imaging. In *Proceedings of the 7th International Seminar on Speech Production*, pages 493–500, 2006.

JHH Bongaerts, D Rossetti, and JR Stokes. The lubricating properties of human whole saliva. *Tribology Letters*, 27(3):277–287, 2007.

Steven K Boyd and Ralph Müller. Smooth surface meshing for automated finite element model generation from 3d image data. *Journal of Biomechanics*, 39(7):1287–1295, 2006.

Matthew Brand. Voice puppetry. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 21–28, 1999.

BA Canning and MF Rose. Clinical measurements of the speed of tongue and lip movements in british children with normal speech. *British Journal of Disorders of Communication*, 9 (1):45–50, 1974.

Chen Cao, Yanlin Weng, Stephen Lin, and Kun Zhou. 3d shape regression for real-time facial animation. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013.

Matthieu Chabanas and Yohan Payan. A 3D Finite Element Model of the Face for Simulation in Plastic and Maxillo-Facial Surgery. *Medical Image Computing and Computer-Assisted Intervention*, pages 1068–1075, 2000. ISSN 16113349.

Matthieu Chabanas, Vincent Luboz, and Yohan Payan. Patient specific finite element model of the face soft tissues for computer-assisted maxillofacial surgery. *Medical Image Analysis*, 7(2):131–151, 2003. ISSN 13618415. doi: 10.1016/S1361-8415(02)00108-1.

Jinxiang Chai and Jessica K Hodgins. Constraint-based motion optimization using a statistical dynamic model. In *ACM SIGGRAPH 2007 papers*, pages 8–es. 2007.

Byoungwon Choe, Hanook Lee, and Hyeong-Seok Ko. Performance-driven muscle-based facial animation. *The Journal of Visualization and Computer Animation*, 12(2):67–79, 2001.

Matthew Cong, Michael Bao, Kiran S Bhat, Ronald Fedkiw, et al. Fully automatic generation of anatomical face simulation models. In *Proceedings of the 14th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 175–183. ACM, 2015.

Matthew Cong, Kiran S Bhat, and Ronald Fedkiw. Art-directed muscle simulation for high-end facial animation. In *Symposium on Computer Animation*, pages 119–127, 2016.

W Cook. Mapping methods for generating three-dimensional meshes. *Comp. Mech. Eng.*, 1: 67–72, 1982.

Martin Cooke, Jon Barker, Stuart Cunningham, and Xu Shao. An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, 120(5):2421–2424, 2006.

Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. In *European conference on computer vision*, pages 484–498. Springer, 1998.

Darren Cosker, Dave Marshall, Paul L Rosin, and Yulia Hicks. Speech driven facial animation using a hidden markov coarticulation model. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 1, pages 128–131. IEEE, 2004.

Darren Cosker, Susan Paddock, David Marshall, Paul L Rosin, and Simon Rushton. Toward perceptually realistic talking heads: Models, methods, and mcgurk. *ACM Transactions on Applied Perception (TAP)*, 2(3):270–285, 2005.

P Del Vigna de Almeida, AM Gregio, MA Machado, AA De Lima, L Reis Azevedo, et al. Saliva composition and functions: a comprehensive review. *J Contemp Dent Pract*, 9(3): 72–80, 2008.

Zhigang Deng, Pei-Ying Chiang, Pamela Fox, and Ulrich Neumann. Animating blendshape faces by cross-mapping motion capture data. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 43–48. ACM, 2006.

Dimitar Dinev, Thabo Beeler, Derek Bradley, Moritz Bächer, Hongyi Xu, and Ladislav Kavan. User-guided lip correction for facial performance capture. In *Computer Graphics Forum*, volume 37, pages 93–101. Wiley Online Library, 2018.

James D Edge. *Techniques for the synthesis of visual speech.* PhD thesis, University of Sheffield, 2004.

Gareth J Edwards, Timothy F Cootes, and Christopher J Taylor. Face recognition using active appearance models. In *European conference on computer vision*, pages 581–595. Springer, 1998a.

Gareth J Edwards, Andreas Lanitis, Christopher J Taylor, and Timothy F Cootes. Statistical models of face images-improving specificity. *Image and Vision Computing*, 16(3):203–212, 1998b.

Pif Edwards, Chris Landreth, Eugene Fiume, and Karan Singh. Jali: an animator-centric viseme model for expressive lip synchronization. *ACM Transactions on Graphics (TOG)*, 35(4):1–11, 2016.

Paul Ekman and Wallace V Friesen. *Manual for the facial action coding system.* Consulting Psychologists Press, 1978.

Irfan Essa, Sumit Basu, Trevor Darrell, and Alex Pentland. Modeling, tracking and interactive animation of faces and heads//using input from video. In *Proceedings Computer Animation'96*, pages 68–79. IEEE, 1996.

Human Facial Expression and New Series. Royal Anthropological Institute of Great Britain and Ireland. 4(4):525–536, 1969.

Faceware. Faceware Analyzer, 2019. URL `http://facewaretech.com/products/software/analyzer/`.

Damian Farnell, Jennifer Galloway, Alexei Zhurov, Stephen Richmond, David Marshall, Paul Rosin, Khtam Al-Meyah, Pertti Pirttiniemi, and Raija Lähdesmäki. What's in a smile? initial analyses of dynamic changes in facial shape and appearance. *Journal of Imaging*, 5 (1):2, 2019.

Philip C Fox. Management of dry mouth. *Dental Clinics of North America*, 41(4):863–875, 1997.

Pablo Garrido, Michael Zollhöfer, Chenglei Wu, Derek Bradley, Patrick Pérez, Thabo Beeler, and Christian Theobalt. Corrective 3d reconstruction of lips from monocular video. *ACM Trans. Graph.*, 35(6):219–1, 2016.

Jorge Gascón, Javier S Zurdo, and Miguel A Otaduy. Constraint-based simulation of adhesive contact. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 39–44. Eurographics Association, 2010.

Gadi Geiger, Tony Ezzat, and Tomaso Poggio. Perceptual evaluation of video-realistic speech. 2003.

Jean-Michel Gérard, Jacques Ohayon, Vincent Luboz, Pascal Perrier, and Yohan Payan. Non-linear elastic properties of the lingual and facial tissues assessed by indentation technique: application to the biomechanics of speech production. *Medical engineering & physics*, 27 (10):884–892, 2005.

Jean-Michel Gérard, Reiner Wilhelms-Tricarico, Pascal Perrier, and Yohan Payan. A 3d dynamical biomechanical tongue model to study speech motor control. *arXiv preprint physics/0606148*, 2006.

E. Gladilin, S. Zachow, P. Deuflhard, and H. C. Hege. Anatomy- and physics-based facial animation for craniofacial surgery simulations. *Medical and Biological Engineering and Computing*, 42(2):167–170, 2004. ISSN 01400118. doi: 10.1007/BF02344627.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Ashley D Gritzman, David M Rubin, and Adam Pantanowitz. Comparison of colour transforms used in lip segmentation algorithms. *Signal, Image and Video Processing*, 9(4):947–957, 2015.

Thierry Guiard-Marigny, Nicolas Tsingos, Ali Adjoudani, Christian Benoit, and M-P Gascuel. 3d models of the lips for realistic speech animation. In *Computer Animation'96. Proceedings*, pages 80–89. IEEE, 1996.

Ahmad BA Hassanat and Sabah Jassim. Color-based lip localization method. In *Mobile Multimedia/Image Processing, Security, and Applications 2010*, volume 7708, page 77080Y. International Society for Optics and Photonics, 2010.

Felix Hausdorff. *Set theory*, volume 119. American Mathematical Soc., 2005.

Volker Helzle and Kai Goetz. Digital albert einstein, a case study. In *ACM SIGGRAPH 2018 Talks*, pages 1–2. 2018.

Andrew Hill, Timothy F Cootes, and Christopher J Taylor. Active shape models and the shape approximation problem. *Image and Vision Computing*, 14(8):601–607, 1996.

Pengyu Hong, Zhen Wen, and Thomas S Huang. Real-time speech-driven face animation with expressions using neural networks. *IEEE Transactions on neural networks*, 13(4):916–927, 2002.

William Graham Hoover and Carol Griswold Hoover. Spam-based recipes for continuum simulations. *Computing in Science & Engineering*, 3(2):78–85, 2001.

Haoda Huang, Jinxiang Chai, Xin Tong, and Hsiang-Tao Wu. Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. In *ACM SIGGRAPH 2011 papers*, pages 1–10. 2011.

Sue P Humphrey and Russell T Williamson. A review of saliva: normal composition, flow, and function. *The Journal of prosthetic dentistry*, 85(2):162–169, 2001.

John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90–95, 2007.

Alexandru-Eugen Ichim, Ladislav Kavan, Merlin Nimier-David, and Mark Pauly. Building and animating user-specific volumetric face rigs. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 107–117. Eurographics Association, 2016.

Alexandru-Eugen Ichim, Petr Kadleček, Ladislav Kavan, and Mark Pauly. Phace: physics-based face modeling and animation. *ACM Transactions on Graphics (TOG)*, 36(4):153, 2017.

Singular Inversions. Facegen modeller (version 3.3)[computer software]. *Toronto, ON: Singular Inversions*, 2008.

Irina Ionescu, James Guilkey, Martin Berzins, Robert M Kirby, and Jeffrey Weiss. Computational simulation of penetrating trauma in biological soft tissues using the material point method. *Studies in health technology and informatics*, 111:213–218, 2005.

Irina Ionescu, James E Guilkey, Martin Berzins, Robert M Kirby, and Jeffrey A Weiss. Simulation of soft tissue failure using the material point method. 2006.

Chenfanfu Jiang, Theodore Gast, and Joseph Teran. Anisotropic elastoplasticity for cloth, knit and hair frictional contact. *ACM Transactions on Graphics (TOG)*, 36(4):1–14, 2017.

Stian F Johnsen, Zeike A Taylor, Matthew J Clarkson, John Hipwell, Marc Modat, Bjoern Eiben, Lianghao Han, Yipeng Hu, Thomy Mertzanidou, David J Hawkes, et al. Niftysim: A gpu-based nonlinear finite element package for simulation of soft tissue biomechanics. *International journal of computer assisted radiology and surgery*, 10(7):1077–1095, 2015.

Grand Roman Joldes, Adam Wittek, and Karol Miller. Non-locking tetrahedral finite element for surgical simulation. *Communications in Numerical Methods in Engineering*, 25(7):827–836, 2009.

Pushkar Joshi, Wen C Tien, Mathieu Desbrun, and Frédéric Pighin. Learning controls for blend shape based realistic facial animation. In *ACM Siggraph 2006 Courses*, pages 17–es. 2006.

Petr Kadleček and Ladislav Kavan. Building accurate physics-based face models from data. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 2(2):1–16, 2019.

K Kahler, J Haber, and H P Seidel. Geometry-based Muscle Modeling for Facial Animation. *Proc of Graphics Interface*, pages 37–46, 2001. ISSN 07135424.

Kolja Kähler, Jörg Haber, and Hans-Peter Seidel. Geometry-based muscle modeling for facial animation. In *Graphics interface*, volume 2001, pages 37–46, 2001.

Erwin Keeve, Sabine Girod, Ron Kikinis, and Bernd Girod. Deformable modeling of facial tissue for craniofacial surgery simulation. *Computer Aided Surgery*, 3(5):228–238, 1998. ISSN 10929088. doi: 10.1002/(SICI)1097-0150(1998)3:5⟨228::AID-IGS2⟩3.0.CO;2-I.

Kimon Matara. Easy Sticky Lips, 2017. URL `http://www.kimonmatara.com/maya-easy-sticky-lips/`.

Rolf M. Koch, Markus H. Gross, Friedrich R. Carls, Daniel F. von Büren, George Fankhauser, and Yoav I. H. Parish. Simulating facial surgery using finite element models. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96*, pages 421–428, 1996. doi: 10.1145/237170.237281. URL `http://portal.acm.org/citation.cfm?doid=237170.237281`.

Yeara Kozlov, Derek Bradley, Moritz Bächer, Bernhard Thomaszewski, Thabo Beeler, and Markus Gross. Enriching facial blendshape rigs with physical simulation. *Computer Graphics Forum*, 36(2):75–84, 2017.

Takaaki Kuratate, Marcia Riley, et al. Building speaker-specific lip models for talking heads from 3d face data. In *AVSP*, page 9, 2010.

John Lasseter. Principles of traditional animation applied to 3d computer animation. In *ACM Siggraph Computer Graphics*, volume 21, pages 35–44. ACM, 1987.

M Leach and SC Maddock. Physically-based sticky lips. In *EG UK Computer Graphics & Visual Computing*. EGUK, 2018.

Matthew Leach and Steve Maddock. An evaluation approach for a physically-based sticky lip model. *Computers*, 8(1):24, 2019.

Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Realistic modeling for facial animation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 55–62. ACM, 1995.

MichaelL J Levine. Salivary macromolecules. a structure/function synopsis. *Annals of the New York Academy of Sciences*, 694:11–16, 1993.

J. P. Lewis and F. I. Parke. Automated lip-synch and speech synthesis for character animation. *ACM SIGCHI Bulletin*, 17:143–147, 1986. ISSN 07366906. doi: 10.1145/30851.30874.

John P Lewis, Ken Anjyo, Taehyun Rhee, Mengjie Zhang, Frederic H Pighin, and Zhigang Deng. Practice and theory of blendshape facial models. *Eurographics (State of the Art Reports)*, 1(8):2, 2014.

Hao Li, Jihun Yu, Yuting Ye, and Chris Bregler. Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph.*, 32(4):42–1, 2013.

Ruilong Li, Karl Bladin, Yajie Zhao, Chinmay Chinara, Owen Ingraham, Pengda Xiang, Xinglei Ren, Pratusha Prasad, Bipin Kishore, Jun Xing, et al. Learning formation of physically-based face attributes. *arXiv preprint arXiv:2004.03458*, 2020.

S-L Liu, Y Liu, L-F Dong, and X Tong. Ras: A data-driven rigidity-aware skinning model for 3d facial animation. In *Computer Graphics Forum*, volume 39, pages 581–594. Wiley Online Library, 2020.

Tiantian Liu, Adam W Bargteil, James F O'Brien, and Ladislav Kavan. Fast simulation of mass-spring systems. *ACM Transactions on Graphics (TOG)*, 32(6):1–7, 2013.

Wing Kam Liu, Sukky Jun, and Yi Fei Zhang. Reproducing kernel particle methods. *International journal for numerical methods in fluids*, 20(8-9):1081–1106, 1995.

Xuecheng Liu, Tianlu Mao, Shihong Xia, Yong Yu, and Zhaoqi Wang. Facial animation by optimized blendshapes from motion capture data. *Computer Animation and Virtual Worlds*, 19(3-4):235–245, 2008.

SH Lo. Finite element mesh generation and adaptive meshing. *Progress in Structural Engineering and Materials*, 4(4):381–399, 2002.

Leon B Lucy. A numerical approach to the testing of the fission hypothesis. *The astronomical journal*, 82:1013–1024, 1977.

Wan-Chun Ma, Yi-Hua Wang, Graham Fyffe, Bing-Yu Chen, and Paul Debevec. A blendshape model that incorporates physical interaction. *Computer Animation and Virtual Worlds*, 23 (3-4):235–243, 2012.

Roy Paul Madsen. *Animated film: concepts, methods, uses*. Interland Publ., 1969.

Charles Malleson, Jean-Yves Guillemaut, and Adrian Hilton. 3d reconstruction from rgb-d data. In *RGB-D Image Analysis and Processing*, pages 87–115. Springer, 2019.

Harry McGurk and John MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588): 746–748, 1976.

Karol Miller, Grand Joldes, Dane Lance, and Adam Wittek. Total lagrangian explicit dynamics finite element algorithm for computing soft tissue deformation. *Communications in numerical methods in engineering*, 23(2):121–134, 2007.

Joseph J Monaghan. An introduction to sph. *Comput. Phys. Comm.*, 48:89–96, 1988.

Mohammad Ali Nazari. *Biomechanical Face Modeling: Control of Orofacial Gestures for Speech Production*. Theses, Université de Grenoble, September 2011. URL `https://tel.archives-ouvertes.fr/tel-00665373`.

Mohammad Ali Nazari, Pascal Perrier, and Yohan Payan. The distributed lambda (λ) model (dlm): A 3-d, finite-element muscle model based on feldman's λ model; assessment of orofacial gestures. *Journal of speech, language, and hearing research*, 2013.

Jun-yong Noh and Ulrich Neumann. A survey of facial modeling and animation techniques. Technical report, USC Technical Report, 99–705, 1998.

Kyle Olszewski, Joseph J Lim, Shunsuke Saito, and Hao Li. High-fidelity facial and speech animation for VR HMDs. *ACM Transactions on Graphics (TOG)*, 35(6):221, 2016.

Eugenio Oñate, Sergio Idelsohn, OC Zienkiewicz, and RL Taylor. A finite point method in computational mechanics. applications to convective transport and fluid flow. *International journal for numerical methods in engineering*, 39(22):3839–3866, 1996.

Agata Opalach and Steve Maddock. Disney effects using implicit surfaces. In *Proc. 5th Eurographics Workshop on Animation and Simulation (Oslo, 17-18th September, 1994)*, 1994.

Romina Palermo, Kirsty B O'Connor, Joshua M Davis, Jessica Irons, and Elinor McKone. New tests to measure individual differences in matching and labelling facial expressions of emotion, and their association with ability to recognise vocal emotions and facial identity. *PloS one*, 8(6), 2013.

Igor S Pandzic, Jörn Ostermann, and David Millen. User evaluation: Synthetic talking faces for interactive services. *The visual computer*, 15(7-8):330–340, 1999.

Frederic I Parke. A model for human faces that allows speech synchronized animation. *Proceedings of the 1st annual conference on Computer graphics and interactive techniques*, page 2, 1974. ISSN 00978493. doi: 10.1145/563182.563183.

Frederic I Parke. Parameterized models for facial animation. *IEEE computer graphics and applications*, 9(2):61–68, 1982.

Jaime Peraire, Joaquin Peiro, Luca Formaggia, Ken Morgan, and Olgierd C Zienkiewicz. Finite element euler computations in three dimensions. *International Journal for Numerical Methods in Engineering*, 26(10):2135–2159, 1988.

Stephen M Platt and Norman I Badler. Animating facial expressions. In *Proceedings of the 8th annual conference on Computer graphics and interactive techniques*, pages 245–252, 1981.

Xavier Provot et al. Deformation constraints in a mass-spring model to describe rigid cloth behaviour. In *Graphics interface*, pages 147–147. Canadian Information Processing Society, 1995.

Daniel Ram, Theodore Gast, Chenfanfu Jiang, Craig Schroeder, Alexey Stomakhin, Joseph Teran, and Pirouz Kavehpour. A material point method for viscoelastic fluids, foams and sponges. In *Proceedings of the 14th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 157–163, 2015.

Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J. Black. Generating 3D faces using convolutional mesh autoencoders. In *European Conference on Computer Vision (ECCV)*, volume Lecture Notes in Computer Science, vol 11207, pages 725–741. Springer, Cham, September 2018.

Shridhar Ravikumar, Colin Davidson, Dmitry Kit, Neill DF Campbell, Luca Benedetti, and Darren Cosker. Reading between the dots: Combining 3d markers and facs classification for high-quality blendshape facial animation. In *Graphics Interface*, pages 143–151, 2016.

Roger Blanco i Ribera, Eduard Zell, JP Lewis, Junyong Noh, and Mario Botsch. Facial retargeting with automatic range of motion alignment. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017.

Rockstar Games. L.A. Noire Trailer - Behind the Scenes, 2016. URL `https://www.youtube.com/watch?v=ZY7RYCsE9KQ`.

Antonio Rodríguez-Ferran, Agustí Pérez-Foguet, and Antonio Huerta. Arbitrary lagrangian–eulerian (ale) formulation for hyperelastoplasticity. *International Journal for Numerical Methods in Engineering*, 53(8):1831–1851, 2002.

P-Y Rohan, Claudio Lobos, Mohammad Ali Nazari, Pascal Perrier, and Yohan Payan. Finite element models of the human tongue: a mixed-element mesh approach. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 5(6):390–400, 2017.

Youcef Saad and Martin H Schultz. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on scientific and statistical computing*, 7(3):856–869, 1986.

Gaizka San-Vicente, Iker Aguinaga, and Juan Tomas Celigueta. Cubical mass-spring model design based on a tensile deformation test and nonlinear material model. *IEEE Transactions on Visualization and Computer Graphics*, 18(2):228–241, 2012. ISSN 10772626. doi: 10.1109/TVCG.2011.32.

Ferdi Scheepers, Richard E. Parent, Wayne E. Carlson, and Stephen F. May. Anatomy-Based Modeling of the Human Musculature. In *SIGGRAPH '97: ACM*, 1997.

Robert Schneiders. A grid-based algorithm for the generation of hexahedral element meshes. *Engineering with computers*, 12(3-4):168–177, 1996.

WH Schwarz. The rheology of saliva. *Journal of dental research*, 66(2_suppl):660–666, 1987.

Andrew Selle, Michael Lentine, and Ronald Fedkiw. A mass spring model for hair simulation. In *ACM SIGGRAPH 2008 papers*, pages 1–11. 2008.

H Sera, Shigeo Morishima, and D Terzopoulos. Physics-based muscle model for mouth shape control. *Robot and Human Communication, 1996., 5th IEEE International Workshop on*, pages 207–212, 1996. doi: 10.1109/ROMAN.1996.568819.

Serkis, Andy. Gollum Performance Capture, 2016. URL `http://www.serkis.com/performance-capture-gollum.htm`.

Jun'ichiro Seyama and Ruth S Nagayama. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and virtual environments*, 16 (4):337–351, 2007.

M Safdari Shadloo, G Oger, and David Le Touzé. Smoothed particle hydrodynamics method for fluid flows, towards industrial applications: Motivations, current state, and challenges. *Computers & Fluids*, 136:11–34, 2016.

Jason F Shepherd and Chris R Johnson. *Topologic and geometric constraint-based hexahedral mesh generation*, volume 68. 2007.

Hang Si. Tetgen, a delaunay-based quality tetrahedral mesh generator. *ACM Transactions on Mathematical Software (TOMS)*, 41(2):1–36, 2015.

Hang Si and Klaus Gärtner. Meshing piecewise linear complexes by constrained delaunay tetrahedralizations. In *Proceedings of the 14th international meshing roundtable*, pages 147–163. Springer, 2005.

Hang Si and A TetGen. A quality tetrahedral mesh generator and three-dimensional delaunay triangulator. *Weierstrass Institute for Applied Analysis and Stochastic, Berlin, Germany*, 81, 2006.

Eftychios Sifakis, Igor Neverov, and Ronald Fedkiw. Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM Transactions on Graphics*, 24 (3):417, 2005. ISSN 07300301. doi: 10.1145/1073204.1073208.

Eftychios Sifakis, Andrew Selle, Avram Robinson-Mosher, and Ronald Fedkiw. Simulating Speech with a Physics-Based Facial Muscle Model. *Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2006)*, pages 261–270, 2006. ISSN 1727-5288.

Ian Stavness, John E Lloyd, and Sidney Fels. Automatic prediction of tongue muscle activations using a finite element model. *Journal of biomechanics*, 45(16):2841–2848, 2012.

Ian Stavness, Mohammad Ali Nazari, Cormac Flynn, Pascal Perrier, Yohan Payan, John E Lloyd, and Sidney Fels. Coupled biomechanical modeling of the face, jaw, skull, tongue, and hyoid bone. In *3D multiscale physiological human*, pages 253–274. Springer, 2014.

Micheal B Stephenson, Scott A Canann, and Ted D Blacker. Plastering: a new approach to automated, 3d hexahedral mesh generation. Technical report, Sandia National Labs., Albuquerque, NM (United States), 1992.

Alexey Stomakhin, Craig Schroeder, Lawrence Chai, Joseph Teran, and Andrew Selle. A material point method for snow simulation. *ACM Transactions on Graphics (TOG)*, 32(4): 1–10, 2013.

Deborah Sulsky, Zhen Chen, and Howard L Schreyer. A particle method for history-dependent materials. *Computer methods in applied mechanics and engineering*, 118(1-2):179–196, 1994.

Satoshi Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing*, 30(1):32–46, 1985.

Timothy J Tautges, Ted Blacker, and Scott A Mitchell. The whisker weaving algorithm: a connectivity-based method for constructing all-hexahedral finite element meshes. *International Journal for Numerical Methods in Engineering*, 39(19):3327–3349, 1996.

Demetri Terzopoulos and Keith Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):569–579, 1993.

Matthias Teschner, Sabine Girod, and Bernd Girod. Direct computation of nonlinear soft-tissue deformation. In *VMV*, pages 383–390, 2000.

Ayush Tewari, Ohad Fried, Justus Thies, Vincent Sitzmann, Stephen Lombardi, Kalyan Sunkavalli, Ricardo Martin-Brualla, Tomas Simon, Jason Saragih, Matthias Nießner, et al. State of the art on neural rendering. *arXiv preprint arXiv:2004.03805*, 2020.

Patrick FA Van Erkel and Peter Thijssen. The first one wins: Distilling the primacy effect. *Electoral Studies*, 44:245–254, 2016.

Janet H Walker, Lee Sproull, and R Subramani. Using a human face in an interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 85–91. ACM, 1994.

Mark Warburton and Steve Maddock. Physically-based forehead animation including wrinkles. *Computer Animation and Virtual Worlds*, 26(1):55–68, 2015.

Keith Waters. A muscle model for animation three-dimensional facial expression. *SIGGRAPH Computer Graphics*, 21(4):17–24, Aug 1987. ISSN 0097-8930.

Keith Waters. Physical model of facial tissue and muscle articulation derived from computer tomography data. In *Visualization in Biomedical Computing*, pages 574–583. International Society for Optics and Photonics, 1992.

Keith Waters and Demetri Terzopoulos. The computer synthesis of expressive faces. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 335 (1273):87–93, 1992.

Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. Realtime performance-based facial animation. *ACM transactions on graphics (TOG)*, 30(4):1–10, 2011.

Lance Williams. Performance-Driven Facial Animation. 24(4), 1990.

Shibiao Xu, Guanghui Ma, Weiliang Meng, and Xiaopeng Zhang. Statistical learning based facial animation. *Journal of Zhejiang University SCIENCE C*, 14(7):542–550, 2013.

Weipeng Xu, Avishek Chatterjee, Michael Zollhöfer, Helge Rhodin, Dushyant Mehta, Hans-Peter Seidel, and Christian Theobalt. Monoperfcap: Human performance capture from monocular video. *ACM Transactions on Graphics (ToG)*, 37(2):1–15, 2018.

Xiangyu You, F Tian, and W Tang. Highly efficient facial blendshape animation with analytical dynamic deformations. *Multimedia Tools and Applications*, 78(18):25569–25590, 2019.

Yonghao Yue, Breannan Smith, Christopher Batty, Changxi Zheng, and Eitan Grinspun. Continuum foam: A material point method for shear-dependent flows. *ACM Transactions on Graphics (TOG)*, 34(5):1–20, 2015.

Yongjie Zhang, Chandrajit Bajaj, and Guoliang Xu. Surface smoothing and quality improvement of quadrilateral/hexahedral meshes with geometric flow. *Communications in Numerical Methods in Engineering*, 25(1):1–18, 2009.

Yu Zhang, Edmond C Prakash, and Eric Sung. Real-time physically-based facial expression animation using mass-spring system. In *Proceedings. Computer Graphics International 2001*, pages 347–350. IEEE, 2001.

Olek C Zienkiewicz, Robert L Taylor, and Jian Z Zhu. *The finite element method: its basis and fundamentals*. Elsevier, 2005.

Olgierd Cecil Zienkiewicz, Robert Leroy Taylor, Robert Leroy Taylor, and Robert Lee Taylor. *The finite element method: solid mechanics*, volume 2. Butterworth-heinemann, 2000.

Michael Zollhöfer, Justus Thies, Pablo Garrido, Derek Bradley, Thabo Beeler, Patrick Pérez, Marc Stamminger, Matthias Nießner, and Christian Theobalt. State of the art on monocular 3d face reconstruction, tracking, and applications. In *Computer Graphics Forum*, volume 37, pages 523–550. Wiley Online Library, 2018.

Appendix A

Perceptual Study Results

This appendix gives a full listing of the results of the perceptual study. For convenience, the videos used in each question are also stated.

A.1 Questions

Question	1	2	3	4	5	6
Videos	os0,os4	os0,os7	os0,os10	os4,os7	os4,os10	os7,os10
Question	7	8	9	10	11	12
Videos	of0,of4	of0,of7	of0,of10	of4,of7	of4,of10	of7,of10
Question	13	14	15	16	17	18
Videos	m0,m4	m0,m7	m0,m10	m4,m7	m4,m10	m7,m10
Question	19	20	21	22	23	24
Videos	ma0,ma4	ma0,ma7	ma0,ma10	ma4,ma7	ma4,ma10	ma7,ma10
Question	25	26	27	28	29	30
Videos	mu0,mu4	mu0,mu7	mu0,mu10	mu4,mu7	mu4,mu10	mu7,mu10
Question	31	32	33	34	35	36
Videos	mu0a,mu4a	mu0a,mu7a	mu0a,mu10a	mu4a,mu7a	mu4a,mu10a	mu7a,mu10a
Question	37	38	39	40	41	42
Videos	p0,p4	p0,p7	p0,p10	p4,p7	p4,p10	p7,p10
Question	43	44	45	46	47	48
Videos	p0a,p4a	p0a,p7a	p0a,p10a	p4a,p7a	p4a,p10a	p7a,p10a
Question	49	50	51	52	53	54
Videos	pu0,pu4	pu0,pu7	pu0,pu10	pu4,pu7	pu4,pu10	pu7,pu10
Question	55	56	57	58	59	60
Videos	pu0a,pu4a	pu0a,pu7a	pu0a,pu10a	pu4a,pu7a	pu4a,pu10a	pu7a,pu10a

A.2 Results

Due to a technical issue, participants 1 and 2 only answered 36 questions each. All other participants answered the full set of 60 questions.

Participant 1

Question	3	11	22	24	10	5	19	18	20	21
Answer	os0	of10	m4a	m10a	of4	os4	m4a	m7	m0a	m0a
Question	35	34	26	8	1	16	4	7	33	2
Answer	mu4a	mu4a	mu0	of7	os4	m4	os4	of4	mu0a	os7
Question	25	9	6	23	14	15	32	13	29	27
Answer	mu4	of10	os7	m10a	m0	m0	mu7a	m4	mu10	mu10
Question	36	17	31	12	28	30				
Answer	mu7a	m4	mu4a	of7	mu7	mu7				

Participant 2

Question	13	27	33	12	2	26	8	16	34	15
Answer	m0	mu0	mu10a	of10	os0	mu7	of0	m4	mu7a	m0
Question	11	6	5	36	9	30	3	29	14	21
Answer	of4	os7	os4	mu7a	of0	mu7	os0	mu4	m7	m10a
Question	32	18	25	1	23	17	4	24	31	19
Answer	mu7a	m7	mu4	os4	m10a	m10	os4	m10a	mu0a	m0a
Question	10	7	28	20	35	22				
Answer	of7	of4	mu4	m0a	mu4a	m7a				

Participant 3

Question	32	58	22	55	42	33	6	34	7	43
Answer	mu7a	pu4a	m7a	pu4a	p10	mu0a	os7	mu7a	of4	p4a
Question	4	12	37	51	41	8	10	25	2	17
Answer	os7	of7	p4	pu10	p10	of7	of7	mu0	os7	m4
Question	46	39	30	3	14	45	59	29	44	19
Answer	p4a	p10	mu10	os10	m7	p10a	pu10a	mu10	p7a	m4a
Question	56	54	47	52	24	23	50	11	26	31
Answer	pu7a	pu10	p10a	pu7	m10a	m10a	pu7	of10	mu7	mu4a
Question	20	18	35	9	1	53	38	21	49	15
Answer	m7a	m7	mu10a	of10	os4	pu10	p7	m10a	pu4	m0
Question	13	28	57	36	27	48	60	40	5	16
Answer	m0	mu7	pu10a	mu10a	mu10	p7a	pu10a	p4	os4	m7

Participant 4

Question	3	60	8	15	29	53	9	19	36	40
Answer	os10	pu10a	of0	m0	mu4	pu4	of0	m0a	mu7a	p7
Question	23	41	54	35	24	42	10	11	4	50
Answer	m4a	p10	pu7	mu10a	m10a	p7	of7	of10	os7	pu0
Question	5	25	33	55	46	14	37	13	52	48
Answer	os4	mu0	mu10a	pu0a	p4a	m7	p0	m4	pu4	p7a
Question	16	57	1	20	45	2	49	32	59	27
Answer	m7	pu0a	os4	m0a	p0a	os7	pu4	mu7a	pu4a	mu10
Question	18	47	31	30	51	7	12	44	58	22
Answer	m7	p4a	mu0a	mu10	pu10	of0	of10	p0a	pu4a	m7a
Question	34	26	43	28	39	21	38	17	6	56
Answer	mu4a	mu7	p4a	mu4	p0	m10a	p7	m10	os7	pu0a

Participant 5

Question	31	52	44	23	11	6	28	58	34	40
Answer	mu4a	pu7	p0a	m4a	of4	os7	mu7	pu4a	mu7a	p4
Question	48	36	4	59	50	14	37	29	41	22
Answer	p10a	mu7a	os7	pu4a	pu7	m7	p4	mu4	p4	m4a
Question	54	47	2	53	3	51	55	20	30	13
Answer	pu10	p10a	os7	pu4	os10	pu10	pu4a	m0a	mu10	m0
Question	7	15	21	1	42	9	46	57	35	60
Answer	of0	m10	m0a	os4	p7	of10	p4a	pu10a	mu4a	pu10a
Question	10	26	39	18	43	8	32	45	49	17
Answer	of7	mu0	p10	m7	p0a	of7	mu0a	p0a	pu0	m4
Question	38	25	19	12	56	24	16	27	33	5
Answer	p7	mu4	m4a	of10	pu7a	m7a	m7	mu0	mu0a	os10

Participant 6

Question	35	51	39	55	58	5	31	17	49	20
Answer	mu10a	pu10	p0	pu4a	pu7a	os4	mu4a	m10	pu4	m7a
Question	13	33	38	19	16	7	52	4	21	56
Answer	m4	mu0a	p7	m0a	m4	of4	pu7	os4	m10a	pu7a
Question	22	8	42	53	41	23	34	50	10	60
Answer	m7a	of0	p7	pu10	p10	m4a	mu4a	pu7	of7	pu10a
Question	27	37	47	44	32	24	30	3	29	6
Answer	mu10	p4	p4a	p0a	mu0a	m10a	mu10	os10	mu4	os7
Question	15	1	48	43	59	2	45	54	26	40
Answer	m0	os0	p7a	p4a	pu4a	os7	p0a	pu7	mu0	p4
Question	18	14	46	12	28	11	36	57	25	9
Answer	m10	m0	p7a	of10	mu7	of10	mu7a	pu10a	mu4	of10

Participant 7

Question	31	19	13	1	20	27	14	50	32	36
Answer	mu0a	m0a	m4	os4	m7a	mu0	m7	pu0	mu0a	mu10a
Question	45	18	26	39	59	38	12	40	55	57
Answer	p10a	m7	mu0	p10	pu4a	p0	of7	p7	pu4a	pu0a
Question	25	48	58	23	44	43	17	30	8	51
Answer	mu0	p7a	pu4a	m4a	p7a	p0a	m4	mu7	of7	pu0
Question	33	42	5	37	15	47	54	53	52	60
Answer	mu0a	p7	os4	p0	m0	p4a	pu7	pu4	pu4	pu7a
Question	16	10	9	35	29	46	11	2	56	49
Answer	m7	of7	of10	mu4a	mu4	p7a	of10	os7	pu7a	pu0
Question	34	41	21	24	4	3	7	28	6	22
Answer	mu7a	p4	m0a	m7a	os4	os0	of4	mu4	os7	m7a

Participant 8

Question	4	27	7	49	14	43	1	12	57	16
Answer	os7	mu10	of0	pu0	m0	p0a	os0	of7	pu0a	m7
Question	51	5	21	13	29	2	40	55	23	48
Answer	pu10	os4	m10a	m4	mu10	os7	p4	pu0a	m10a	p7a
Question	17	32	39	20	33	41	34	11	56	46
Answer	m4	mu0a	p10	m7a	mu0a	p4	mu4a	of10	pu0a	p4a
Question	10	38	26	45	28	22	24	52	25	53
Answer	of4	p0	mu7	p0a	mu7	m4a	m7a	pu7	mu4	pu4
Question	60	58	42	6	30	36	15	59	47	9
Answer	pu7a	pu4a	p7	os7	mu7	mu7a	m0	pu4a	p4a	of10
Question	8	44	19	54	3	31	35	50	37	18
Answer	of7	p7a	m4a	pu7	os0	mu0a	mu10a	pu0	p4	m10

Participant 9

Question	54	53	40	52	24	29	34	56	4	18
Answer	pu10	pu10	p7	pu4	m7a	mu4	mu7a	pu0a	os7	m7
Question	48	60	13	12	22	35	41	51	59	30
Answer	p10a	pu7a	m4	of10	m7a	mu10a	p4	pu0	pu4a	mu7
Question	36	57	20	42	50	11	33	1	27	55
Answer	mu7a	pu0a	m7a	p7	pu0	of4	mu0a	os0	mu0	pu0a
Question	25	5	58	23	9	21	44	6	37	46
Answer	mu4	os10	pu7a	m4a	of10	m10a	p7a	os7	p0	p7a
Question	49	2	15	43	26	32	45	17	7	16
Answer	pu4	os7	m10	p0a	mu7	mu7a	p0a	m10	of4	m4
Question	14	28	19	38	39	10	3	47	8	31
Answer	m0	mu7	m0a	p0	p0	of7	os10	p4a	of7	mu4a

Participant 10

Question	32	47	19	21	57	13	60	49	17	16
Answer	mu0a	p4a	m4a	m10a	pu10a	m0	pu7a	pu0	m10	m7
Question	46	5	27	36	58	10	29	55	59	20
Answer	p7a	os10	mu0	mu10a	pu4a	of4	mu4	pu4a	pu4a	m7a
Question	11	50	26	7	4	2	22	41	56	39
Answer	of10	pu0	mu0	of0	os4	os7	m7a	p4	pu0a	p10
Question	30	45	53	8	38	51	9	44	6	54
Answer	mu10	p0a	pu10	of7	p7	pu10	of10	p0a	os10	pu7
Question	43	31	18	37	35	34	25	48	15	1
Answer	p0a	mu0a	m7	p4	mu4a	mu7a	mu4	p10a	m10	os4
Question	14	24	12	3	33	28	52	42	40	23
Answer	m0	m7a	of10	os10	mu10a	mu7	pu7	p7	p4	m10a

Participant 11

Question	48	16	46	9	12	53	25	8	32	44
Answer	p10a	m7	p4a	of10	of10	pu4	mu0	of7	mu0a	p0a
Question	41	7	60	14	23	18	43	45	56	20
Answer	p4	of0	pu7a	m7	m10a	m7	p0a	p10a	pu0a	m7a
Question	2	42	19	55	29	59	58	28	51	36
Answer	os7	p10	m4a	pu4a	mu4	pu4a	pu4a	mu7	pu10	mu7a
Question	5	47	54	3	57	27	39	37	35	24
Answer	os10	p4a	pu10	os10	pu0a	mu10	p10	p4	mu10a	m7a
Question	22	31	6	34	52	1	13	40	49	4
Answer	m7a	mu4a	os7	mu4a	pu7	os4	m0	p7	pu0	os7
Question	38	11	26	10	21	17	15	30	33	50
Answer	p0	of10	mu0	of7	m10a	m10	m10	mu10	mu0a	pu0

Participant 12

Question	47	32	7	30	46	29	33	4	12	52
Answer	p4a	mu0a	of0	mu10	p7a	mu4	mu10a	os7	of7	pu4
Question	56	34	22	42	5	17	8	59	2	15
Answer	pu7a	mu4a	m7a	p7	os4	m4	of7	pu10a	os7	m10
Question	35	20	14	1	54	23	3	28	26	43
Answer	mu10a	m7a	m7	os4	pu10	m10a	os10	mu7	mu7	p4a
Question	41	38	13	58	37	21	55	16	60	19
Answer	p4	p7	m4	pu7a	p4	m10a	pu4a	m4	pu10a	m0a
Question	9	40	24	50	45	6	10	36	18	31
Answer	of10	p4	m7a	pu7	p10a	os7	of7	mu7a	m7	mu0a
Question	11	49	57	39	53	25	44	51	27	48
Answer	of10	pu0	pu10a	p10	pu10	mu4	p7a	pu10	mu10	p10a

Participant 13

Question	8	11	35	38	13	5	51	32	27	10
Answer	of0	of4	mu10a	p0	m0	os4	pu10	mu7a	mu10	of7
Question	47	36	17	37	60	18	50	43	53	46
Answer	p4a	mu10a	m10	p4	pu7a	m7	pu7	p0a	pu10	p4a
Question	9	22	52	29	59	39	33	24	14	55
Answer	of10	m7a	pu7	mu4	pu10a	p10	mu10a	m10a	m7	pu4a
Question	21	25	40	12	44	16	30	7	49	28
Answer	m10a	mu4	p4	of7	p0a	m7	mu7	of4	pu0	mu7
Question	57	4	2	42	56	19	20	45	34	23
Answer	pu0a	os4	os7	p7	pu7a	m4a	m0a	p10a	mu7a	m10a
Question	26	58	1	6	15	48	41	31	3	54
Answer	mu7	pu4a	os4	os7	m10	p10a	p4	mu0a	os10	pu7

Participant 14

Question	4	56	28	36	40	31	59	45	14	22
Answer	os7	pu7a	mu7	mu7a	p4	mu0a	pu10a	p0a	m0	m7a
Question	16	19	1	25	8	44	35	51	29	32
Answer	m7	m0a	os4	mu4	of7	p7a	mu10a	pu10	mu4	mu0a
Question	57	13	38	24	60	21	49	30	42	46
Answer	pu0a	m4	p7	m7a	pu10a	m10a	pu0	mu7	p7	p7a
Question	5	7	9	39	26	2	27	41	33	55
Answer	os4	of4	of10	p10	mu0	os0	mu10	p10	mu0a	pu4a
Question	18	34	17	3	15	37	6	58	20	12
Answer	m7	mu4a	m10	os0	m10	p4	os7	pu7a	m7a	of7
Question	43	52	50	23	10	11	54	47	48	53
Answer	p0a	pu7	pu7	m10a	of4	of10	pu10	p10a	p10a	pu10

Participant 15

Question	49	16	59	19	25	30	4	8	52	40
Answer	pu4	m7	pu4a	m4a	mu0	mu10	os7	of7	pu7	p7
Question	46	21	37	9	60	41	57	2	18	50
Answer	p7a	m10a	p4	of10	pu10a	p4	pu0a	os7	m7	pu0
Question	3	11	26	42	22	38	34	28	29	14
Answer	os10	of4	mu0	p7	m7a	p0	mu4a	mu7	mu4	m0
Question	54	39	43	20	17	58	33	27	55	15
Answer	pu7	p10	p4a	m0a	m4	pu4a	mu10a	mu10	pu4a	m10
Question	24	31	44	53	1	6	5	23	32	35
Answer	m7a	mu4a	p7a	pu4	os4	os7	os10	m10a	mu0a	mu10a
Question	45	12	36	56	10	47	13	7	51	48
Answer	p0a	of10	mu7a	pu0a	of4	p4a	m0	of4	pu10	p7a

Participant 16

Question	6	35	33	29	5	34	56	10	22	45
Answer	os7	mu10a	mu10a	mu4	os4	mu4a	pu7a	of7	m7a	p10a
Question	21	39	27	25	31	8	1	18	26	55
Answer	m10a	p0	mu0	mu4	mu4a	of7	os4	m7	mu0	pu4a
Question	28	46	32	47	9	49	20	4	48	16
Answer	mu4	p7a	mu7a	p4a	of10	pu4	m7a	os7	p10a	m7
Question	36	42	58	30	38	43	59	50	44	54
Answer	mu7a	p10	pu7a	mu7	p0	p0a	pu4a	pu0	p0a	pu10
Question	7	13	15	24	2	17	23	60	14	41
Answer	of4	m0	m10	m7a	os7	m10	m10a	pu7a	m0	p4
Question	3	37	12	40	51	52	11	57	19	53
Answer	os0	p4	of10	p4	pu10	pu7	of10	pu0a	m0a	pu10

Participant 17

Question	4	28	7	60	5	53	29	34	3	23
Answer	os7	mu7	of4	pu10a	os10	pu4	mu4	mu7a	os10	m10a
Question	52	51	57	6	18	20	58	1	35	48
Answer	pu7	pu10	pu0a	os7	m7	m0a	pu7a	os4	mu4a	p10a
Question	38	24	39	32	22	27	41	49	44	12
Answer	p7	m10a	p0	mu0a	m7a	mu10	p4	pu0	p7a	of10
Question	2	19	30	56	21	13	14	46	54	26
Answer	os7	m0a	mu10	pu0a	m10a	m0	m7	p7a	pu10	mu7
Question	36	16	25	59	40	9	17	45	33	47
Answer	mu7a	m4	mu4	pu4a	p4	of10	m10	p0a	mu0a	p4a
Question	42	50	8	43	10	37	55	31	11	15
Answer	p7	pu7	of7	p4a	of7	p4	pu4a	mu4a	of4	m0

Participant 18

Question	7	23	48	30	18	4	47	49	20	2
Answer	of4	m4a	p10a	mu10	m7	os4	p4a	pu0	m7a	os7
Question	54	55	15	52	25	43	32	6	50	42
Answer	pu10	pu4a	m10	pu7	mu4	p4a	mu0a	os7	pu7	p7
Question	3	16	8	58	31	41	45	56	11	22
Answer	os10	m4	of7	pu4a	mu4a	p4	p0a	pu0a	of10	m4a
Question	28	37	26	33	14	10	39	17	34	21
Answer	mu7	p4	mu7	mu10a	m0	of7	p10	m10	mu4a	m10a
Question	9	40	27	24	36	38	53	5	19	29
Answer	of10	p4	mu10	m10a	mu7a	p7	pu10	os4	m0a	mu4
Question	46	51	60	35	13	1	12	57	59	44
Answer	p7a	pu10	pu10a	mu10a	m4	os0	of7	pu0a	pu4a	p7a

Participant 19

Question	44	4	56	60	55	39	27	9	45	50
Answer	p0a	os4	pu7a	pu10a	pu0a	p10	mu0	of0	p0a	pu0
Question	16	51	26	37	18	42	23	2	33	43
Answer	m7	pu10	mu0	p4	m7	p7	m4a	os7	mu10a	p4a
Question	53	29	7	12	58	28	48	32	5	59
Answer	pu4	mu10	of4	of7	pu4a	mu7	p10a	mu7a	os10	pu10a
Question	46	57	10	49	11	15	35	38	52	21
Answer	p7a	pu0a	of4	pu4	of4	m0	mu10a	p7	pu4	m10a
Question	8	3	22	24	25	19	41	40	13	20
Answer	of7	os0	m7a	m7a	mu0	m0a	p4	p4	m0	m7a
Question	30	47	54	6	36	17	1	31	34	14
Answer	mu10	p4a	pu7	os10	mu7a	m10	os0	mu4a	mu4a	m7

Participant 20

Question	48	14	27	10	37	20	42	28	39	29
Answer	p7a	m0	mu10	of7	p4	m7a	p7	mu7	p10	mu4
Question	34	55	9	17	47	59	30	24	45	13
Answer	mu7a	pu0a	of10	m10	p4a	pu4a	mu10	m7a	p10a	m4
Question	57	15	32	41	25	1	50	56	51	23
Answer	pu10a	m0	mu0a	p4	mu0	os0	pu7	pu7a	pu10	m4a
Question	11	7	8	12	35	46	43	60	52	22
Answer	of10	of4	of7	of10	mu4a	p4a	p0a	pu10a	pu4	m7a
Question	54	26	38	58	6	18	5	33	44	36
Answer	pu7	mu0	p0	pu4a	os7	m7	os4	mu0a	p0a	mu7a
Question	16	31	49	19	2	40	3	21	53	4
Answer	m4	mu4a	pu4	m0a	os7	p4	os0	m0a	pu4	os7

Appendix B

Ethics Applications and Approval

Application 023822

Section A: Applicant details

Date application started:
Wed 21 November 2018 at 09:05

First name:
Matthew

Last name:
Leach

Email:
mileach1@sheffield.ac.uk

Programme name:
PhD Computer Science

Module name:
PhD
Last updated:
23/01/2019

Department:
Computer Science

Applying as:
Postgraduate research

Research project title:
Sticky Lips (Video capture)

Has your research project undergone academic review, in accordance with the appropriate process?
Yes

Similar applications:
N/A

Section B: Basic information

Supervisor

Name	Email
Steve Maddock	s.maddock@sheffield.ac.uk

Proposed project duration

Start date (of data collection):
Thu 13 December 2018

Anticipated end date (of project)
Fri 28 February 2020

3: Project code (where applicable)

Project code
- not entered -

Suitability

Takes place outside UK?
No

Involves NHS?
No

Health and/or social care human-interventional study?
No

ESRC funded?
No

Likely to lead to publication in a peer-reviewed journal?
Yes

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial or a medical device study?
No

Involves social care services provided by a local authority?
No

Involves adults who lack the capacity to consent?
No

Involves research on groups that are on the Home Office list of 'Proscribed terrorist groups or organisations?
No

Indicators of risk

Involves potentially vulnerable participants?
No

Involves potentially highly sensitive topics?
No

Section C: Summary of research

1. Aims & Objectives

This research aims to enhance the computer graphics animation of the human mouth through physically based modelling of the ‘sticky lip’ effect. This describes the phenomenon where the lips become slightly stuck together as a result of dried or drying saliva. As the lips are drawn apart, the saliva bond causes them to remain stuck together momentarily, before the bond breaks and the lips separate properly. This effect is being studied as the mouth is currently the least convincing component of facial animation in films and games. Humans are particularly perceptive of mouth movements and it is believed modelling such effects will improve the believability of the animation. In order to better understand the dynamics of the effect and provide source material for evaluation, we are seeking approval to capture video of the mouth performing a set of movements with varying levels of lip stickiness.

2. Methodology

The objective of this work is to capture the effects caused by stickiness between the human lips. To capture this, participants will be recorded using a video camera speaking a set of words. Participants will also be asked to recreate some expressions and mouth movements shown to them in video on a laptop, desktop computer or tablet. These speech segments and movements will be recorded using a high speed camera targeting the participant's mouth area. Participants will be asked to repeat these speech segments/expressions a number of times. Before the first recording, the participants will be asked to lick their lips to wet them and then recordings will be repeated at set time intervals. The camera will be affixed to a tripod and the participant will be seated. The room will be darkened and the participant's face will be lit by a light from either side to provide contrast between the outside and inside of the mouth, as this is necessary

for good evaluation of the data.

3. Personal Safety

Have you completed your departmental risk assessment procedures, if appropriate?

Yes

Raises personal safety issues?

No

Participants will be comfortably seated for the duration of the recording. There is no dangerous equipment involved in the process. If participants eyes or mouth become uncomfortable, they will be free to take a break or leave at any time.

Section D: About the participants

1. Potential Participants

Participants will be people who have a high colour contrast between their natural lip colour and the surrounding skin as the evaluation technique which will be used depends on computationally detecting differences in colour between the lips and their surroundings. This is more effective when there is a greater contrast in colour. This must be the participants natural lip colour, as lipstick etc. may affect the results.

2. Recruiting Potential Participants

Potential participants for the video recordings will include friends and acquaintances with a high colour contrast between their lips and surrounding skin. They will be asked and invited through private messages, emails and social media. The message will include a summary of the study, what the data collection will involve and what they will be asked to do (i.e. be recorded while they read a list of words and form certain facial expressions). They will be required to read the information sheet and sign the consent form before recording commences.

2.1. Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by CiCS? No

- not entered -

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

Prior to the video recording, the participants will be provided with an information sheet along with the consent form. The information sheet notes that the session is expected to take 30 - 40 minutes. They will be given the time to read the information sheet and to ask any questions. Once potential participants are satisfied with their understanding of the research and are willing to participate, they will be asked to complete and sign the consent form. The process will also be explained to them in full detail and discussed with them before beginning.

4. Payment

Will financial/in kind payments be offered to participants? No

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

Participants may find it slightly uncomfortable refraining from licking/wetting their lips.

How will this be managed to ensure appropriate protection and well-being of the participants?

Consent will be taken from the participants and the process will be explained beforehand so that they know what to expect. The participants will be shown a demonstration video to show the process in advance. They will also be allowed to pause anytime should they feel the need to take a break.

Section E: About the data

1. Data Processing

Will you be processing (i.e. collecting, recording, storing, or otherwise using) personal data as part of this project? (Personal data is any information relating to an identified or identifiable living person).
Yes

Which organisation(s) will act as Data Controller?

University of Sheffield only

2. Legal basis for processing of personal data

The University considers that for the vast majority of research, 'a task in the public interest' (6(1)(e)) will be the most appropriate legal basis. If, following discussion with the UREC, you wish to use an alternative legal basis, please provide details of the legal basis, and the reasons for applying it, below:

- not entered -

Will you be processing (i.e. collecting, recording, storing, or otherwise using) 'Special Category' personal data?
No

3. Data Confidentiality

What measures will be put in place to ensure confidentiality of personal data, where appropriate?

No names or identifiable personal information will be collected. Each participant will be assigned an arbitrary ID. A list of names and associated IDs will be securely stored and only accessed for the purposes of deleting data if a participant requests this. All the data collected will be stored on a password-protected Google Drive folder, which is only accessible by the student researcher and their supervisors. Videos will be immediately cropped to only show the lower part of the face, showing only showing the bottom of the nose, mouth, chin and face outline, ensuring that the person stays anonymous. The data will be deleted when the project ends unless the participant agrees on the consent form to have their data saved for future research. If participants consent, data will also be stored within the ORDA system at the University of Sheffield.

4. Data Storage and Security

In general terms, who will have access to the data generated at each stage of the research, and in what form

The student researcher and supervisor, Dr Steve Maddock are the only people who will have access to the password protected Google drive folder on which the data will be stored. This password protected folder will contain the cropped versions of the recorded videos and the completed consent forms. Cropped versions of the videos will only show the lower half of the face of the participants who give permission to have their data saved. The original recorded videos showing the entire face will be deleted as soon as the cropped versions have been generated.

What steps will be taken to ensure the security of data processed during the project, including any identifiable personal data, other than those already described earlier in this form?

Participants will not be asked to give any further identifiable personal data.

Will all identifiable personal data be destroyed once the project has ended?
Yes

Please outline when this will take place (this should take into account regulatory and funder requirements).

Data will be destroyed upon completion of the PhD unless the participant agreed to have their data saved for future research. Participants may request at any time that their data be deleted. This includes after project completion if they have agreed to have their data on ORDA.

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project?
Yes

Document 1053490 (Version 3) All versions

Consent forms relevant to project?
Yes

Document 1053424 (Version 2) All versions

Additional Documentation

External Documentation

- not entered -

Section G: Declaration

Signed by:
Matthew Leach
Date signed:
Tue 22 January 2019 at 17:23

Offical notes

- not entered -

Matthew Leach
Registration number: 150146922
Computer Science
Programme: PhD Computer Science

Dear Matthew

PROJECT TITLE: Sticky Lips (Video capture)
APPLICATION: Reference Number 023822

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 23/01/2019 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 023822 (form submission date: 22/01/2019); (expected project end date: 28/02/2020).
- Participant information sheet 1053490 version 3 (22/01/2019).
- Participant consent form 1053424 version 2 (22/01/2019).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Alice Tucker
Ethics Administrator
Computer Science

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Application 030947

Section A: Applicant details

Date application started:
Mon 16 September 2019 at 16:13

First name:
Matthew

Last name:
Leach

Email:
mileach1@sheffield.ac.uk

Programme name:
PhD Computer Science

Module name:
PhD
Last updated:
22/11/2019

Department:
Computer Science

Applying as:
Postgraduate research

Research project title:
Sticky Lips User Evaluation

Has your research project undergone academic review, in accordance with the appropriate process?
Yes

Similar applications:
- not entered -

Section B: Basic information

Supervisor

Name	Email
Steve Maddock	s.maddock@sheffield.ac.uk

Proposed project duration

Start date (of data collection):
Wed 6 November 2019

Anticipated end date (of project)
Sat 29 February 2020

3: Project code (where applicable)

Project code
- not entered -

Suitability

Takes place outside UK?
No

Involves NHS?
No

Health and/or social care human-interventional study?
No

ESRC funded?
No

Likely to lead to publication in a peer-reviewed journal?
Yes

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial or a medical device study?
No

Involves social care services provided by a local authority?
No

Is social care research requiring review via the University Research Ethics Procedure
No

Involves adults who lack the capacity to consent?
No

Involves research on groups that are on the Home Office list of 'Proscribed terrorist groups or organisations?
No

Indicators of risk

Involves potentially vulnerable participants?
No
Involves potentially highly sensitive topics?
No

Section C: Summary of research

1. Aims & Objectives

This research aims to conduct a user evaluation of human mouth animations. The animations will differ in the strength of the stickiness of the sticky-lip effect, which has been the focus of my PhD work. This study will aim to establish whether there is a perceptible improvement in realism due to the inclusion of the sticky lip effect in animations.

2. Methodology

Each participant will be shown a short series of videos. Participants will be shown pairs of videos of computer-produced animations of the mouth. The participants will be asked to select the more realistic animation in an A/B test. The two videos in each pair will show the same mouth movement/action. The stickiness level will range from none to high using parameters in the algorithm. Five mouth movements will be used: the mouth opening, producing an 'm' sound, a 'p' sound and speaking the words 'mummy' and 'puppy'. Pairs will be shown more than once to increase the reliability of the tests.

In order to avoid first-answer bias, the order of the two animations will be randomised. The order in which the animations are shown to participants will also be randomised. The results of the study will be analysed using standard statistical measures.

3. Personal Safety

Have you completed your departmental risk assessment procedures, if appropriate?

Yes

Raises personal safety issues?

No

There are no dangerous activities or environments involved in this research. There is also no dangerous equipment involved.

Section D: About the participants

1. Potential Participants

There are no special requirements for the participants other than being able to clearly see the videos.

2. Recruiting Potential Participants

There will be two sets of participants. Set 1 will include friends, acquaintances and students at the university. Friends and acquaintances will be asked and invited through private messages, emails and social media. The message will include a summary of the study, what the process will involve and what they will be asked to do (i.e. view sets of videos and choose the most realistic). Students will be invited via the CiCS mailing list. Set 1 will conduct the test in person. Set 2 will consist of internet users who will take an online version of the test. This will be advertised on social media websites. All participants will be required to read the information sheet and sign the consent form before the study commences.

2.1. Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by CiCS? Yes

Other methods will also be employed, however we would like to reach the greatest number of people possible to strengthen the data.

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

The participants of set 1 will be verbally briefed before beginning the process. In addition, they will be provided with an information sheet and consent form explaining to them how their data will be used. Participants in set 2 will be provided with an electronic copy of the information sheet and consent form which must be accepted before the process can be begun.

4. Payment

Will financial/in kind payments be offered to participants? Yes

The study will take 45-60 minutes to complete. A cash payment of £10 will be made to participants upon completion of the experiment. The payment will not apply to users of the online study.

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

There is no probable potential for harm to the participants.

How will this be managed to ensure appropriate protection and well-being of the participants?

Participants will be informed they may leave at any time during the process.

Section E: About the data

1. Data Processing

Will you be processing (i.e. collecting, recording, storing, or otherwise using) personal data as part of this project? (Personal data

is any information relating to an identified or identifiable living person).
No

Please outline how your data will be managed and stored securely, in line with good practice and relevant funder requirements

Participants will be assigned an ID number which will be used for all future processing of the data. The only personally identifiable data will be the participants' names in the consent forms, which will be stored securely. This data will be analysed for use in the researcher's PhD and possibly for subsequent publications. The data will be destroyed within 3 years of the completion of the researcher's PhD.

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project?
Yes

Document 1070759 (Version 4) All versions

Consent forms relevant to project?
Yes

Document 1070760 (Version 3) All versions

Additional Documentation

External Documentation

- not entered -

Section G: Declaration

Signed by:
Matthew Leach
Date signed:
Thu 7 November 2019 at 12:35

Offical notes

- not entered -

Matthew Leach
Registration number: 150146922
Computer Science
Programme: PhD Computer Science

Dear Matthew

PROJECT TITLE: Sticky Lips User Evaluation
APPLICATION: Reference Number 030947

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 22/11/2019 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 030947 (form submission date: 07/11/2019); (expected project end date: 29/02/2020).
- Participant information sheet 1070759 version 4 (07/11/2019).
- Participant consent form 1070760 version 3 (05/11/2019).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Com Ethics
Ethics Administrator
Computer Science

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.