Dynamic Real-Time 3-Dimensional Model of the Tongue's Motion

BY

DAVIDE LACCA
B.S., Politecnico di Milano, Milan, Italy, 2018

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Science in Bioengineering
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Hananeh Esmailbeigi, Chair and Advisor

Muge Karaman

Cristian Luciano

Andrea Aliverti, Politecnico di Milano

ACKNOWLEDGMENTS

First, I want to thank my beautiful family, Mamma, Papà and Chiara, who gave me the opportunity to live this great experience. During these months far from home I always felt your love and support even if we were distant. Thank you for always believing in me.

A special thank you goes to my advisor, Professor Hananeh Esmailbeigi, who always helped me and pushed me to give my best every day. Even if I did not talk much, I really appreciated working with you. I learnt a lot during this time together. I am also very grateful for all the support you gave to me and to all the other guys of the WTSE Lab (besides the academic "resources").

I also want to thank Professor Muge Karaman for helping me with the acquisition of MR images. Her contribution has been essential to the outcome of my work.

Another felt thank you goes to Professor Cristian Luciano, who gave me really helpful tips about how to improve my project.

Thank you to all the great friends I met during this experience. You have been like a big family to me and I will never forget all the good times we had together.

Thank you, Camilla. Thank you for always being there for me. I know that these months have been hard. We went through a lot and I have not always been good to you. Thank you for believing in us and in our future together. I am sure that the best is yet to come. I love you.

DL

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE **PAGE**

LIST OF FIGURES (continued)

FIGURE **PAGE**

LIST OF ABBREVIATIONS

3D	Three-Dimensional
cm	Centimeters
cm^3	Cubic Centimeters
CN	Cranial Nerve
EMA	Electromagnetic Articulography
EPG	Electropalatography
mm	Millimeters
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
ms	Milliseconds
UI	User Interface
US	Ultrasound

SUMMARY

Patients undergoing traditional speech therapy often have difficulty grasping the therapist's audible cues for the correct intended movement of the tongue. This lack of clarity is attributed to the absence of visual biofeedback. Therefore, a model to visualize the tongue's real-time motion could assist in providing the necessary biofeedback required in speeding up the process of the therapy. Hence, we have developed a dynamic 3-Dimensional (3D) model of the human tongue using the Unity software platform (Unity Technologies, San Francisco). This model is a standalone platform that could also be used in conjunction with the discreet oral wearable device developed by our team. The model was developed from static Magnetic Resonance (MR) images of the tongue, collected at UIC Center for Magnetic Resonance Research. 34 series of MR images of the static position of the tongue during various static tasks were captured. In order to create the 3D model, first the MR images were segmented. Then, a mesh of the tongue was developed and optimized for all the shapes that the tongue could assume. The meshes were then exported to Unity and were animated. The model of the motion of the tongue was enabled by moving a set of splines attached to the mesh. Furthermore, a User Interface (UI) was developed to interact with the model. The resulted model allows for reproduction of movements associated with 24 English phonemes and the words composed of these 24 phonemes. The model could interact in real-time with the oral wearable device in order to create live models of the motion. The model is also capable of reproducing the tongue's motion using offline data from the wearable device.

CHAPTER 1

INTRODUCTION

Communication is a fundamental part of our daily lives and constitutes the basis of social relationships. We usually associate communication with speech and language, as we consider the auditory channel as the primary way of interaction. However, the visual components of speech, such as facial expressions or body gestures, play a very important role in comprehending each other. This is true especially for people who have any speech-related disorders or hearing impairments. Patients undergoing traditional speech therapy often have difficulty grasping the therapist's audible cues for the correct intended tongue's movement [1]. This lack of clarity is attributed to the absence of visual biofeedback. Therefore, the goal of this thesis is to develop a realistic 3D model of the human tongue's motion in order to provide an augmented visual biofeedback. We believe that the 3D model, alongside with the wearable oral device that we developed, could activate neuromodulation and improve therapy outcomes. Our hypothesis is that by providing this kind of biofeedback, patients will learn faster and the rehabilitation process will be more effective. The 3D model of the tongue can be useful for several applications. First, it allows to visualize how the tongue moves during the articulation of different English phonemes and words. The model, in conjunction with the wearable device, could also be used to show the tongue's motion in real time. This model could be useful as well for free exploration and error augmentation studies.

1.1 Thesis Structure

Chapter 2 presents an overview of the anatomy of the tongue and its functions, the main technologies to quantify the tongue's motion, the importance of visual feedback in speech therapy and existing 3D models of the human tongue and their applications.

Chapter 3 describes in detail all the steps involved in the process of development of the tongue's 3D model.

Chapter 4 illustrates the main results obtained in this thesis project.

Chapter 5 provides a summary of the present work, discussing the advantages and innovations that can be obtained by using this model and possible future developments of this project.

CHAPTER 2

BACKGROUND AND RELATED WORKS

This chapter contains the background for this thesis project. First, the anatomy of the tongue and its functions are presented. Then, we will talk about speech-related disorders and their causes. The main technologies to quantify the tongue's activity are presented. After that, the importance of visual biofeedback in speech therapy is discussed. Finally, various similar works are described.

2.1 The Tongue

The tongue is a muscular organ located on the floor of the mouth. It has several functions such as sucking, chewing and swallowing. It is the principal organ related to the sense of taste and has a fundamental role in speech production [2]. The tongue is formed by four intrinsic muscles, that are the Inferior and Superior Longitudinal muscles, the Transverse muscle and the Vertical muscle. There are also four extrinsic muscles, the Genioglossus, the Styloglossus, the Palatoglossus and the Hyoglossus. These muscles connect and anchor the tongue to the surrounding bones [3] [4]. The Palatoglossus has vagal innervation (CN X), while all the other muscles (both intrinsic and extrinsic) are innervated by the hypoglossal nerve (CN XII) [5].

2.2 Speech Disorders

Speech disorders constitute an alteration or impairment to the normal speech function, and they can be of different kinds. The most common speech disorders are stuttering, apraxia and dysarthria. The last two are considered motor speech disorders. Apraxia is due to damages to the cortical areas related to speech, whereas dysarthria is caused by weakness of the muscles involved in speaking. Speech disorders could be caused by strokes or brain damages [6]. Other pathologies, such as autism or hearing impairments, can generate speech disorders as secondary side effects [7].

2.3 Technologies for Monitoring the Tongue's Motion

The tongue is a very active organ, involved in several actions throughout everyday life. Quantifying the activity of the tongue, however, is not so trivial. Even if the tongue seems to be easily accessible through the mouth, there are some aspects that have to be considered to perform measurements on this organ. First of all, the oral cavity is an environment characterized by the presence of saliva. Secondly, the tongue is not completely visible, since it is hidden inside the jaw. The technologies to quantify tongue's activity also have to take into account the safety of the subjects. If instrumentation has to be placed inside the mouth, the risks related to swallowing or choking should be considered very carefully. Moreover, the measurements devices should not be an impairment to the normal functionality of the tongue in order to collect consistent data.

2.3.1 Electromagnetic Articulography (EMA)

Electromagnetic articulography (EMA) is a technique that uses coils to measure the position of the tongue during motion. A variable number of small coil sensors are placed using adhesives on the tongue and inside the mouth to interact with the electromagnetic field generated by the external magnet. The current induced on the sensors is related to the distance between coils, hence enabling the measurement of the position of the tongue over time. This method is mainly used for studies of speech or swallowing.

2.3.2 Electropalatography (EPG)

Electropalatography (EPG) uses a grid of electrodes placed on the hard palate of the patients to monitor the contacts between the tongue and the palate during articulation and speech. Every time a contact is detected, a signal is sent to a processing unit to assess where that touch event has happened.

2.3.3 Magnetic Resonance Imaging (MRI)

MRI is a non-invasive imaging technique based on the interaction between strong magnetic fields and hydrogen atoms inside the human body. MRI is mainly used to obtain images of soft tissues (since they contain more water, hence more hydrogen atoms), such as muscles, tendons or organs. MRI could be used to obtain static tongue's images in order to study the shape and position of the tongue's different articulations. For example, Liu et al. [8] showed how to track the tongue's motion in three dimensions space using MRI. They used tagged MR images and a technique called thin-plate spline interpolation to characterize the movements of the tongue during speech. Woo et al. [9] created a statistical multimodal atlas of 4D tongue motion. They combined cine- and tagged-MRI to provide both anatomical references and dynamic information of the tongue during speech.

2.3.4 Ultrasound Imaging (US)

US is another non-invasive imaging technique used to study soft tissues. An image can be reconstructed by measuring the time interval between the emission and the successive echo collection of an ultrasounds beam. The human tissues have different physical properties, hence they could be distinguished by the delays associated to the reflection of soundwaves.

Other methods and technologies to quantify and visualize the tongue's activity are electromyography (EMG), fluoroscopy and X-rays. However, they are not so commonly used.

2.4 Visual Biofeedback in Speech Therapy

Several studies discuss the importance of utilizing visual biofeedback to improve speech ability. In 2007, Engwall and Bälter [10] conducted a study in order to identify the most effective feedback characteristics for language leraning. In this study language teachers and students were interviewed and observed. The study suggested that the most effective feedback should provide a visual representation to the student, for example *showing* an animation of a particular articulation. Engwall and Bälter then applied these findings to one of the first animated models of the oral cavity, ARTUR (ARticulation TUtoR). A user testing of the model was performed, and subjects' opinions were obtained through a survey. Qualitative results suggested that ARTUR was considered an intuitive and effective tool in learning new sounds.

In a preliminary study, Morgan et al. [11] showed that the visual biofeedback provided by EPG could be effectively used to treat patients with dysarthria caused by a traumatic brain injury. After ten weeks of treatment (once per week) with EPG biofeedback, the results showed that the three tested adolescents improved their precision of articulation. The effects were visible not only for single phonemes production, but also for words and entire sentences.

Gibbon and Wood [12] described how to use EPG as an intuitive and interactive visual feedback to help children affected by speech disorders. EPG can be used efficiently both for diagnosis and therapy. By looking at the binary display showing the correct articulation, children can visualize the correct contact patterns and reproduce them.

Katz et al. [13] studied the effects that an EMA-based visual feedback had on the treatment of patients with apraxia of speech. They provided the feedback for three groups of speech motor targets (/j/, /θ/ and /ʧ/). They also examined how changing the feedback frequency influenced the rehabilitation process. They concluded that the use of a kinematic biofeedback improved the articulation of most of the phonemes. Moreover, the improvements were maintained one month after the end of the treatment.

Bacsfalvi and Bernhardt [14] showed how the use of visual feedback provided by EPG and US in speech therapy had not only short-term, but also long-term beneficial effects on hearing impaired adolescents. In fact, five out of seven subjects maintained or improved their level of performance after 2-4 years since the end of the therapy with visual feedback.

Shtern et al. [1] combined an EMA-based system with the Unity Game Engine (Unity Technologies, San Francisco) to create a videogame that provided augmented visual feedbacks during speech therapy. The game represented common scenarios in which the user had to control elements or characters using the tongue. The decision of not using an anatomical model of the tongue was based on the belief that a more competitive environment, like a videogame, could be more intuitive and bring faster improvements to the user.

In 2014, Ouni [15] claimed that even a short training session (15-20 minutes) with visual feedback is helpful to gain awareness about tongue's gestures. He compared pronunciation trainings with and without providing US real-time imaging of the tongue and concluded that the feedback improved the results.

More recently, Preston et al. [16] proposed a tutorial on how to use ultrasound imaging during speech therapy to provide biofeedback. They explained all the necessary steps that therapists and patients have to follow to collect images, interpret and use them to improve and to correct pronunciation errors.

Heng et al. [17] published a pilot study that suggests the effectiveness of using ultrasound visual feedback to treat velar fronting sounds (/k/ and /g/) in preschool children. The subjects showed improvements both during the therapy and the follow-up sessions.

2.5 Existing 3D Models of the Tongue

To date, several 3D models of the human tongue motion have been created. They were developed by using different acquisition techniques of the tongue's activity and they were built for various purposes.

Pelachaud et al. [18] proposed one of the first 3D model of the tongue to enhance digital facial animation systems. They described how to create a geometric and kinematic model of the tongue that could be animated to show speech production. The model was based on a mesh composed by nine 3D triangles. By moving the vertices of the mesh and by changing the length of triangles' edges, the model could assume various shapes.

In 2000, Engwall [19] developed a 3D model of the tongue based on MR images acquired during the production of Swedish articulations. He manually extracted the contours of the tongue and defined six linear parameters to characterize the different positions and reconstruct the shape of the tongue. In a successive work, Engwall [20] assessed the accuracy of his previous model by comparing it with real-time MRI. He showed that it was possible to develop models that would accurately recreate real-time articulations form static images of the tongue.

Engwall and Beskow [21] investigated if the tongue's motion can be synthesized from facial data using an optical tracking system. They extracted parameters to animate a 3D model of the tongue. Results showed that facial data were only sufficient for recreation of only a few movements of the tongue.

A very detailed morphological analysis of the tongue was performed by Takemoto [4]. He described the macroscopic and microscopic aspects of the tongue's musculature and structure. He defined the basic elements that should be included in the reconstruction of a realistic biomechanical 3D model of the tongue. He showed how the tongue's muscles are structured and organized from the inner part to the most external layer of the tongue. This study was fundamental to understand tongue's anatomy and physiology and established precise guidelines for future research about 3D modeling of the tongue's musculature. In fact, this kind of tongue's modeling is based on muscle contraction parameters.

Badin et al. [22] extended articulatory midsagittal (2D) models to three-dimensional models. They extracted the contours of the vocal tract from MR images and videos and defined a set of articulatory parameters for controlling the model. They successfully reconstructed 3D linear models of the tongue, lips and face during the articulation of French sounds.

Gérard et al. [3] [23] proposed a specific approach to 3D modeling of the tongue, by considering the biomechanical aspects of the tongue. They developed a dynamic model that took into account the mechanical properties of the tongue, such as force generation mechanisms and tissue's physical behavior. Their work resulted in creating a Finite Element Modeling (FEM) structure that represented all the tongue's muscles. They run simulations by changing the parameters of the model and compared the results in terms of tongue's motion with real MR images. Their model was accurate for specific movements but required improvements.

Katz and Mehta [24] investigated how a visual 3D model of the tongue could improve speakers' learning of non-native speech sounds. This experiment showed that 3D tongue's models can be useful not just in speech therapy but also for language training. The results showed that a visual feedback of the speech articulation process could aid in learning and improving faster. In fact, the accuracy of articulation was measured before, during and after the training with biofeedback. The subjects went from an average accuracy of 12.6% ± 14.1% during the pre-training to an average accuracy of 74.9% ± 15.6% when the biofeedback was provided. The accuracy gain was also maintained during the post-training phase, since the average accuracy was 85.3% ± 12.8%. The accuracy was measured as the number of correct articulations over the number of total attempts. The biofeedback was provided by an interactive EMA-based system that allows for visualization of a 3D model of the tongue. This system is called *Opti-Speech* and was developed by Vulintus LLC (Sachse, Texas, USA). In a previous study, Katz et al. [25] have evaluated the *Opti-Speech* system, concluding that it could be used as a very helpful biofeedback tool in speech therapy since it provided real-time biofeedback of the motion of the tongue.

CHAPTER 3

METHODS AND MATERIALS

This chapter illustrates the process of creation of the 3D model of the tongue, as well as the development of the final application. The whole procedure can be divided into consecutive steps. After the collection of MR images of the tongue, the 3D model was created through segmentation and then optimized. We performed some measurements on the 3D tongue to assess the accuracy of the model. Finally, the 3D model was animated to recreate realistic tongue's movements and it was integrated with a User Interface (UI).

3.1 MRI Acquisition and Segmentation

The first step to create the model was to collect MR images of the tongue (Figure 1). A set of 34 series has been recorded from a single, healthy, female subject at the UIC Center for Magnetic Resonance Research. The scanner used was the GE DISCOVERY MR750 (3.0 T). A T2-weighted Single-Shot Fast Spin Echo (SSFSE) MRI sequence was used. Repetition time (TR) and echo time (TE) were 1492.4 ms and 88.7 ms, respectively. The thickness of each slice was 2 mm, while the in-plane resolution was 0.5469 mm × 0.5469 mm. Each series was composed of 30 sagittal images. The time needed to complete a single scan (one series) was 45 seconds. The field-of-view was 28 cm × 28 cm and the reconstruction matrix size was 512 × 512.

The first scan position corresponded to the tongue at the rest state. Then, the subject was instructed to place the tongue at the 8 positions corresponding to the electrodes' positions on the oral wearable device (Figure 1D). The session was followed by 25 scans where the subject was asked to pronounce 25 static phonemes of the English language, reported in Table I. The static phonemes are characterized by a fixed tongue's position during the pronunciation. The 19 dynamic phonemes in English language require tongue active movement during pronunciation which hinders the ability to collect MR images, hence they were excluded from this study.

TABLE I: STATIC PHONEMES WITH EXAMPLES

Phoneme	Example	Phoneme	Example	Phoneme	Example
/k/	**c**at	*/l/*	**l**eg	*/s/*	**s**un
/r/	**r**un	*/j/*	**y**es	*/t/*	**t**op
/n/	**n**o	*/m/*	**m**ad	*/æ/*	h**a**t
/ɛ/	b**e**d	*/b/*	**b**ig	*/ʤ/*	**j**et
/f/	**f**ish	*/a/*	h**o**t	*/v/*	**v**et
/z/	**z**ip	*/w/*	**w**et	*/ɪ/*	**i**f
/u/	**u**se	*/d/*	**d**og	*/g/*	**g**o
/p/	**p**ie	*/θ/*	**th**umb	*/ʃ/*	**sh**ip
/ʧ/	**ch**ip				

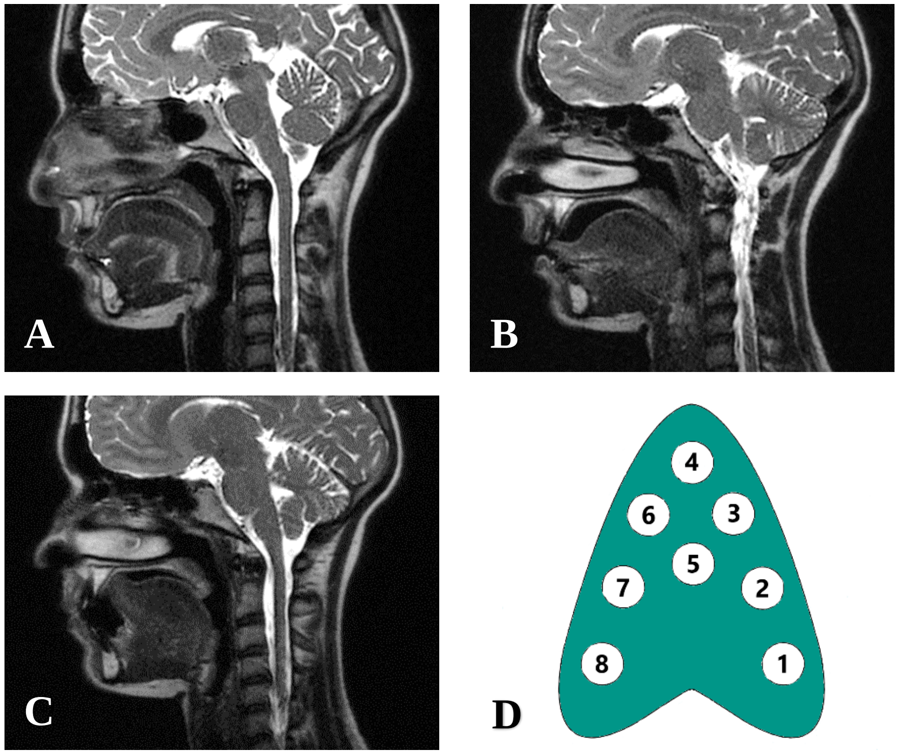

Figure 1: MRI of the tongue at rest (A), MRI of the tongue during the articulation of the phoneme /d/ (B), MRI of the tongue touching the center of the palate (C), and map of the 8 positions on the palate (D).

In order to segment and to create the 3D tongue model, the MR images were exported into 3D Slicer. 3D Slicer is “an open source software platform for medical image informatics, image processing, and three-dimensional visualization” [26] [27] [28]. For each series of images, two segments were created, one for the tongue and one for the contour around the tongue. The segmentation was performed manually, by “painting” the tongue shape on each slice. The software allowed to visualize the images on all the three planes, sagittal, axial and coronal

(Figure 2). After the tongue and tongue's contour were identified, the 3D model was composed semi-automatically by using the *Watershed* method, contained in the free *SegmentEditorExtraEffects* extension of the software [29]. A 3 mm Gaussian filter was applied to the 3D model for smoothing effect (Figure 3). The models were saved as a 3D objects (.*stl* or stereolithography file). We segmented 33 out of 34 tongue's acquisitions. It was not possible to correctly segment the remaining series (corresponding to the phoneme /ʤ/) since the tongue's shape was altered between slices, probably because the subject was not completely still during the scan.

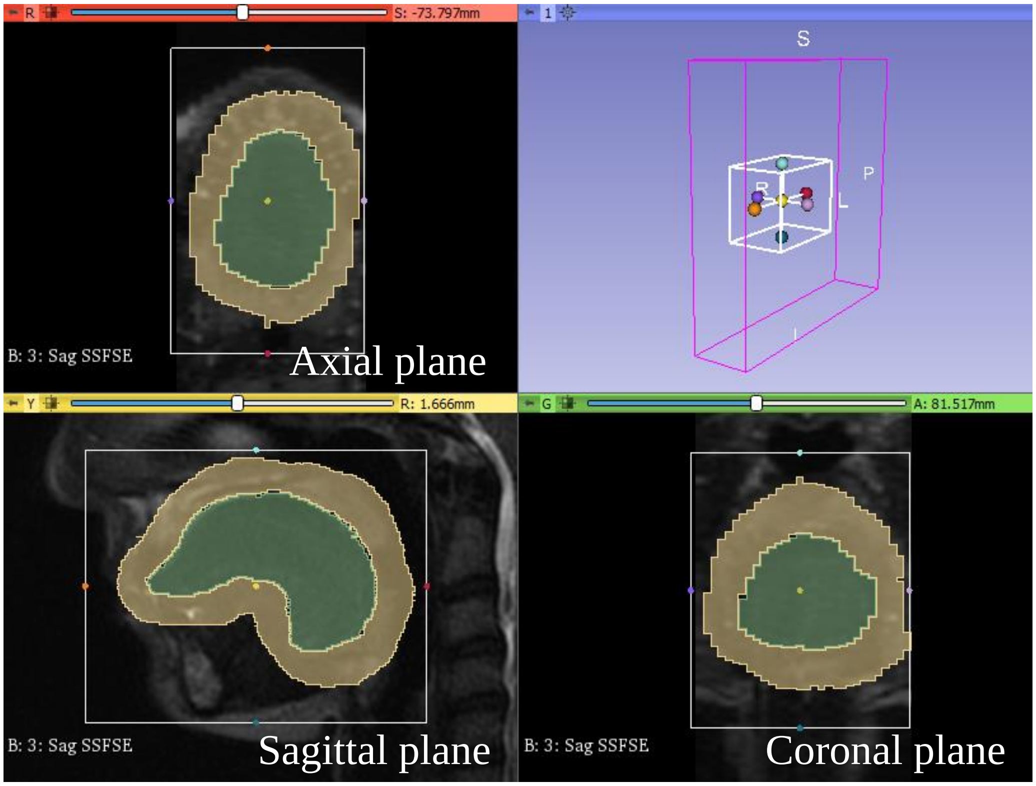

Figure 2: Slicer editor: the tongue (green) and its contour (yellow) are painted on the three planes. The white box in the top right window represents the region of interest (ROI).

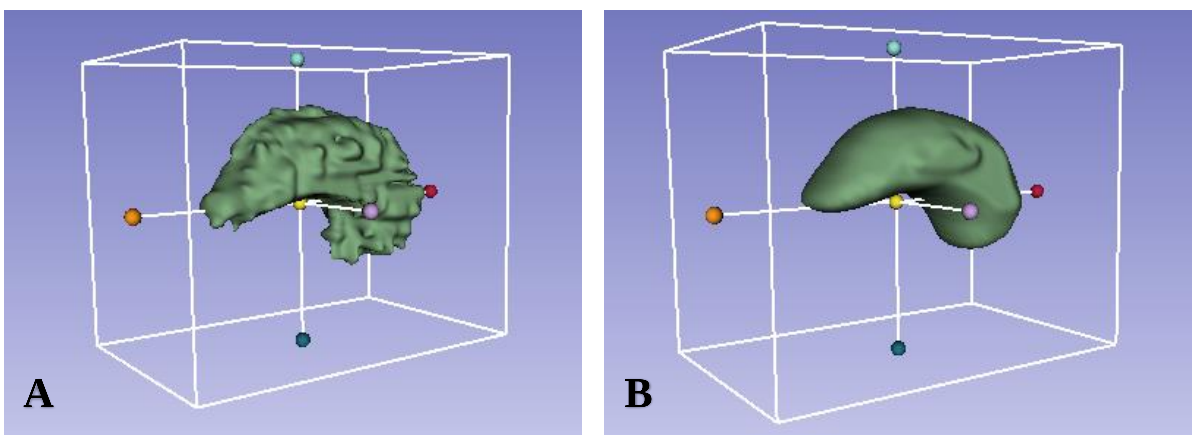

Figure 3: Segmentation results before (A) and after (B) the smoothing effect.

3.2 Mesh Measurements

To quantify the accuracy of the segmentation process, the dimensions of the 3D model at rest were measured using 3D Slicer (Figure 4). Length, width and thickness of the 3D tongue were measured. We utilized the same measurement procedure outlined in [30]. The thickness was measured in the central and thickest part of the tongue. Length and width were measured by following the surface of the tongue, taking into account its natural curvature. Twelve fiducial points were created and placed on the 3D model to perform the anatomical measurements. All the measurements were determined by computing the distances between the fiducial points and summing them up. To obtain consistent results, the positioning of the fiducial points and the measurements were repeated 10 times. The average value and the standard deviation were calculated.

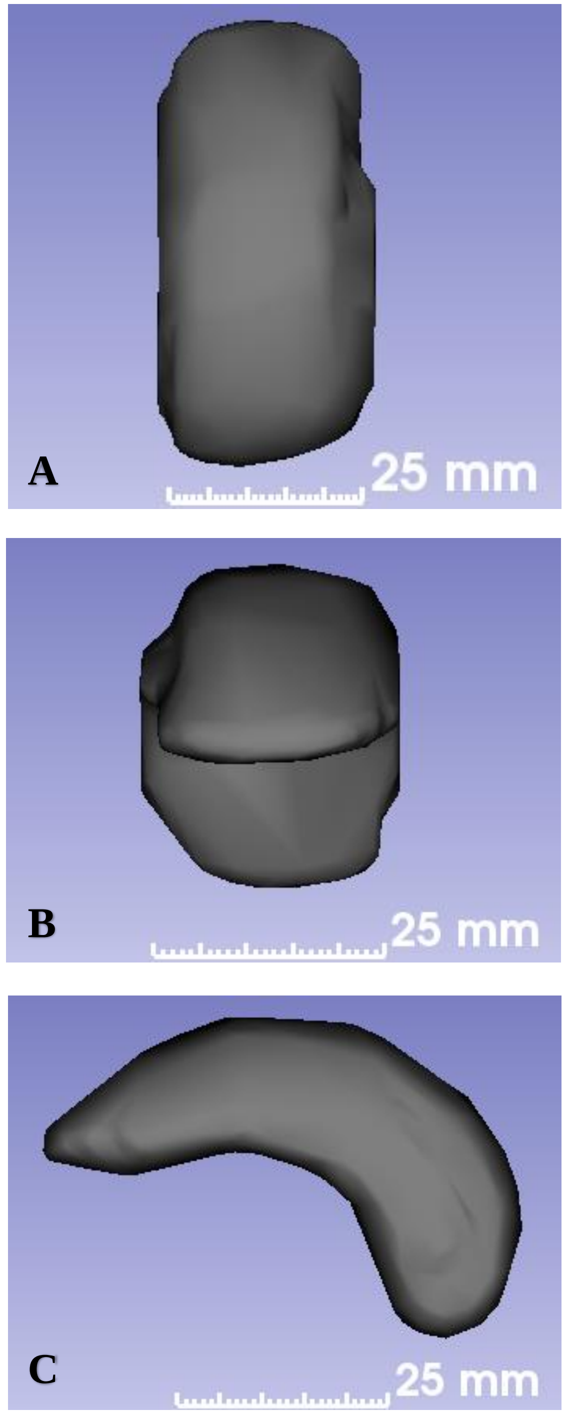

Figure 4: Tongue’s model at rest. Axial (A), Coronal (B) and Sagittal (C) views.

Another series of measurements were conducted using 3ds Max (Autodesk®, San Rafael, California). All 33 models of the tongue's shapes (Figure 5) were imported into the software. The linear dimensions and the volume of the model were automatically calculated. Average and standard deviation were computed over all the shapes. We also tried to assess if there was some degree of asymmetry during the articulation of the different phonemes with respect to the axial plane. To do so, we divided each model into two halves along the longitudinal axis and quantitatively compared the total surface of the right and left sides. We determined the presence of asymmetry by computing the difference between right and left surfaces. If the difference was above an arbitrary threshold, the model was considered asymmetric. In particular, the asymmetry was classified as *right* asymmetry or *left* asymmetry depending on which of the two halves had the larger surface.

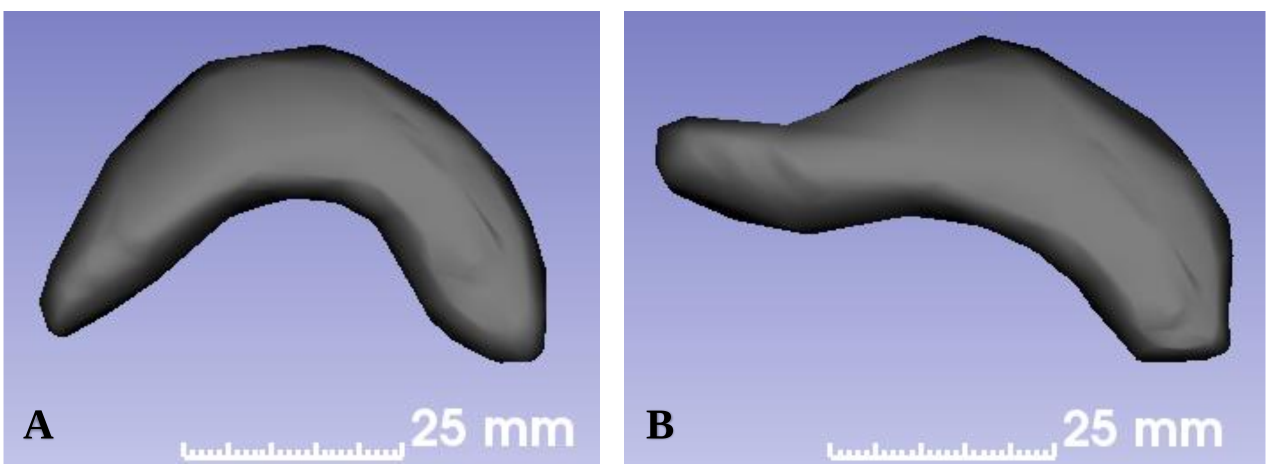

Figure 5: Examples of 3D models of the tongue at different positions. /k/ (A) and /d/ (B) phonemes.

3.3 Mesh Optimization in 3ds Max

The 3D models of the tongue were imported into 3ds Max (Autodesk®, San Rafael, California) to be optimized and prepared for the animation. The O*ptimize* modifier was applied to each mesh in order to reduce the number of vertices and triangles (Figure 6). The goal of the optimization process was to reduce the mesh complexity in order to obtain computationally lighter models while maintaining the realistic shape of the tongue. These computationally lighter models result in smoother animations. The tongue's mesh at rest was merged with a jaw and skull model (Figure 7) [31]. All the models were placed inside a human head mesh to obtain the realistic representation (Figure 8) [32].

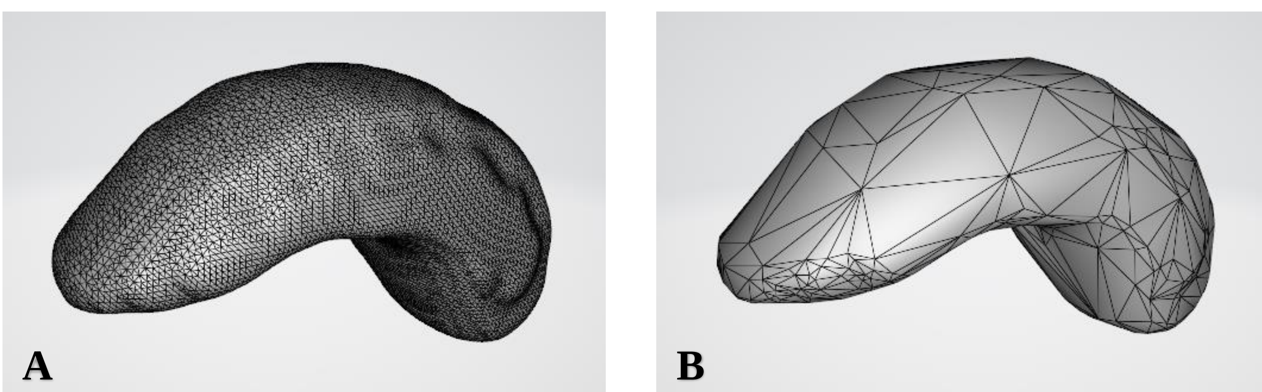

Figure 6: Mesh before (A) and after (B) optimization.

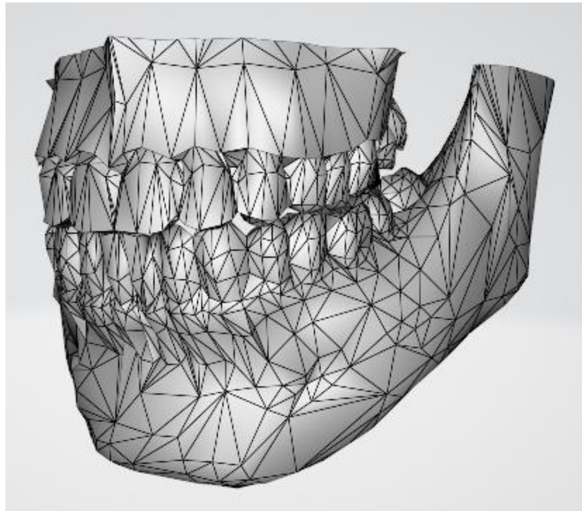

Figure 7: Models of the bones to complete the oral cavity.

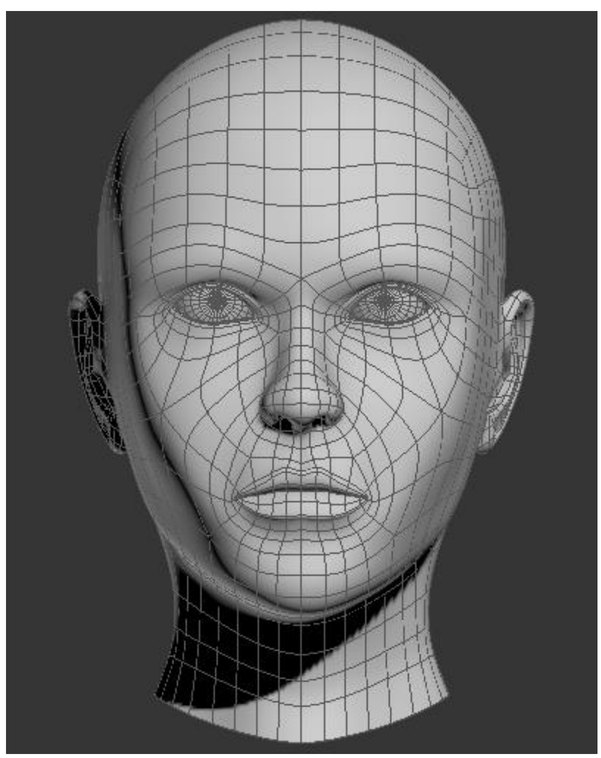

Figure 8: Female head mesh.

3.4 Lips Animation

The lips animation was created into 3ds Max (Autodesk®, San Rafael, California) using the *Morpher* modifier. This process consisted of making seven copies of the head's mesh and manually modifying the lips to the desired shape (Figure 9). Once these new shapes were associated to the original mesh, the modifier automatically created the animations starting from the original lips' position and finishing at the new configurations. Since MR images of the lips were not available, digital pictures of the lips while assuming various shapes were used to create the model. 12 animations were created to represent the phonemes and the words. All the objects were then exported into the Unity software platform (Unity Technologies, San Francisco).

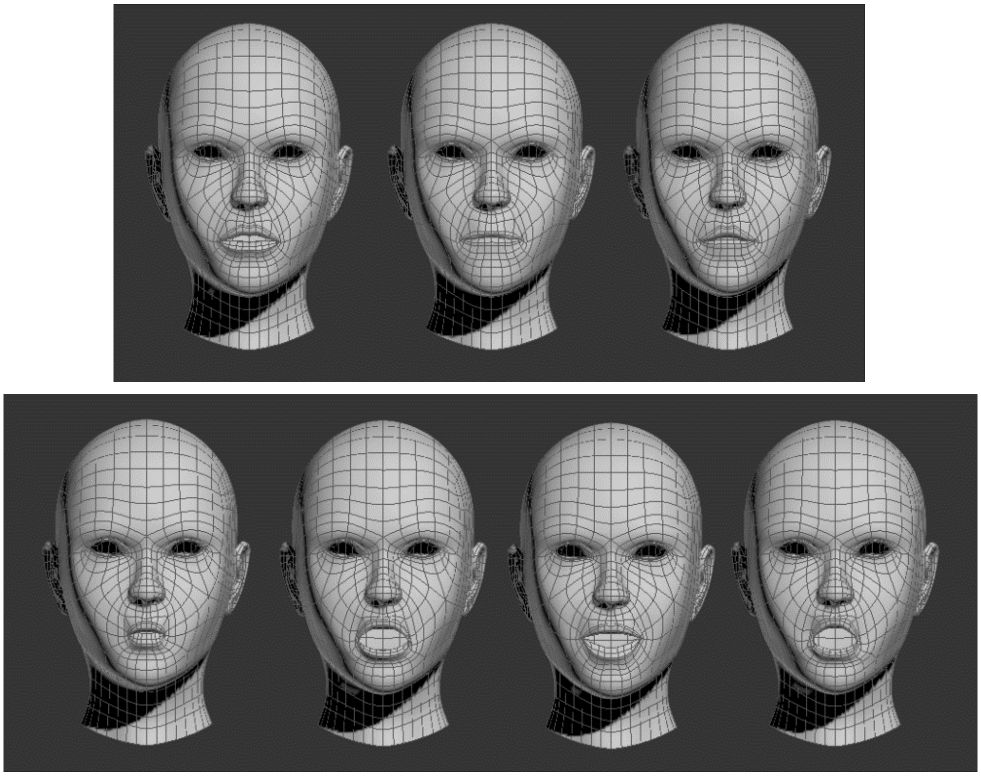

Figure 9: Head models with seven lips shapes.

3.5 Model Animation in Unity

Unity software platform (Unity Technologies, San Francisco), is mainly a game engine, that could be utilized for simulations of 3D objects. Due to versatility of the Unity Game Engine, it could be used for various modeling purposes. In this case, the aim was to create a scalable method to animate the tongue. To do so, an external plugin called *SplineMesh* was used, which allows for creation of splines [33]. Splines are polynomial functions defined piecewise. The splines are composed of a variable number of nodes, whose position and direction determine the overall shape of the curve. We used splines as an internal structure for the mesh. Hence, by changing the shape of the spline, the mesh attached to that spline would change. The advantage of this approach is that it allows for the motion of the nodes to be controlled by a script, making this approach versatile and scalable. A 4-nodes spline was created and the mesh of the tongue at the rest position was assigned to the spline (Figure 10). When a node is selected, it can be moved to shape the curve. Moreover, once a node is selected, two additional points appear in order to adjust the curvature of the corresponding segment of the spline. The position and direction of each node are the parameters used for controlling the model. Each mesh representing the tongue at different positions is animated according to a specific spline shape. In total 34 splines were created. The splines were shaped in order to match the structures of the meshes that resulted from the segmentation (Figure 11). The models created with splines were visually compared to those obtained from the segmentation process in order to assure the maintenance of the realistic shapes (Figure 12). Position and direction of the 4 nodes of each spline was saved. Each spline was associated to a button on the User Interface (UI), corresponding to the representative phoneme or position. The code to animate the splines is based on a *switch-case* statement. When a button is pressed, the controlled spline moves towards and assumes the shape associated to that button. Five complete words (‘cat’, ‘leg’, ‘no’, ‘top’, ‘yes’) were also created by combining the phonemes splines. Jaw motion was implemented with three levels of mouth opening, depending on the pronounced sound. Alongside with tongue and jaw movements, lips animations were used to make the model more realistic. The animations were imported from 3ds Max and embedded into the head model through the *Animator* component (Figure 13). The component enables controlling of different clips. When the corresponding motion is selected, the animation is activated through a transition from the *Idle* state. Transitions are conditions that determine when the animation clips start. We implemented the transitions using Boolean variables. When the motion is selected, the status of the corresponding variable changes and the relative animation is played.

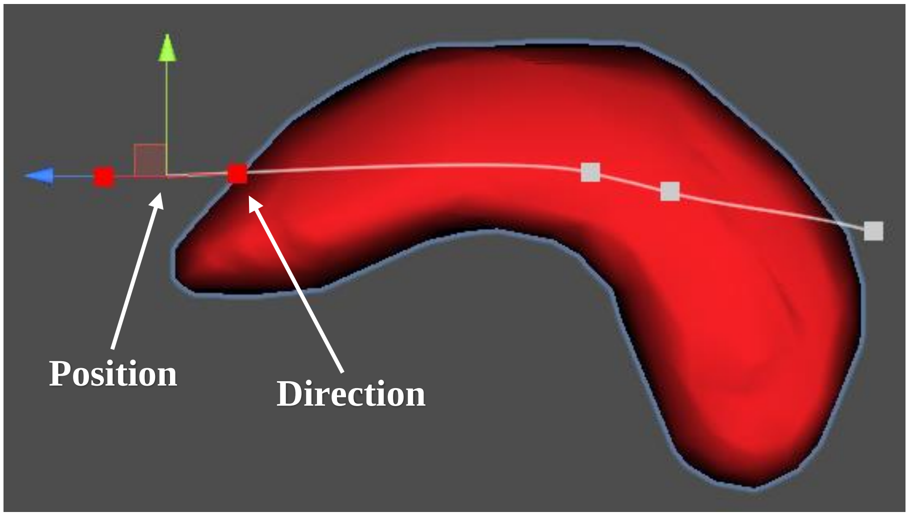

Figure 10: The spline associated to the tongue at the rest position.

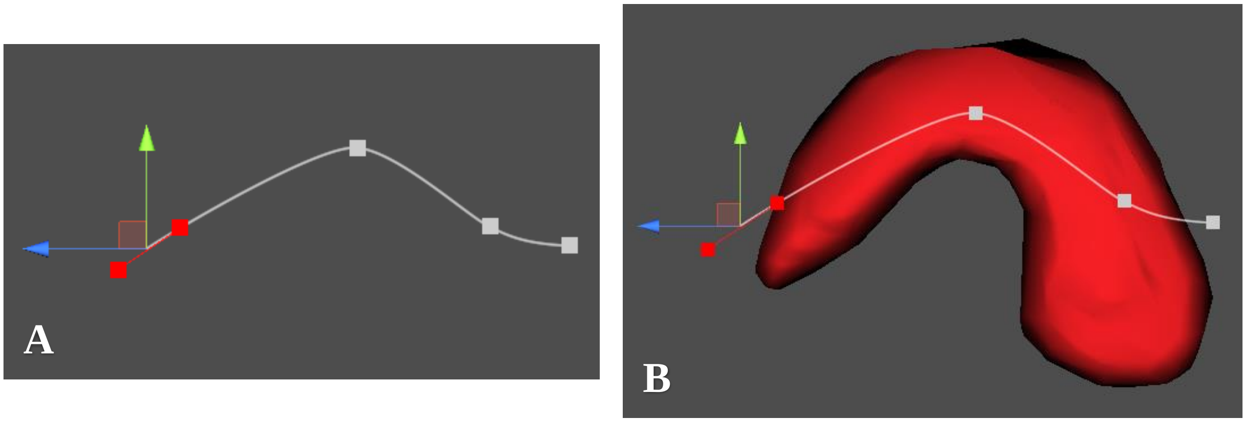

Figure 11: The spline representing the phoneme /k/ (A) and the associated mesh (B).

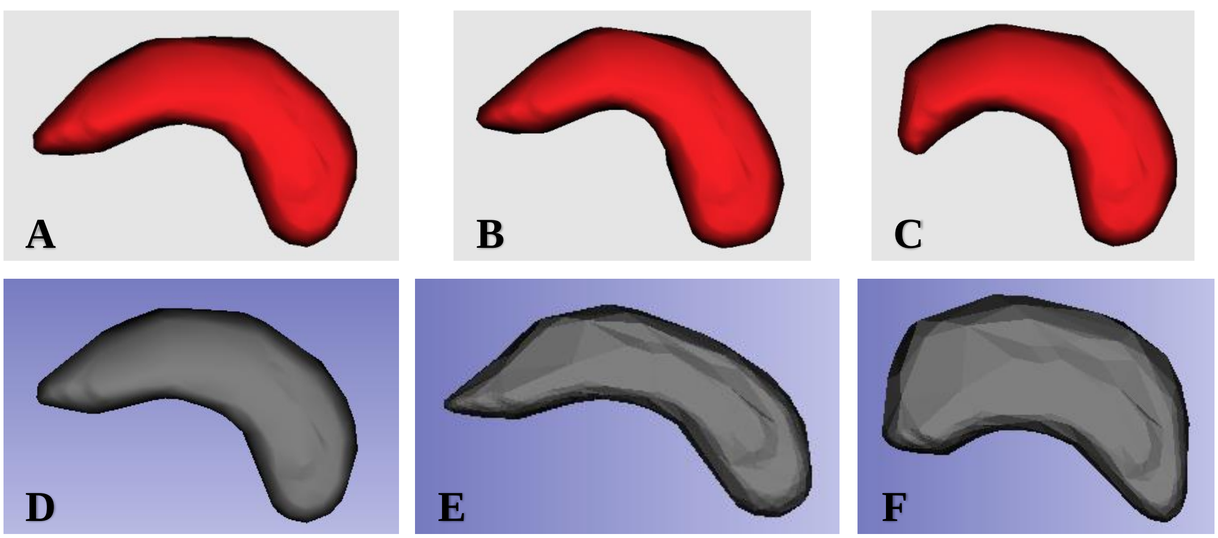

Figure 12: Models obtained from the splines (A, B, C) and the segmentations (D, E, F). Rest position (A, D), phoneme /z/ (B, E) and phoneme /ɪ/ (C, F).

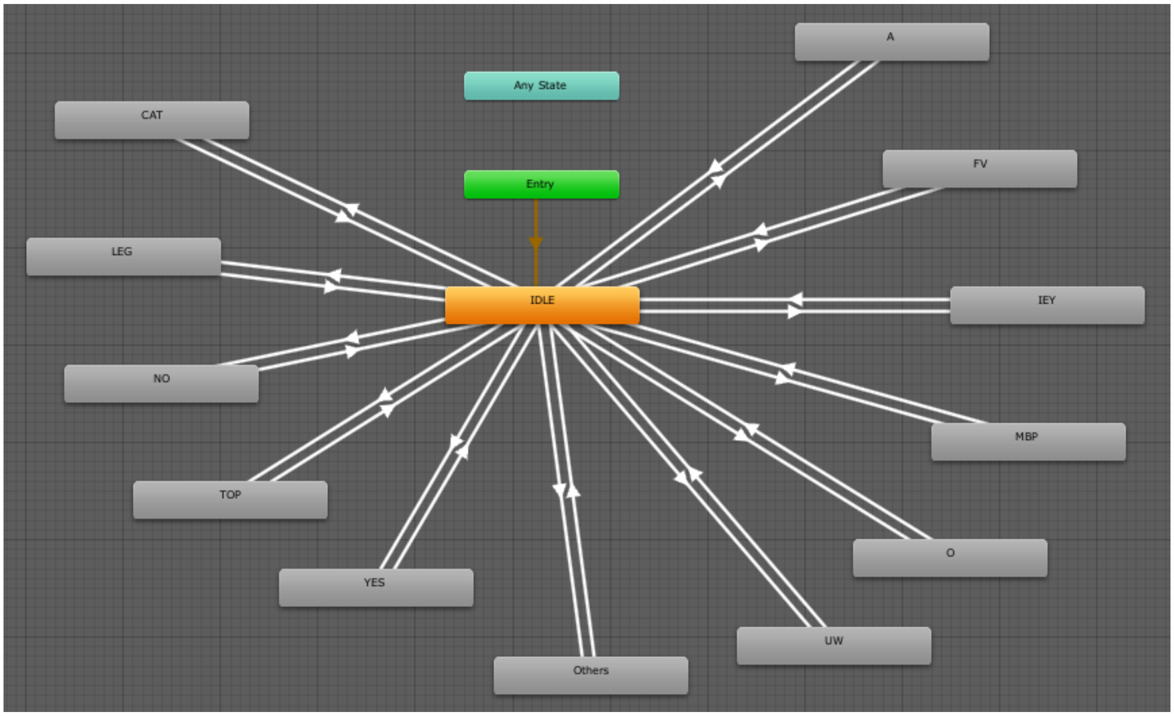

Figure 13: The Animator editor. The grey boxes are the animation clips, while the white arrows represent transitions from and to the *Idle* state.

The model allows for selection of three projection planes, top-view, front-view, and side-view. These three views correspond to the three different *camera* objects in Unity that were placed around the head in parallel with the orthogonal planes. The *cameras* can be activated by selecting the corresponding buttons. There can be only one active camera at the time. Two degrees of transparency were associated to the head mesh in the three views. While the front view shows a 'solid' mesh for clear visualization of the lips' motion, the top and side views display a more transparent head for clear visualization of the tongue's motion (Figure 14). The transparency is a property of the material associated to the objects. The transparency is modifiable through a script that manages the *Skinned Mesh Renderer* component of the head model.

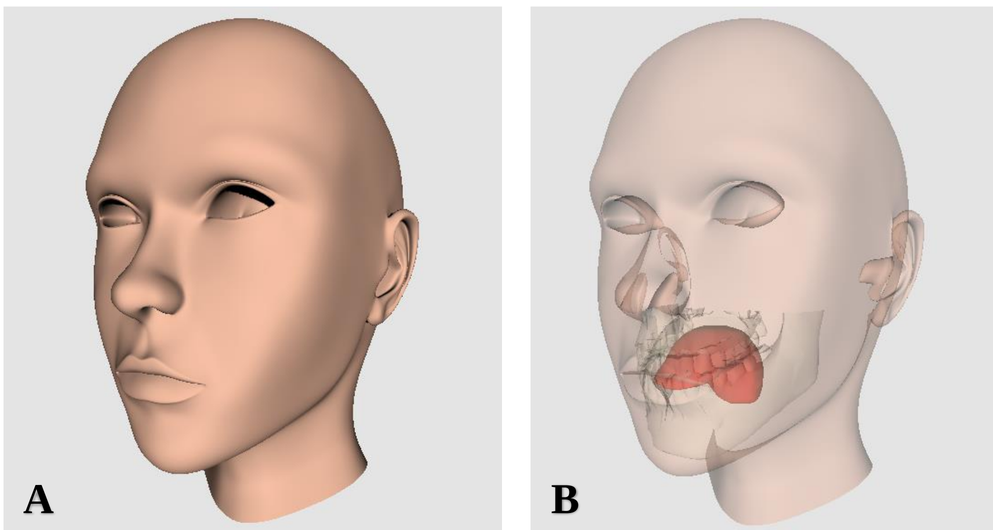

Figure 14: Solid (A) and transparent (B) head models.

Another important aspect of this application is the possibility to show real-time motion of the tongue. To do so, a custom wireless oral wearable device developed by the Wearable Technology and Sensory Enhancement Lab was used (Figure 15). This device is used as a computer trackpad that tracks the location of the tongue on the roof of the palate and moves the cursor accordingly. The device communicated live tongue location data to the application which in turn controlled the position of the 3D tongue model. The two cartesian coordinates corresponding to the cursor position on the screen were associated to the motion of the first node of the spline, which defines the tip of the tongue in the horizontal plane. The movement of the tongue is constrained to the boundaries of the wearable device (3.4 cm × 4.4 cm). These boundaries are represented in the scene view. The trajectory of the tip of the tongue is displayed in order to guide the user while using the application.

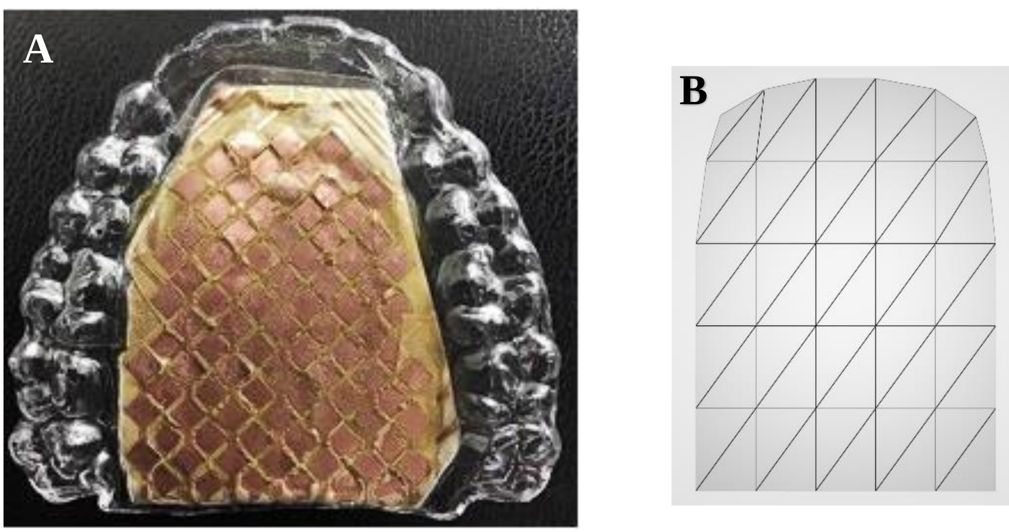

Figure 15: The wearable oral trackpad device (A) and the corresponding trackpad plane used in the model (B).

The application also features a ‘Free Exploration’ mode. In this mode the data corresponding to the tongue’s free exploration experiments (collected off-line using the wearable device) could be displayed. For the free exploration experiments the subjects were asked to move their tongue randomly and without any constraint while trying to cover the entire surface of the wearable device. The device collected the tongue’s location data. The model is then able to read the x and y coordinates of the tongue on the wearable device, that are saved into an Excel file, and recreate the pattern of the movement in the scene view.

The entire application is controlled using the User Interface (UI) in Unity. Unity contains tools that allow for building UIs, making it simple and straight forward. After defining a *Canvas*, we developed the starting screen and the main menu, through which the user can access all the functionalities of the model (Figure 16). Each command is associated to a button on the screen. When a button is clicked, the relative function *On Click* is executed. All the buttons are associated to particular cases in the script that controls the spline motion. The communication between different scripts was implemented through a static (global) variable. The UI was created to be simple and user-friendly.

Figure 16: User Interface with interactive elements.

CHAPTER 4

RESULTS

This chapter shows the results obtained in this project. First, the anatomical validation performed on the 3D model is presented. Then, the functions of the 3D model and the application are described.

4.1 Anatomical Data

The accuracy of the 3D tongue model was compared to the dimensions described in literature by Hopkin [30]. He reported an average adult tongue's length of 79.83 ± 7.57 mm, a width of 51.90 ± 4.24 mm and a thickness of 16.13 ± 2.51 mm. The 3D tongue model length is 75.27 ± 2.90 mm, the width is 40.69 ± 0.99 mm and the thickness is 19.33 ± 0.73 mm. We repeated the measurements 10 times in order to obtain more consistent data (Table II).

TABLE II: DIMENSIONS OF THE 3D TONGUE AT REST

	Width [mm]	Length [mm]	Thickness [mm]
Trial 1	40,8	77,6	20,5
Trial 2	38,3	69,2	18,2
Trial 3	40,8	75,1	20,2
Trial 4	41,4	73,9	19,1
Trial 5	41,7	73,6	19,4
Trial 6	39,8	80	18,8
Trial 7	41,3	75,8	19,8
Trial 8	40,8	76,7	19,6
Trial 9	40,6	73,9	18,5
Trial 10	41,4	76,9	19,2
Avg. (± SD)	40.69 (± 0.99)	75.27 (± 2.90)	19.33 (± 0.73)

The volume of the 3D tongue at rest is 22.26 cm^3 and this value is close to the volume reported by Tamari et al., which is 22.6 ± 4.1 cm^3 [34]. The average volume of all the other shapes that the model acquires is 21.85 ± 2.87 cm^3. This confirms the fact that the volume of the tongue is maintained almost constant, as reported by Rezende et al. [35]. As for the linear dimensions during the articulation of different phonemes, most of the variability is associated to the changes in tongue's length. In fact, the model shows an average linear length of 57.24 ± 6.78 mm. The phoneme /l/ is associated to the largest tongue's length, which is 70.47 mm. The minimum length, which is 45.53 mm, corresponds to the vowel /æ/. The difference between these two values (24.94 mm) represents the range of motion of the tongue along the longitudinal direction. In order to analyze asymmetry in the axial plane, the 3D models were separated in the two sections along the longitudinal axis. The surfaces of the two parts were compared and if the difference was greater than an arbitrary threshold the model was considered asymmetric. In this way 15 out of 24 tongue's models (62.5%) showed asymmetric. In particular, 9 out of 15 models presented larger right side, whereas 6 out of 15 models had larger left side. All these results seem to differ from those found by Miller et al. [36]. By analyzing a large dataset of palatograms of English phonemes, they found that 97.5% presented some level of asymmetry. Moreover, most of the subjects showed left side asymmetry. The difference in results could be mainly due to the different methods of quantifying symmetry and to the small number of data that we considered.

4.2 The Final Application

The 3D model is included within an application through which the user can switch between menus representing each aspect of the model. After clicking on the screen to start the application (Figure 17), the main menu is displayed (Figure 18). The user can access five different menus. We will go through these features in the following sections. The 'Exit' button allows to quit the interface when using the mobile application.

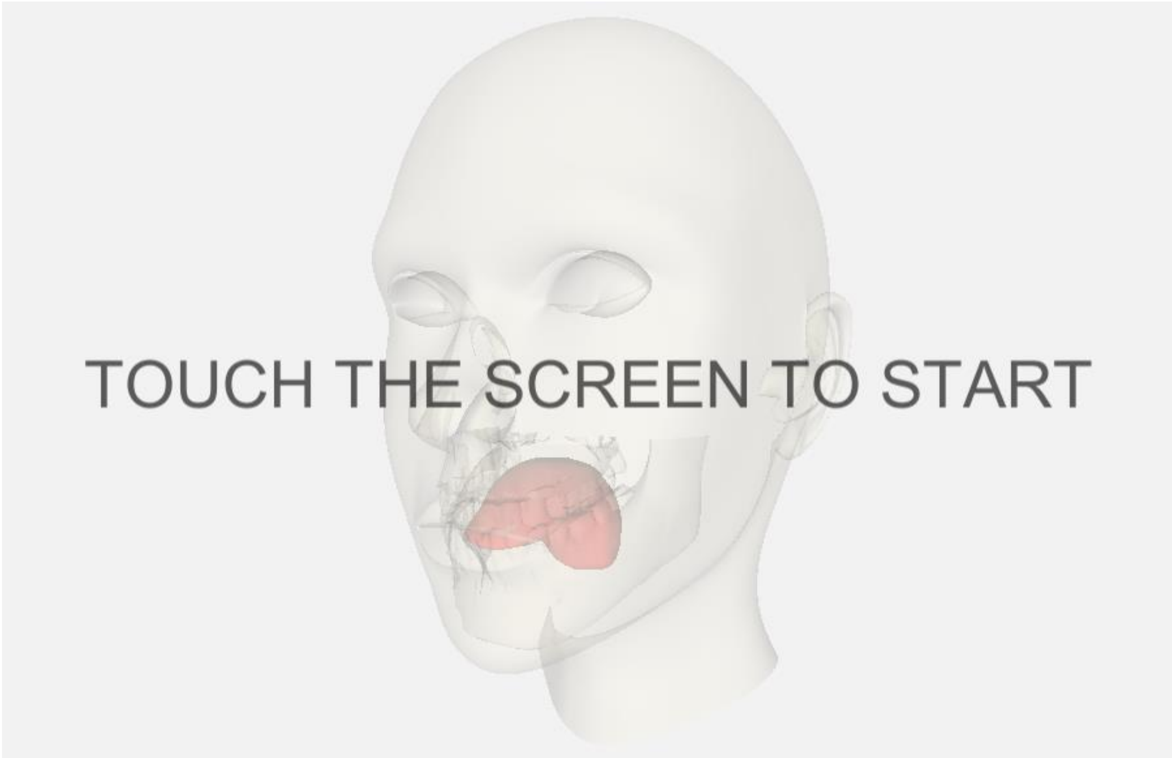

Figure 17: Start screen of the application.

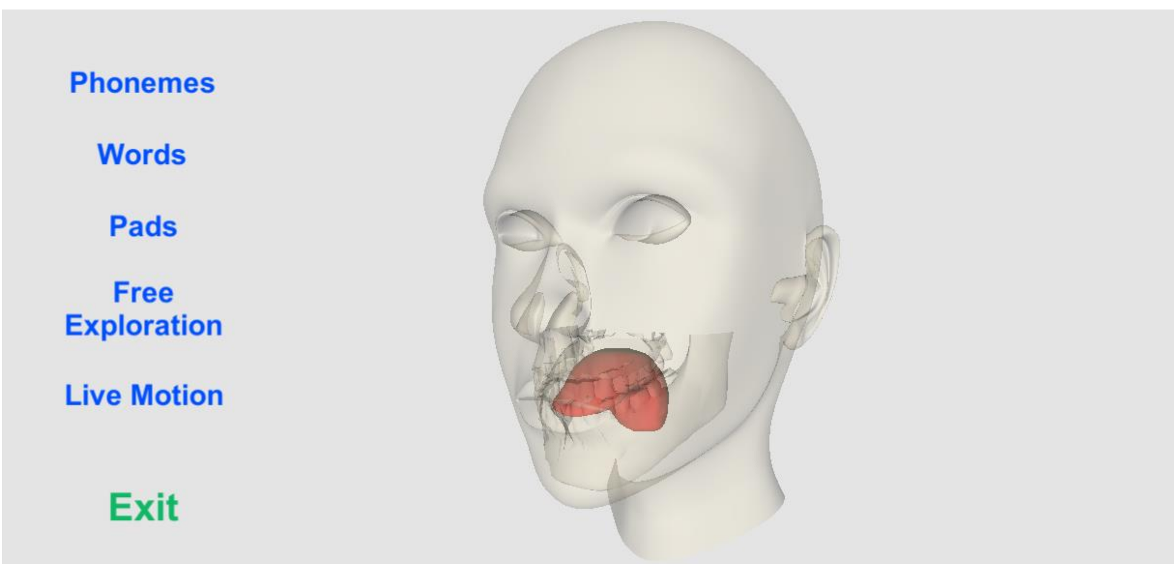

Figure 18: Main menu of the application.

4.2.1 Phonemes Menu

This menu implements the tongue's movements associated to the pronunciation of the 24 English static phonemes. Once a button is pressed, the tongue changes its shape and assumes the position representing the phoneme (Figure 19, Figure 20). The motion of the jaw and the lips are also implemented for each phoneme. While the tongue is moving, a message showing which phoneme has been selected is displayed to the user together with an example of a word that contains the phoneme. Whenever the 'Rest' button is pressed, the tongue comes back to its initial position (Figure 21, Figure 22). The user could also change the scene view during the motion, to better visualize how the movement is conducted by looking at it from different angles. The three possible cameras represent the model along the three orthogonal projection planes (Top, Front and Side views). Finally, the 'Back' button is embedded to return to the main menu.

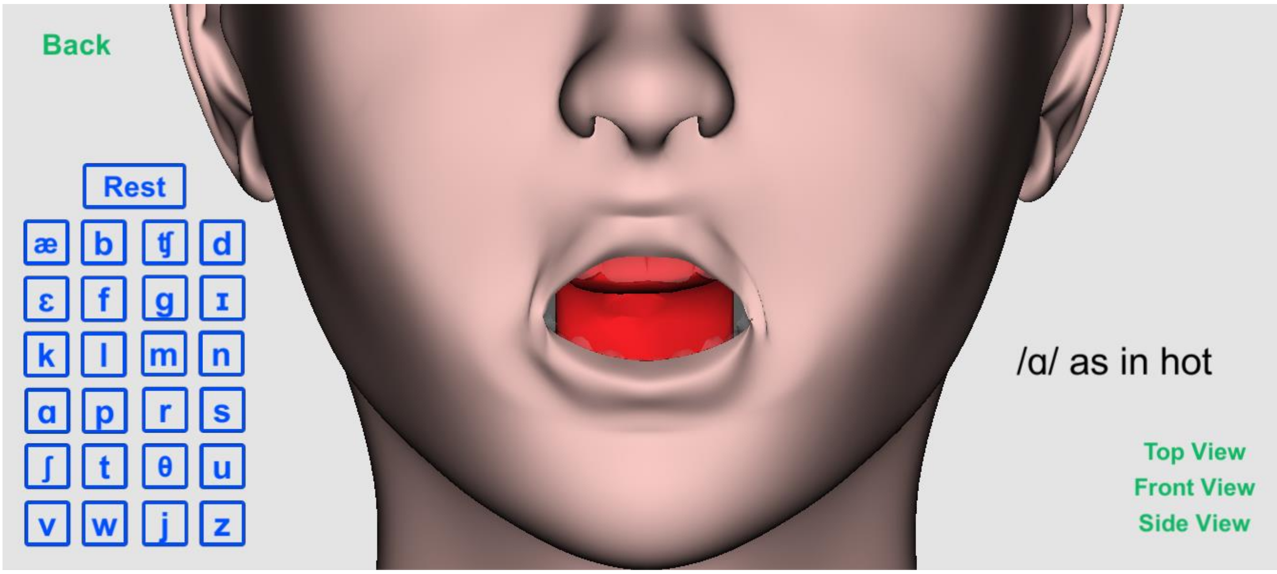

Figure 19: Tongue's model representing the /ɑ/ phoneme (Front View).

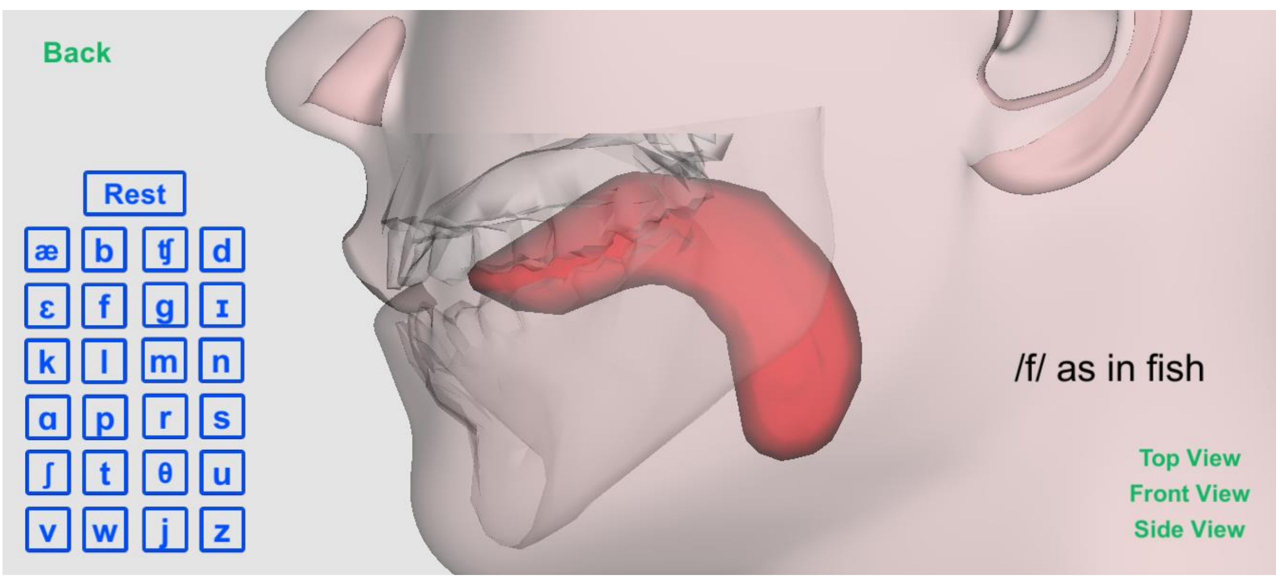

Figure 20: Tongue's model representing the /f/ phoneme (Side View).

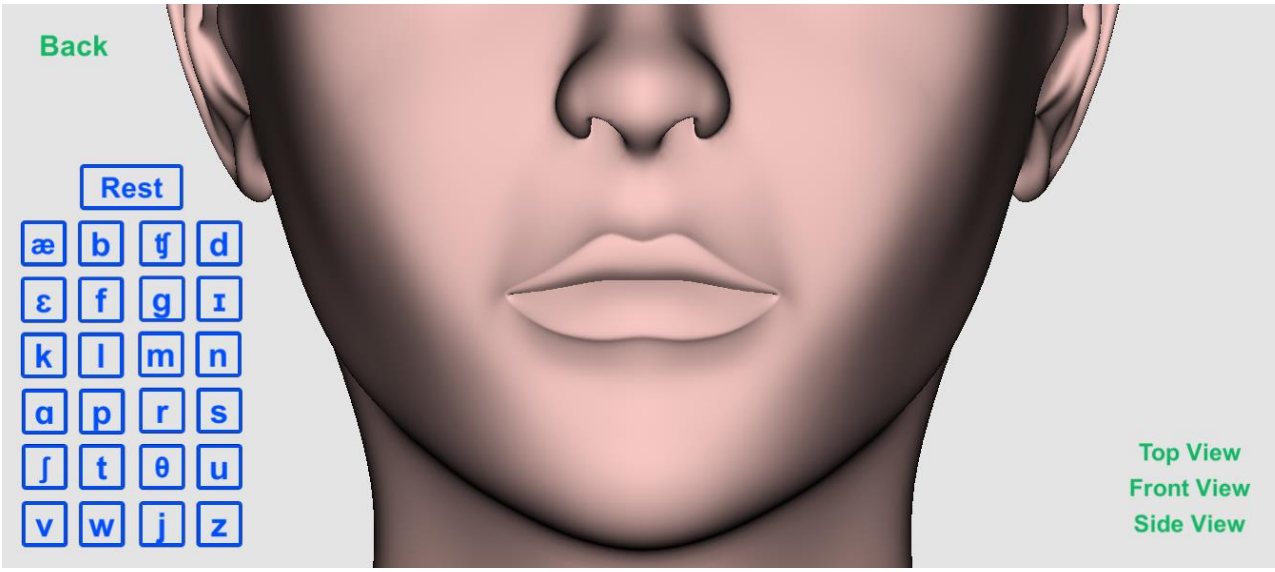

Figure 21: Tongue's model at the rest position (Front View).

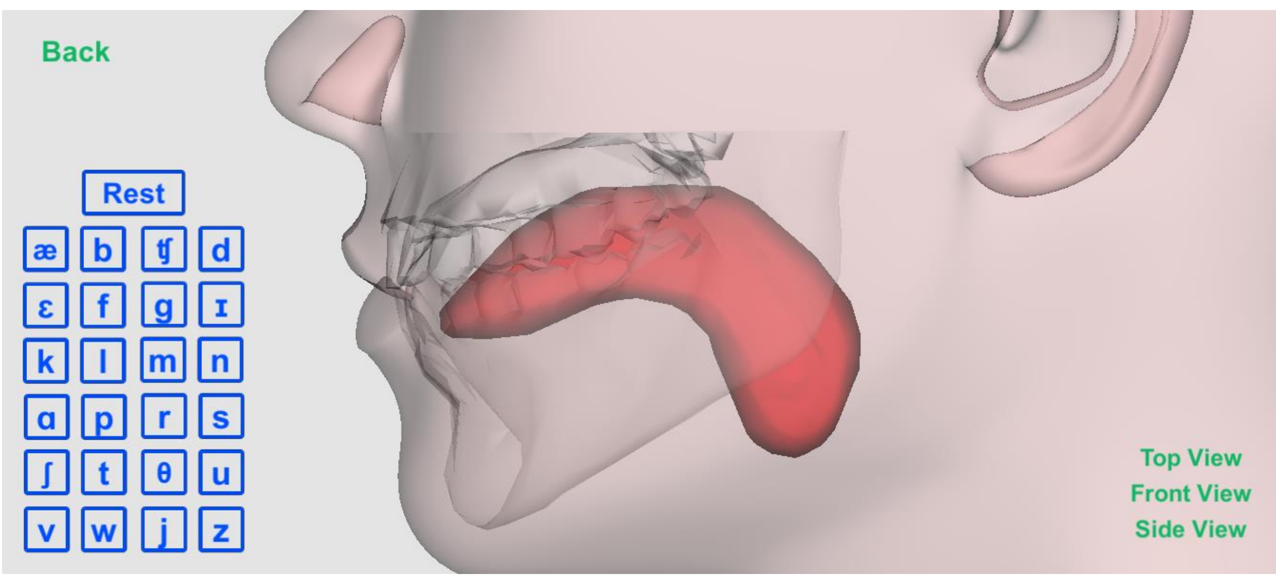

Figure 22: Tongue's model at the rest position (Side View).

4.2.2 Words Menu

In this menu phonemes are combined to recreate an entire word. Five words have been implemented, ‘cat’, ‘leg’, ‘no’, ‘top’ and ‘yes’. When the respective button is clicked, the movements of the tongue, jaw and lips associated to the phonemes constituting the word are represented in a sequence. ‘Back’ button and different scene views are also present (Figure 23).

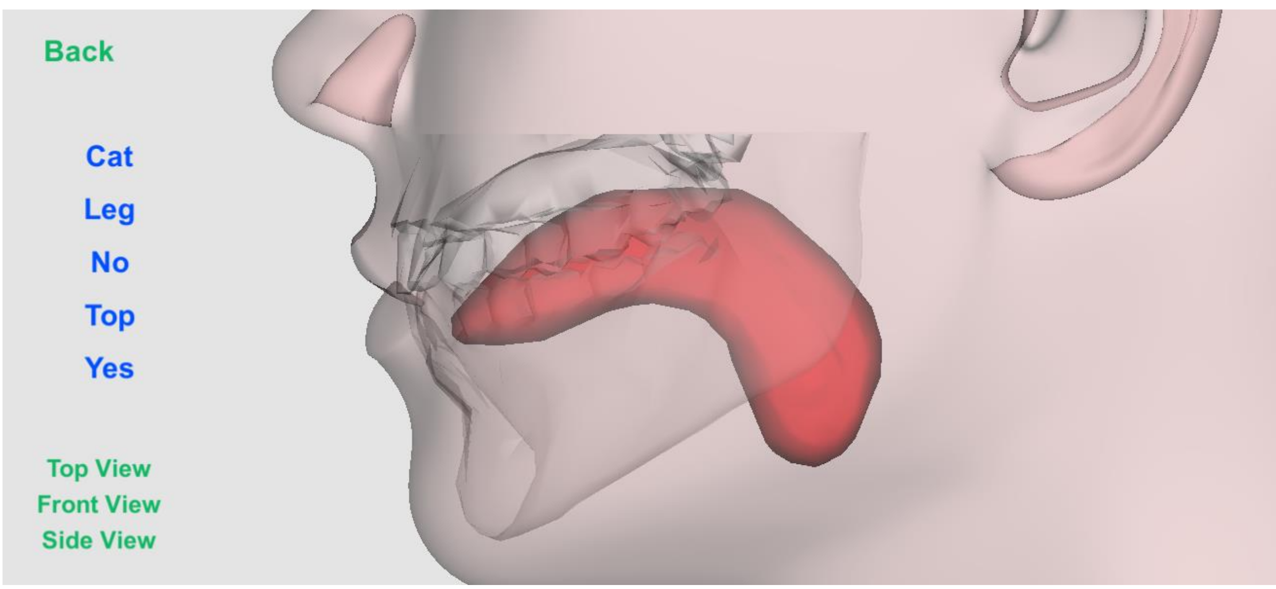

Figure 23: Words menu (Side View).

4.2.3 Pads Menu

Within this menu, a map of the palate is shown, and the 8 points representing the pads positions on the oral wearable device are highlighted (Figure 24). Clicking on one of these pads will make the 3D tongue move towards that specific position. A message tells the user which position has been selected. Also in this case the motion can be viewed from different points of view (Figure 25, Figure 26).

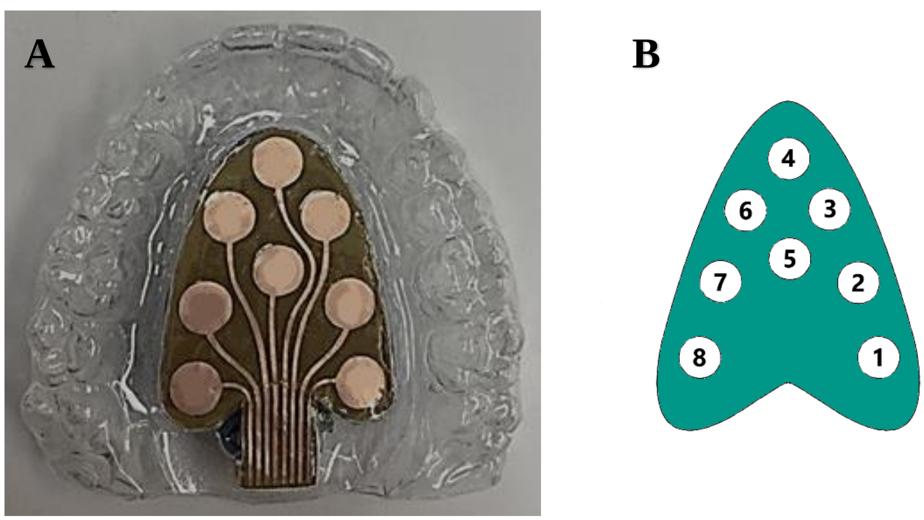

Figure 24: Oral wearable device (A) and map of the palate (B).

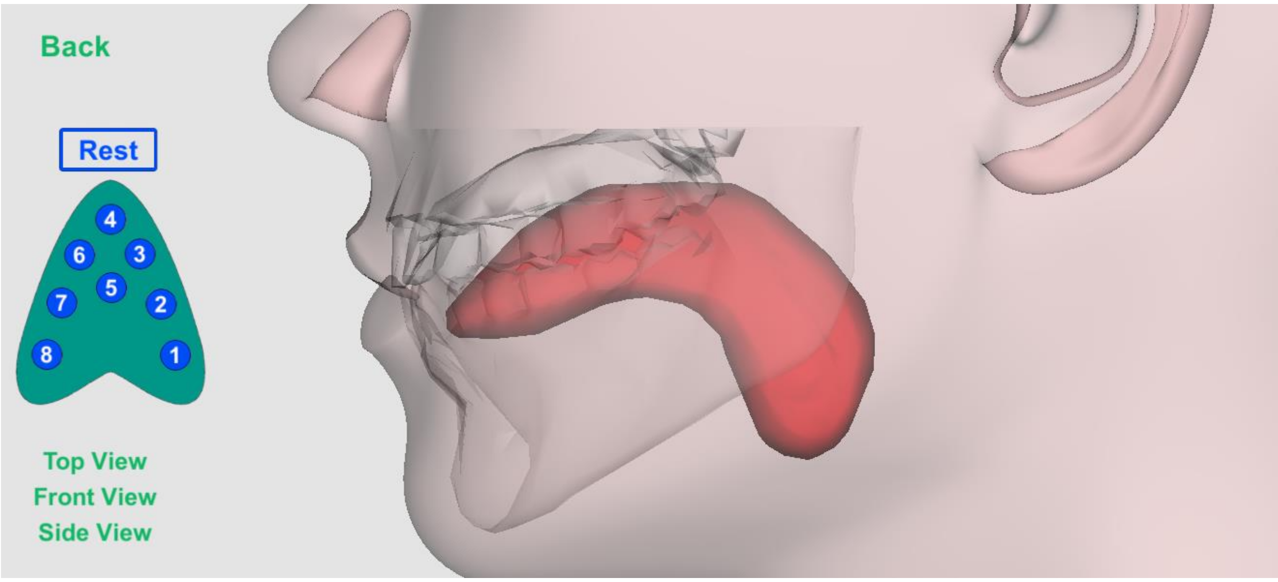

Figure 25: Pads menu (Side View).

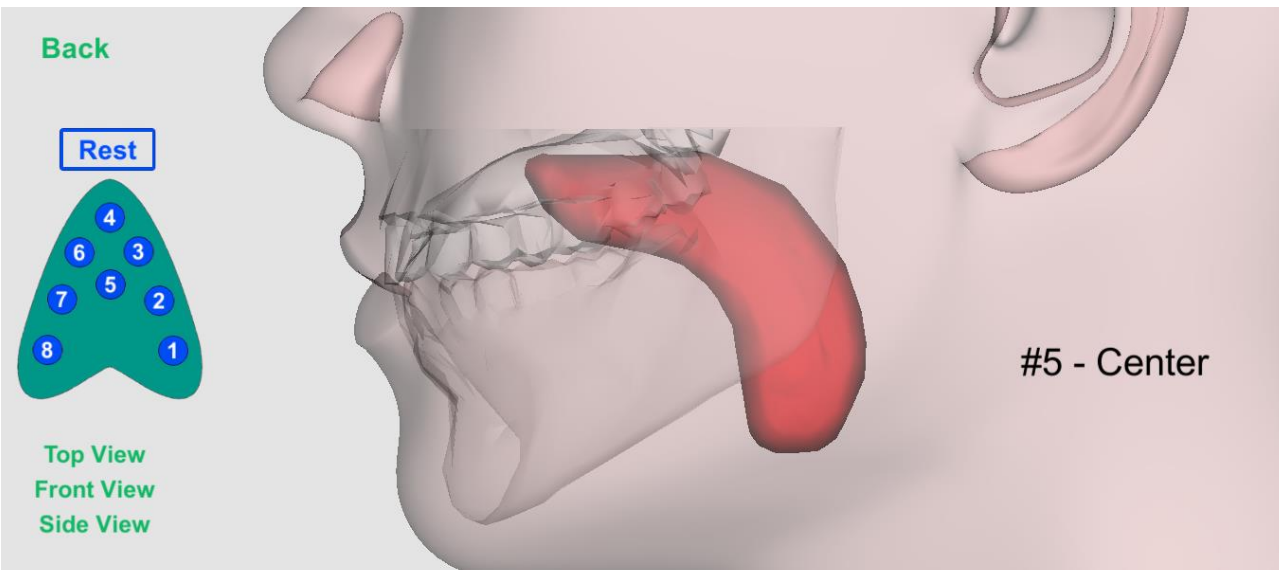

Figure 26: Tongue's model touching the center of the palate (Side View).

4.2.4 Free Exploration Menu

This menu has been implemented to show the tongue's motion during a free exploration task. Free exploration is basically a random motion of the tongue that is recorded and then analyzed. To do so, we have collected data with the wearable oral trackpad. The model is able to read the coordinates of the tongue on the trackpad and reproduce them. As an example, five trials have been collected and each of them can be visualized in this section. The motion is only viewable from the top view and the camera is fixed because all the movements takes place on one plane (the trackpad). The wearable's area is highlighted on the screen and the trajectory of motion is displayed (Figure 27).

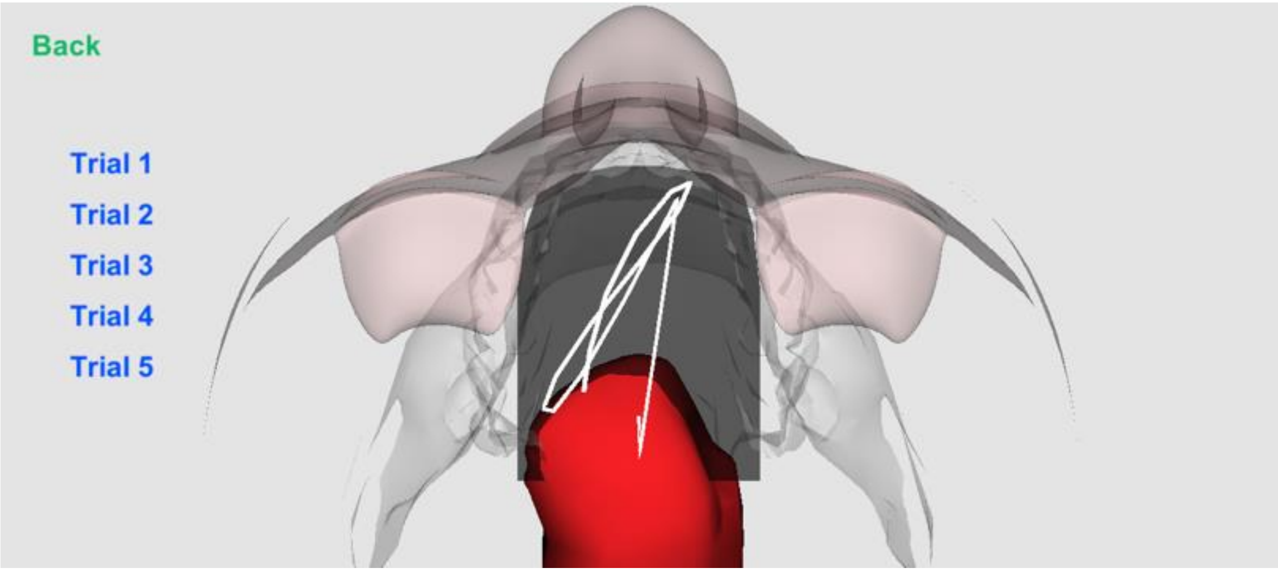

Figure 27: Free exploration menu.

4.2.5 Live Motion Menu

This menu allows the user to visualize the motion of their own tongue in real-time. To do so, the user has to wear the wearable oral retainer and use it as a computer's mouse. While the tongue slides on the trackpad, the model recreates the motion on the screen (Figure 28). The movements are limited to the trackpad plane only. The trajectory of the tongue is displayed to the user. To improve the user's feedback, in this mode the mouse cursor is not visible.

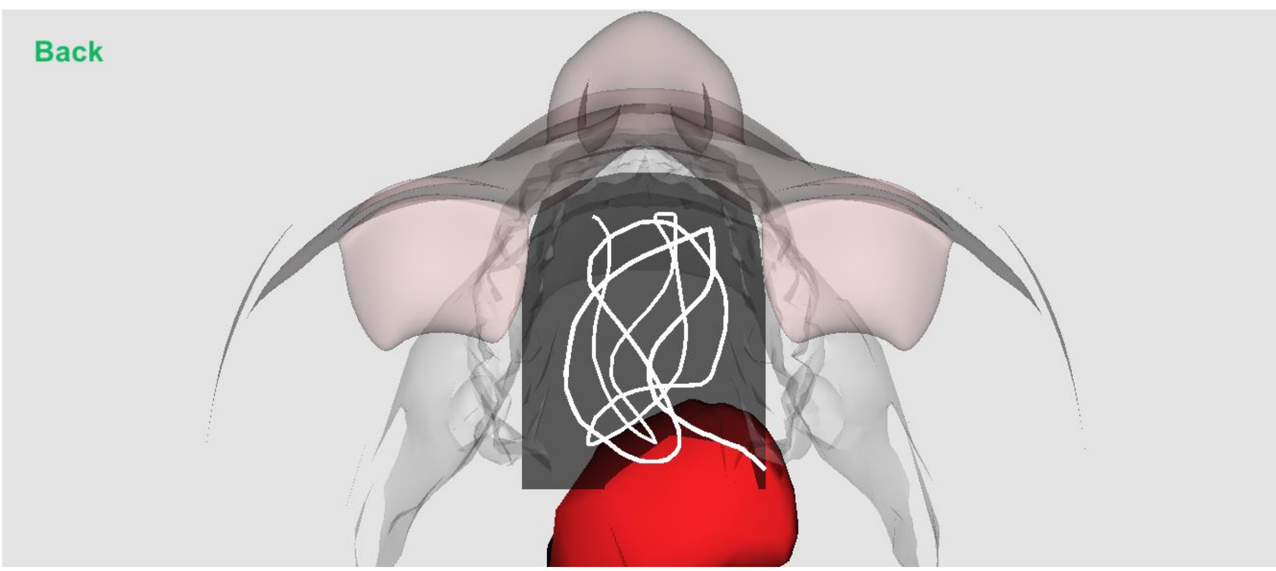

Figure 28: Tongue's real-time motion.

4.2.6 Mobile App Version

Unity allows for exportation of the application into various platforms. We exported our model into an Android phone (Samsung Galaxy A50). The application runs smoothly on the phone (Figure 29). In addition, if the mobile device is paired with our trackpad, the user could interact with all the application's functions through the wearable. This constitutes a great advantage for the patients interested in practicing at home.

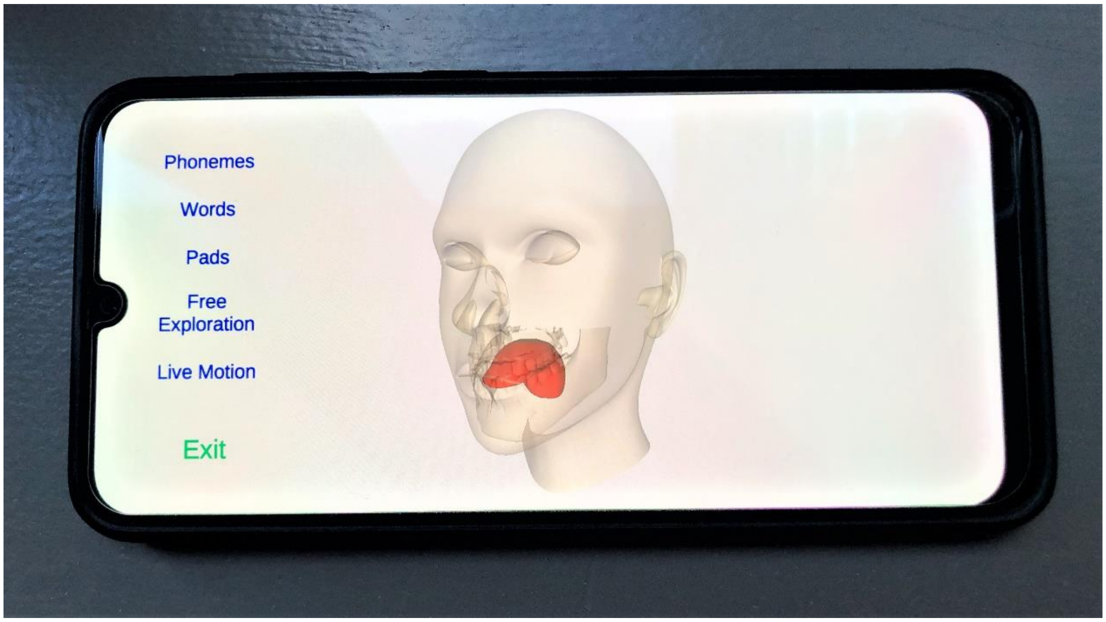

Figure 29: Mobile version of the application (Android).

CHAPTER 5

DISCUSSION AND CONCLUSION

Patients undergoing traditional speech therapy often have difficulty grasping the therapist's audible cues for the correct intended movement of the tongue [1]. This lack of clarity could be attributed to the absence of visual biofeedback. Hence, a model that would provide an augmented visual biofeedback to the patients during the therapy sessions could be vital in improving the efficiency of the therapy. We hypothesize that the visual biofeedback provided by the model is the neuromodulation component required in therapy sessions that would improve the effectiveness of the therapy. In the current work we presented on the development of a dynamic 3-Dimensional (3D) model of the human's tongue motion. These models are recreated via the Unity software in the framework of an avatars face, making them realistic and easy for the patients to comprehend. This model is a standalone platform that could also be used in conjunction with the discreet oral wearable device developed by our team. While the patient is wearing the device, they would receive live feedback about the position of the tongue. The models will also allow the therapist the opportunity to highlight points or gestures of interest to the patient. The model is compatible with both smartphones and computers.

The developed model consists of various menus. In the 'Phonemes' and 'Words' menus, the patients could learn the pronunciation of specific sounds by watching the correct tongue's motion. Also, the platform allows for the visualization of the patient's own tongue's motion in real-time.

The ‘Free Exploration’ menu offers a visual instrument for in depth analysis of the tongue’s motion. Even though we had intended to conduct user testing, unfortunately due to the Covid-19 pandemic the model was never tested on patients in the presence of a therapist. Future work will address this issue and validate the effectiveness of the model. In order to validate the motion sequence created by our model, the results could be compared to dynamic MRI or other image acquisition techniques. The model could be advanced by adding an auditory feedback to the model and playing the sounds associated with the visualized model. The model could also incorporate error augmentation of the tongue’s motion, in order to prove the effectiveness of this method in tongue’s therapy. Also, we propose that this model could be utilized in improving pronunciation of foreign languages.

CITED LITERATURE

1. Shtern, M., Haworth, M. B., Yunusova, Y., Baljko, M., and Faloutsos, P.: A game system for speech rehabilitation. *Motion in Games. MIG 2012. Lecture Notes in Computer Science*, 7660: 43–54, 2012.

2. The Editors of Encyclopaedia Britannica: Tongue Anatomy. *https://www.britannica.com/science/tongue.* [Online; accessed 10/1/19].

3. Gérard, J. M., Wilhelms-Tricarico, R., Perrier, P., and Payan, Y.: A 3D dynamical biomechanical tongue model to study speech motor control. *Research Developments in Biomechanics*, 1:49-64, 2003.

4. Takemoto, H.: Morphological Analyses of the Human Tongue Musculature for Three-Dimensional Modeling. *Journal of Speech, Language, and Hearing Research,* 44(1): 95–107, 2001.

5. Kenning, M.: The Tongue. *https://teachmeanatomy.info/head/muscles/tongue.* [Online; accessed 10/1/19].

6. Kahn, A.: Speech Disorders. *https://www.healthline.com/health/speech-disorders*. [Online; accessed 2/1/20].

7. 10 Most Common Speech-Language Disorders: *https://www.speechpathology-graduateprograms.org/2018/01/10-most-common-speech-language-disorders,* 2018. [Online; accessed 2/1/20].

8. Liu, X., Stone, M., and Prince, J. L.: Tracking tongue motion in three dimensions using tagged MR images. *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro*, 1372-1375, 2006.

CITED LITERATURE (continued)

9. Woo, J., Xing, F., Stone, M., Green, J., Reese, T. G., Brady, T. J., Wedeen, V. J., Prince, J. L., and El Fakhri, G.: Speech Map: a statistical multimodal atlas of 4D tongue motion during speech from tagged and cine MR images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, 7(4): 361–373, 2017.

10. Engwall, O., and Bälter, O.: Pronunciation feedback from real and virtual language teachers. *Computer Assisted Language Learning*, 20(3): 235–262, 2007.

11. Morgan, A. T., Liegeois, F., and Occomore, L.: Electropalatography treatment for articulation impairment in children with dysarthria post-traumatic brain injury. *Brain Injury*, 21(11): 1183–1193, 2007.

12. Gibbon, F. E., and Wood, S. E.: Visual feedback therapy with electropalatography. *Interventions in Speech Sound Disorders*, 509-536, 2010.

13. Katz, W. F., McNeil, M. R., and Garst, D. M.: Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. *Aphasiology*, 24(6–8): 826–837, 2010.

14. Bacsfalvi, P., and Bernhardt, B. M.: Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics and Phonetics*, 25(11–12): 1034–1043, 2011.

15. Ouni, S.: Tongue control and its implication in pronunciation training. *Computer Assisted Language Learning*, 27(5): 439-453, 2014.

16. Preston, J. L., Byun, T. M., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., and Whalen, D. H.: Ultrasound Images of the Tongue: A Tutorial for Assessment and Remediation of Speech Sound Errors. *Journal of Visualized Experiments*, 119: 1–10, 2017.

CITED LITERATURE (continued)

17. Heng, Q., McCabe, P., Clarke, J., and Preston, J. L.: Using ultrasound visual feedback to remediate velar fronting in preschool children: A pilot study. *Clinical Linguistics and Phonetics*, 30(3–5): 382–397, 2016.

18. Pelachaud, C., van Overveld, C. W. A. M., and Seah, C.: Modeling and animating the human tongue during speech production. *Proceedings of Computer Animation*, 40-49, 1994.

19. Engwall, O.: A 3D tongue model based on MRI data. *6th International Conference on Spoken Language Processing*, 901-904, 2000.

20. Engwall, O.: From real-time MRI to 3D tongue movements. *8th International Conference on Spoken Language Processing*, 1109–1112, 2004.

21. Engwall, O., and Beskow, J.: Resynthesis of 3D tongue movements from facial data. *8th European Conference on Speech Communication and Technology*, 2261–2264, 2003.

22. Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C., and Savariaux, C.: Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*, 30(3): 533–553, 2002.

23. Gérard, J. M., Perrier, P., and Payan, Y.: 3D biomechanical tongue modeling to study speech production. *Speech Production: Models, Phonetic Processes, and Techniques,* 6: 85–102, 2006.

24. Katz, W. F., and Mehta, S.: Visual Feedback of Tongue Movement for Novel Speech Sound Learning. *Frontiers in Human Neuroscience,* 9:612, 2015.

25. Katz, W., Campbell, T., Wang, J., Farrar, E., Eubanks, J. C., Balasubramanian, A., Prabhakaran, B., and Rennaker, R.: Opti-Speech: A real-time, 3D visual feedback system for speech training. *15th Annual Conference of the International Speech Communication Association*, 1174–1178, 2014.

CITED LITERATURE (continued)

26. 3D Slicer: *https://www.slicer.org.* [Online; accessed 10/1/19].

27. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin. J-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F. M., Sonka, M., Buatti, J., Aylward, S. R., Miller, J. V., Pieper, S., and Kikinis R.: 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30(9):1323-41, 2012.

28. Kikinis, R., Pieper, S. D., and Vosburgh, K.: 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. *Intraoperative Imaging Image-Guided Therapy*, 3(19):277–289, 2014.

29. Lasso, A., MacNeil, K., and Fedorov, A.: SegmentEditorExtraEffects. *http://slicer.kitware.com/midas3/slicerappstore/extension/view?extensionId=293582*, 2019. [Online; accessed 10/1/19].

30. Hopkin, G. B.: Neonatal and Adult Tongue Dimensions. *The Angle Orthodontist*, 37(2): 132-133, 1967.

31. Free3D: Human Skull. *https://free3d.com/3d-model/skull-human-anatomy-82445.html*. [Online; accessed 10/1/19].

32. Turbosquid: Young Female Head. *https://www.turbosquid.com/3d-models/free-young-female-head-3d-model/914517*. [Online; accessed 10/1/19].

33. Unity: SplineMesh asset. *https://assetstore.unity.com/packages/tools/modeling/splinemesh-104989*. [Online; accessed 2/1/20].

34. Tamari, K., Murakami, T., and Takahama, Y.: The dimensions of the tongue in relation to its mobility. *American Journal of Orthodontics and Dentofacial Orthopedics*, 99(2): 140–146, 1991.

CITED LITERATURE (continued)

35. Rezende, B. A., Furlan, R. M. M. M., de Las Casas, E. B., and Motta, A. R.: Clinical assessment of the tongue in healthy young adults. *Revista CEFAC*, 18(3): 559–567, 2016.

36. Miller, N. R., Reyes-Aldasoro, C. C., and Verhoeven, J.: Asymmetries in tongue-palate contact during speech. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1734-1738, 2019.

VITA

NAME	Davide Lacca
EDUCATION	
2019 – Current	Master of Science in Bioengineering, University of Illinois at Chicago, USA
2018 – Current	Master’s Degree in Bioengneering, Politecnico di Milano, Italy
2015 – 2018	Bachelor’s Degree in Biomedical Engineering, Politecnico di Milano, Italy
LANGUAGE SKILLS	
Italian	Native speaker
English	Full working proficiency
	A.Y. 2019/20 Double degree program abroad in Chicago, Illinois
	A.Y 2018/19 – 2019/20 Lessons and exams attended exclusively in English
	August 2018 – TOEFL iBT examination
SCHOLARSHIPS	
A.Y. 2019/2020	Italian scholarship for excellent students at Politecnico di Milano
A.Y. 2018/2019	Italian scholarship for excellent students at Politecnico di Milano
TECHNICAL SKILLS	
	3ds Max, Slicer, Unity, C, C#, Matlab, Arduino, Processing, EAGLE