3D Tongue Reconstruction from Two Orthogonal Ultrasound Images

Ava Ahadipour

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements for the degree

Master of Applied Science in Electrical and Computer Engineering

Ottawa-Carleton Institute for Computer Science

School of Information Technology and Engineering

University of Ottawa

Ottawa, Ontario, Canada

May 2013

Abstract

Tongue is the most important articulator of speech. Given the dominant role of speech in human interactions, a large proportion of the research on tongue movement focuses on its role in speech, specifically vowel production. A three-dimensional (3D) tongue model could provide important visual information that would help an individual to correct the tongue position during speech.

Researchers have been tried using different imaging techniques such as MRI, CT and Ultrasound imaging to capture the tongue information. Among them ultrasound imaging is, recently, getting popular in the field of linguistic, phonetic research and medical imaging. Having access to real-time display of the movement of the tongue, non-invasive measurement with relatively cheap equipment and also avoiding risks of x-rays or other types of radiation made us use this method for capturing set of required data.

We introduce a new method of reconstructing the 3D generic tongue model. The novelty of our method is employing only two ultrasound images for tongue reconstruction purpose. We presented a feature-based deformation method that could represent the different shapes of the tongue during speech with no computationally expensive requirements.

Tongue is a difficult target for deformation due to its muscle complexity. The fact that tongue is concealed in the oral cavity can also provide additional challenges to capture the whole image of it using available methods such as MRI, X-Ray and ultrasound imaging. We overcome these difficulties by: (a) Analysis of tongue surface data using ultrasound imaging; Capturing tongue shape and tracking the tongue surface data for different sets of

pronunciations from obtained ultrasound images (b) Deformation of 3D generic tongue model using deformation method named Radial Basis Functions(RBFs) based on two orthogonal (side and front view) ultrasound images.

We tested our methodology on a set of Ultrasound images received from Department of linguistics. We selected different sample of Ultrasound frames from a set of mid-coronal and corresponding mid-sagittal images. The Ultrasound data is collected while a male person was pronouncing 1 to 10 and then sustained "i", "u", "e", "o", "a", "s", "sh", "l", "n", and "r".

Our model is capable of representing tongue shapes during the production of vowels and also the general motion of tongue. A set of control points of the 3D tongue neutral model, allows each tongue segment to be easily captured.

In order to test the capability and sufficiency of the proposed method in terms of deformation, we have compared our results (deformed tongue models based on arbitrary set of control points on mid-sagittal and the mid-coronal line of the 3D tongue model) with the results from an open-source tongue simulation system named “Artisynth” (deformed tongue models obtained from different activations of tongue muscles).

Acknowledgements

This dissertation would not have been possible without the help of so many people in so many ways.

First and foremost, my utmost gratitude goes to my supervising professor, Dr. Won-Sook Lee for sharing her invaluable knowledge, advice and guidance.

I would also like to thank Dr. Jeff Mielke (Department of Linguistics) supplying us with our data and for being all-around helpful.

I thank my lab-mates, Nima, Mauricio, Gabriel, Adriana, Ali, Emre, Di, Saleh, Muyiwa, Zhongrui, Iman, Sadaf and Ramin who helped to make my time in graduate school a wonderful experience.

Elena, Lillian, Judy, Ravi, Sunset, Devon, many thanks for being there through it all.

Many thanks to:

Peyman - Through good times and bad, your kindness and extensive support have been ever-present in this important time of my life, for which I am eternally grateful.

Aida - Thank you for being a terrific friend.

and

Seyed - For making so many things possible.

Special mention goes to my mother and father whose gifts I can never fully repay- and yet shall forever endeavor to try.

My sister and my little brother, Nava and Amin, you have been my best friends all my life and I love you dearly. Thank you for all your support.

This thesis is dedicated to the memory of my grandmother. I miss you every day….

Table of Contents

List of Figures

List of Tables

Glossary of Terms

2D Two Dimensional

3D Three Dimensional

CT Computed Tomography

MRI Magnetic Resonance Imaging

RBFs Radial Basis Functions

ACMs Active Contour Models

Chapter 1 Introduction

1.1 Motivation

Ultrasound (Ultrasonic) imaging also called ultrasound scanning or sonography involves the use of a small transducer (probe) and ultrasound gel to expose the body to high-frequency sound waves.

Ultrasound imaging is getting popular in the field of linguistic, phonetic research and medical imaging [Sonies et al. 1981] [Keller & Ostry 1983] [Stone et al. 1983] [Stone & Shawker 1986] [Stone & Lundberg 1996] [Hiiemae & Palmer 2003] [Lundberg & Stone 1999] as it produces real-time display of the structure and movement of the internal organs. It includes non-invasive measurement with relatively cheap equipment. It does not have the same risks as x-rays or other types of radiation since it does not have ionizing radiation [Watkin 1999] [Tang & Hamarneh 2010]. Reasons mentioned above make ultrasound imaging an appropriate approach of acquiring required information.

Figure 1 shows a typical ultrasound machine and transducer. Analysis of tongue ultrasound images is helpful to extract information about tongue shape and dynamics. The achieved information can be used for studying the effects of aging on speech, tongue modelling, speech perception and speech therapy [Watkin 1999] [Akgul et al. 1998] [Tang & Hamarneh 2010]. Also tongue shape modeling and tracking its motion can provide useful information for medical diagnosis. Based on the advantages about Ultrasound imaging

mentioned above, ultrasound imaging started to be considered to be one of the most capable and practical methods to capture the tongue information [Davidson 2006] [Davidson 2005] [Tang & Hamarneh 2010].

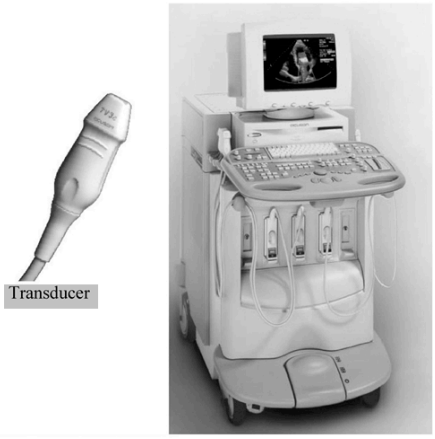


Figure 1: Sample Ultrasound machine and transducers [Honggang et al. 2011].

Our research has focused on 3D tongue reconstruction method that uses only two orthogonal ultrasound images (One mid-coronal image and one mid-sagittal one (Figure 2)). The most popular systems apply mid-sagittal ultrasound images to study the tongue shape, although coronal slices have also been analyzed [Davidson 2006]. However, 3D shape is more understandable than 2D images without doubt. Thus we proposed 3D tongue reconstruction method based on two orthogonal ultrasound images. Using only two images for the purpose of 3D reconstruction of the tongue is considered as minimal set of information [Lundberg & Stone 1999].

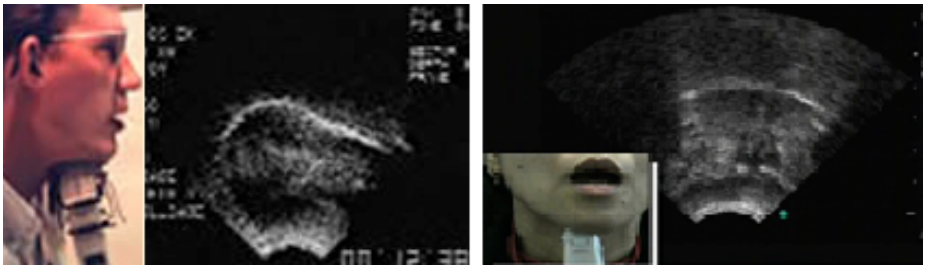


Figure 2: Captured ultrasound images and face direction. (Left) sagittal/side View [Vocal Tract Visualization Laboratory, n.d.]; (Right) the coronal/front view. The right image (coronal view) is selected from our own data set.

The advantage of using the proposed method of using two ultrasound images for 3D tongue reconstruction is the capability of producing 3D tongue deformed models using the ultrasonic machine quick capturing using two ultrasound views at the same time, which is very useful for real-time capturing and later on for real-time applications. The proposed 3D tongue deformation method can be the essential component in constructing a unique design of ultrasound devices which are capable of providing a 3D model of the human tongue merely by exposing the patient to the ultrasound rays once. Current setting makes it possible to use two scanning probes without interfering each other to capture two views simultaneously in order to produce 3D model. A possible commercial product can provide a simple portable capturing ultrasound machine to capture once and produce a 3D model. Apart from the safety advantages, and considering the fact that Ultrasound portable devices are the fastest growing industry in the medical ultrasound industry [Thompson 2010] the model produced, can make the future ultrasound machine product competitive in the market as well as extending the research applications in various fields. Having a real-time tongue

simulation system also expands the range of possibilities for exploring "what-if" scenarios for pre-operative planning. Also 3D deformation of tongue during speech production can be really useful in training users about correct pronunciation of different phonemes in different languages.

Our method is motivated by well-known face modeling method using two orthogonal set of pictures [Lee 2000].

1.2 Problem Statement

This thesis addresses the problem of developing an algorithm of 3D reconstruction of tongue shape using 2 orthogonal ultrasound images as an input.

In the field of 3D modeling of tongue there are some elements that complicate reconstruction.

These complications include:

- Capturing suitable set of 2D ultrasound tongue images
- Tongue surface tracking from noisy [Stone & Lundberg 1996] ultrasound images (correct surface diagnosing) which is a crucial component to tongue shape analysis
- Reliable reconstruction method capable of modeling a shape of tongue

In the face of these difficulties, the primary question this research seeks to answer is:

- How can the 3D model of the tongue be reconstructed and deformed from two 2D ultrasound images?

  In addition, to the driving question of this research, we seek to answer the following secondary question:

- Are the 3D models created from our method acceptable enough to aid in the analysis of tongue shape and dynamics for applications such as visual speech training systems?

1.3 Proposed Solution

This section details the methodology we use to solve our primary question in the face of the previously listed complications:

1. Since we only need one mid-coronal and one mid-sagittal 2D ultrasound images, selecting only two ultrasound images from a sequence of frames in which tongue surface is easy to track is noticeably an easier task compared to the methods that need multiple Ultrasound slices.
2. After selecting the 2D images, the next step is tracking the surface of the tongue from those. For this purpose we use a manual extraction of tongue surface. We have a user (linguistic specialist) select points on the surface of the tongue from tip to root on two ultrasound images in order to determine the surface of the tongue.

   Linear interpolation of these points entails the surface of the tongue in two ultrasound images. One represents the tongue's surface in coronal view while the other covers the

surface from sagittal view. This concludes the manual portion of our tongue surface tracking strategy.

Although manual surface extraction is time consuming, it is helpful in terms of avoiding the hardness of making mistake in diagnosing the surface of the tongue in an ultrasound image with unrelated high-contrast edges. Another benefit is having an accurate result and in this case the accuracy is validated by a linguistic with relevant expertise. These points we called them "control points" of Ultrasound images.

In case of having the sequence of Ultrasound images, we have benefited from the famous automatic tongue surface tracking software named "EdgeTrak" which is helpful and necessary to avoid time consuming manual surface extractions [Vocal Tract Visualization Laboratory, n.d.].

3. Deform the neutral tongue mesh model into the shape of provided ultrasound input images using RBFs (a well-established method of interpolating scattered data) [Botsch & Kobbelt 2005] [Wright 2003] [Matsopoulus et al. 2005].

1.4 Thesis Organization

The rest of this thesis is organized as follows: Chapter 2 contains literature presenting background information of ultrasonic imaging, ultrasound image features, tongue surface tracking and 3D tongue reconstruction methods.

Chapter 3 performs an overview of our objective, the complications associated with that objective and our proposed solution.

Chapter 4 describes our method in detail.

Chapter 5 lists our qualitative and quantitative experimental results followed by evaluation and discussions.

Chapter 6 marks our conclusion along with the direction of future research.

Chapter 2 Literature Review

2.1 Tongue Anatomy

Muscle fibers, internal fats, nerves and vessels constitute tongue organ. Styloglossuns, Hyoglossus, genioglossus, palataglossus and chondrglossus are the extrinsic muscles and they are in charge of placing the tongue in the oral cavity. Superior longitudal, inferior longitudal, transverse lingualis and vertical lingualis are intrinsic muscles and tongue shape deformation is a result of their activation. Tongue floor muscles are mylohyoid and geniohyoid [King & Parent 2001] [Fang et al. 2008] [Takemoto 2001] [Bateman & Mason 1984] [Gray 1977].

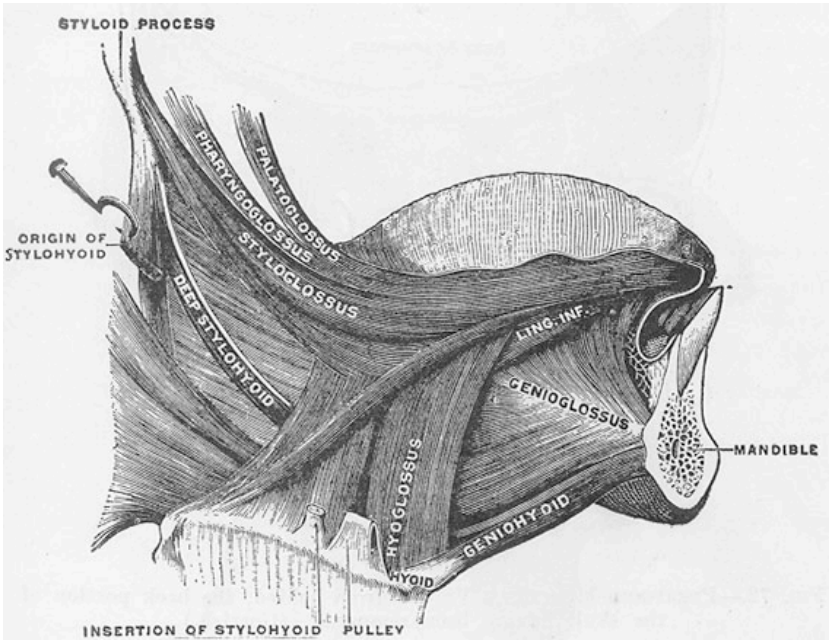


Figure 3: Tongue Muscles [Hager, n.d.].

Complex muscles of tongue (extrinsic and intrinsic [Kent 1997]) enable it to create diverse shapes. Figure 3 represents position of some tongue muscles (Stylohyoid, Styloglossus, Hyoglossus, Genioglossus, Pharyngoglossus, Palatoglossus and Deep Stylohyoid).

Mid-sagittal and mid-coronal perspective of the tongue and hard palate from MRI images are shown in Figure 4 (a) and (b).

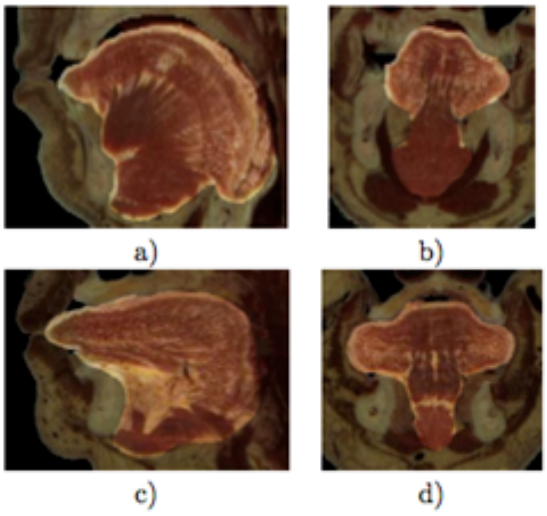


Figure 4: Tongue Images from the visible human project National Library of Medicine. a) A mid-sagittal view of the man. b) A coronal cross-section of the man. c) A mid-sagittal view of the woman. d) A coronal cross-section of the woman.

As it is observable in the figure, tongue has different shapes and volumes among individuals [King & Parent 2001].

2.2 Ultrasonic Imaging

In this section we will discuss some of the fundamentals of ultrasonic imaging along with their application to the imaging of tongue.

Real-time 2D images of tongue gesture can be evaluated using medical ultrasound imaging [Watkin 1999] [Tang & Hamarneh 2010] [Engwall 2002] [Keller & Ostry 1983]. There are several reasons that emphasis the significant role of ultrasound imaging in the assessment and evaluation of tongue information [Tang & Hamarneh 2010] [Li et al. 2004] [Akgul et al. 1998] [Watkin 1999]. Benefits comprise real-time 2D images (real time tongue visualization), having safe and harmless ray and capability of being repeatedly utilized, being accessible, portable and user friendly [Chi-Fishman 2005]. Distinct advantages mentioned above, make ultrasound imaging the reasonable and convenient choice for the tongue information assessment process.

Ultrasound images of the tongue are obtained by placing the ultrasound transducer in different positions such as under the chin projecting the ultrasound signals toward the surface of the tongue. The positions of ultrasound probe associated with corresponding sagittal and coronal ultrasound images are depicted in Figure 5.

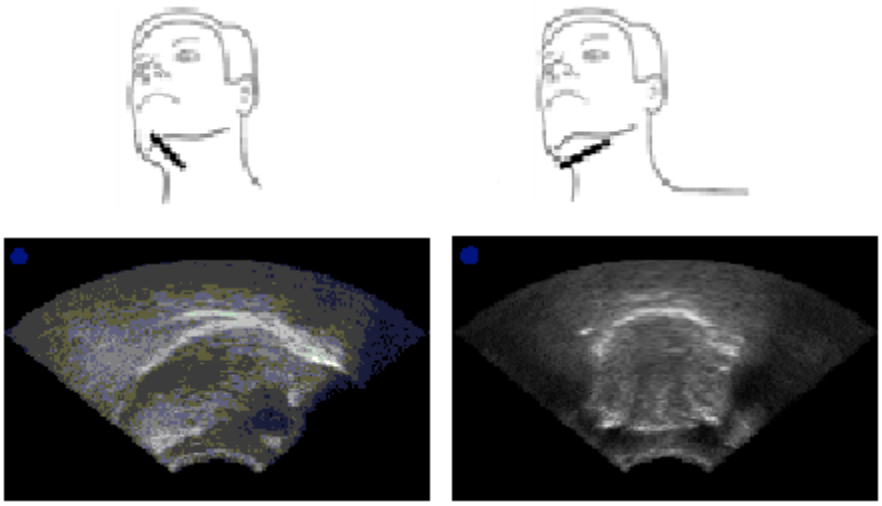


Figure 5: Captured ultrasound images: (Left) sagittal ultrasound acquisition (Right) the coronal view and position of ultrasound probe for each of them. In the left image the root area is toward the left side of image and the tip of the tongue is in on the right side.

Tongue information obtained from ultrasound images can be employed for different purposes such as speech therapy, (mend incorrect tongue movement in speaking), speech research [Kelsey et al. 1969] [Engwall 2002], swallowing research [Chi-Fishman 2005], tongue animation [Ilie et al. 2012] [King & Parent 2001] and other applications.

2.3 Tongue Surface Tracking

In the direction of 3D reconstruction of tongue model from 2D ultrasound image sequences, extraction and tracking of the tongue surface is an essential step.

There are several approaches to accomplish this task.

One of the complications of automatic extraction of tongue surfaces from ultrasound frames is the noise level of ultrasound images which makes tongue surface detection a very difficult task. In order to solve this problem snake method (Active Contour Models (ACMs)) has been utilized [Kass et al. 1998].

In the snake method the energy minimization approach has been used. There, two terms were defined: internal energy and external energy. Minimization aim for internal energy is to attain smooth and continuous curves (surfaces). This step is a crucial component in a process of estimating contour (this contour can be surface of the tongue) positions in a noisy and full of decorrelated edges image of ultrasound. The external energy is related to the gradient of the image. Since ultrasound images are noisy it is not possible to use only gradient information to extract the tongue surface. The method of ACMs [Chan et al. 2000] is utilized to evolve the initial contour to attach to the maximum homogenous region or to constrain the homogeneity of intensity in a region.

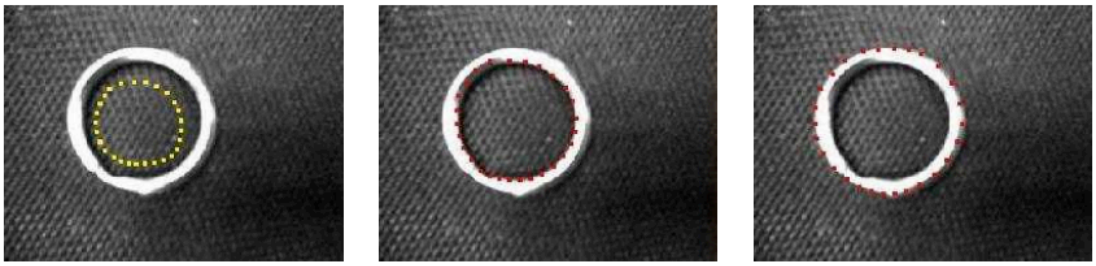


Figure 6: Contour extraction in a key-chain ring using ACMs. (a) Snake initialization. (b) Red dots show extracted edge using ACMs without band energy. (c) Extracted edge using band energy method proposed in [Li et al. 2004].

However ACMs [Chan et al. 2000] have some limitations:

1. Useful for closed contours (ring example shown in Figure 6). For tongue surface tracking there is an open contour and since there is no enclosed region, it is not possible to apply constraint of homogeneity of intensity.
2. Obtaining the exterior (outer) edge (Figure 6 (c)) in some cases like closed ring (Figure 6) will be failed since the internal region is more homogenous.

Later, combination of edge (contour) gradient and intensity in specific regions (not only enclosed regions) is proposed as a solution to the problem mentioned above [Li et al. 2003] [Li et al. 2004]. The intensity difference is a key for distinguishing inner or outer edges, which in tongue surface tracking, these terms are respectively related to lower surface or upper surface of the tongue (Figure 7) in an ultrasound image.

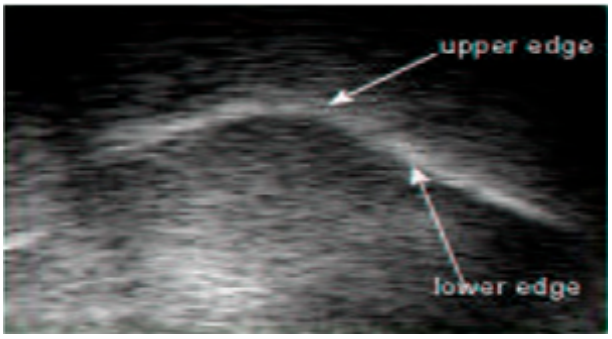


Figure 7: Lower edger and upper edge in a sample ultrasound image of the tongue.

A region based band energy is introduced which benefits from new external energy (gradient as a snake element). There is also intensity information (neighbors around the snake element). In this approach the unrelated contours and edges will be omitted.

Figure 8 illustrates the performance of the region based band energy method.

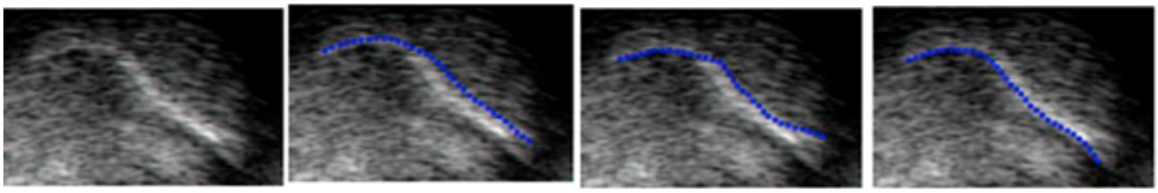


Figure 8: Result of tongue surface tracking using (b) snake initialization (c) surface extraction without using band energy (d) surface extraction with band energy (a) in an ultrasound image.As it is observable from the Figure 8, some snake elements are attracted to the uninteresting high gradient upper edge of the air reflection.

Another approach of tongue surface tracking is a system known as "EdgeTrak" [Li et al. 2004] which employs ACMs with and intensity of local regions so that the partial tongue surfaces can also be tracked. The contour orientation is also considered in order to avoid unrelated high intensity edges in ultrasound images. EdgeTrak introduces an automatic approach for tongue surface extraction from ultrasound images. Applying this system, tracking the surface of the tongue in a sequence of ultrasound images is accessible. In EdgeTrak system, user needs to delicate initial points on a tongue surface in a single frame (initial frame) of ultrasound image. Using B-spline interpolation of the selected points, a

contour along the surface of the tongue can be obtained. Then an optimization process is utilized to approximate the obtained contour to the tongue surface in a given ultrasound frame. Concerning the approximation of the contour to the tongue surface in adjacent frames, the optimized contour of a current frame can be exploited and for this purpose the optimization process needs to be applied afresh in adjacent frames. The outcome is series of tongue surfaces in a sequence of ultrasound images (Figure 9).

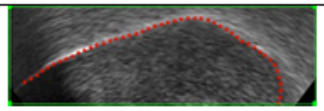
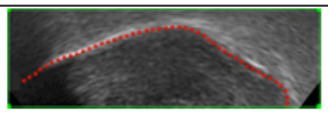
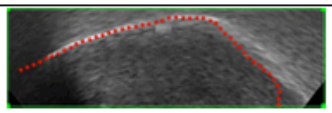
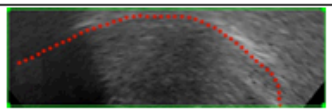
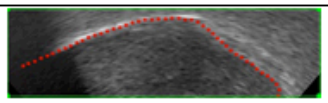
FRAME NUMBER	RESULT FROM EDGETRAK
1	
2	
3	
4	
5	

Figure 9: Tongue surface tracking using EdgeTrak for an ultrasound sequence of 5 frames.

EdgeTrak has the following benefits:

1. Available for public.
2. Useful for comparative analysis [Aron et al. 2008] [Tang & Hamarneh 2010].

3. Validated by specialists as an accurate system for different applications.

There are also some limitations about EdgeTrak. There is a need for manual refinement due to use of local approach (using the current contour for initialization in next frame).

Another method of tongue surface tracking is proposed by Tang et al. [Tang & Hamrneh 2010]. The semi-automatic approach named "graph-based" is proposed for tracking of the human tongue in an ultrasound image sequence. The tracking problem was solved as a graph-labeling problem. The user selects points (control points) on the ultrasound image. In a graph-labeling approach each control point is correspondent to a vertex in a graph and the goal is to find the displacement value of the vertex in adjacent frames. In construction of the graph they have defined vertices and edges [Tang & Hamrneh 2010]. Vertices are the contours of frames and edges of the graph are defined as a correspondence between points in consecutive frames. Thus inter-frame and intra-frame (spatial and temporal) connectivity is considered.

Adjustment of the determined displacement values is evaluated for neighboring control points in a current contour and current frame (spatial) and also for correspondent control points across different frames (temporal).

One of the ideal results expected from each tongue surface tracking method is to attach the contour obtained by control points to the surface contour. In [Tang & Hamarneh 2010] this goal is considered as data energy. The other aim is to keep the achieved contour smooth and continuous, which is referred as regularization energy. The reason of introducing

the"data energy" and "regularization energy" is to find a way of optimization to attribute translocation vector to a vertex. Data energy term applies force to control point in order to get it close to image edge and regularization energy arranges the smoothness of the contour and ensure sticking and clinging movement of segemented contour between neighbor frames.

A comparison between tongue surface tracking using graph-based method and EdgeTrak system is illustrated in Figure 11.

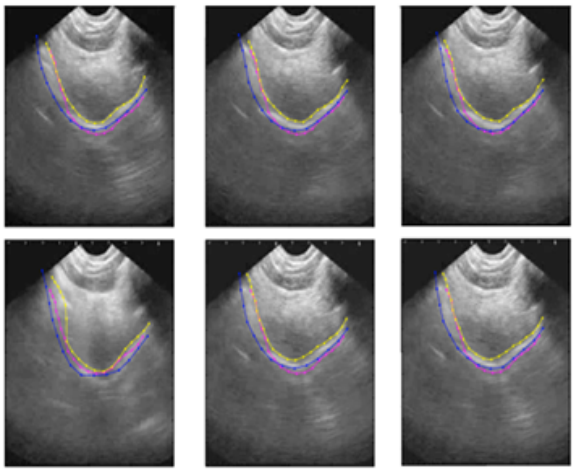


Figure 10: Results generated using EdgeTrak (blue line), graph-based method (pink line) and ground truth (yellow line) [Tang & Hamarneh 2010].

As it is observable in Figure 11, graph-based method (pink lines) provides more accurate segmentations than those obtained with EdgeTrak (blue lines). However graph-based method is a semi-automatic method and there is a need for manual adjustment of parameters

which makes it hard to use in cases of having multiple ultrasound frames. For this reason we use “EdgeTrak” as an automatic approach of tongue surface tracking.

2.4 3D Tongue Reconstruction

Generally there are two methods for creating 3D ultrasound volumes:

1. Attainment of 3D volume from 2D images (using 2D ultrasound probe for image production and employing different ultrasound image algorithms on 2D images to achieve 3D volume).
2. Usage of 3D mechanical probes (specific 3D ultrasound probes).

The choice of a rational set of coronal and sagittal contours is a critical step in sparse reconstruction of tongue [Lundberg & Stone 1999]. The goal of our work is to obtain a reasonable 3D tongue model from only one mid-coronal and one mid-sagittal slice. In the following we explained some suggestions regarding usage of different number of ultrasound slices for the purpose of 3D reconstruction of tongue [Lundberg & Stone 1999] [Stone 2005] [Badin et al. 2002] [Stone & Lundberg 1996].

Based on the work done by Stone and Lundberg [Stone & Lundberg 1996] and Lundberg and Stone [Lundberg & Stone 1999], exploiting the information of 5 to 6 coronal slices qualified as a sufficient number of 2D Ultrasound slices for 3D reconstruction of tongue surface with 80% surface coverage. Adding more Ultrasound slices will increase the reconstructed surface coverage [Bressmann 2008]. Figure 11 depicts reconstruction

outcomes [Lundberg & Stone 1999] from compressed data set and also from six coronal slices for vowel /I/.

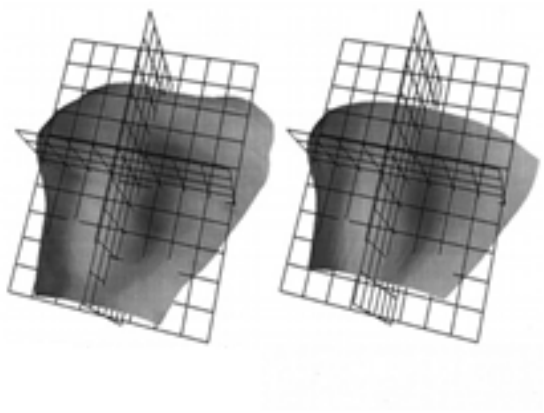


Figure 11: Comparison **[**Lundberg & Stone 1999] of surface coverage from compressed (full) data set (left) and from six coronal slices (sparse data set) (right) for vowel /I/.

Lundberg et al. [Lundberg & Stone 1999] also mentioned that applying linear interpolation method on obtained mid-coronal and mid-sagittal slices is a possible approach for 3D tongue surface reconstruction [Lundberg & Stone 1999]. Figure 12 represents tongue surface coverage based on the number of Ultrasound slices [Lundberg & Stone 1999].

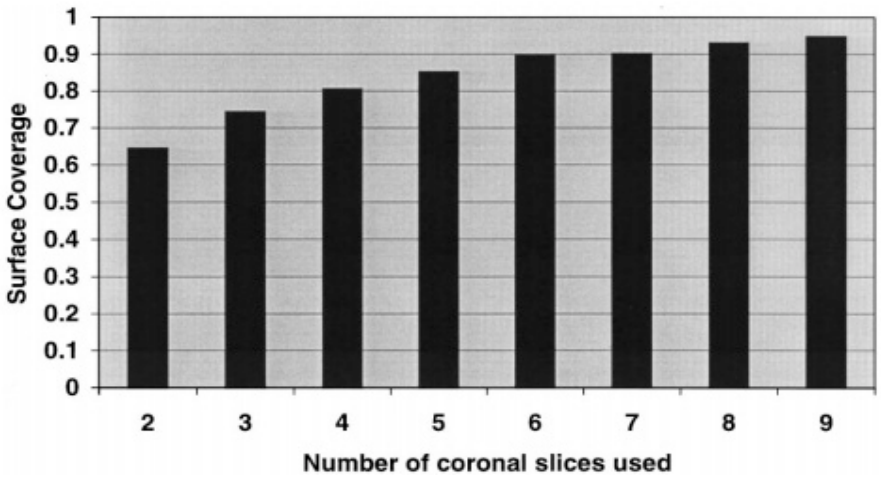


Figure 12: Tongue surface coverage based on the number of 2D Ultrasound slices [Lundberg & Stone 1999].

Olov Engwall [Engwall 2000] used MRI images of different articulations to fulfill 3D tongue reconstruction task. MRI images acquired in three groups (Figure 13) [Engwall 2000]:

1. Horizontal images of pharynx
2. Diagonal images of velar region
3. Coronal slices covering oral cavity

The outcome of manual contour derivation process is a set of 42 contours. In the next phase, remodeling of resulted contours to provide the spaced points was completed. Connection of spaced points provided a 3D Mesh model (each point connected to the adjacent point) as shown in Figure 14.

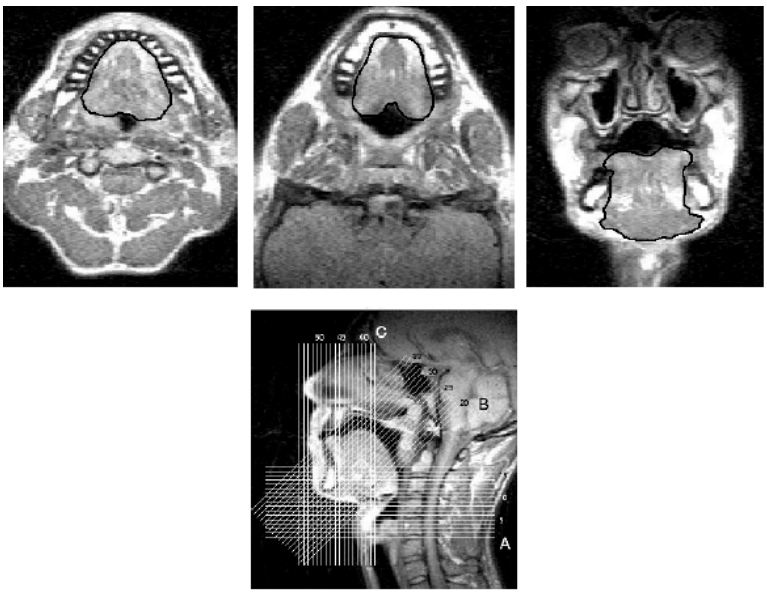


Figure 13: Three sets of MRI images (upper row) provided from three different directions (lower image) [Engwall 2000].

The neutral 3D model for subject at rest position was considered as reference model and articulations are modeled as deformation from this model. In order to fulfill this modeling there are some parameters that are defined as following: “The parameters are defined using the activation A (-1<A<1) of the movement of a prototype vertex (P) towards a target vertex (T) and a weight vector, determining the influence of the parameter on every vertex of the mesh” [Engwall 2000].

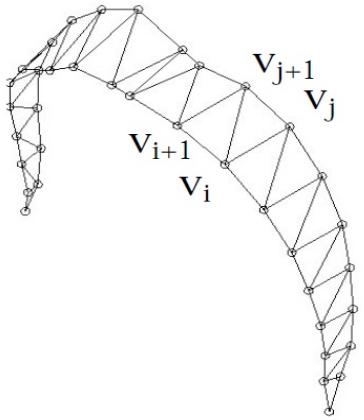


Figure 14: Reconstruction of the 3D polygon mesh [Engwall 2000].

Translation of mesh vertices is a deformation method [Engwall 2000] and this deformation is defined as a displacement of each vertex $(\Delta x_i, \Delta y_i, \Delta z_i)$ due to an activation parameter (A) and a weighting parameter for that vertex (i):

$$(\Delta x_i, \Delta y_i, \Delta z_i) = A \bullet W_i \bullet (T_x - P_x, T_y - P_y, T_z - P_z) \qquad \textit{Eq. 2.1}$$

Six tongue parameters are included in Olov's model [Engwall 2000]: Jaw Height (JH), Tongue Body (TB), Tongue Dorsum (TD), Tongue Tip (TT), Tongue Advance (TA) and Tongue Width (TW) (Tongue width is orthogonal to mid-sagittal slice). Mid-sagittal deformation is covered by first five tongue parameters. Using each tongue parameter, certain set of vertices of the 3D mesh model are displaced from the prototype to the target based on Eq. 2.1.

Another 3D reconstruction approach was introduced by Fang et. al. [Fang et al. 2008]. Fang et. al have benefited from the Physiological articulatory models to achieve muscles'

activations information for specific articulation. The mesh model they have used which is presented in Figure 15 is obtained from volumetric MRI images with respect to the method proposed in [Dang & Honda 2001].

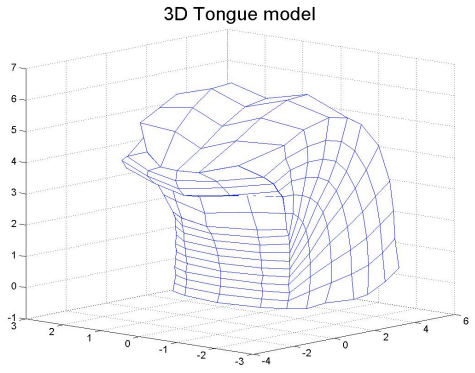


Figure 15: 3D Neutral Tongue Model [Fang et al. 2008].

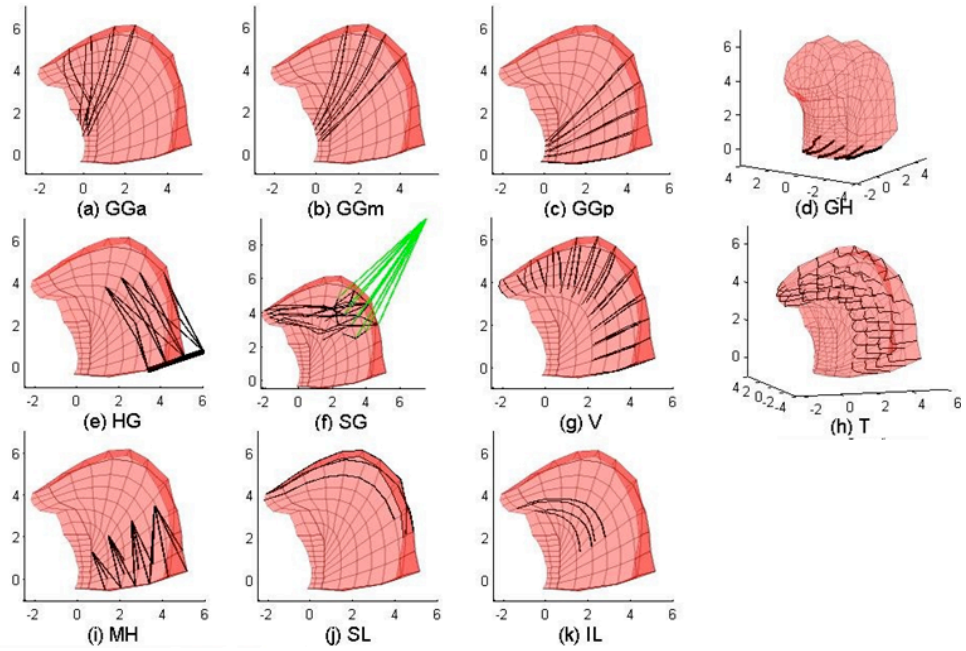


Figure 16: Representation of tongue muscles' positions (black lines) in 3D mesh model [Fang et al. 2008].

Nine tongue muscles were included in their method (Figure 16) [Fang et al. 2008]:

1. Exterior muscles: Genioglossus (GG), Styloglossus (SG), Hyogossus (HG).
2. Intrinsic muscles: Superior Longitudinal (SL), Inferior Longitudinal (IL), Transvers (T), Verticalis (V).
3. Tongue floor muscles: Mylohyoid (MH), Geniohyoid (GH).

Also effective muscles related to jaw movement are introduced:

1. The jaw-closer group (jawCL)
2. The jaw opener group (jawOp)

There, tongue deformation is defined by "functional relationship" between defined muscles. The 3D difference between the real shape of the tongue and the simulated tongue model for the specific vowel needs to be minimized and this task is done under the "cost function" in terms of muscle activation. Basically for any vowel production, related muscle activations are regularized automatically using an optimization method named "cost function" (Fang et al. 2008).

Manual adjustment of activation parameters for tongue muscles is employed as a method for the minimization of "cost function". The proper muscle activation is the one which minimizes the difference between lengths for the real shape and the simulated one. Also, In order to find the similarity between the simulated model and the real tongue shape curve

fitting for the tongue surfaces in both models in the mid-sagittal palne using a grid line system has been used.

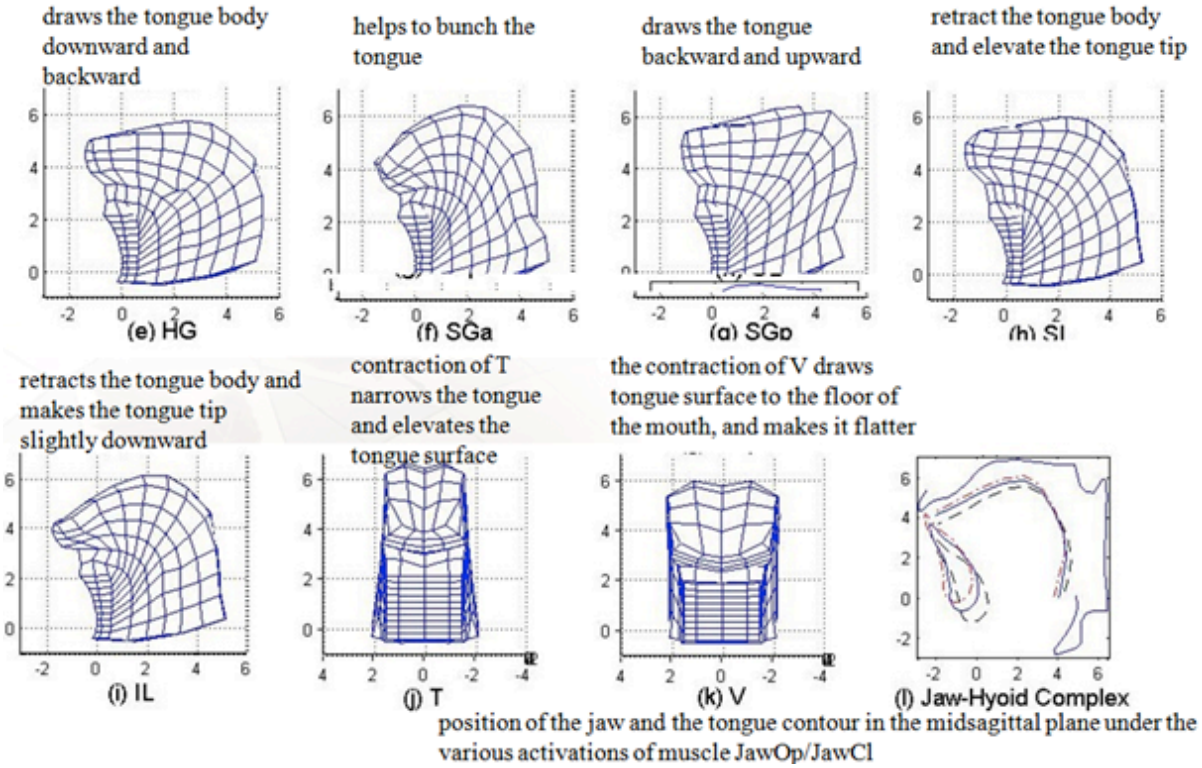


Figure 17: Representation of the effect of tongue muscles' activation on the deformation of the tongue [Fang et al. 2008].

Figure 17 represents the effect of selcting different values for activation parameters in different tongue sections [Fang et al. 2008].

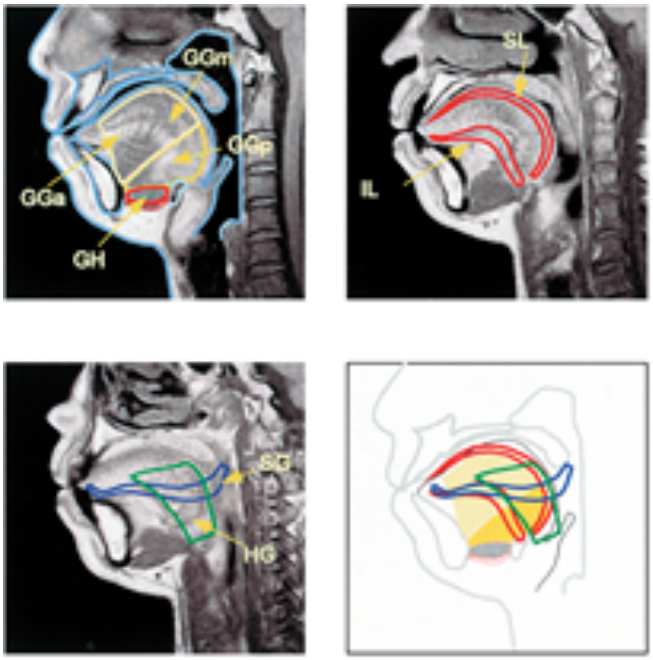


Figure 18: Muscles extraction from MRI images [Dang & Honda 2001].

Different combination of muscles' activation can be used to create different tongue shapes. Extraction of muscles' boundaries (colored lines in Figure 18 represents different boundaries for different tongue muscles) was completed using MRI images and combination of those resulted in a major muscle [Dang & Honda 2001].

The number of tongue muscles used for 3D reconstruction of tongue is different based on the application. In [Dang & Honda 2001] eleven muscles has been selected and two forces for tongue deformation are defined: (a) muscle contraction and (b) contact of tongue with palate and surrounding anatomical structures .The software of this method is PHAROS (Physiological Articulatory Organ Simulator) which is available for research applications. Investigation about tongue model, physical properties of the tongue and its muscles and

adjustment of tongue model are some of the available options in this software. In order to simulate tongue behavior during speech, 3D articulatory model (Figure 19) has been introduced where tongue considered as a symmetric model. The proposed model includes all the surrounding organs that have effects on the tongue deformation (Jaw, Hyoid bone and vocal tract wall) [Dang & Honda 2001].

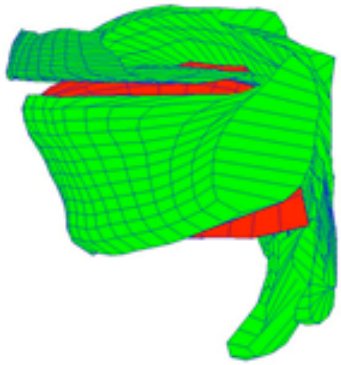


Figure 19: 3D articulatory model [Dang & Honda 2001].

3D modelling of tongue was fulfilled by a famous method of Finite Element Method (FEM) [Kakita et al. 1985] [Wilhelms-Tricarico 1995] [Gerard et al. 2003]. Using FEM analysis tongue is modeled as a set of detached finite elements. At the boundary of each section or finite elements, there are nodal points and masses are distributed at these nodal points. In this way tongue tissue is modeled as a "mass-spring network" [Dang & Honda 2001].

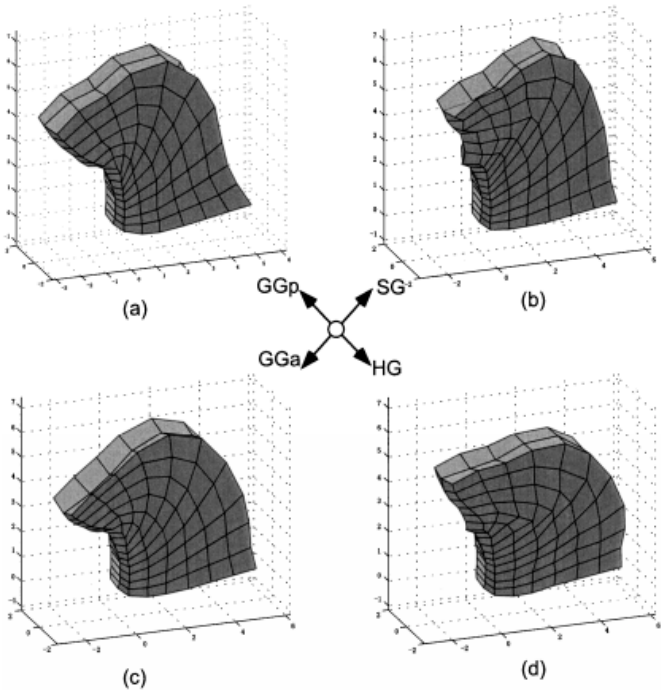


Figure 20: Tongue deformations as a result of exciting 4 muscles (a) Tongue dorsum moves forward and rises up by exciting GGp (b) rises up and sets back by SG (c) lowers and groove BY GGa (d) sets back and lower by HG [Dang & Honda 2001].

The tongue tissue and the surrounding structure were included in a motion-equation system and as a prerequisite of using a motion equation system for tongue deformation, tongue is considered as a mass spring network. Solving the motion equation system, they find the external forces on the nodal points (vertices of the 3D mesh model) of the mass spring network. Forces generated by tongue muscles and the one produced by the collision of the tongue with the hard surrounding structures were also considered. In this work they have benefited from the concept of "target-based control strategy" [Dang & Honda 2001] which is a strategy of investigating the effect of each muscle on the tongue deformation.

Selecting a control point on the 3D model, evaluation of muscle activation can be achieved from the displacement of that control point form a prototype to a given target. Figure 20 represents tongue deformation by exciting four tongue muscles [Dang & Honda 2001]. Basically, excitement of each muscle motives the control point displacement from its initial point to a new target position.

Based on FEM method explained above an open-source simulation system named "ArtiSynth" which is capable of tongue simulation (considering hard and soft tissue effect on tongue deformation) has been presented [Stavness et al. 2011] [Buchaillard et al. 2009].

2.5 3D Tongue Animation

There are different methods of creating 3D tongue animation.

In [Ilie et al. 2012], a skeletal structure for tongue has been considered for tongue animation applications. In this structure different features such as tongue tip and tongue body are defined and based on these features a bone anatomy has been used. The 3D tongue model (Figure 21) is designed manually. The mesh model includes 5 control nodes with 4 edges in-between. Deformation is calculated for each section of this model. Normalization of the 3D tongue model is done by aligning it to the Z axis and placing the center of mesh in the origin.

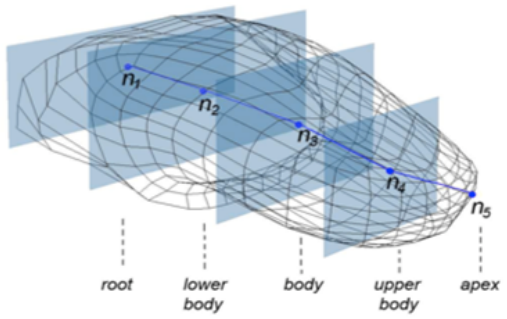


Figure 21: 3D Parametric Tongue Model [Ilie et al. 2012].

Bony structure has been used in order to control the tongue deformation. The linking edges have been used to show effect of the bones on the control nodes (blue circles in Figure 21). Rotation of each bone and also combination of bones' rotations will result in a 3D tongue deformation. These rotations can be obtained by choosing different rotation angles for each bone axis (X, Y, Z) and finally interpolation between rotation angles results in a tongue motion [Ilie et al. 2012]. Figure 22 is a representation of 3D tongue deformed models based on different rotation angles [Ilie et al. 2012].

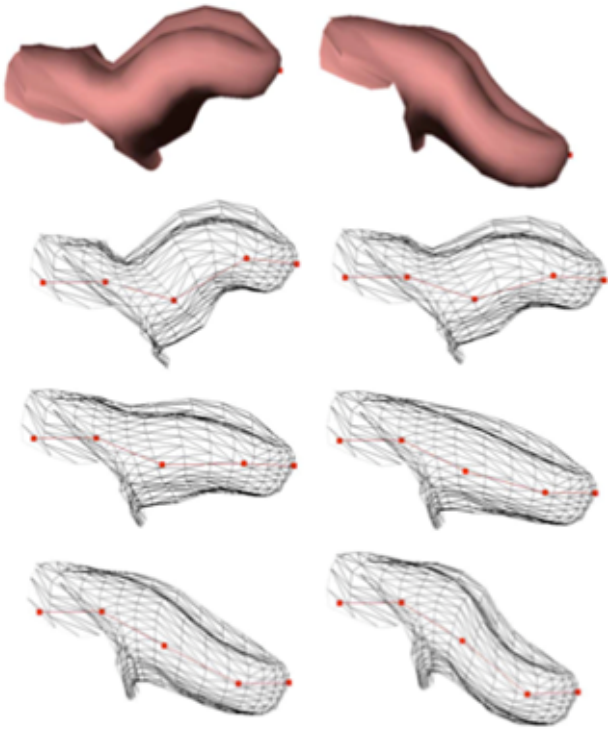


Figure 22: Tongue Animation sequence using skeletal model [Ilie et al. 2012].

Parameterization has also been proposed as a method of deforming the tongue model in [King & Parent 2001]. From [Lundberg & Stone 1999] it can be concluded that only few parameters are enough to define different tongue shapes. In the proposed parameterization method [King & Parent 2001] only six parameters has been utilized (control points) to complete the tongue deformation task. Parameters are defined as the following:

1. P1 (Tongue Tip X)
2. P2 (Tongue Tip Y)
3. P3 (Tongue Tip Z)
4. P4 (Tongue Dorsal Y)

5. P5 (Tongue Lateral X)
6. P4 (Tongue Lateral Y)

Then, new control points were obtained as following:

$$P^{'}=P_{i,j} + \sum W_{k,j} P_k \alpha_{k,i}$$

where $W_{k,j}$ and $\alpha_{k,i}$ are the weights defined respectively for row j and column i of parameter k. Tongue tip movement toward front- back (X), up-down (Y) and sides (Z) direction. Tongue dorsal responsible for back raising (Y) movement and tongue lateral for grooving (X and Y) of the tongue. Different weights can be considered for X, Y, Z in order to change the shape of the tongue based on the obtained values of defined parameters [King & Parent 2001].

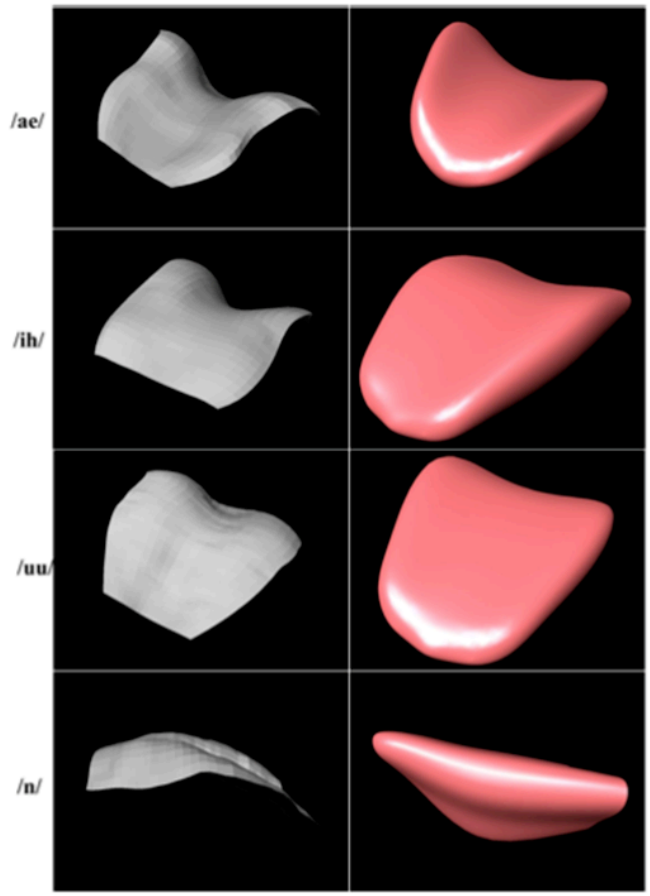


Figure 23: The left column is the surface reconstruction from Ultrasound images during production of specified phonemes and the right column is the 3D deformed tongue model based on the parameterization method proposed in [King & Parent 2001].

Figure 23 shows deformed tongue shape models based on the method proposed in [King & Parent 2001] for some phonemes.

Chapter 3 Overview of the Proposed Solution

3.1 Objective

As the purpose of this research is to develop a method of representing tongue shapes for applications such as speech therapy and 3D facial animation a biomechanical model of tongue including different groups of muscles and their role in tongue movement during speech were not considered in the proposed method. We rather focus on shape itself. The aim in this work is to introduce a model that covers wide variety of tongue shapes based on two views. Generating such models is an essential component of speech production demonstration.

We required a 3D deformed tongue model that is efficient and capable of providing visual perspective of correct gesture of the tongue for different articulations. In the other words, we present a method that is capable of producing a deformable tongue anatomy based on only 2 Ultrasound images.

This work benefits from a deformation method based on Radial Basis Functions (RBFs) applying on control points on 2 orthogonal tongue ultrasound images and also predefined control points on the generic 3D tongue mesh model.

3.2 Complications

Several factors complicate any attempt to create a sophisticated method that can describes modeling of 3D tongue, explains tongue behavior and can control tongue model during speech production, especially when it comes to using 2D Ultrasound images as an input. It is not always feasible to cover all details of the tongue movement during speech.

These complications can be classified as either a technical limitation associated with tongue surface tracking in ultrasound images, a subject –specific tongue shape complication, or a complication associated with introducing a method capable of producing realistic, acceptable and accurate tongue models without computationally expensive requirements.

3.2.1 Complications Associated with Tongue Surface Tracking and Ultrasound Imaging

Accurate tongue surface tracking in a noisy full of artifact ultrasound images is a difficult task. Several approaches tried to fulfill this task [Li et al. 2005] [Akgul et al. 1998] [Yu et al. 2011]. Ultrasound images contain high-contrast unrelated edges which are unrelated to the tongue surface and these features complicate tongue surface tracking. As a solution to this problem we deploy EdgeTrak software [Vocal Tract Visualization Laboratory, n.d.] in order to track the tongue surface based on an ideal number of control points on the ultrasound images. Manual tongue surface tracking in ultrasound images is time consuming, but in our method in cases of having only two orthogonal images (one frame), user performs control

point selection task. This provides an accurate tongue surfaces and it avoids any mistake in the surface extraction task.

3.2.2 Subject-Specific Tongue Shape Complication

There are noticeable differences of tongue shapes (the length and the volume) among individuals and different shapes must be considered for a detailed presentation of tongue deformation. Figure 4 illustrates how length and the volume of the tongue are different for different genders. We propose a method capable of modeling different tongue shapes with respect to general tongue anatomy. Based on the proposed method variety of tongue shapes can be modeled.

3.2.3 Method Overview

A flawless tongue model must be capable of modeling accurate shape of the tongue based on the input images and this is crucial component in applications such as medical simulations that require accurate and trustable display of the tongue. However, for facial animation, computer game industry and creating realistic animations, the accuracy can be lightened and only modeling the important visible tongue section should be enough. Thus, different tongue shapes and deformations resulted from interaction of tongue with surrounding anatomical structures that are mostly invisible can be declined for these kinds of applications.

Our proposed method summarizes the characteristic of tongue shape away from actual geometry. This enables us to model the particular tongue shape easily.

A robust method of free form deformation based on RBFs has been introduced to assign control points' effect on the surrounding vertices of mesh model and also it has been utilized to deform the tongue model based on transformation (translation, rotation and scaling) of control points defined on different tongue sections.

3.3 Solution Overview

Chapter 4 covers the full details of the tongue deformation method we have developed to meet our objective. A brief summary follows:

User interaction and Surface determination– We first have a user select number of control points on 2D ultrasound images: One set includes control points on the mid-coronal ultrasound image and another set on the mid-sagittal one (Figure 24).

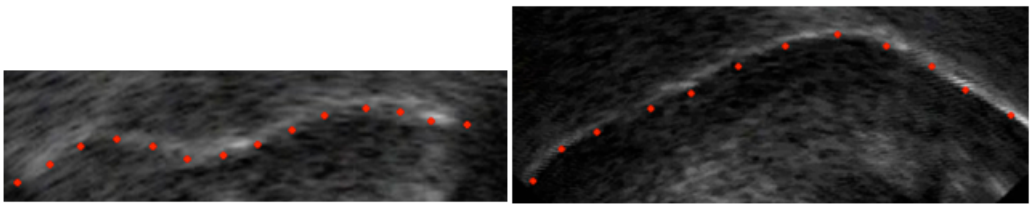


Figure 24: Control point selection on Ultrasound images: 12 control points has been selected on the mid-sagittal Ultrasound image (Right image) and 14 control points on the mid-coronal image (Left image).

Also in the cases where there is sequence of ultrasound images instead of few frames, EdgeTrak has been utilized to perform automatic tongue surface tracking. There, after selecting the interesting area on the first frame and setting initial points on it, the total number of control points has been set. After optimization of the first frame, EdgeTrak optimizes the rest of the frames automatically. The result of this step is a set of tongue surfaces defined with control points (same number of control points for all ultrasound images) on selected ultrasound frames.

Registration – Next, each control point is registered to the corresponding control point on the 3D generic mesh model and this will be continued until finding the corresponding set of control points on both 2D ultrasound images and the 3D generic tongue model. However the increase of dimensions from 2D to 3D should be considered. Control points' registration in both 3D generic model and orthogonal 2D ultrasound images has been completed using Procrustes analysis [Berretti et al. 2010] [Ansari 2003] [Everson 1998]. Using this method, best set of translation, Rotation and scaling will be calculated in order to find the best acceptable distance between set of control points that are from different dimensions.

Transformation value - The comparison of the coordinates of correspondent registered control points (x, y, z) provides the transformation value.

Deformation – Once the value of transformation is obtained a free form deformation method based on Radial Basis Functions (RBFs) will be applied and it affects the in-between areas of the control points based on the transformation value.

3.4 Radial Basis Functions

In our method we have deformed the 3D generic tongue mesh model using an approach based on RBFs. Our aim is to use a method capable of controlling the shape of the tongue and also having high quality of the deformation using displacement from a known vertex position (Control points on the 3D model) to the target position (control points on the Ultrasound image).

Radial Basis Functions have become a well-established tool to interpolate scattered data [Botsch & Kobbelt 2005]. Smooth interpolation between known vertex positions (control points) at a discrete set of positions can be evaluated based on RBFs [Wright 2003]. Basically, the displacement value between a vertex position in 3D space and another vertex position at which we like to obtain is an input argument and this explains the phrase “radial” in RBFs. Using RBFs any displacement to map a vertex to the target position can be evaluated [Botsch & Kobbelt 2005] and this distance value between a vertex position in 3D space and another vertex position explains the phrase “Radial” in RBFs.

3.4.1 Interpolation using RBFs

Having a value of a function $F: R^3 \rightarrow R$ in M vertices in 3D space, using RBFs, smooth interpolation function of FF can be evaluated in the whole domain of R^3. This interpolation function is the sum of M assessment of RBFs function $g(r_i): R \rightarrow R$. r_i is the distance between known vertex position X and vertex to be evaluated X_i.

$$F(x) = \sum_{i=1}^{M} a_i g(|x - x_i|) + c_0 + c_1 x + c_2 y + c_3 z, \quad x = (x, y, z) \qquad Eq.\ 3.1$$

where a_i is the coefficient and $c_0 + c_1x + c_2y + c_3z$ constitute first degree polynomial with coefficients c_0 to c_3 and all these terms are describing affine transformation that cannot be achieved by radial basis functions alone.

Having M known function values $F(x_i, y_i, z_i) = F_i$ we can collect a system of $M+4$ linear equations of $GA = F$ *where* $F = (F_1, F_2, \ldots, F_M, 0, 0, 0, 0)$, $A = (a_1, a_2, \ldots, a_M, c_0, c_1, c_2, c_3)$ and G is a $(M+4) \times (M+4)$ matrix:

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1M}\, 1\, x_1 y_1 z_1 \\ <b_r/> & g_{21} g_{22} & \ddots & g_{2M}\, 1\, x_2 y_2 z_2 \\ <b_r/> & \bullet\bullet & \cdots & \bullet\bullet\bullet\bullet\bullet \\ <b_r/> & \bullet\bullet & \cdots & \bullet\bullet\bullet\bullet\bullet \\ <b_r/> & \bullet\bullet & \cdots & \bullet\bullet\bullet\bullet\bullet \\ <b_r/> & g_{M1} g_{M2} & \cdots & g_{MM}\, 1\, x_M y_M z_M \\ <b_r/> & 1\ 1 & \cdots & 1\ 0\ 0\ 0\ 0 \\ <b_r/> & x_1 x_2 & \cdots & x_M\ 0\ 0\ 0\ 0 \\ <b_r/> & y_1 y_2 & \cdots & y_M\ 0\ 0\ 0\ 0 \\ <b_r/> & z_1 z_2 & \cdots & z_M\ 0\ 0\ 0\ 0 \end{bmatrix} \quad Eq.3.2$$

Here $g_{ij} = g(|| x_i - x_j ||)$. For g there are multiple solutions and we utilized Shifted Log Function which is defined as following:

$$g(t) = \sqrt{\log(t^2 + k^2)}, \quad k^2 \geq 1$$

t is a time parameter to change the displacement with time and K controls displacement magnitude. Solving the equation system for A provides the coefficients required in Eq. 3.1 between known vertex positions.

3.4.2 Deformation using RBFs

Suppose we have the deformation value (3D displacement vector U_i) for M control points and X_i vertex position needs to be moved to position $X_i + U_i$. Interpolation of displacements to other vertices is evaluated by RBFs.

Using three linear equations we are able to find values in the following equation using U_i displacement.

$$GA_x = (u_1^x, u_2^x, \dots, u_M^x, 0,0,0,0)^T \quad Eq.\,3.3$$

$$GA_y = (u_1^y, u_2^y, \dots, u_M^y, 0,0,0,0)^T \quad Eq.\,3.4$$

$$GA_z = (u_1^z, u_2^z, \dots, u_M^z, 0,0,0,0)^T \quad Eq.\,3.5$$

Solving for A_x, A_y and A_z involves a single matrix inversion and three matrix-vector multiplications we can evaluate the coefficients for interpolation of displacements in 3D space specified in Eq. 3.1.

Chapter 4 3D Tongue Reconstruction from Two Orthogonal Ultrasound Images

4.1 3D Generic Mesh Model

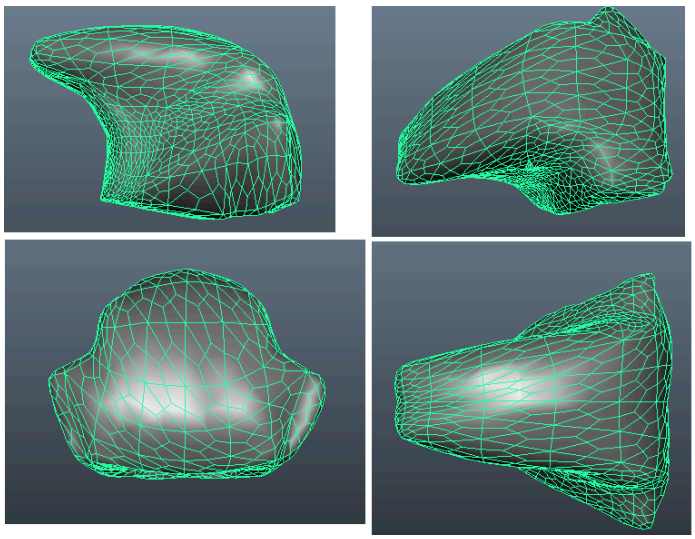


Figure 25: 3D generic Tongue Mesh Model from different perspectives.

A 3D generic model required must be modifiable so that it can cover all different tongue shapes. The 3D tongue model used for testing in this work was manually modeled using Maya, a commercial 3D modeling software package. The 3D tongue model is regarded as a quadrilateral mesh and it is designed with respect to the HexTongue mesh model used in ArtiSynth software [Buchaillard et al. 2009] [Stavness et al. 2011]. This model was based on

a reference tongue model proposed in [Buchaillard et al. 2009]. Figure 25 shows the generic mesh model to be deformed.

4.2 Ultrasound Imaging

Ultrasound images not only benefit studying information about tongue shape and its dynamics, but also help researchers who investigate tongue shapes information via ultrasound imaging. However, the ultrasound images are noisy and are full of artifacts [Li et al. 2004] [Yu et al. 2011] [Berretti et al. 2010].

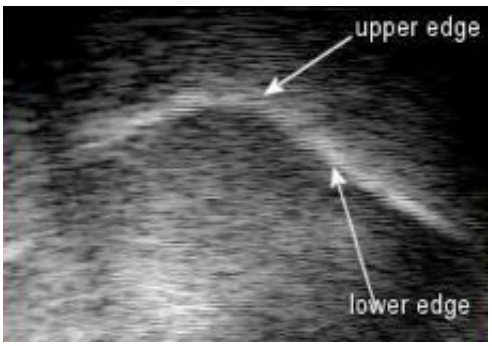


Figure 26: Sample ultrasound images (Sagittal view).

Most of the current generation of portable Ultrasound machines can decompose the results of a scan as a set of 2D slices or video frames, rather than the totality of the scan as a volume. Slices can be acquired from coronal or sagittal view.

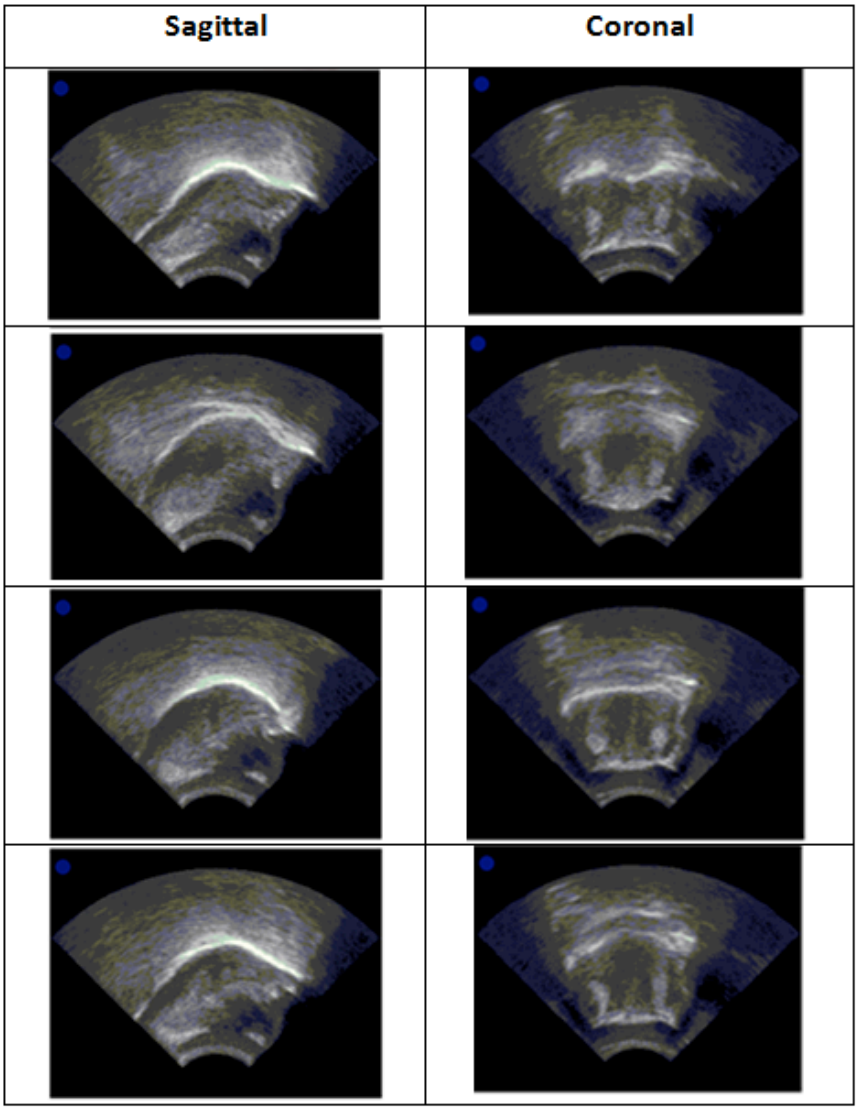


Figure 27: sample set of sagittal and coronal ultrasound images.

Professor Jeff Mielke from Department of Linguistics kindly provided the Ultrasound images of a male subject, an essential input data required for this research. Two sets of ultrasound images were used: (1) A set of coronal (front view) ultrasound images (2) a set of sagittal (side view) of ultrasound images (Figure 27). Using the mid-sagittal images as a

reference, Two coronal set were provided: coronalA is pointed about 5 degrees to the right (anterior) of what would appear vertical in the mid-sagittal images, and coronalB is pointed about 25 degrees to the left (posterior) of vertical. The Ultrasound device used for this purpose is the Terason Ultrasound System.

In the first step, initial number of ultrasound slices captured from sagittal and coronal view (Figure 28) during English vowel production, have been selected as an input. Once ultrasound slices have been loaded, the next step is to extract the tongue surface from them.

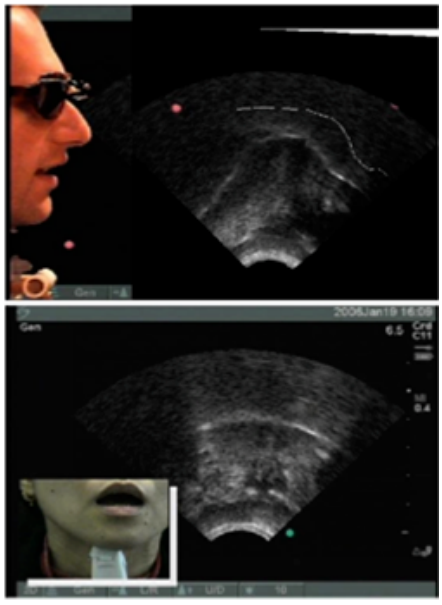


Figure 28: Captured Ultrasound images. The upper image shows sagittal Ultrasound acquisition and the lower image represents the coronal view.

4.3 Tongue Surface Tracking from Ultrasound Images

Tongue surface determination from Ultrasound images is the first step of the proposed method in this work in order to change the neutral shape model from the shape of the provided ultrasound images.

Manual tongue surface determination is done by a user on two orthogonal ultrasound images. User selects arbitrary number of control points on the mid-sagittal ultrasound image and also on the mid-coronal slice. Coronal control points include two extreme points corresponding to the both sides of the tongue. Linear interpolation between selected control points on each ultrasound image entails the tongue surface. Results from the interpolation process provide us tongue surface shape from mid-coronal and mid-sagittal ultrasound images.

In the cases where we have several ultrasound frames we deployed EdgeTrak software. Using sequence of 8 images from the Ultrasound video file [Coupling electromagnetic sensors and ultrasound images for tongue tracking: acquisition set up and preliminary results 2006], we tested EdgeTrak method on some sample Ultrasound. First we needed to adjust the target area (green region) in a sample Ultrasound image for first frame and then set the image gradients for it. Figure 30 illustrates determination of initial points and specifying the total number of them for final contour extraction (50 points for this experiment). Figure 31 represents Initial contour for frame 1, before and after optimization.

Figure 29: Adjustment of the target area (green region) and setting the image gradients.

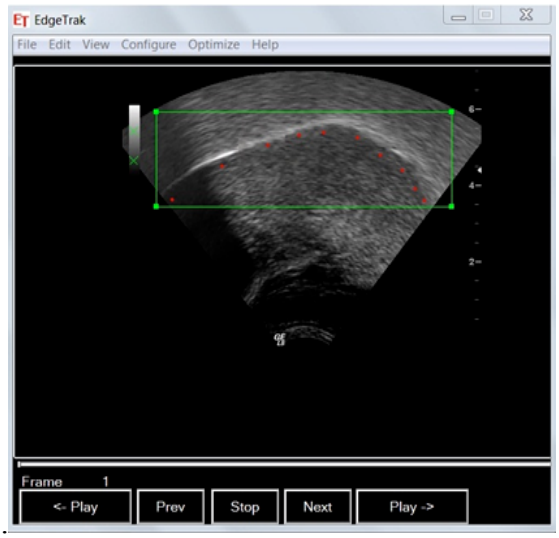


Figure 30: Setting initial points.

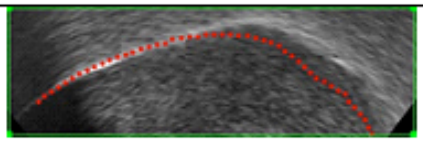
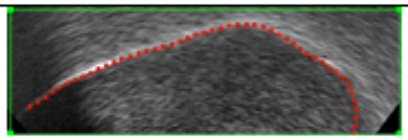
BEFORE OTIMIZATION	AFTER OPTIMIZATION
	

Figure 31: Contour determination using EdgeTrak system for frame 1.

After this step, the optimization of an Ultrasound sequence was automatically completed which resulted in an optimal set of contours for the input Ultrasound sequence (In this experiment 5 frames are presented).

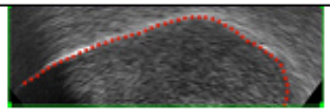
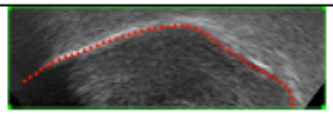
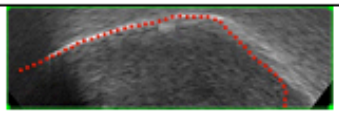
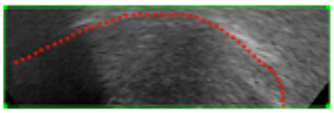
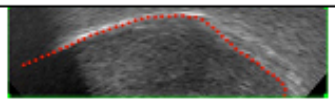
FRAME NUMBER	RESULT FROM EDGETRAK
1	
2	
3	
4	
5	

Figure 32: Tongue surface tracking using EdgeTrak for an ultrasound sequence of 5 frames.

Figure 33 illustrates the tongue surface obtained from EdgeTrak for some ultrasound frames captured while the subject saying the numbers 1-10 and then sustained “i”, “u”, “e”, “o”, “a”, “s”, “sh”, “l”, “n” and “r”.

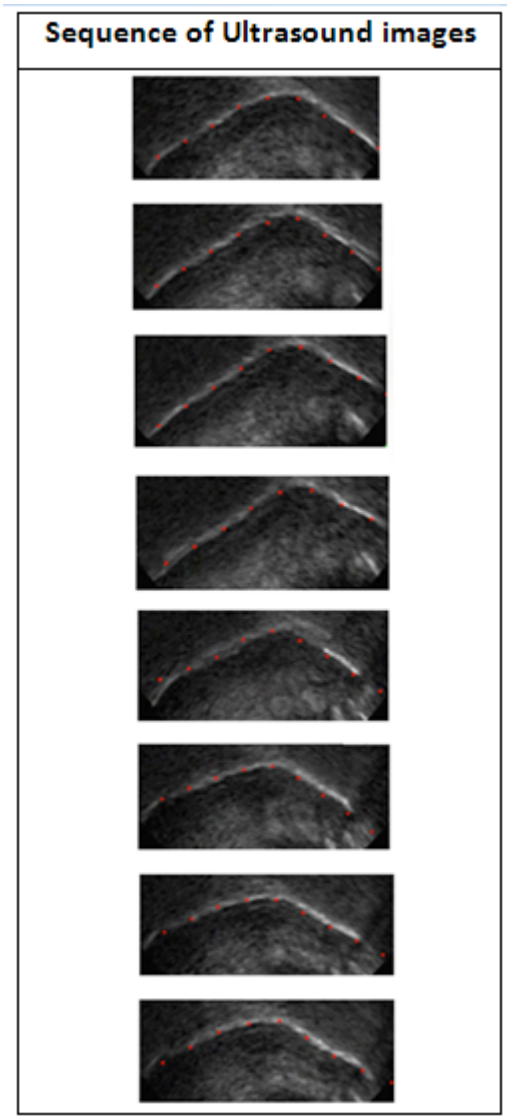


Figure 33: Tongue surface tracking using EdgeTrak; Ultrasound images are randomly selected from a recording of a subject saying /ta/.

4.4 Generic Model Deformation

The generic model deformation is a problem about 3D space deformation. Displacement of a set of control points is calculated. Using an interpolation function of RBFs, position of rest of vertices of the 3D mesh model will be evaluated.

4.4.1 Control Points' Registration

The next step is to determine the correspondence between control points on the 3D mesh model and control points on the surfaces obtained from the ultrasound images. In order to complete this task, same number of control points were defined on the mid-sagittal line of the 3D tongue model and the same number of control points on the mid-sagittal Ultrasound image (same number of control points on the mid-coronal line of 3D model and mid-coronal Ultrasound image have also been considered).

Figure 34 illustrates the control points selected on a 3D mesh model on both mid-sagittal line and mid-coronal line and also corresponding control points on orthogonal Ultrasound images.

At this point there are two sets of control points: mid-sagittal control points obtained from mid-sagittal Ultrasound image and mid-sagittal points of 3D mesh model. Also mid-coronal control points from mid-coronal ultrasound image and mid-coronal points of a 3D mesh.

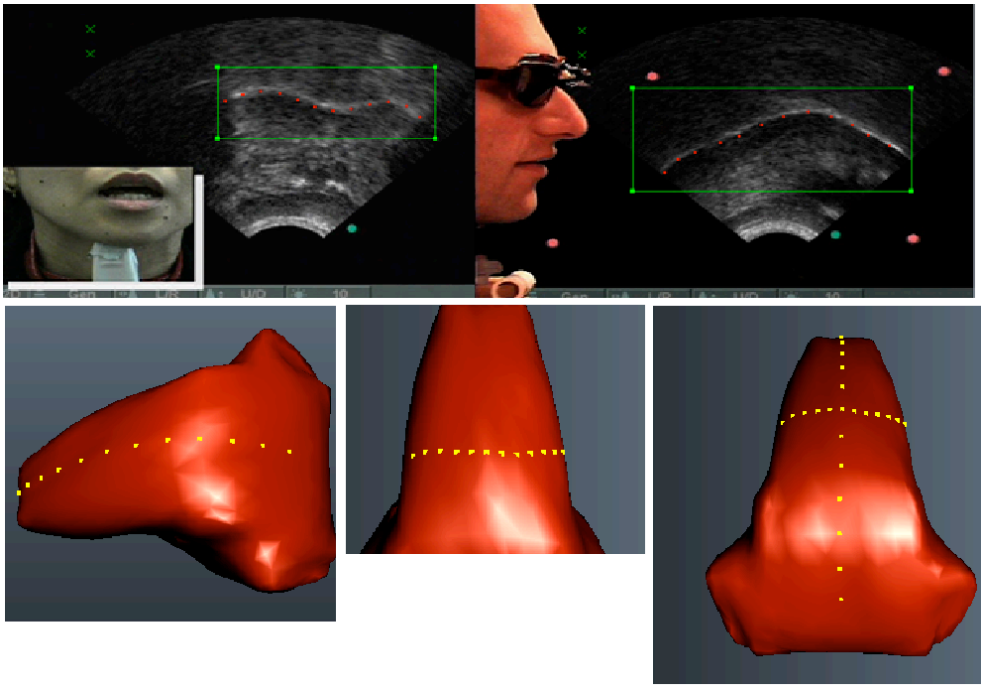


Figure 34: Set of control points on orthogonal sample ultrasound images (Upper row) and correspondent control points on the mid-coronal and mid-sagittal line of 3D mesh model (Lower row).

To achieve the goals of smooth deformation and having the neutral 3D mesh deformed with respect to ultrasound images provided, the control points from orthogonal ultrasound images (which we will call P_{Model}) needs to be translated, scaled and rotated to match the control points of 3D tongue model (which we will call $P_{calculated}$). However the increase in dimensions from 2D to 3D (Control points of Ultrasound image and points on a 3D mesh model) must be considered [Berretti et al. 2010].

This Dimension incensement was evaluated using Procrustes Analysis method [Ansari 2003] [Everson 1998]. Following is the detailed description of Procrustes Analysis:

$$min\{f(s,R,T)\} = min\left\{\sum_{i=1}^{23}(P_{model} - P_{calculated})^2\right\} \quad Eq.4.1$$

where n is number of points on both set of control points and P_{Model} is defined as following:

$$P_{Model} = \begin{bmatrix} X_m \\ Y_m \\ Z_m \end{bmatrix} = s.R \begin{bmatrix} X_{m0} \\ Y_{m0} \\ Z_{m0} \end{bmatrix} + T \quad Eq.4.2$$

where s is scale matrix, R is Rotation Matrix and T is a column vector for translation.

$P_{calculated}$ is a matrix of control points from ultrasound images and P_{Model} stands for control points of the 3D tongue mesh. X_m, Y_m *and* Z_m are coordinates of a point in P_{Model} (a point from set of control points on ultrasound images) after transformation and X_{m0}, Y_{m0} *and* Z_{m0} are initial coordinates of a point in P_{Model} before transformation (a point from set of control points of 3D tongue model).

We first calculate the mean of P_{Model} and $P_{calculated}$ and then we place both set of control points at their origins using:

$$P_{C0} = P_{calculated} - mean(P_{calculated}) \quad Eq.4.3$$

$$P_{m0} = P_{Model} - mean(P_{Model}) \quad Eq.4.4$$

We call centered control points of $P_{calculated}$, P_{c0} and for P_{Model}, we call them P_{m0}. Then the normalization of P_{Model} and $P_{calculated}$ is evaluated which have been called $||P_{c0}||^2$ and $||P_{m0}||^2$ respectively.

In order to normalize each set:

$$P_{c0} = \frac{P_{c0}}{\|P_{c0}\|^2} \qquad Eq.4.5$$

and

$$P_{m0} = \frac{P_{m0}}{\|P_{m0}\|^2} \qquad Eq.4.6$$

After evaluating the normalization the Singular Value decomposition (SVD) should be calculated.

For an arbitrary variable A:

$$A = P_{c0}^{t} P_{M0} \qquad Eq.4.7$$

Computing the SVD of A provide three essential matrices named L,D and M needed for evaluation of rotation *(R)*, scaling *(S)* and translation *(T)* matrix.

$$LDM = SVD(A) \qquad Eq.4.8$$

Rotation matrix *(R)* is computed as:

$$R = ML' \quad Eq.4.9$$

Scaling factor *(S)* is calculated as:

$$S = Tr(D) * \frac{\|P_{c0}\|^2}{\|P_{M0}\|^2} \qquad Eq.4.10$$

And finally Translation vector is computed as:

$$T = mean(P_c) - s.mean(P_M).R \qquad Eq.4.11$$

In the next step transformation from P_{Model} to $P_{calculated}$ (control points on the ultrasound images) is done using achieved Transformation values (*s*, *R* and *T*) as following:

$$P_{Model.Final} = s.P_{Model.Initial}.R + T \qquad Eq.4.12$$

$P_{Model.initial}$ is initial coordinates of ultrasound control points and $P_{Model.Final}$ is coordinates of ultrasound control points after transformation.

Each control point on mid-sagittal ultrasound image, depending on its location on the tongue surface, is registered with its corresponding point on the 3D mesh model. Mid-coronal control points on ultrasound images are also registered with their corresponding

points on the mid-coronal line of the 3D mesh model. Finally, comparison between registered control points is done as described in following:

The difference between coordinates of each control point on an ultrasound image and coordinates of corresponding mid-sagittal control point of 3D mesh is calculated. It can be either *X* or *Y* or *Z* coordinates depending on the mesh orientation. The result is difference values (deformation value for each control point of 3D mesh model) that will be employed for sagittal and coronal deformation. The deformation value determines the target position for each control point which is used as an input for RBFs.

4.4.2 Mesh Deformation using RBFs

In this step a realistic deformed model is obtained where the control points of 3D mesh model are mapped to the target position. Using RBFs, interpolation is evaluated by the displacements of the control points. The rest of the vertices of 3D mesh model are mapped using RBFs.

Utilizing RBFs method, each control point (on the 3D mesh model) is mapped from its initial position to the target position (correspondent control point on the Ultrasound image) with the distance values obtained from the previous stage. The translation of the root area of the neutral 3D model set to be zero so that the root of the tongue stays fixed in its current position.

Table **1** represents 3D deformed models obtained from corresponding ultrasound slices.

Table 1: 3D deformed tongue models based on the input ultrasound images.

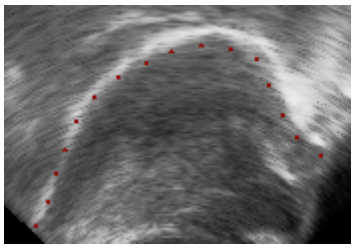
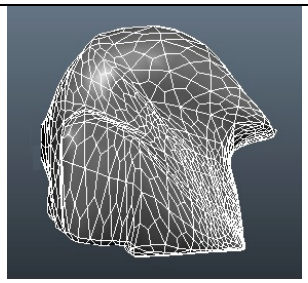
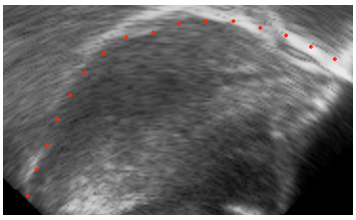
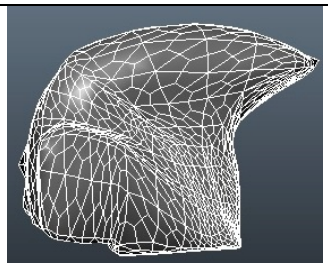
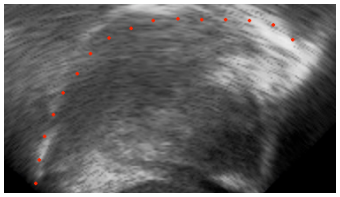
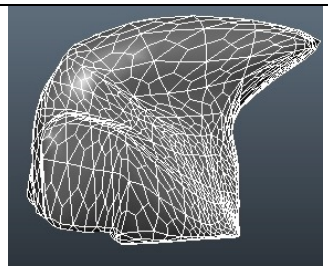
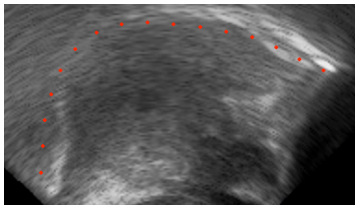
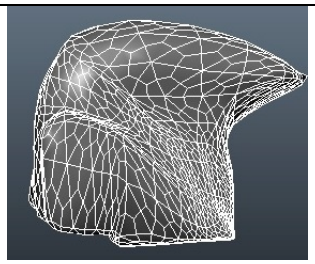
ULTRASOUND IMAGES (SAGITTAL VIEW)	3D DEFORMED TONGUE MODELS
	
	
	
	

Figure 35 shows two sample input ultrasound images on the mid-coronal and mid-sagittal lines of the resulted deformed tongue model.

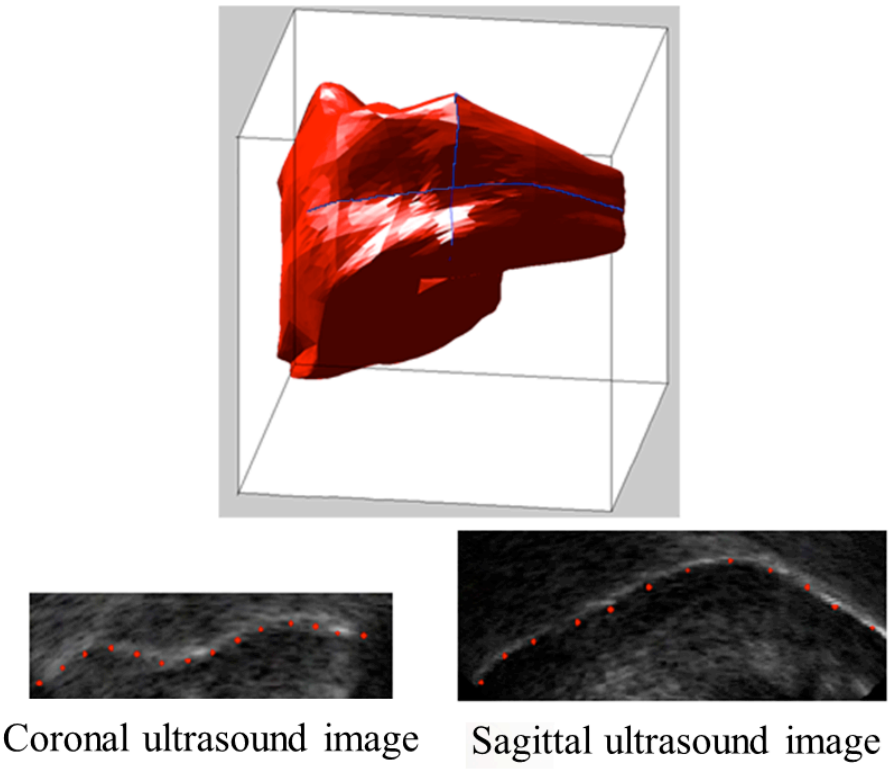


Figure 35: Deformed 3D tongue model from different perspectives (Upper row) using mid-coronal and mid-sagittal input ultrasound images (blue lines on the 3D model).

The blue lines on the top image in Figure 35 represents the tongue surface obtained from orthogonal Ultrasound images.

Chapter 5 Experimental Results and Evaluations

Our aim was to use a method capable of controlling the shape of the tongue while having a high quality of deformation for applications such as language learning or realistic 3D Facial Animation. We have selected ultrasound images (corresponding coronal and sagittal for one particular pronounced vowel) as an input to our system and a deformation technique RBFs is used to produce smooth interpolation between control points at a discrete set of positions.

5.1 RBFs Validation (Qualitative and Quantitative Comparison)

In order to test the capability and sufficiency of RBFs method, we have compared our results (deformed 3D tongue models based on control points on mid-sagittal and the mid-coronal line of the 3D tongue model) with Artisynth results (deformed tongue models based on tongue muscles activation). For this purpose, we selected the "HexTongue" model from the Artisynth. In the next step we set the activation value of each muscle (a value between 0 and 1) to get the desired tongue model. Genioglossus Posterior (GGP), Genioglossus Middle (GGM), Genioglossus Anterior (GGA), Styloglossuns (STY), Geniohyoid (GH), Mylohyoid (MH), Hyogossus (HG), Verticalis (V), Transvers (T), Inferior Longitudinal (IL) and Superior Longitudinal (SL) are the muscles considered in the ArtiSynth simulation.

Table 2: Muscles activations set (ArtiSynth Tongue Model).

Muscles' name	0<Muscle activation values <1
GGP	0.235
GGM	0.11
GGA	0
STY	0
GH	0.085
MH	0
HG	0
VERT	0
TRANS	0.17
IL	0
SL	0

The deformed tongue model from ArtiSynth after activating muscles (with activation values presented in Table 2) is shown in Figure 36.

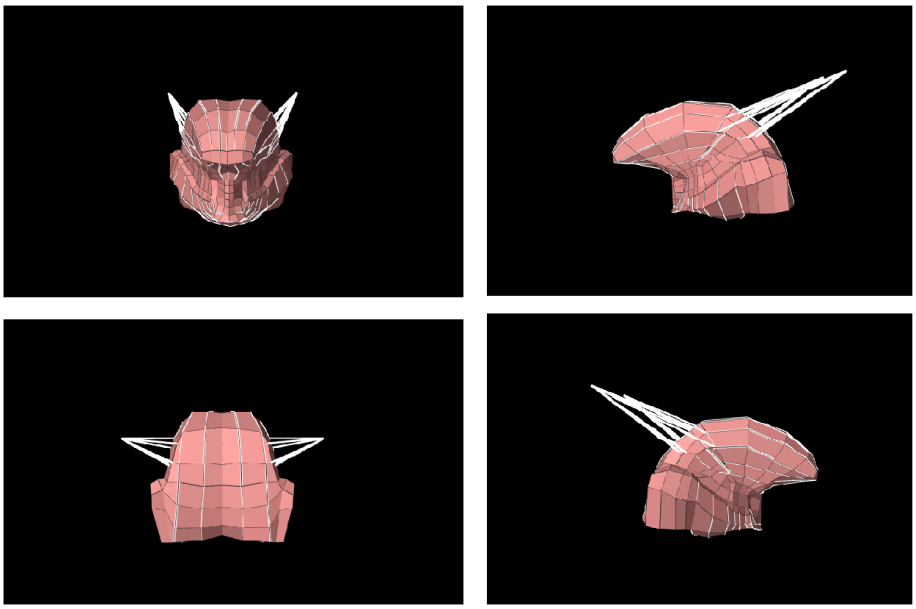


Figure 36: HexTongue model from Artisynth after setting activation parameter for GGP, GGM, GH and TRANS muscles.

In the next step using mid-sagittal and mid-coronal points from the deformed model obtained from ArtiSynth (Figure 36), deformation of the exact same neutral 3D tongue model was completed this time using the proposed method of deformation based on RBFs (Figure 37).

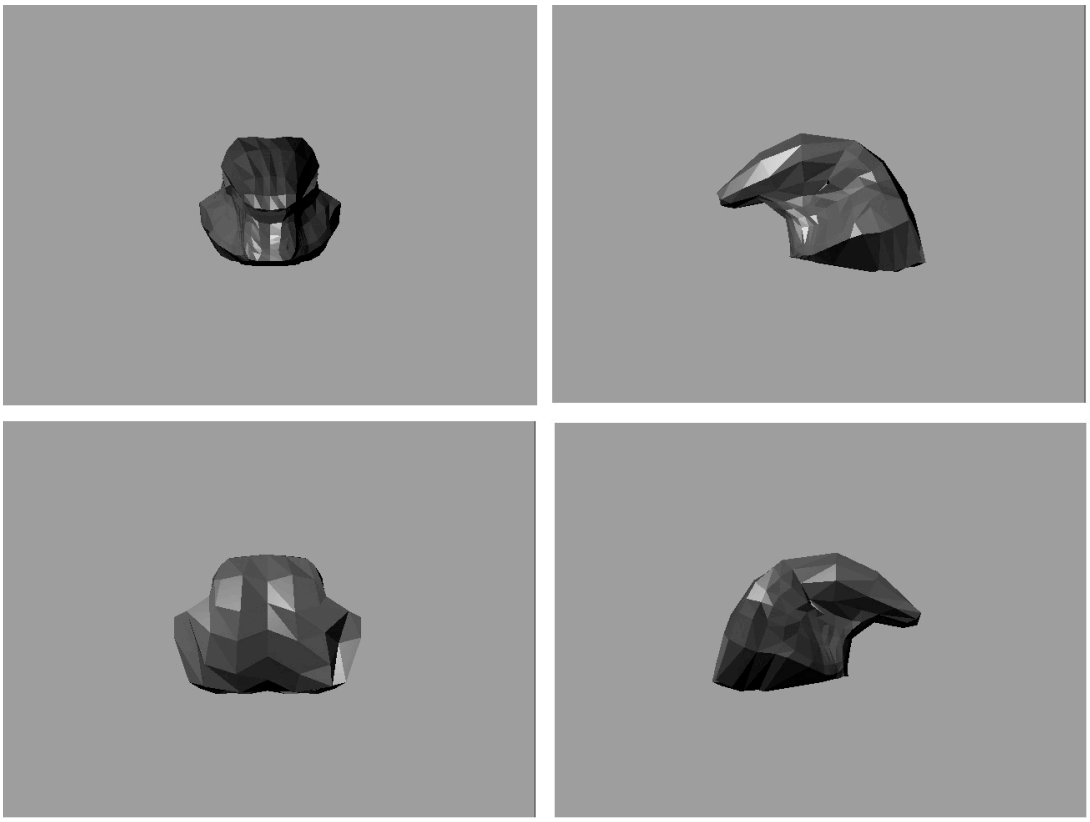


Figure 37: Resulted 3D tongue model based on RBFs method.

Table 3 illustrates the maximum and average Euclidian Distance between set of 3D vertices of 3D deformed tongue model achieved from RBFs and the one obtained from Artisynth. The Euclidean distance (d) is the length of the line segment connecting 2 points $p= (p_1, p_2, \ldots, p_n)$, and $q= (q_1, q_2, \ldots, q_n)$ and it is defined as follows:

$$d(p,q) = \sum_{i=1}^{n}(p_i - q_i)^2 \qquad Eq.5.1$$

Table 3: 3D distance between deformed models obtained from RBFs and muscle-based method of Artisynth.

Average Distance mm	**Maximum Distance** mm
0.0027	0.0963

5.2 Validation

5.2.1 Qualitative Comparison

Comparison between 3D deformed models obtained from our method and 5 tongue shapes including /i/ and /a/ vowels and 3 distinctive tongue shapes (resulted from setting different values for muscles' activation parameter)obtained from the method proposed in [Stavness et sl. 2012] and ArtiSynth were done.

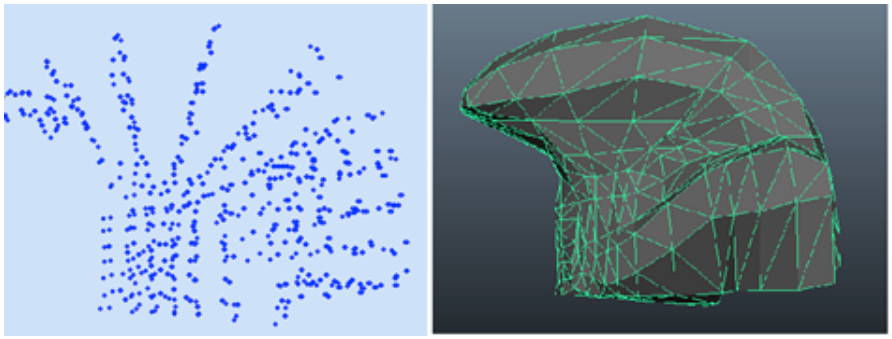


Figure 38: Representation of 3D point cloud (left image) of the neutral tongue model and the actual mesh model (neutral tongue) used in [Stavness et al. 2012].

Figure 38 demonstrates the point cloud of 3D tongue model used in Artisynth and also the corresponding 3D mesh model of it in the rest position.

Table 4 shows muscles activation values, used for simulating the tongue model for vowel /i/ in [Stavness et al. 2012]. They have used a dynamic 3D jaw-tongue-Hyoid model introduced in the ArtiSynth [Yu et al. 2011] to represent the deformed tongue model.

Table 4: Muscles activation values for simulating vowel /i/.

Tongue muscles	Activation values for vowel /i/	Activation values for vowel /a/
GGp	15	-
GGm	15	10
GGa	5	15
SLa	5	-
Sla_lat	60	-
ILa	60	-
HG	-	15
TRANSp	-	-
TRANSa	5	-
VERTp	10	-
VERTa	-	5
Jaw_open	-	4
Jaw_close	3	-

The 3D vertices' data for vowel /i/ and /a/ are provided from [Stavness et al. 2012] project. Figure 39 is an illustration of 3D point cloud and the mesh model of vowel /i/ and /a/.

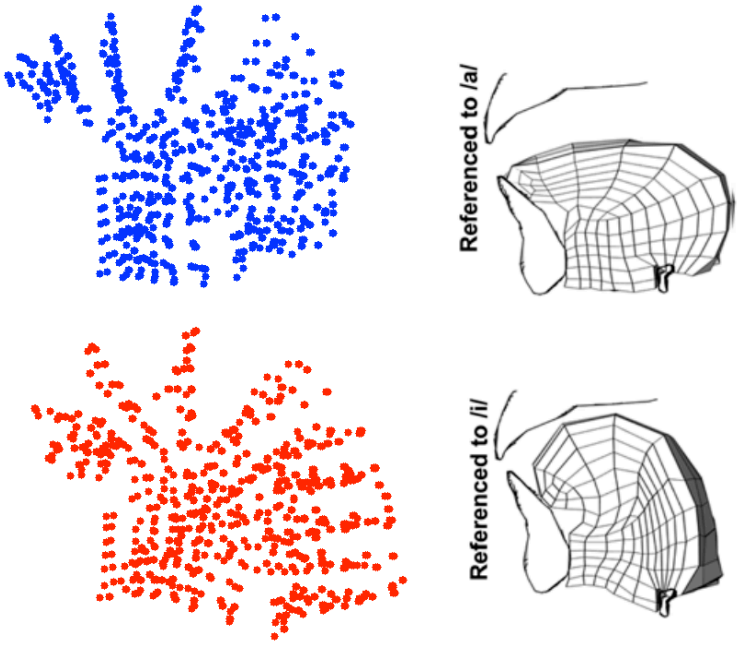


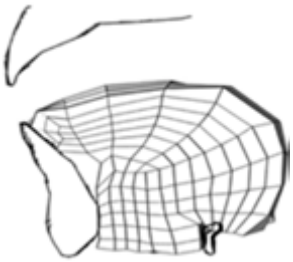
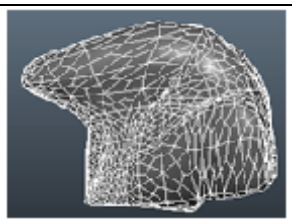
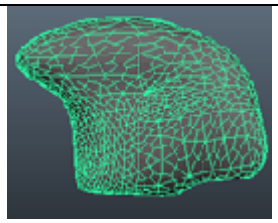
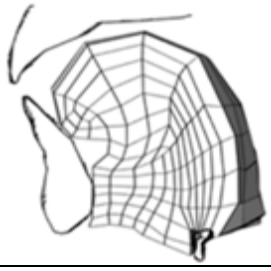
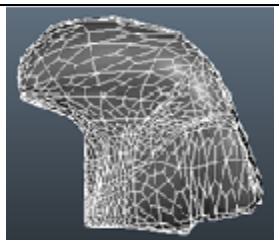
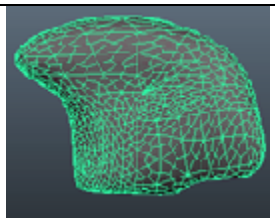
Figure 39: Point clouds (left column) and 3D tongue mesh model (right column) for vowel /a/ (upper row) and vowel /I/ (lower row) [Stavness et al. 2012].

The mid-sagittal and mid-coronal lines (as we required an orthogonal input in order to deform the neutral shape of the generic tongue model) were determined by selecting control points on 3D deformed models produced by ArtiSynth. Deformation of 3D generic model was done using the deformation method based on RBFs and orthogonal data as an input.

We have also tested the proposed method on 3 arbitrary mesh models obtained from Artisynth. These 3D models are results of changing the muscles' activation parameters. Figure 40 illustrates the activation parameters value and the resulted 3D tongue model related to them obtained from ArtiSynth.

Error! Not a valid bookmark self-reference. illustrates the 3D reconstructed tongue models for vowels /i/ and /a/ obtained from our method.

Table 5: 3D reconstructed tongue shapes for vowel /i/ and vowel /a/ using the proposed method.

Vowels to be reconstructed	***3D corresponding models from ArtiSynth***	***3D corresponding models from proposed method***	***Neutral tongue mesh model***
/a/			
/i/			

Having the mid-coronal and mid-sagittal points from the 3D models from ArtiSynth (Figure 40) and applying the proposed method, the 3D deformed models corresponding to the models from ArtiSynth have been obtained (Table 6) (Table 7).

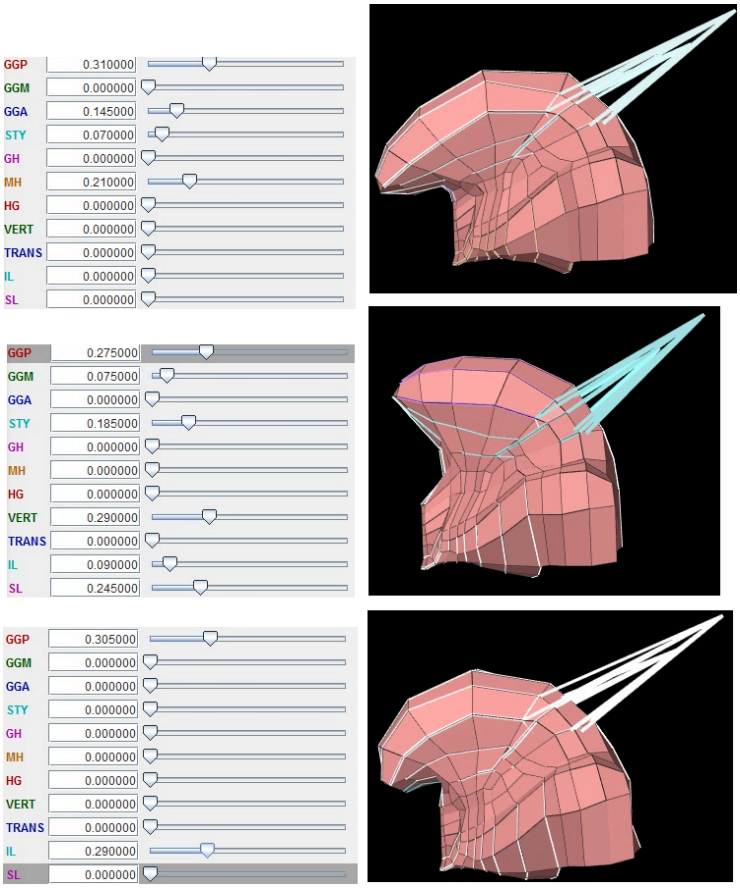


Figure 40: 3D tongue models obtained from ArtiSynth by changing the values of muscles' activation parameters.

Figure 41 represents control points on the mid-sagittal and mid-coronal lines of tongue models of Figure 40.

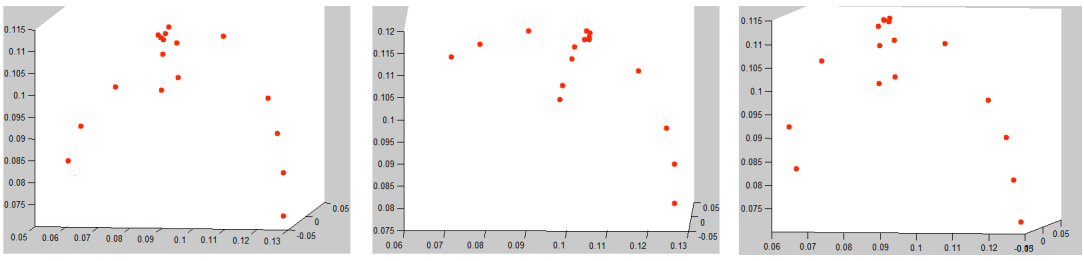


Figure 41: Control points obtained from 3D tongue models of ArtiSynth.

Table 6: Comparison of 3D tongue models obtained from ArtiSynth and the proposed deformation method (sagittal view).

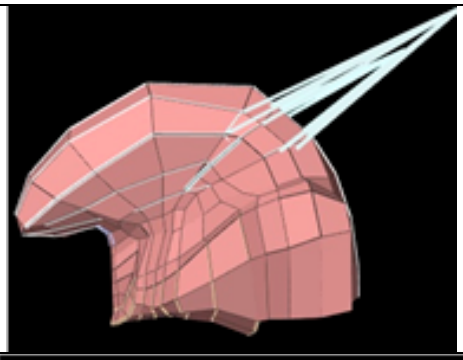
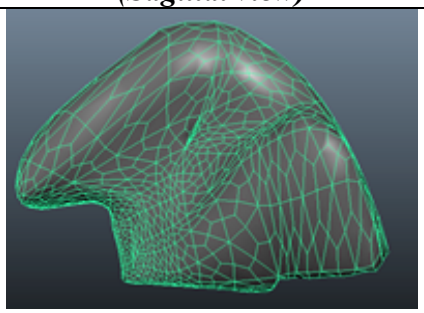
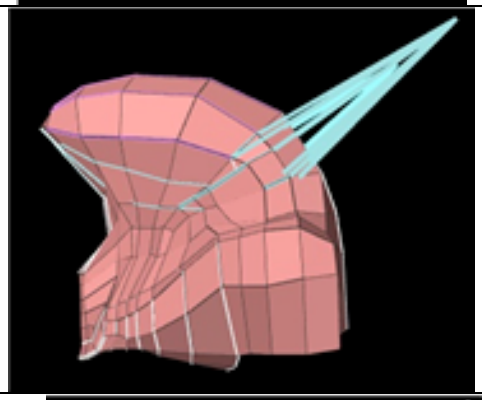
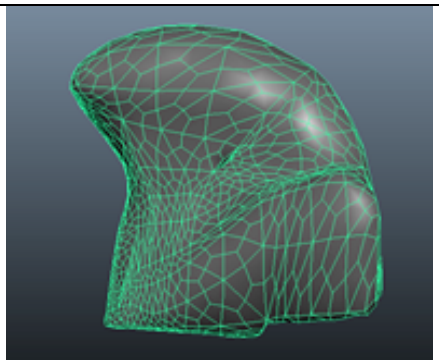
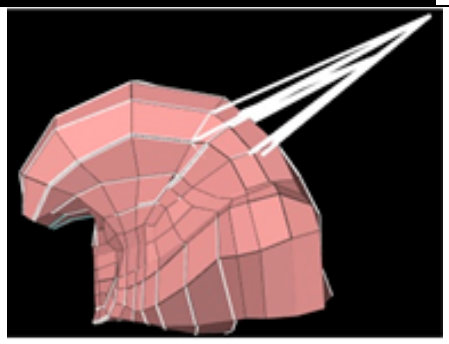
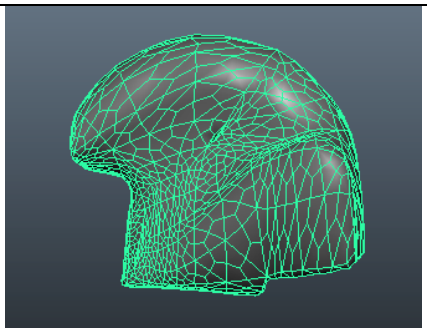
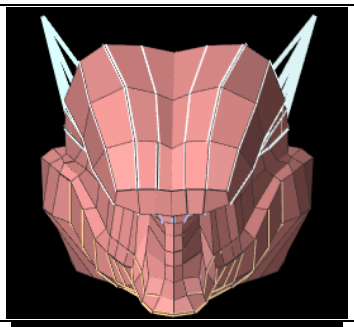
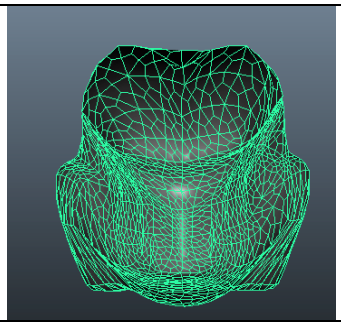
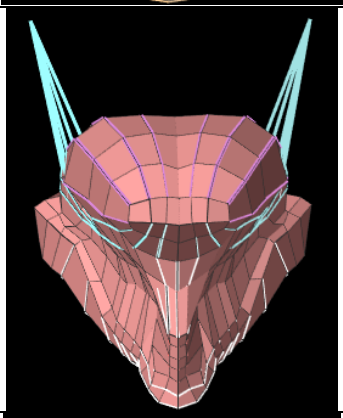
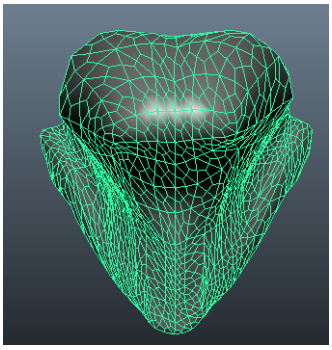
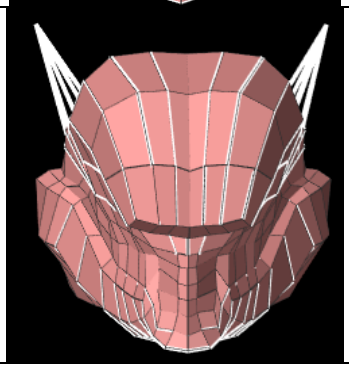
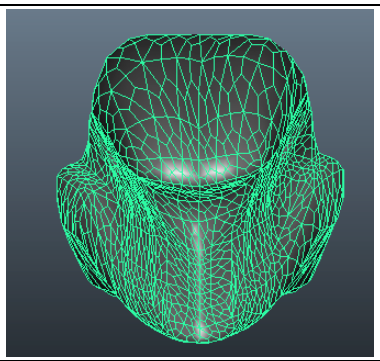
Models	***3D resulted model from ArtiSynth (Sagittal view)***	***3D resulted model using the proposed method (Sagittal view)***
Model 1		
Model 2		
Model 3		

Table 7: Comparison of 3D tongue models obtained from ArtiSynth and the proposed deformation method (coronal view).

Models	*3D resulted model from ArtiSynth (Sagittal view)*	*3D resulted model using the proposed method (Sagittal view)*
Model 1		
Model 2		
Model 3		

5.2.2 Quantitative Comparison

3D distance between vertices of the reconstructed tongue models of our method and the resulted models from ArtiSynth has been measured. For each point in our 3D deformed models, the closest surface point was found for the ArtiSynth model. The 3D distances between these point sets over all points provided a set of distances. To measure the accuracy of the proposed tongue reconstruction using RBFs which we will call TRRBF from now on in the thesis, we compare it with ArtiSynth tongue simulation system.

The ArtiSynth tongue simulation system has been shown to generate highly accurate tongue models, the details regarding the ArtiSynth tongue simulation system is discussed in chapter 2. The ArtiSynth deforms tongue model by setting different activation values for the muscles of the tongue. The detailed muscles' information is required in order to achieve the accurate tongue model. The FEM based method used in ArtiSynth makes it computationally expensive. The advantage of the proposed TRRBF when compared to ArtiSynth tongue simulation system is that it performs tongue reconstruction using only two orthogonal ultrasound views and therefore is quicker and computationally less expensive. This thesis considers the tongue models achieved from the ArtiSynth tongue simulation system as ground truth and compares the tongue models of the proposed TRRBF with those of the ArtiSynth tongue simulation system. Table 8 shows the quantitative results of the proposed TRRBF. The inter-rater reliability measures, calculated for two vowels and three arbitrary tongue models, are: Average Euclidean distance, maximum Euclidean distance, minimum Euclidean distance and standard deviation of the Euclidean distance.

The Euclidean distance between two corresponding vertices in the deformed models, one model achieved from the proposed TRRBF and the other from the ArtiSynth, is calculated and averaged for the total number of vertices. The mean of the average distances, the distances in the first column of Table 8 , for 2 vowels (/i/ and /a/) and three arbitrary tongue models is 0.222 mm and is considered as an average distance error. The mean standard deviation of 0.1304 depicts that the proposed TRRBF offers a small variability in the measured Euclidean distances. Table 8demonstrates that the proposed TRRBF performs favorably with a small average distance error of 0.222 mm.

The total dimensions of the 3D tongue model are:

- Anterior-posterior length = 92.892 mm
- Width = 71.783 mm
- Height = 70.384 mm

Table 8: 3D distances between deformed models for /a/ and /i/ vowels and 3 arbitrary tongue models, resulted from the proposed method and ArtiSynth method.

Tongue models	***Average Distance***	***Maximum Distance***	***Minimum Distance***	***Standard Deviation Distance***
/i/	*0.264*	*0.500*	*0.004*	*0.106*
/a/	*0.274*	*0.551*	*0.007*	*0.131*
Arbitrary model 1	*0.225*	*0.722*	*0*	*0.156*
Arbitrary model 2	*0.131*	*0.44*	*0*	*0.082*
Arbitrary model 3	*0.216*	*1.011*	*0*	*0.177*

Chapter 6 Conclusions and Future Work

Tongue is the most important organ for speech. However details about the tongue movement during speech are hard to observe because the tongue has a complex anatomy and it is masked in oral cavity which makes it inaccessible. Ultrasound imaging has been recently a popular tool to capture tongue data on speech such as speech sciences and speech-language pathology.

The aim of this study was to obtain a reasonable 3D tongue model from only two ultrasound slices (one mid-coronal image and one mid-sagittal one), which is considered to be an easy and practical in data acquisition. Our tongue model is capable to represent the different shapes of the tongue during speech with no computationally expensive requirements.

6.1 Contributions

Our main contributions can be summarized as follow:

- Proposing a mathematical model for performing 3D tongue reconstruction by assigning translation value to control points and having smooth deformation for surrounding tongue mesh vertices.
- Proposing a fast method with no computationally expensive requirements

- Creating tongue shapes that are acceptable enough to use in speech training systems (helpful in learning second language by providing user with a general shape of the tongue for different vowels).
- Representing a method to reconstruct any tongue model that user might need.

The limitations of representing a method based on ultrasound images are quite apparent especially in cases as sticking out the tongue or licking the lips where Ultrasound cannot capture the tongue surface. An extrapolation technique on video tracking might help to solve this limitation.

6.2 Future Research

Our tongue deformation method is able to approximate possible tongue shapes only with two input Ultrasound images. This work can be improved and continued for future studies in several frontiers.

Tongue muscles' information can be included to produce more physically based deformation. Also the effect of tongue interactions with surrounding anatomical structure such as teeth or the palate area can be included. This is suitable for medical applications where more accurate result is needed especially in modeling of complex characteristics of tongue tissue.

References

Akgul, Y. S., Kambhamettu, C., & Stone, M. 1998, 'Automatic motion analysis of the tongue surface from ultrasound image', *In Proceedings of the IEEE workshop on biomedical image analysis*, pp. 126–132.

Akgul, Y., Kambhamettu, C., and Stone, M. 1998, 'Extraction and tracking of the tongue surface from ultrasound image sequences', *In Proceedings of the IEEE on Computer Vision and Pattern Recognition*, vol. 124, pp. 298- 303.

Ansari, A. N. 2003, '3D face modeling using two views and a generic face model with application to 3d face recognition', paper published at *IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 37- 44.

Aron, M., Roussos, A., Berger, M. O., Kerrien, E., and Maragos, P. 2008, 'Multimodality Acquisition of Articulatory Data and Processing', *In Proceeding of* European Signal Processing Conference.

Badin, P., Bailly, G., Reveret, L., Baciu, M., Segerbarth, C., and Savariaux, C. 2002, 'Three dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images', *Journal of Phonetics*, vol. 30, no. 3, pp. 533-553.

Bateman, H. E. and Mason, R. M. 1984, 'Applied Anatomy and Physiology of the Speech and Hearing Mechanism', Charles C. Thomas, Springfield, IL. viii, pp. 636.

Berretti, S., Del Bimbo,A., Pala, P. 2010, '3D Face Reconstruction fromTwo Orthogonal Images for Face Recognition Applications', *International Journal of Digital Library Systems*, vol. 1, no. 3, pp. 41- 57.

Botsch, M., and Kobbelt, L. 2005, 'Real-time shape editing using radial basis functions', *Computer Graphics Forum*, vol. 24, no. 3, pp. 611- 621.

Bressmann, T. 2008, 'Quantitative Assessment of Tongue Shape and Movement Using Ultrasound Imaging', paper published at selected Proceedings of the 3rd *Conference on Laboratory Approaches to Spanish Phonology*, Cascadilla Proceeding Project, Somerville, MA, USA, pp. 101-106.

Bressmann, T., Thind, P., Uy, C., Bollig, C., Gilbert, R., and Irish, J. 2005, 'Quantitative three-dimensional ultrasound analysis of tongue protrusion,grooving, and symmetry: Data from 12 normal speakers and a partial glossectomee', *Clinical Linguistics and Phonetics*, vol. 19, no. 6-7, pp. 573-588.

Buchaillard, S., Perrier, P., & Payan, Y. 2009, 'A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning', *Journal of the Acoustical Society of America*, vol. 126, no. 4, pp. 2033- 2051.

Chan, T., Sandberg, B.Y., and Vese, L. 2000, 'Active Contours without Edges for Vector-Valued Images', *Journal of Visual Communication and Image Representation*, vol. 11, no. 2.

Chi-Fishman, G. 2005, 'Quantitative lingual, pharyngeal and laryngeal ultrasonography in swallowing research: A technical review', *International Journal of Clinical Linguistics and Phonetics*, pp. 589-604.

Coupling electromagnetic sensors and ultrasound images for tongue tracking: acquisition set up and preliminary results 2006, viewed 19 June 2012, < http://magrit.loria.fr/Confs/Issp06/>.

Dang, J., and Honda, K. 2001, 'A physiological articulatory model for simulating speech production process', *Acoustical Science and Technology*, Vol. 22, pp. 415- 425.

Dang, J., and Honda, K. 2004, 'Construction and control of a physiological articulatory model', *Journal of the Acoustical Society of America*, vol. 115, no. 2, pp. 853- 870.

Davidson, L. 2005, 'Addressing phonological questions with ultrasound', Clinical Linguistics & Phonetics, pp. 619-633.

Davidson, L. 2006, 'Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance', *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 407-415.

Engwall, O. 2000, 'A 3D Tongue Model Based on MRI Data', *In Proceeding of the International Conference of Spoken Language (ICSLP)*.

Engwall, O. 2002, 'Tongue Talking - Studies in Intraoral Speech Synthesis', PhD Thesis, KTH, Sweden.

Everson, R. 1998, 'Orthogonal but not orthonormal, Procrustes problems', *In Advances of Computational Mathematics*.

Fang, Q., Fujita, S., Lu, X., and Dang, J. 2008, 'A model based investigation of activation patterns of the tongue muscles for vowel production', *in Proceedings of InterSpeech*, pp. 2298-2301.

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

Gray, H. 1977, *Anatomy, Descriptive and Surgical*, Gramercy Books, New York.

Hiiemae, K. M., and Palmer, J. B. 2003, 'Tongue movements in feeding and speech'. *Critical Reviews in Oral Biology and Medicine*, vol. 14, no. 16, pp. 413-429.

Honggang, Y., Pattichis, M. S., Agurto, C., and Goens, M. B. 2011, 'A 3D Freehand Ultrasound System for Multi-view Reconstructions from Sparse 2D Scanning Planes', *BioMedical Engineering OnLine*, vol. 10, no. 1.

Ilie, M. D., Negrescu, C., and Stanomir, D. 2012, 'An efficient parametric model for real-time 3D tongue skeletal animation', paper published at *6th international conference on Communications and Information Technology*, pp. 129- 132.

Joseph C. Hager 2003, DataFace, Psychology, Appearance, and Behavior of the Human Face, http://face-and-emotion.com/dataface/anatomy/tonguemuscles.jsp.

Kakita, Y., Fujimura, O., and Honda, K. 1985, 'Computation of mapping from the muscular contraction pattern to formant pattern in vowel space', *The Journal of the Acoustical Society of America*, vol. 74, no. S1, pp. S117-S117.

Kass, M., Witkin, A., Trezopoulos, D. 1998, 'Snakes: Active contour models', *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321- 331.

Keller, E., and Ostry, D. 1983, 'Computerized measurement of tongue dorsum movements with pulsed echo ultrasound'. *Journal of the Acoustical Society of America*, vol. 73, no. 4, pp. 1309-1315.

Kelsey, C. A., Minifie, F. D., & Hixon, T. J. 1969, 'Applications of ultrasound in speech research', *Journal of Speech and Hearing Research*, vol. 12, pp. 564-575.

Kent, RD 1997, 'The speech sciences' Singular, San Diego.

King, S. A., and Parent R. E. 2001, 'A 3D parametric tongue model for animated speech', *Journal of Visualization and Computer Animation*, vol. 12, no. 3, pp. 107-115.

Lee, W., Magnenat-Thalmann, N. 2000, 'Fast head modeling for animation', *Image and Vision computing*, vol. 18, no. 4, pp. 355- 364.

Li, M., Kambhamettu, C., & Stone, M. *2003, 'Snake for Band Edge Extraction and Its Applications', Paper presented at 6th IASTED International Conference on Computers, Graphics, and Imaging*, August, Honolulu, HI, USA.

Li, M., Kambhamettu, C., & Stone, M. 2005. 'Automatic contour tracking in ultrasound images', *Clinical Linguistics and Phonetics*, vol. 19, pp. 545–554.

Li, M., Kambhamettu, C., and Stone, M. 2004, 'Automatic contour tracking in ultrasound images', *Clinical Linguistics & Phonetics*, pp. 545- 554.

Li, M., Kambhamettu, C., and Stone, M. 2004, 'Automatic contour tracking in ultrasound images', *Clinical Linguistics & Phonetics*, pp. 545- 554.

Lundberg, A. J., and Stone, M. 1999, 'Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound data'. *Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2858–2867.

Matsopoulos, G. K., Mouravliansky, N. A., Asvestas, P. A., Delibasis, K. K., and Kouloulias, V. 2005, 'Thoracic non-rigid registration combining self-organizing maps and radial basis functions', Medical Image Analysis, pp. 237– 254.

Olov Engwall 2001, Centre for Speech Technology/TMH, viewed 15 January 2012, http://www.speech.kth.se/~olov/tng_width.html.

Sonies, B. C., Shawker, T. H., Hall, T. E., Gerber, L. H., and Leighton, S. B. 1981, 'Ultrasonic visualization of tongue motion during speech'. *Journal of the Acoustical Society of America*, vol. 70, pp. 693–696.

Stavness, I., Gick, B., Derrick, D., and Fels, S. 2012, 'Biomechanical modeling of English /r/ variants', *Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. 355- 360.

Stavness, I., Lloyd, J., Payan, Y., and Fels, S. 2011, 'Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics', *International Journal for Numerical Methods in Biomedical Engineering*, vol. 27, no. 3, pp. 367- 390.

Stone, M. 2005, 'A Guide to Analyzing Tongue Motion from Ultrasound Images', *Clinical Linguistics & Phonetics*, vol. 19, no. 6-7, pp. 455-501.

Stone, M., and Lundberg, A. 1996, 'Three-dimensional tongue surface shapes of English consonants and vowels', *Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3728-3737.

Stone, M., and Shawker, T. H. 1986, 'An ultrasound examination of tongue movement during swallowing'. *Dysphagia*, vol. 1, no. 2, pp. 78-83.

Stone, M., Sonies, B., Shawker, T. H., Weiss, G., and Nadel, L. 1983, 'Analysis of real-time ultrasound images of tongue configuration using a grid digitizing system'. *Journal of Phonetics*, vol. 11, pp. 207-218.

Takemoto, H. 2001, 'Morphological analyses of the human tongue musculature for three-dimensional modeling', *Journal of Speech, Language, and Hearing Research,* vol. 44, pp. 95-107.

Tang, L., and Hamarneh, G. 2010, 'Graph-based tracking of the tongue contour in ultrasound sequences with adaptive temporal regularization', *Mathematical Methods for Biomedical Image Analysis (MMBIA)*, pp. 1-8.

Thompson, M. R. 2010, 'The Future of Portable Ultrasound: Business Strategies for Survival', *Massachusetts Institute of Technology*.

U.S. National Library of Medicine 2012, National Institutes of Health, Health & Human Services, viewed 15 February 2012, <http://www.nlm.nih.gov>.

Vocal Tract Visualization Laboratory n.d., viewed 15 January 2012, <http://speech.umaryland.edu/edgetrak.html >.

Watkin, K. L. 1999, 'Ultrasound and Swallowing'. *International Journal of Phonetics, Speech Therapy and Communication Pathology*, vol. 51, no. 4-5, pp. 183-198.

Wilhelms-Tricarico, R. 1995, 'Physiological modeling of speech production: Methods for modeling soft-tissue articulators', *Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3085-3098.

Wright, G. B. 2003, 'Radial Basis Interpolation: Numerical and Analytical Developments', PhD Thesis, University of Colorado, Boulder.

Yu, *H.,* Pattichis, *M. S.,* Agurto, *C., and* Goens*, M. B.* 2011, 'A 3D Freehand Ultrasound System for Multi-view Reconstructions from Sparse 2D Scanning Planes', *Biomechanical Engineering Online*, vol. 10, no. 1.