A Robust Tongue Shape Model for Ultrasound Recordings of Normal and Impaired Speech

by

Sahba CHANGIZI

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER'S DEGREE
WITH THESIS IN SOFTWARE ENGINEERING
M.A.Sc.

MONTREAL, JULY 11, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Ms. Catherine Laporte, Thesis supervisor
Department of Electrical Engineering at École de technologie supérieure

Ms. Lucie Ménard, Co-supervisor, Thesis Co-Supervisor
Directrice du Laboratoire de phonétique at Université du Québec à Montréal

Ms. Sylvie Ratté, Chair, Board of Examiners
Department of Electrical Engineering at École de technologie supérieure

Ms. Rachel E. Bouserhal, External Examiner
Department of Electrical Engineering at École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON JUNE 23, 2022

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

First of all, I would like to thank Professor Catherine Laporte, my supervisor at ÉTS, who helped me throughout my studies, especially when I faced problems. She provided consistent support during the running of this project. She trained me well and was also an excellent mentor to me. I am grateful that I had the chance to learn from her.

I would like to thank Professor Lucie Ménard, who agreed to co-supervise me throughout my studies, and her willingness to impart her knowledge. Thanks to her insightful remarks, I always had access to the data I needed to complete my project. I also have to thank Fonds de Recherche Québécois – Nature et Technologies for their financial support through the REPARTI strategic cluster. I also must thank Compute Canada for the computational resources I used in my experiences.

My heartfelt gratitude goes to my family for their kind support during this research. I apologize to my children for being grumpier than usual while I worked on this project. Without the encouragement of my wife, Leva, I would have given up on this research long ago. They have been a tremendous support during my studies.

Un modèle de forme de langue robuste à partir d'enregistrements ultrasons de la parole normale et altérée

Sahba CHANGIZI

RÉSUMÉ

L'imagerie par ultrasons est un outil utile pour observer les mouvements de la langue en interférant minimalement avec la parole naturelle. Il existe une variété de modèles pour quantifier la forme de la langue à partir de contours extraits d'images échographiques. Cependant, ceux-ci peuvent être affectés par une mauvaise qualité d'image, par exemple, lorsque des parties de la langue manquent sur les images en raison d'artefacts d'imagerie. Dans cette étude, nous étudions les effets de diverses erreurs d'extraction de contour sur la précision et la cohérence de différentes mesures de forme.

Nous avons développé des modèles de perturbation de contour exponentiels et polynomiaux, puis simulé la pointe et la racine de la langue manquantes, et étudié l'impact de ces perturbations sur les mesures de forme basées sur la transformée de Fourier discrète (DFT), l'indice de courbure modifié (MCI) et l'ajustement triangulaire. Ceci a été appliqué à un ensemble d'énoncés CV collectés auprès de locuteurs sains et déficients. Les résultats démontrent l'efficacité de la DFT et de l'ajustement triangulaire dans le regroupement de différents phonèmes malgré le bruit ajouté.

Il existe également un compromis entre la robustesse du modèle et la sensibilité aux différences réelles mineures dans la forme de la langue. Parfois, ces légères différences aident à regrouper les formes de langue qui diffèrent, par exemple en raison d'effets de coarticulation. Par conséquent, nous avons tenté d'améliorer la précision du modèle DFT en ajoutant des informations de contact palatal. Notre expérience montre que le nouveau modèle de forme est robuste au bruit et peut classer avec succès les énoncés CV et augmenter le score de classification de 23% pour les locuteurs qui ont des troubles de la parole.

Mots-clés: Quantification de la forme de la langue, Ultrason, Modélisation robuste de la forme de la langue

A Robust Tongue Shape Model for Ultrasound Recordings of Normal and Impaired Speech

Sahba CHANGIZI

ABSTRACT

Ultrasound imaging is a helpful tool to observe tongue movements while minimally interfering with natural speech. There exists a variety of models to quantify tongue shape based on contours extracted from ultrasound images. However, these can be affected by poor image quality, e.g., when parts of the tongue are missing from the images due to imaging artifacts. In this study, we investigate the effects of various contour extraction errors on the accuracy and consistency of different shape measures.

We developed exponential and polynomial contour perturbation models, then simulated missing tongue tip and root, and investigated the impact of these perturbations on shape measures based on the discrete Fourier transform (DFT), modified curvature index (MCI), and triangular fitting. This was applied to a set of CV utterances collected from healthy speakers and speakers who were diagnosed with speech deficits. Results demonstrate the effectiveness of DFT, MCI and triangular fitting in clustering different phonemes despite the added noise. There is also a trade-off between the robustness of the model and sensitivity to minor actual differences in tongue shape. Sometimes, these slight differences help group tongue shapes that differ, e.g., due to coarticulation effects. Therefore, we have attempted to improve the precision of the DFT model by adding palatal contact information. Our experiment shows that the new shape model is robust to the noise and can successfully classify our target CV utterances and increase the classification score by 23% for speakers with speech deficits.

Keywords: Tongue shape quantification, Ultrasound, Robust tongue shape modeling

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

CV	Word starting with a consonant sound and followed by a vowel sound
CI	Curvature Index
DEI	Dorsum Excursion Index
DFT	Discrete Fourier Transform
EMA	Electromagnetic Articulography
EPG	Electropalatography
ÉTS	École de Technologie Supérieure
EXP	Exponential
LDA	Linear Discriminant analysis
LOC	Location of Curvature
MCI	Modified Curvature Index
MRI	Magnetic resonance imaging
NINFL	Number of Inflections
PLY	Polynomial
SS	Smoothing Spline
UQAM	Université du Québec à Montréal
US	Ultrasound

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

mm	millimeters

INTRODUCTION

The study of tongue motion can help us gain knowledge about movements of the tongue during speech articulation. This can be applied to the various contexts of speech-related research on topics such as second language acquisition, speech therapy, and biomechanical modeling of the tongue. Ultrasound imaging is one of the most common imaging techniques used to observe tongue movements during speech, because it provides real-time images of the tongue movements in the oral cavity. It is also relatively cheap and non-invasive (Karimi, Ménard & Laporte, 2019).

The main objective of this project is to construct a concise model of tongue shape from ultrasound images across a variety of short utterances. However, the direct application of such a model to tongue contours extracted from ultrasound images is a challenging task. For example, due to the reflections of ultrasound waves, some parts of the tongue might be missing from the US image. In scanning the tongue with ultrasound, we are primarily interested in the tongue's upper surface. But in this analysis, various sources of noise might exist in the US image. Backscattered echoes reflected from tongue structures, including vessels and tendons, might create high contrast edges that might be mistaken with the actual tongue surface. Other scenarios might result in images with poor contrast. For example, the tissue between the probe and the tongue might weaken the ultrasound reflection, which results in a lower contrast of the US image. And finally, because of the trapped air beneath the tongue and especially on the tip section, the ultrasound waves might not penetrate through the tongue tissue, and a shadow appears in the tip section of the tongue (Karimi *et al.*, 2019).

Because of such scenarios, it is crucial to consider the impact of noise and missing tongue sections on the accuracy of tongue shape models. As part of this project, we will review several tongue shape models and present an innovative shape model that combines the palatal information with the tongue shape model. In the next step, we will survey the impact of different imaging artefacts and analyze how robustly each measure behaves in their presence. We will

apply perturbation functions to the tongue contours (mainly to the root and tip of the tongue) to simulate various tongue contour extraction errors, including those arising from missing parts of the tongue in the US image. Then we will analyze the robustness of each measure to see how it behaves in these scenarios. By understanding the limitations and strengths of each measure, we can obtain suitable measures for specific contexts and use the concepts behind each of them to propose new shape measures that can quantify the most important characteristics of tongue shape and perform less sensitive to the missing tip and root tongue sections in the US frames.

CHAPTER 1

BACKGROUND AND LITERATURE REVIEW

This chapter provides background material and a literature review supporting the new tongue shape model presented in this thesis. The structure of this chapter is described below: Section 1.1 discusses anatomical concepts; section 1.2 addresses the application of ultrasound in speech disorders. Section 1.3 discusses some principles related to ultrasound imaging and how the ultrasound device works. Lastly, section 1.4 through 1.6 discuss the tongue shape quantification methods and outlines the described modeling techniques and the strengths and weaknesses of each model.

1.1 Anatomy of the pharyngeal-oral apparatus

The upper portion of the vocal tract consists of the oral cavity and the nasal cavity. As is illustrated in figure 1.1, the oral cavity is bounded by the lips, tongue, palate, and epiglottis.

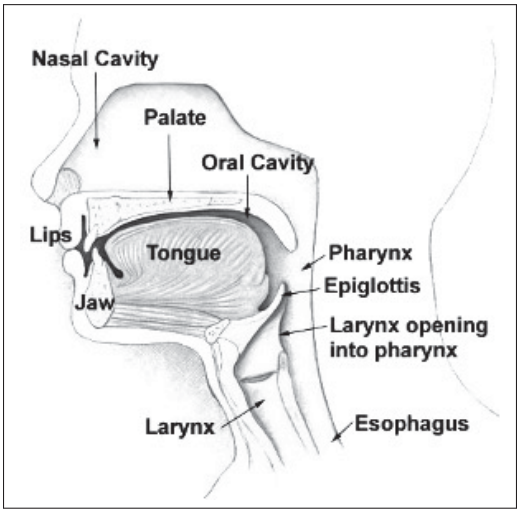

Figure 1.1 Illustration of the vocal tract and related structures Taken from Andrade-Miranda (2017)

As air passes through the vocal tract, the vocal folds in the larynx come together, and sound is produced. In oral sounds, the sound travels through the oral cavity, and the movements of the tongue, jaws, and lips reshape the vocal tract to create understandable speech sounds. The tongue, in particular, modifies the produced sound by modulating the width of the oral cavity in various locations.

An essential component of the oral cavity is the tongue. In the literature, it is often subdivided into separate regions. Figure 1.2 illustrates these parts of the tongue:

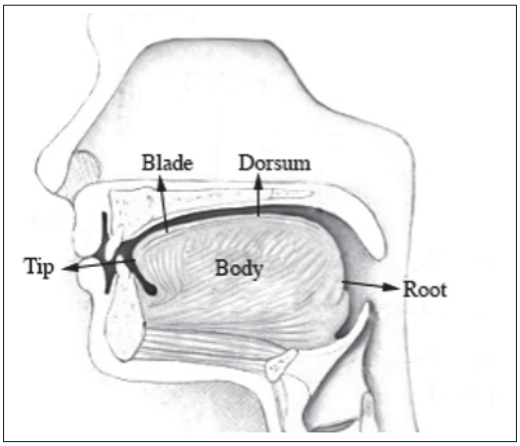

Figure 1.2 Five components of the tongue structure
Taken from Hixon (2014)

Depending on their intent, the divisions may have anatomical or functional bases. The literature usually identifies four different sections of the tongue: Blade, dorsum, root, and tip.

1.2 Ultrasound imaging in speech research

In this section, first, we compare different measurement techniques used for speech research and then compare the advantages and disadvantages of each method. The first method we survey is X-ray microbeam imaging. X-ray radiation is passed through the jawbone and tongue soft tissue

in X-ray microbeam imaging, and received by a detector on the other side (Zharkova, 2018). The different amounts of radiation absorbed by various organ densities create the resulting images. Electromagnetic articulography (EMA) measures the position of different flesh points in a subject's mouth during speech (Lancia & Tiede, 2012). Electromagnetic coils are inserted inside the subject's mouth and tracked by an electromagnetic sensor to measure the tongue, jaw, and lip movements.

In Electropalatography (EPG), an artificial palate is molded and attached below the speaker's hard palate. Upon contact between the hard palate and the tongue, electrodes on the artificial palate send electrical signals to the processing unit, and the contact information is recorded. Magnetic resonance imaging with tagging (Tagged MRI) is another imaging technique that uses strong magnetic fields with tissue tagging to measure the motion and deformation of tongue muscles during speech (Chen, Lee, Byrd, Narayanan & Nayak, 2020). And lastly, in the ultrasound imaging method, the ultrasound probe is positioned below the subject's chin. It transmits high-frequency sound waves emitted through the tongue soft tissue and received by the transducer, resulting in real-time imaging of the shape and movement of the tongue during speech.

In comparison to other measurement methods like X-ray imaging, which has radiation risks; MRI which is expensive; EMA, which needs placing intrusive coils and sensors inside the mouth; and EPG, which is also invasive and only provides limited data about contact points of the tongue and hard palate; ultrasound imaging is safe, non-invasive, easy to use and provides an image of the whole tongue surface at reasonably good frame rates. This makes ultrasound an ideal tool for imaging tongue gestures (Stone, 2005). In addition, ultrasound is better suited for imaging in small children than EPG because patients are not expected to wear artificial dental palates. Children are not comfortable talking and having artificial palate in their mouths (Lee, Wrench & Sancibrian, 2015). Ultrasound can also provide visual feedback to patients with speech disorders related to their articulatory movements. Traditionally a mirror was used in front of the patient during the speech to obtain visual feedback (Stone, 2005). But with the new method, the patient and speech therapist observe real-time movements of different tongue

parts, overall tongue shape, and parts of the hard palate using the ultrasound device, which is not possible when using a mirror in the traditional method. Ultrasound is also used to assess treatment efficiency, in speakers who undergo long-term speech therapy.

In phonetics research, ultrasound can help characterize tongue shapes and the relationship between tongue shapes and speech acoustics. For example, some specific measures classify the tongue shapes for correct and incorrect speech productions in sagittal or coronal US tongue images (Dawson, Tiede & Whalen, 2016). Portable ultrasound systems are also particularly well suited for the field work when analyzing under-documented languages for instance. In sections 1.5 through 1.7, we will look at some important measuring methods suitable for analyzing the tongue shape. We compare them and survey the strengths and weaknesses of each method.

1.3 Concepts of Ultrasound Imaging

Ultrasound uses short-wavelength acoustical waves to produce the image of an object. Ultrasound is a high-frequency sound wave, and it is typically produced using a piezoelectric crystal. A piezoelectrical crystal is a material that transforms electrical signals into mechanical vibrations and vice-versa. An ultrasound transducer, which is indeed a transceiver, is made of piezoelectric crystals that both emit and receive ultrasound waves. The density variation between an object (e.g., the tongue surface) and its surroundings is the cause of an acoustic impedance mismatch, which in turn causes the partial reflection of ultrasound waves towards the transducer. Figure 1.3 demonstrates the propagation of ultrasound waves in objects with different densities. The left side image shows acoustic impedance mismatch ΔZ when density is not equal, and the right side illustrates equal impedances in the same densities:

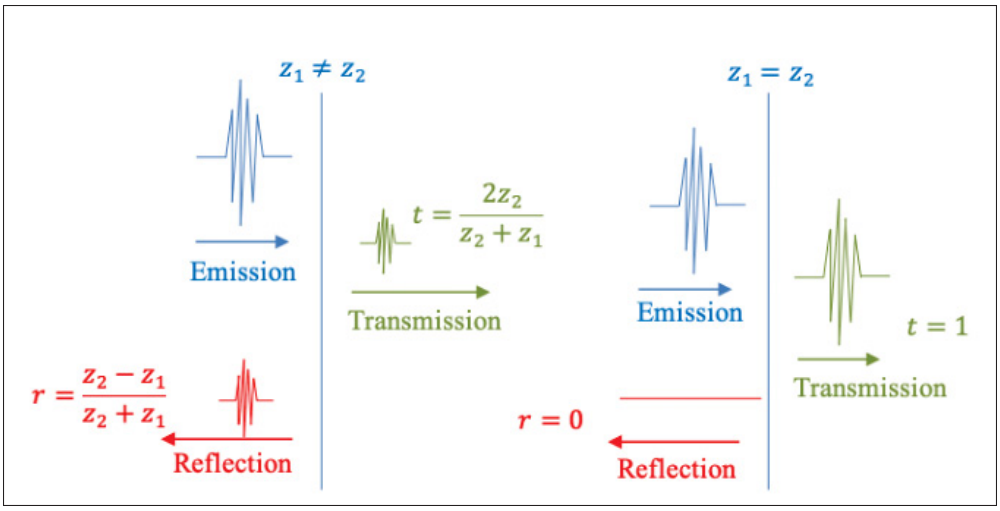

Figure 1.3 The propagation of ultrasound waves between two densities
Taken from Guillermic *et al.* (2019)

To image an object with ultrasound, the transducer is mounted on one side of the object. The sound travels through the object until the impedance mismatch at the opposite edge or surface creates a reflection. The reflected echo will then be received by the transducer and translated into an image of that object. The reflected sound returns to the source when the reflecting surface is perpendicular to the ultrasound probe. A significant change in density produces a strong echo, as seen in tissue-to-air (e.g., tongue to air) and tissue-to-bone interfaces. Weaker echoes can be created at interfaces with similar densities like tissue-to-tissue and tissue-to-water objects. Figure 1.4 illustrates the vocal tract anatomy on the right side and a superimposed ultrasound image on the left side:

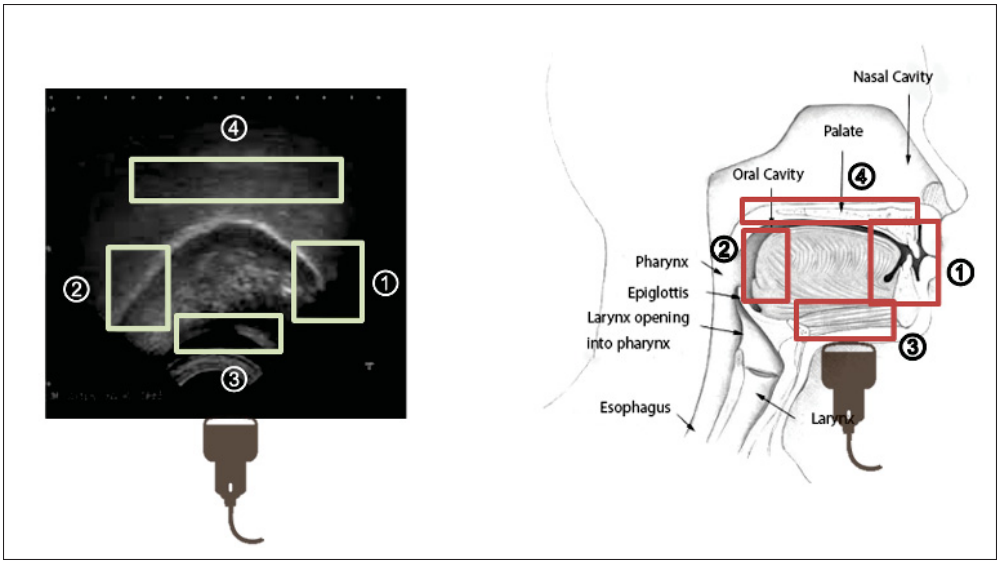

Figure 1.4 The anatomy of the vocal tract on the right. Ultrasound image of the tongue on the left
Taken from Gosztolya *et al.* (2019)

For scanning the tongue surface, the transducer is placed beneath the chin, and the ultrasound waves travel upward and are reflected by the top surface of the tongue. Usually, air or the palate above the tongue restricts the transmission of ultrasound waves and forces the waves to reflect towards the transducer.

1.4 Review of US tongue shape measures

This section presents existing measures for quantifying tongue shape in US images. This section describes tongue shape measures currently documented in the literature and then discusses the weaknesses and strengths of each of those measures (Table 1.1). Quantifying and measuring the tongue shape is a challenging task because of the muscular hydrostat characteristic of the tongue (Dawson *et al.*, 2016). Deformations like twisting, bending, shortening, and elongation can happen simultaneously at multiple organ points in a muscular hydrostatic organ. These deformations can occur independently concerning other parts of the structure at each point.

This characteristic is also responsible for the various shapes of the tongue during articulation (Kier & Smith, 1985).

The complexity of the tongue shape is a parameter that describes how an individual can independently control their tongue muscles (Dawson *et al.*, 2016). Other than the shape complexity, extra parameters can be quantified by the shape measures described in this thesis. For example, absolute shape measures quantify parameters like curvature degree and asymmetry index that pertain to a single shape. On the other hand, relative shape measures quantify differences between the two tongue shapes.

1.5 Absolute shape measures

The absolute shape measures are suitable for describing the tongue shape in one US frame. In sections 1.5.1 through 1.5.7 of this chapter we will review a set of absolute shape measures and explain how they are used for summarizing individual tongue contours in a single US frame.

1.5.1 Lingua

The Lingua method was developed by Ménard, Aubin, Thibeault & Richard (2012) to address a metric for measuring tongue asymmetry and tongue curvature parameters. The tongue asymmetry index specifies the bunching position along the tongue, and the curvature index identifies the amount of bunching in tongue shape. These parameters are applied to vowels. In the first step, a 19-segment radial grid with an angular distance of 5° is superimposed onto the tongue contours. The intersection point of these segments is considered as the transducer's location. The intersection point between the tongue contour and each grid segment is also located on the tongue curvature. The radial distances between this point and the origin of the grid are calculated, yielding a collection of parameters that quantifies the tongue position during articulation.

A triangle is fitted to each contour in the second step, as shown in Figure 1.5. The first and last points of the tongue contour are linked together and considered as the triangle base. Point C on

the triangle is the highest point of the tongue contour relative to the triangle base. The curvature index and tongue asymmetry index are calculated using the following equations:

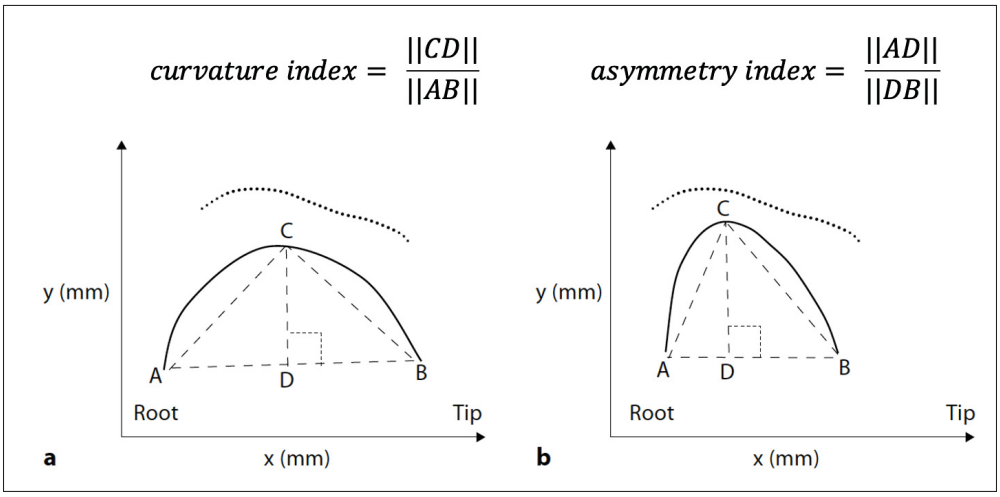

Figure 1.5 Reshaping tongue contours into a triangle shape

Considering the Lingua shape indexes, we can visualize how high curvature occurs, with the sounds like [u] and low tongue curvature corresponds to the sounds like [a]. The tongue asymmetry index is a measure of the position of the tongue's mass with respect to the whole tongue. For instance, sound [u] results in higher values for tongue asymmetry index which means bunching occurs at back of the tongue. On the other hand, low values for asymmetry index related to the sound [i] indicate that bunching happens at front of the tongue. One advantage of the Lingua shape measure is that it is easy to implement and ideal for describing vowel sounds. On the other hand, Lingua cannot precisely model tongue shapes with multiple inflection points, so it is not ideal for quantifying more complex tongue shapes like those related to consonant sounds.

1.5.2 Dorsum excursion index

The dorsum excursion index (DEI) is an essential measure for differentiating velar and alveolar phonemes, presented by (Zharkova, 2013). In the English language, the consonants [k], [g], and [ŋ] are examples of velar sounds. The consonants [t], [n] and [d] are examples of alveolar sounds. DEI is calculated using the equation below:

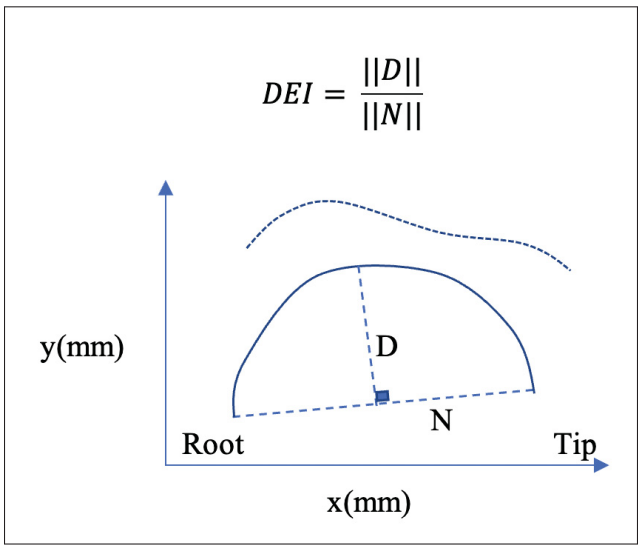

Figure 1.6 Annotation of DEI parameters

In the first step of DEI computation, a line connects the beginning and end of the tongue curve, named with the letter N (see Figure 1.6). Perpendicular lines to N from each point on the tongue contour are drawn. Line D denotes the perpendicular line that crosses line N at the midpoint. The DEI method is sensitive to covert contrast, which refers to sounds that are heard similarly but produced differently (Cleland & Scobbie, 2020). An instance of covert contrast happens in children with late developed motor control abilities that substitute the sound [t] with the sound [d], but the sound that is heard is [t] for both cases. The calculation of the DEI is similar to calculating the curvature value in the Lingua shape model. One weakness of the DEI method is

that it can be affected when some parts of the anterior section of the tongue are missing from the ultrasound image.

1.5.3 Loc a-i measure

The LOCa-i, which stands for Location Of Curvature, describes bunching along the tongue curve (Zharkova, 2018). The difference between the sound [i] and [a] is that excursion happens more at the back section of the tongue curve for sound [a] in comparison with the sound [i]. The letters “a” and “i” in this shape model mean that high index values are similar to tongue shapes producing the sound [i], and low index values are similar to tongue shapes making the sound [a]. This index quantifies the ratio of the excursion of the anterior section of the tongue relative to the posterior section. Figure 1.7 illustrates the calculation steps for this index:

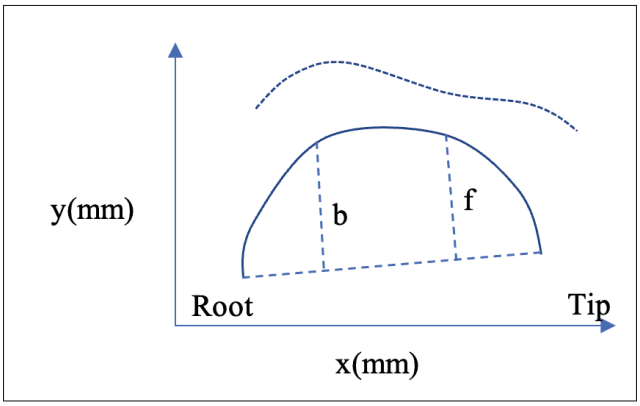

Figure 1.7 Annotation of the LOCa-i parameter

Two points are selected along the tongue surface. The two points are set at one-third and two-thirds from the front of the tongue curve. Then two lines are drawn from two points perpendicular to the tongue curve, named f and b in the figure. The length ratio of lines f to b yields the LOCa-i index. The advantage of the LOCa-i shape measure is that it is highly effective in quantifying bunched tongue shapes.

1.5.4 Modified curvature index (MCI)

Stolar & Gick (2013) suggested the curvature index (CI) measure to characterize the tongue shape in any plane (midsagittal, coronal, or transverse). This shape measure is intended as a measure of shape complexity. This method is suitable for comparing tongue shapes with significant curvature variations, image quality, orientation, and different tongue sizes. These variations are crucial as they can be associated with speech disorders. The CI is a scalar index derived from a seventh-order polynomial fit applied to a tongue contour. The modified curvature index (MCI) measure (Dawson *et al.*, 2016) complements the CI method.

The original CI method first calculates a seventh-order polynomial fit along the tongue contour. In the next step, the integral of the absolute curvature is calculated. Equation 1.1 and 1.2 presents the initial and modified equations:

CI (Stolar & Gick, 2013):

$$\int_a^b |k| dx \tag{1.1}$$

MCI (Dawson *et al.*, 2016):

$$\int_\alpha^\beta |k| ds \tag{1.2}$$

In equation 1.1:

- k denotes the curvature value.
- a and b represent the x coordinates of the beginning and end of the tongue contour.

In equation 1.2:

- α and β represent the x coordinates of the beginning and end of the tongue contour.

The MCI measure is modified in two aspects compared to the earlier version of the shape measure:

The modified version integrates the curvature values concerning the tongue arc length instead of the x-axis. Dawson *et al.* (2016) applied this modification because x-axis integration causes variation in the metric when the ultrasound probe is rotated.

A seventh-order polynomial fit was used in an earlier version of this measure. The application of polynomial fitting is to obtain less noisy curves. However, high order polynomials cause undesired results at the tongue boundaries. So, in the modified version of the measure, a 5th-order low-pass filter is applied to the tongue contour to eliminate the noise. Figure 1.8 illustrates the calculation steps:

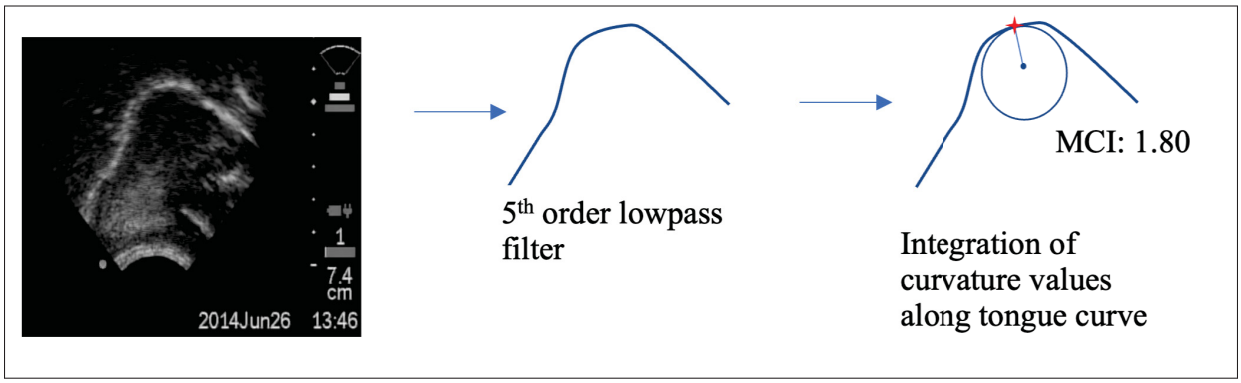

Figure 1.8 Illustration of processing steps to calculate the MCI shape measure

In comparison, the CI measure utilizes the polynomial function that fits a smooth curve along the tongue curve. However, as a weakness, using the polynomial fit might result in unexpected results on tongue contour boundaries. The modified version (MCI) uses an enhanced approach to eliminate the noise and yields a continuum index that can be translated to low and high-complexity tongue shapes. One weakness of the MCI measure is that it is susceptible to the noise sources produced during contour extraction.

1.5.5 Principal component analysis

Principal component analysis (PCA) is a general-purpose statistical analysis technique used to reduce the dimensionality of a population of high-dimensional data points. A multidimensional coordinate vector stores the coordinates of points along tongue contours. PCA can be used to find the most important modes of variation in a population of tongue contour shapes. PCA decreases the dimensionality of the tongue contour by determining the orthogonal directions of data variability. The input data for PCA can be a population of tongue contours extracted from different recording sessions. Each contour can then be individually represented by the weights associated with each PC to reconstruct the shape of the tongue.

This method generates a collection of vectors named as principal components (PCs) containing the most important tongue shape features. The importance of features is determined by the magnitude of the corresponding values in eigenvectors. Meaning higher magnitude values are translated into more important features. First, the covariance matrix is calculated, the covariance matrix demonstrates the covariance between two data point coordinates and is a $p \times p$ symmetric matrix, p is the number of datapoint dimensions multiplied by the number of points. The covariance matrix demonstrates the covariance between two data point coordinates. This step aims to identify the relationship between original data points. In the next step, the covariance matrix is used to calculate the eigenvectors and eigenvalues. Eigenvectors are the direction of axes containing the most variance, and eigenvalues are the coefficients related to eigenvectors. Eigenvalues help calculate the amount of variance that exists in each principal component. Principal components are a set of transformed variables that are uncorrelated and contain most of the features from the original data points. However, the most important features of the original datapoints exist in the first principal component. Finally, the principal components relative to their degree of significance will be obtained by sorting the eigenvalues from the most significant to minor values.

PCA can be used to extract shapes in different designs, which can be helpful in a variety of domains like nanobiology and signal processing (Tuta, Nicolaescu, Mariescu-Istodor & Digulescu,

2019). PCA can also help analyzing the velocity patterns in high dimensional velocity vectors superimposed onto the tongue shape (Dawson *et al.*, 2016). These patterns help cluster tongue shapes without prior knowledge about tongue shapes for specific sounds. Although the PCA model is easy to calculate (which can be identified as an advantage of this shape measure), the results produced by this shape measure are not readily interpretable. Meaning they can not be directly translated into the motor control abilities of the experiment subject.

1.5.6 NINFL measure

The number of inflections (NINFL) measure represents the number of tongue curvature changes whose values are more significant than a threshold (Preston, McCabe, Tiede & Whalen, 2019). This method is developed to quantify the shape complexity of a tongue contour. The sound [ʃ] is an example sound that represents tongue shape with moderate complexity; also, the shape of the tongue for the sound [ɹ], e.g., in the word "bird" is an example of a complex tongue shape. Figure 1.9 demonstrates how inflection points are detected for a tongue curve:

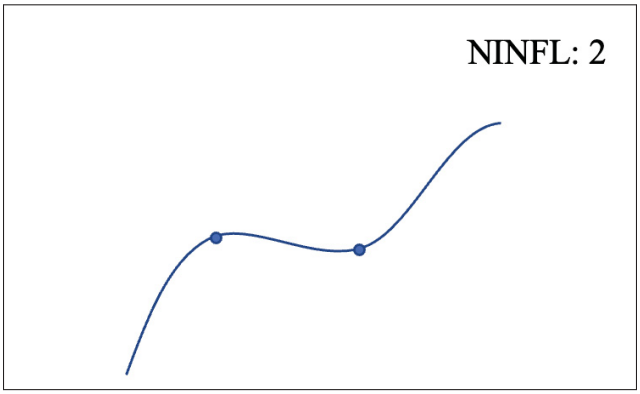

Figure 1.9 Demonstration of inflection points on tongue shape

Equation 1.3 yields the degree of inflection at a particular point x, y. The primes denote derivatives with respect to offsets along the curve:

$$k = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{\frac{3}{2}}} \tag{1.3}$$

The curvature changes are calculated using the derivatives. An inflection point is a point in a curve where the concavity changes. As an advantage of this shape measure Preston *et al.* (2019) showed that NINFL is a good measure for characterizing individuals who are having distorted [ɹ] productions and can effectively describe the tongue shape complexity in US images. On the other hand, complex tongue shapes with high NINFL values do not yet guarantee the precise speech quality of the experiment subject, which can be identified as a weakness for this measure.

1.5.7 Discrete Fourier transform measure

In this method, the tongue shape is translated into the domain of spatial frequency. The main concept of the Discrete Fourier transform (DFT) algorithm is to transform a function into a sum of sine and cosine statements. The discrete Fourier transform, transforms a sequence of N discrete sample points denoted $\{x_n\} := x_0, x_1, ..., x_{N-1}$ in cartesian space into a vector of coefficients represented as $\{C_n\} := C_0, C_1, ..., C_{N-1}$. Equation 1.4 demonstrates the calculation formula of coefficient (C_k) :

$$C_k = \sum_{n=1}^{N-1} x_n . e^{\frac{-i2\pi}{N} kn} \tag{1.4}$$

In the method presented by Dawson *et al.* (2016), tongue shape is characterized by the first three Fourier coefficients. The first Fourier coefficient (C1) contains the largest-scale feature set of the tongue curve, and higher coefficients represent the smaller-scale features. Liljencrants (1971) also offered a method to quantify the tongue shape obtained by X-ray imaging using Fourier transformation. The differences between Liljencrant's and Dawson's methods are that Liljencrant's method is focused on profiling tongue contours extracted from X-ray images and a static coordinate system with two coefficients calculated for the Fourier series. The function transformed using the Fourier series is the tongue shape relative to the fixed coordinate system.

In Dawson's implementation, there is no fixed coordinate system defined. The Fourier series transforms tangent angles related to each point on the tongue contour as a function of arc length (Dawson *et al.*, 2016). Liljencrants (1971) observed that the maximum constriction point in the tongue shape moves from the dental toward the pharyngeal position in different vowels. Thus, he concluded that the phase of the spatial fundamental identifies the constriction location in tongue shape and that magnitude identifies the extent of this constriction.

The advantage of the Fourier transform is that it is relatively insensitive to the rotation of the probe, which can be the source of error in US recordings. Dawson *et al.* (2016) showed that the discrete Fourier transform is a good classifier for delineating tongue complexity groups. As a weakness, this shape measure does not provide a continuum index that can be interpreted in terms of complexity, and it is not easy to directly translate this shape measure to the motor control abilities of the experiment subject.

1.6 Relative Measures

All measures described in section 1.5 were categorized as absolute shape measures. In this section, we describe the category of relative shape measures. This category of tongue shape measures is suitable for identifying the differences between two tongue shapes for a given acoustic target, making it ideal for monitoring speech therapy or second language acquisition.

1.6.1 KT measures

The KTmax and KT crescent area measures were first proposed by Cleland & Scobbie (2020). These two measures are suitable for differentiating between tongue shapes by quantifying the magnitude of this differentiation into absolute spatial terms (Cleland & Scobbie, 2020). These models can be used to measure characteristics of tongue shape for typical children and then comparing them with children diagnosed with speech disorders before, during, and after the intervention. These measures can also be used to classify children's speech with velar fronting history as they develop contrast longitudinally (Cleland & Scobbie, 2020). The KTmax measure

quantifies the difference in constriction magnitude between two target tongue shapes. It is a comparative measure that superimposes 42 radials on two overlapped tongue contours. The overlapped tongue contours intersect in two points near the root and tip of the tongue, with the area in between forming a crescent shape. Knot points are defined as the intersection points of the two tongue contours and the radials. The radial difference is also defined as the distance between two knot points placed on a single fan line. The maximum radial difference yields the KTmax index.

In the KT crescent area measure, same as the KTmax, 42 radials are superimposed onto overlapped tongue contours. The intersection of radials and tongue contours forms several annular sectors in the area between two tongue shapes. Figure 1.10 demonstrates the annular sectors created between two tongue contours:

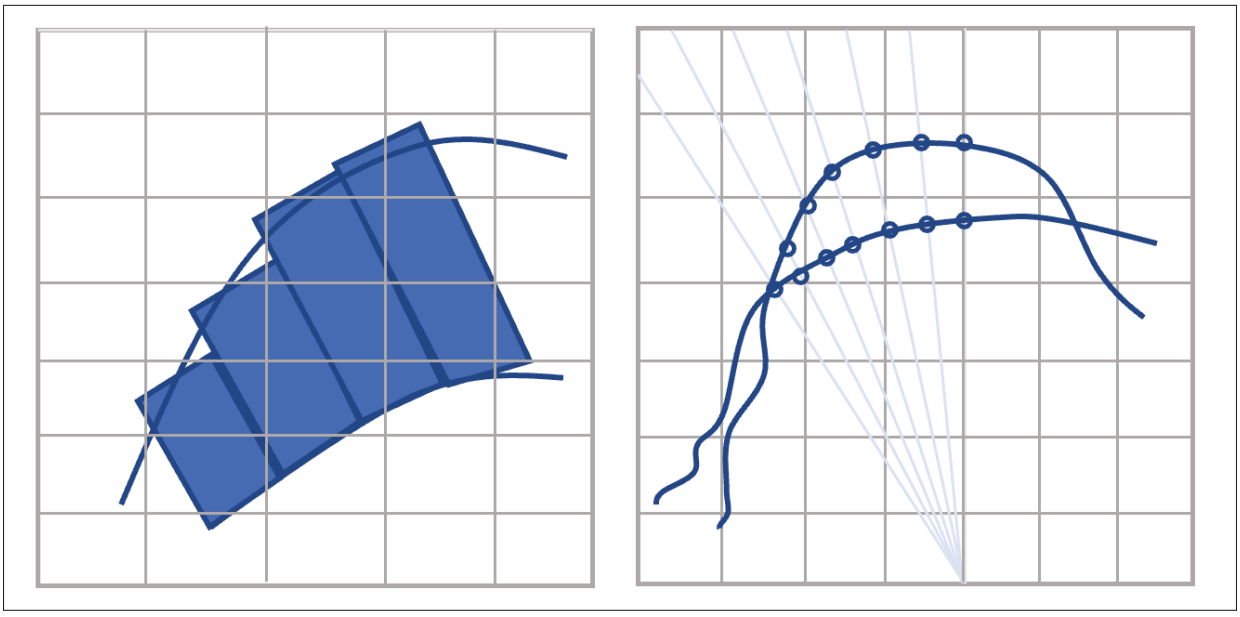

Figure 1.10 Left side: Annular sectors for the KT crescent area measure, Right side: Illustration of crescent shape between two contours. KTMax refers to the maximum radial difference between knot points
Taken from Cleland & Scobbie (2020)

By summing up the area of these sectors, we can estimate the crescent area. Other than KTmax, which calculates the maximum radial difference, KT crescent, calculates the midsagittal crescent area in each vowel context. Equation 1.5 is used to calculate the crescent area:

$$A = (\frac{\theta}{360}) \times \pi(k^2 - T^2) \tag{1.5}$$

where A denotes the crescent area, θ represents the angle between radials. K represents the radial distance to the knot on the upper curve, and T indicates the radial distance to the knot on the lower curve.

An advantage of the KTmax measure is that it is relatively easy to calculate. Still, both KTmax and KTCrescent measures cannot accurately classify complex tongue shapes with more than one constriction point along the tongue curve.

1.6.2 The $RD\Sigma$ measure

Summed radial difference ($RD\Sigma$) is a shape measure presented by Havenhill (2019) and is suitable for quantifying differences in tongue fronting between two tongue contours. Figure 1.11 illustrates the calculation process for the $RD\Sigma$ measure:

Figure 1.11 Illustration of how the difference of radials is calculated for the $RD\Sigma$ measure Taken from Havenhill (2019)

As with the KT crescent measure, 42 radials are superimposed on the overlapped tongue contours. The intersection points of radials and two tongue curves are inspected. The $RD\Sigma$ value for a pair of tongue curves is then defined as the sum of differences in radius between two tongue curves. One weakness of this shape measure is that it cannot accurately classify complex tongue shapes with more than one constriction point along the tongue curve.

1.6.3 Smoothing spline analysis of variance

The smoothing spline analysis of variance (SS ANOVA) is a tongue shape analysis method used for tongue shape comparison proposed by Davidson (2006). In this method, tongue contours are approximated by smoothing splines. A smoothing spline is a smoothing function $G(x)$ that fits a smooth piecewise polynomial curve to discrete tongue contour data points. The smoothing spline function (Eq. 1.6) has two terms. The first term tries to fit the curve into the discrete

points, and the second one assesses the roughness of the curvature to make sure that the resulting curve is smooth:

$$G(x) = \frac{1}{n}\sum_{all\ i}(y_i - f(x_i))^2 + \lambda \int_a^b (f''(u))^2 du \tag{1.6}$$

In equation 1.6:

- λ denotes the smoothing parameter
- a is the x coordinate of the start point of the spline
- b represents the x coordinate of the endpoint of the spline
- n is the number of discrete data points

The SS ANOVA model is a statistical model that can tell if two sets of smoothing splines are significantly different. Applying this model to the tongue contours can determine if two groups of tongue shapes are statistically different with a specified confidence level. When using this technique, we must ensure that the head and jaw movements relative to the ultrasound probe are limited or corrected. Otherwise, this measure does not yield accurate results. Figure 1.12 illustrates the results of SS-ANOVA comparing the sound [k] in two different contexts:

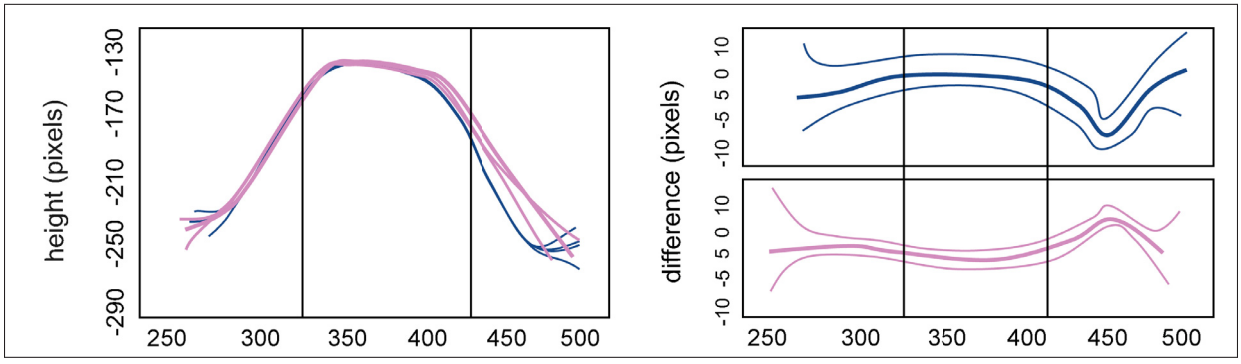

Figure 1.12 The SS-ANOVA comparison of tongue shapes for /k/ in words “black top” (dark blue line) and “blacktop” (light pink line). Left side is the original contour information and right side is the distance value plot of the co-registered points Taken from Davidson (2006)

The resulting output of this model is a single number that quantifies the mean squared distance between the co-registered points of the two curves. A greater value indicates significant differences between two shapes and a lower value represent further similarities between the two contours. Barbier *et al.* (2020) suggested an extension of the SS-ANOVA model in which the area between both average splines is calculated and used to quantify the differences between two tongue shapes. Figure 1.13 demonstrates the average splines and the 95% confidence interval related to different experiment repetitions:

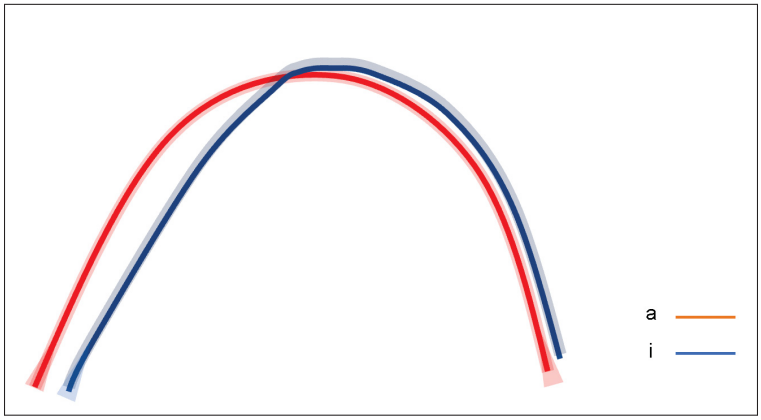

Figure 1.13 Representation of average splines plotted with 95% confidence interval for two different sounds. The unit for x and y axis is mm
Taken from Barbier *et al.* (2020)

The advantage of this shape measure is that changes in the shape of the tongue, rotation, or translation are considered in the statistical analysis for this model. However, as a weakness, this requires head and jaw movement to be constrained or compensated during the analysis of US images.

1.7 Discussion

From the list of shape measures described in sections 1.5 and 1.6, each has advantages and disadvantages, making them suitable for studies in specific domains. Each technique might describe certain characteristics of the tongue shape but have weaknesses in certain circumstances. For instance, extracting smoother and less noisy curvature values is the crucial objective of applying the polynomial fit. On the other hand, using high order polynomial fits might lead to unexpected results on the boundaries of the tongue shape (Dawson *et al.*, 2016). Table 1.1 summarizes the characteristics of each shape measure:

Table 1.1 Summary of tongue measuring methods with related advantages and disadvantages

Method	Notes
Lingua (Ménard *et al.*, 2012)	**Description:** Fits a triangle to the tongue contour; curvature and asymmetry measures are derived from the triangle **Advantages:** • Easy to implement and ideal for describing vowels **Weaknesses:** • Not able to correctly model tongue shapes with multiple inflections • Not suitable for complex tongue shapes related to consonants

Method	Notes
DEI (Dorsum Excursion Index) (Zharkova, 2013)	**Description:** The degree of the tongue dorsum excursion conceptually represented by the point on the curve of the tongue, positioned across the center of the straight line between the start and end of the curve line **Advantages:** • Sensitive to covert contrast **Weaknesses:** • Cannot differentiate consonants effectively in different vowel contexts • Highly affected by the portion of tongue curve included in the US image
CI (Curvature index) (Stolar & Gick, 2013)	**Description:** The measure of the degree of convexity of the tongue over its entire volume **Advantages:** • The polynomial function helps to fit a smooth curve along tongue contours **Weaknesses:** • Using the polynomial fit might result in unexpected results on tongue contour boundaries

Method	Notes
MCI (Modified Curvature Index) (Dawson *et al.*, 2016)	**Description:** Quantifies the degree of curvature **Advantages:** • Relatively insensitive to the rotation of the probe • Forms a continuum where low and high complexity shapes are less likely to be confused **Weaknesses:** • Highly sensitive to the noise produced by the contour tracking method
LOCa-i (Zharkova, Gibbon & Hardcastle, 2015)	**Description:** Quantifies the degree of tongue excursion compared to the root of the tongue. **Advantages:** • Can effectively classify bunched tongue shapes
PCA (Principal component analysis) (Stone, Liu, Chen & Prince, 2010)	**Description:** Transforms a set of tongue contour data points into a smaller collection of vectors that describe the most important features of the original contours **Advantages:** • Easy to calculate **Weaknesses:** • Results are not always readily interpretable

Method	Notes
NINFL (Number of inflections) (Preston *et al.*, 2019)	**Description:** Counts the number of curvature sign changes **Advantages:** • A good measure for characterizing individuals who are having distorted [ɹ] productions • An excellent measure to quantify tongue shape complexity **Weaknesses:** • Complex tongue shapes with high NINFL values do not yet guarantee precise speech quality
Discrete Fourier transformation (Dawson *et al.*, 2016)	**Description:** Transform tongue contour into Sine and Cosine components **Advantages:** • Relatively insensitive to the rotation of the probe • Good parameter classifier for delineating tongue complexity groups **Weaknesses:** • It does not provide a continuum that can be interpreted in terms of complexity • Not easy to interpret from the motor control aspect
Maximal KT crescent radial differences (Cleland & Scobbie, 2020)	**Description:** Quantifies the maximum radial difference in the dorsal crescent **Advantages:** • Easy to calculate **Weaknesses:** • Not able to accurately classify complex tongue shapes with more than one constriction point

Method	Notes
KT Crescent (Cleland & Scobbie, 2020)	**Description:** Quantifies the spatial dorsal velar constriction of the alveolar baseline **Advantages:** • Relatively insensitive to the rotation of the probe **Weaknesses:** • Not able to accurately classify complex tongue shapes with more than one constriction point
$RD\Sigma$ (Summed Radial Distances) (Havenhill, 2019)	**Description:** Sum of the differences in radius between the two mean tongue contours. **Advantages:** • Suitable for quantification of tongue fronting **Weaknesses:** • Not able to accurately classify complex tongue shapes with more than one constriction point
SS ANOVA (Davidson, 2006)	**Description:** Assesses data variation in the form of tongue contour categories, which reflect assignment to a specific category. **Advantages:** • All changes in shape, rotation, or translation of the tongue are considered in the statistical analysis **Weaknesses:** • Requires head and jaw movement to be constrained or compensated for

Method	Notes
Procrustes analysis (Dawson *et al.*, 2016)	**Description:** Summarizes the squared difference between two tongue shapes after a sequence of translations and alignments **Weaknesses:** • Relatively insensitive to the rotation of the probe

1.8 Conclusion

This section represents a conclusion comparing different tongue quantification methods and sheds light on our proposed shape model, described in the next chapter. The shape measures discussed in this chapter allow us to quantify one tongue shape in an ultrasound frame or quantify the differences between two or more contours. There are also other experiments that compare shape models. For example Dawson *et al.* (2016) showed that the first Fourier coefficient has a higher classification rate in comparison with the MCI and Procrustes analysis. However, they also concluded that some tongue shapes might be less reliable than the others because the tongue parts that contain more complexities of the tongue contour (e.g., tongue tip) are not completely imaged in the ultrasound recording. One improvement to the experiment of Dawson is to include the palatal distance information, which is less sensitive to the missing tongue components in the ultrasound image. This thesis presents a novel shape model that addresses the limitations of shape models discussed in this chapter. Those limitations are mostly related to the sensitivity against noise sources and probe rotation. The main topics in this thesis related to the novel shape measure are summarized below:

- Using the data related to the hard palate in the shape measure
- Quantifying the most relevant characteristics of tongue contour
- Comparing the sensitivity of our novel shape measure to the noise sources and comparing it with other shape measures

These topics address the questions about the usage of the hard palate in determining speech sounds and the sensitivity of the novel shape measure to the noise sources and probe misalignment. We were able to fill these gaps by proposing an approach that combines the DFT shape measure and palatal information that can be used to quantify the tongue contours in US video recordings.

CHAPTER 2

NEW SHAPE MODEL WITH PALATE INFORMATION

This chapter introduces a new approach to quantifying tongue shape during articulation. Like many of the shape measures reviewed in chapter 1, this model can be used to cluster different tongue shapes within and between speakers. The difference between the new shape model and other models is that the new model utilizes tongue to palate distance information. Palatal information is vital because it allows us to measure the shape of the oral cavity rather than only the shape of the tongue and ultimately, the shape of the oral cavity determines what speech sound is produced. In such a case, the relevance of a tongue shape measure can be increased by adding the palatal information. Lu, Hsu, Goldstein & Toutios (2021) have also presented a novel shape model which combines the palatal information with the tongue shape metrics in MRI images to construct a robust predictor of inter-speaker variation related to American English /ɹ/ sound, which is similar to the idea of this chapter. However, in this thesis, we focus on ultrasound images. We also analyze the effect of simulated noise on the presented shape model, which is missing from the experiment of Lu *et al.* (2021). This chapter is organized as follows. Section 2.1 presents an abstract overview of the steps required to calculate the new shape measure. In sections 2.2 through 2.5, we propose the detail steps of formulating the new shape measure.

2.1 Tongue shape quantification using Fourier transformation

Figure 2.1 illustrates the proposed method for quantifying vocal tract shape. First, we extract the tongue contours from short ultrasound recordings. In section 2.2 this step is described in detail. The tongue contour is then converted into the Fourier coefficients based on Liljencrants (1971) method. The transform produces three Fourier coefficients. For this experiment, we only use the first coefficients.

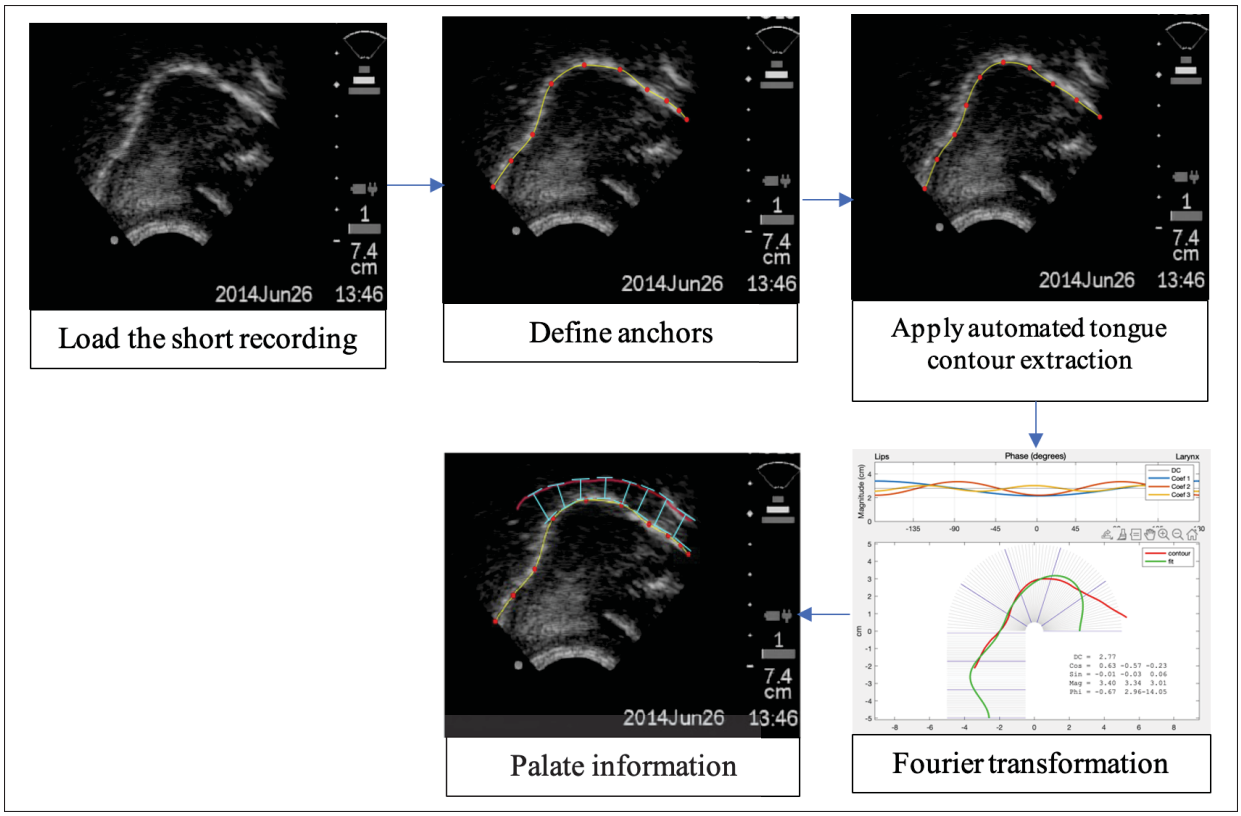

Figure 2.1 The diagram demonstrates the steps for the presented tongue shape index

Some parts of the hard palate are visible as bright regions during the swallowing sequence, recorded at the beginning of each session for each experiment subject. We detect the palate position and draw orthogonal lines from the tongue contour to the hard palate. Next, we measure the length of each line and calculate the variance of the length array. We use the variance value as an index for quantifying the tongue to palate position in the US image; details are described in sections 2.4 and 2.5. We use the combination of Fourier coefficients and tongue-palate distance to quantify the tongue shape in a given frame of a US video recording.

2.2 Tongue contour extraction

The first step in quantifying tongue shape is to extract the tongue contours from a particular frame of the US recording. To extract the contours, we manually define ten anchor points on

the bright curve showing the tongue surface in the first frame of the US image. Next, we use the automated tongue tracker to detect and track the tongue curve through all frames of the US recording. We use the GetContours and SLURP software (Tiede, 2020) for that purpose. We use the default tracking algorithm, which utilizes a combination of snakes and particle filtering to track the tongue contour through successive frames; for details of implementation, refer to Laporte & Ménard (2018). Figure 2.2 illustrates the result of applying snake fitting to the annotated contours:

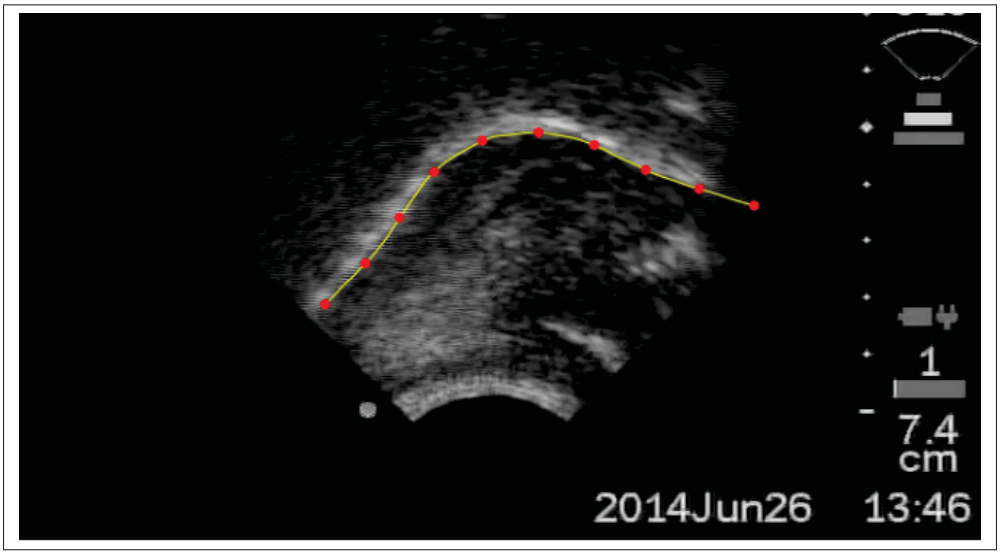

Figure 2.2 The result of applying snake fitting to the annotated contours

For snake fitting, SLURP uses the approach of Li, Kambhamettu & Stone (2005). In this approach, two types of energy are defined. The external energy describes how accurately the snake matches the tongue shape in the US image. The internal energy measures the degree of complexity and rigidity of the snake. The goal of this approach is to minimize total energy, which is defined as the sum of internal energy terms (E_i) and external energy using the following equation:

$$E = \sum_{i=1}^{n} \alpha E_i(v_i) + \beta E_g(v_i)(E_b)(v_i) \tag{2.1}$$

In the equation above, V is the set of 39 vertices that are interpolated based on the manually defined set of anchor points. E_i denotes the internal energy of the snake. E_g represents the gradient energy of the snake, and the goal is to reposition the snake to a region with the highest intensity variation. E_b denotes the contrast level between the areas above and below the contour.

2.3 Fourier transformation calculation in GetContours

The Fourier transform analysis available in GetContours (Tiede, 2020) is based on the method described by Liljencrants (1971). In this implementation, the contour is mapped to a semi-polar grid which approximates the vocal tract shape (See figure 2.3 for an illustration of the semi-polar grid). Firstly, the millimeter to pixel ratio is determined by manually drawing a rectangle that bounds at least five calibration markers in the US image. Then static semi-polar gridlines are plotted with a fixed resolution separation value. The formulation of gridlines does not require palate tracing and is fixed for different tongue shapes and different subjects. The segmented tongue contour will be superimposed on the gridlines. This semi-polar grid is implemented in GetContours software to approximate the vocal tract shape. In the next step, the software formulates a cross-sectional distance function from this semi-polar grid. This distance function will then be approximated by the coefficients calculated by the GetContours Fourier transformation. See figure 2.3 for an illustration of cross-section distance:

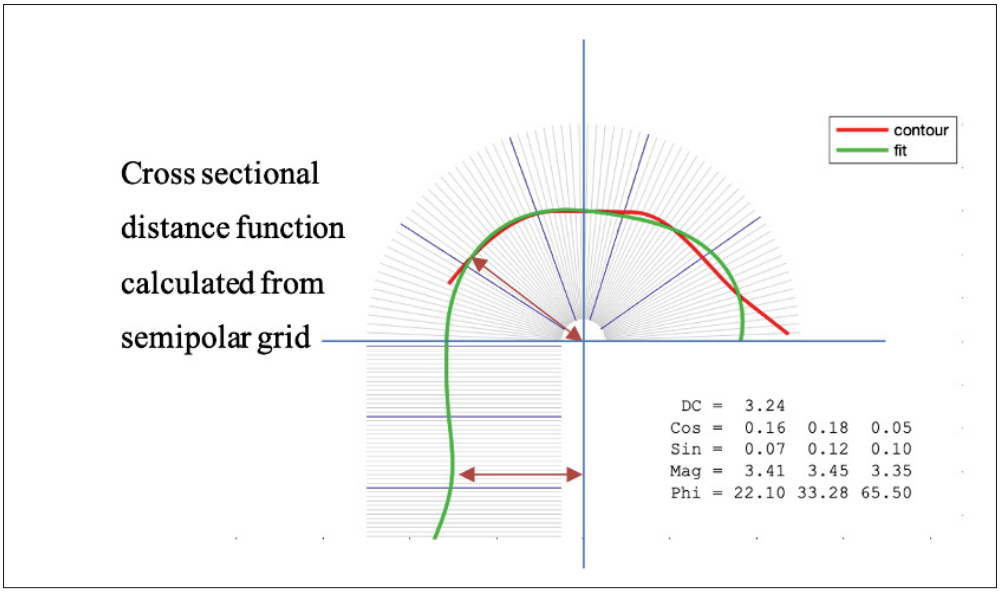

Figure 2.3 Example of Fourier coefficient approximation calculated by GetContours, blue lines are semi-polar gridlines that approximate the vocal tract shape

To calculate the cross-section distance in GetContours, the input contour points are first interpolated into 100 uniformly sampled points. The intersection between the semi-polar grid lines and the contour points from the previous step is then examined. The distance function consists of a polar component that converts the oral cavity and a Cartesian element that transforms the pharyngeal region. The distance is the radius in the polar system of the mouth and the distance from the vertical axis in the pharyngeal system (Liljencrants, 1971). In this model, the Fourier coefficients can alternatively be denoted using the following equations:

$$Mag_r = \sqrt{CC_r^2 + CS_r^2} \qquad \varphi_r = atan(\frac{CS_r}{CC_r}) \tag{2.2}$$

In equation 2.2, the CS_r and CC_r symbols denote sine and cosine Fourier coefficients.

2.4 Palatal information

Palate position measurement can be helpful in US-based studies of articulatory tongue movements since it provides additional information about the constrictions of the oral cavity (Faucher, Karimi, Ménard & Laporte, 2019). The palate is usually not completely visible during the ultrasound recording because of the air between the tongue and palate. Most parts of the palate are visible during swallowing. But even during the swallow, the palate details cannot consistently be recognized in a single frame. In this experiment, we used the swallowing recording that comes with the speech recording for each participant. This is to make sure that the palate and tongue (as seen during speech) belong to the same speaker and recording session and are probably in the same coordinate system so that it makes sense to combine both types of information. The steps we use to detect and track the palate are based on the work presented by Faucher *et al.* (2019), which is a method for extracting the hard palate contour from short swallowing recordings. The procedure involves computing the cumulative echo skeleton image of the recording. The cumulative skeleton image is the brightest section in regions that consistently contain bright structures throughout the recording. The cumulative skeleton is constructed based on the processing of individual frames related to swallowing recording and leads to the extraction of line drawings called Skeletons, which characterize the image's brightest ridge-like structure. In the next step, a threshold is applied to the skeleton images to construct a concise shape of the palate. The widest section of the bright pixels is then extracted from the US image. Next, a cubic spline is fitted into the points from the previous step. And finally, a snake is applied to the fitted spline to form the palate structure.

2.5 New shape model

In the previous section, we discussed how to detect the position of the hard palate in ultrasound recording. In this section, we discuss how we use the position of the hard palate to construct our new shape model. First, we sample 10 uniformly distanced points on the extracted tongue contour. Next, we draw orthogonal lines from each point on the curve to the hard palate. Then we measure the length of each line that has an intersection with the extracted hard palate. Figure

2.4 illustrates the calculation of distance values on an actual tongue shape. The new shape measure quantifies the following frame with the following indices:

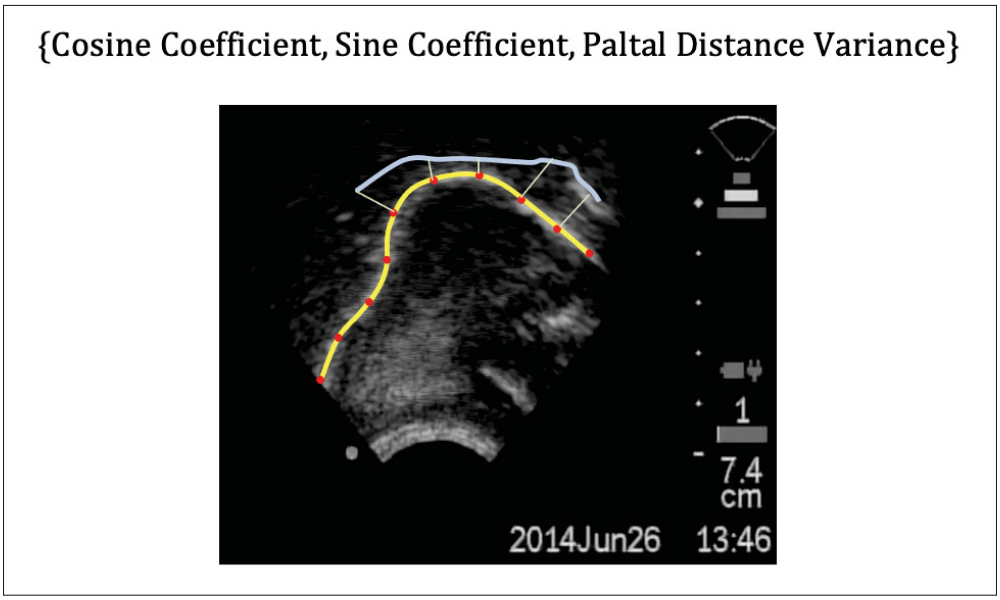

Figure 2.4 Illustration of the calculation results for palatal information for the middle frame of a sample sound [nu], collected from a speaker diagnosed with Steinert's disorder

We use a (MathWorks, Natick, Massachusetts, United States) script to obtain the distance values automatically. Next, we calculate the variance of those measurements as a new dimension with the Fourier coefficients to describe the tongue shape in a target US frame. We calculate the distances in pixel units.

2.6 Conclusion

This chapter explained how we use the palatal distance information as a feature alongside the calculated Fourier transform of the tongue shape to obtain a new vocal tract shape model. We use the Faucher *et al.* (2019) method to automatically detect and accumulate the components of the hard palate during ultrasound recording. Then we use the tongue to palate distance information

as a new dataset feature combined with the Fourier coefficients. This new feature acts as a new dimension for the dataset. We rely on this assumption that the position of the hard palate is fixed during the US recording session and is not affected by probe or head movement. Adding this new dimension as a descriptor to the tongue shape model reduces the model's sensitivity to the noise sources and yields a robust vocal tract shape measure.

CHAPTER 3

EXPERIMENTS

This experiment investigates the robustness and discriminative power of various shape measures discussed in Chapter 1 and the proposed shape model described in Chapter 2, which combines palatal information with the Fourier coefficients. In this experiment, we simulate perturbation and analyze the effect of such a noise on candidate shape models. The diagram below demonstrates an abstract flow of the experiment:

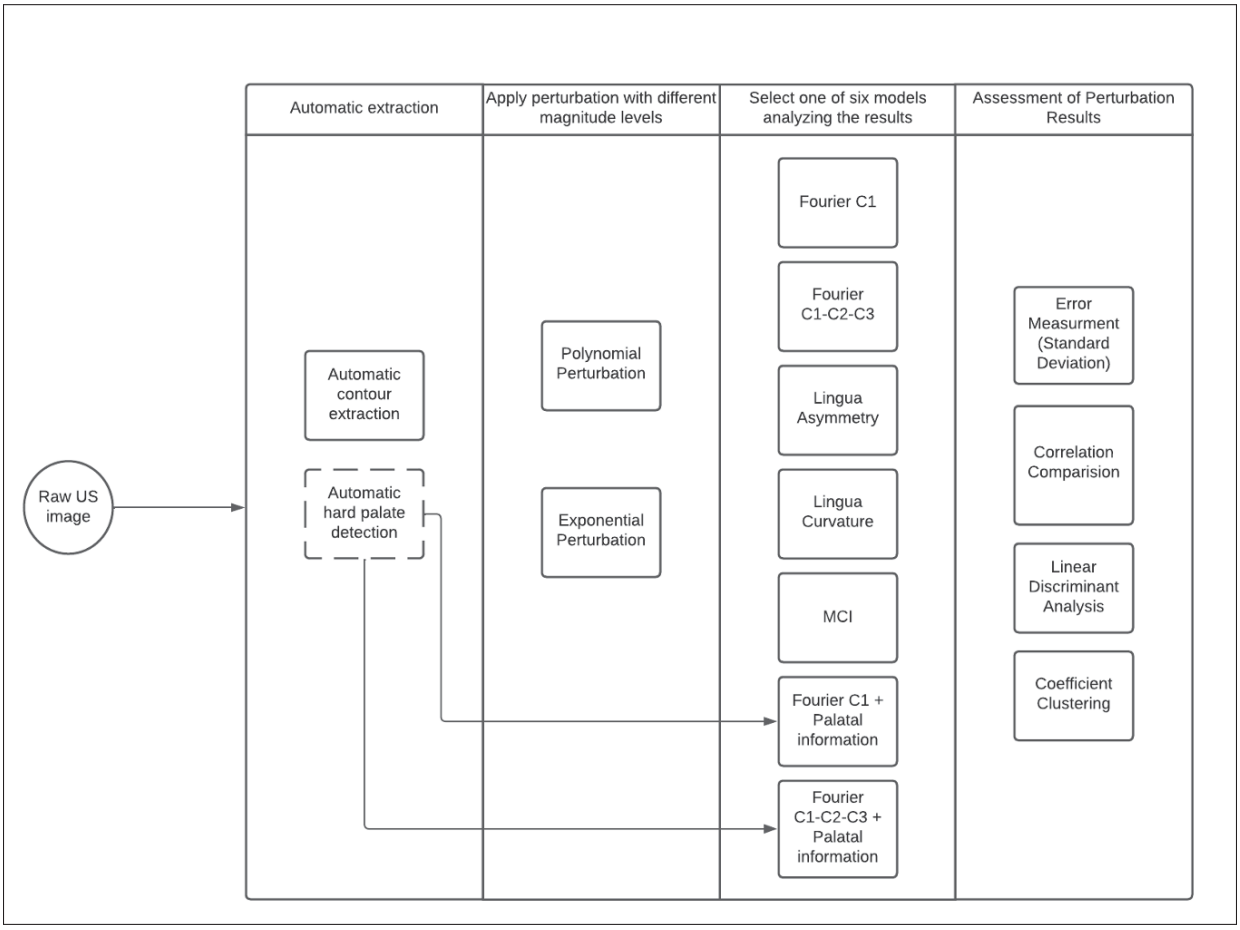

Figure 3.1 The process of the experiment. The dashed line and related entities indicate the process that yields the palatal information, which is a complementary data feature for our new shape measure

Section 3.1 describes the data acquisition step in detail. In section 3.2, we review the perturbation methods used to measure the robustness of the shape measures to the sources of error. Sections 3.6 and 3.7 present some metrics for assessing the outcomes and evaluating how each metric reacts to the noise sources. Using those metrics, we will then assess the robustness of several of the shape measures reviewed in Chapter 1 and the new shape measure introduced in Chapter 2. Section 3.7 analyzes how adding palatal information increases clustering accuracy for healthy and clinical subjects.

3.1 Data acquisition

For the experiments of this thesis, we extracted short segments from recordings acquired at the UQAM phonetics laboratory. The participants in this study were 10 French Canadian speakers aged from 10 to 14 years old. Six speakers were healthy with no speech impairment, and four subjects were diagnosed with Steinert's disease. Steinert is a disease that engages muscles, and it is mainly recognized with myotonia and multiorgan impairment that combines numerous muscle weakness and cardiac conduction disorders (Bouhour, Bost & Vial, 2007). Steinert's disease leads to slow speech, false impression of consonants and vowels, and low speech intelligibility (Laporte & Ménard, 2018). In the remainder of this thesis, healthy subjects are labeled with an alias starting with the letter H (Healthy), and participants with Steinert disease are marked with an alias beginning with the letter C (Clinical). This experiment is focused on two sounds [nu] and [bu] recorded through short ultrasound recordings. We analyzed 18 repetitions of each sound for healthy subjects and 18 repetitions of each sound for subjects with Steinert disease. Figure 3.2 demonstrates the tongue shapes related to sound [nu] and [bu]:

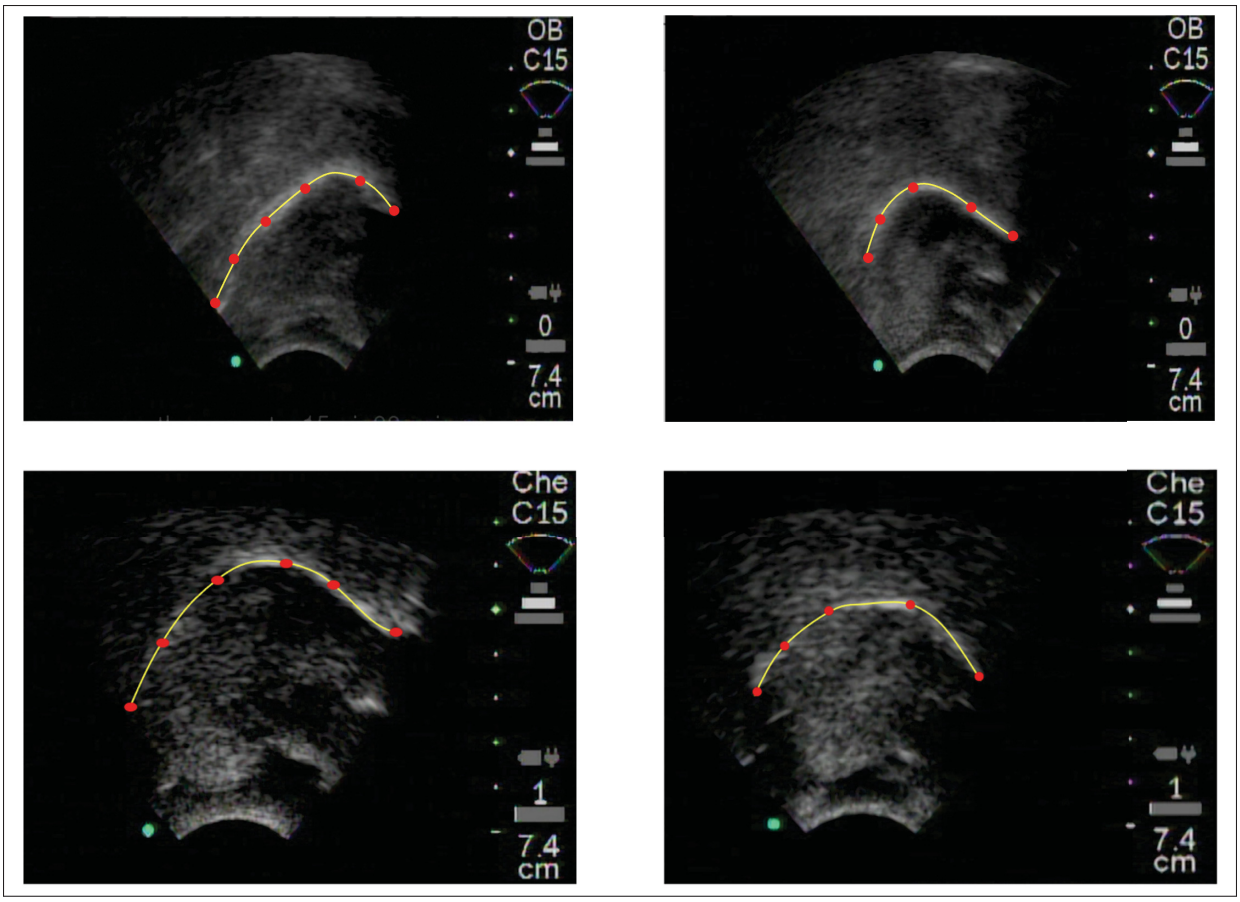

Figure 3.2 Comparison of tongue shapes for two experiment subjects. The left side demonstrates the tongue shape related to sound [nu], and the right side reflects the tongue shape related to sound [bu]. The top row represents a healthy subject, and the bottom row represents a clinical subject with Steinert's disease. The middle frame of the recording was selected for both sounds

We choose those sounds because they contain both consonant and vowel contexts. The selected sounds also demonstrate the coarticulation effect, which is the acoustic or visual modification in one speech segment because of the neighboring sounds. In this case, the sounds [n] and [b] are pronounced differently because of the adjacent vowel. Also, the contrast between the complete palatal contact observed when producing the sound [n] and the bunched tongue shape when producing the sound [b] is an excellent example of the effectiveness of palatal information in classifying the candidate CV utterances. As one of the results of this experiment, we analyzed how the shape measure (mean value) shifts as the perturbation varies from zero to

more significant levels. We investigated the perturbation effect on the Fourier coefficient, MCI shape measure, and Lingua model.

3.2 Perturbation functions

The trapped air around the tongue might cause missing tongue portions in the resulting US image. Misalignment of the ultrasound probe might also result in poor image quality or missing tongue components. Partial mislabeling of the tongue contour may lead to an error in contour extraction. Also, speckle noise can be mistaken with the actual tongue surface. Figure 3.3 illustrates one example of such a loss in tongue contour data.

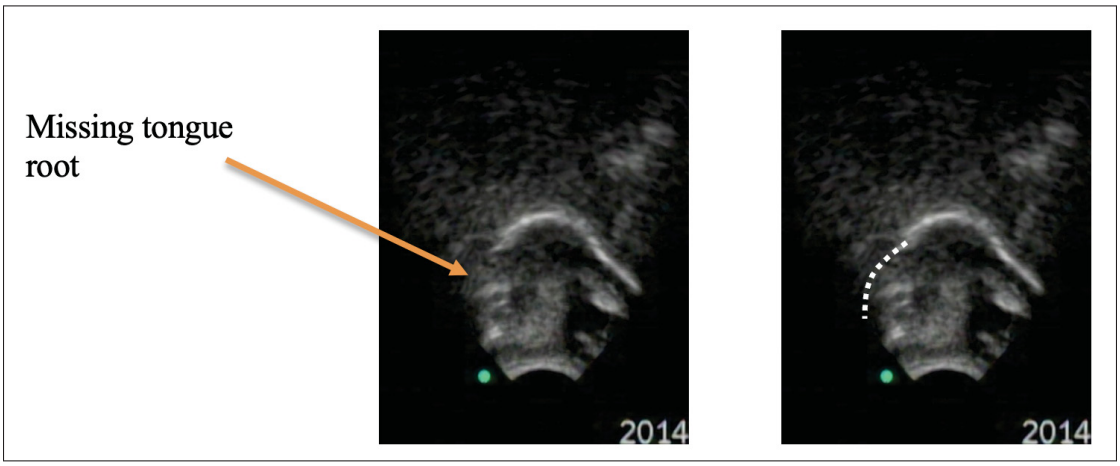

Figure 3.3 Unclear tongue root

We applied two different functions to simulate the effect of poor image quality on the extracted tongue contours. These perturbation functions modify the root and tip sections of the tongue and are used to assess the robustness of the candidate shape measures discussed in this thesis. The missing tongue sections usually affect the shape measure value, and we try to simulate the same impact with the perturbation functions proposed in next section.

3.3 Exponential perturbation function

We used an exponential root/tip displacement function to simulate the effect of missing tongue components on tongue contour extraction and to investigate how the displacement of the tongue root or tip can affect the different shape models discussed in this thesis. The exponential function models tracking errors where the tracked tongue contour departs from the actual contour as one gets closer to the tip or root. Figure 3.4 illustrates the exponential function for tongue displacement on an extracted contour:

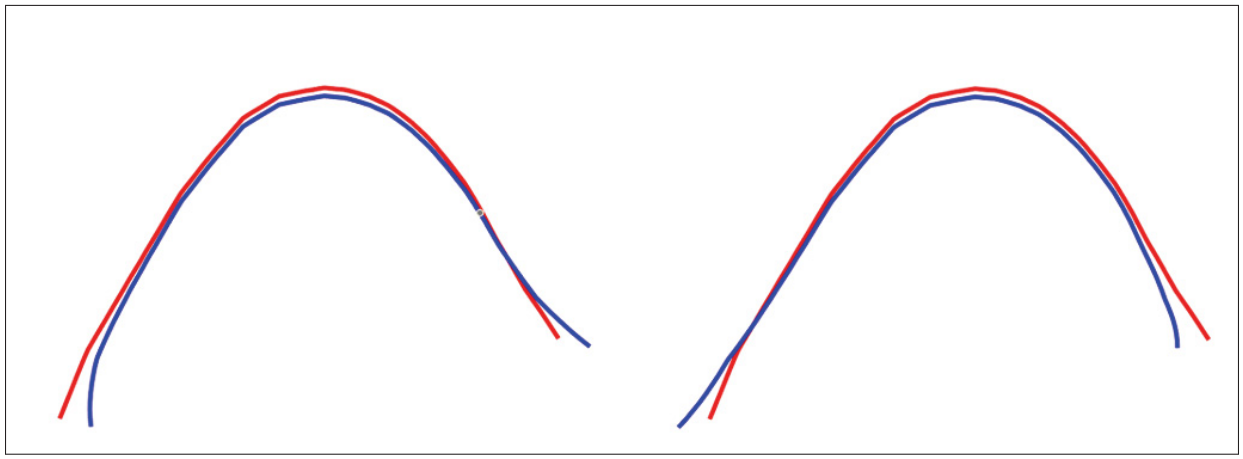

Figure 3.4 Exponential root/tip displacement. Left: original and perturbed contour using the exponential function with the parameter -3 for both α and β. Right: shows the actual and perturbed contour using the exponential function with the parameter +3 applied to both α and β. The red curve is the original contour; the blue curve is the perturbed contour

This type of error frequently occurs with automated contour tracking at the tip or root sections of the tongue if they are simply not as bright as the middle body of the tongue. The displacement is applied along the original tongue curve. Equation 3.1 shows how the displacement is calculated:

$$x_{perturbed} = x + (\alpha \times e^{\theta \times l} + \beta \times e^{\theta \times (1-l)}) \tag{3.1}$$

Consider having contour points $P = (x, y)$ based on equation 3.1, $x_{perturbed}$ represents the updated values for x coordinates of the contour points. l is the original contour length and θ is a constant value to control the amount of perturbation on each side of the tongue. α represents the coefficient for displacing the points placed close to the root of the tongue, and β represents the coefficient for a displacement of the points placed close to the tip of the tongue. The sign of the coefficients identifies the bending direction associated with the perturbation. The instances of such a bending is demonstrated in figure 3.4 of this section.

3.4 Polynomial extension or shortening

In this type of simulation, we extend or shorten each side of the tongue curve to model the changes in extracted tongue contour due to missing extremities of the tongue on either side of the US image. Figure 3.5 demonstrates the polynomial perturbation applied on an extracted contour:

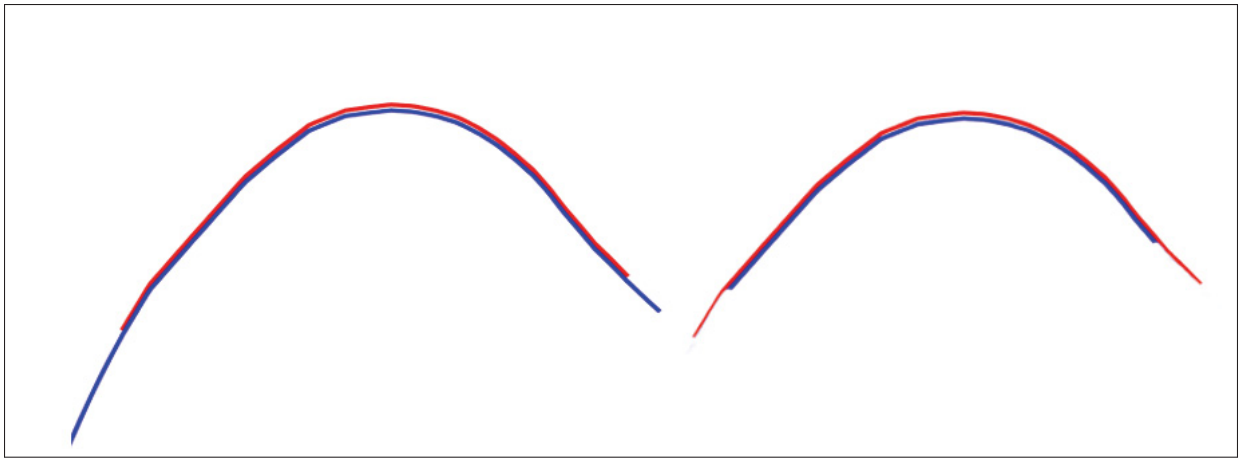

Figure 3.5 Polynomial perturbation illustration, left: (10% extension) and right: (10% shortening). The red curve is the original contour; the blue curve is the perturbed contour

We apply a 5th order polynomial fit to the tongue contour and then extend/shorten the contours based on the n percentile of the beginning/end of the curve.

3.5 Perturbation magnitudes

For our experiments, we applied perturbations with magnitudes that resulted in a realistic tongue contour. When talking about perturbation magnitudes, we mean replacing the values α and β in equation 3.1 related to the exponential perturbation method and also setting the percentage of data points considered for shortening/extending in polynomial perturbation. Out of all possible perturbation magnitude values and their effect on tongue contours, we only selected magnitude levels resulting a realistic tongue shapes. Excessively high or low magnitude levels produce an unrealistic tongue contour, which is out of the scope of this experiment. We also classified the perturbation magnitudes into two categories of Extreme and Moderate levels. Table 3.1 summarizes the perturbation magnitudes used in this experiment:

Table 3.1 List of perturbation methods and related magnitudes used in this experiment

Method	**Category Name**	**Magnitude**
Exponential	Extreme	$\alpha = -4 \quad \beta = -4$
Exponential	Moderate	$\alpha = -2 \quad \beta = -2$
Exponential	Moderate	$\alpha = +2 \quad \beta = +2$
Exponential	Extreme	$\alpha = +4 \quad \beta = +4$
Polynomial	Extreme	$Root = 0.25 \quad Tip = 0.25$
Polynomial	Moderate	$Root = 0.1 \quad Tip = 0.1$
Polynomial	Moderate	$Root = -0.1 \quad Tip = -0.1$
Polynomial	Extreme	$Root = -0.25 \quad Tip = -0.25$

Regarding the perturbation magnitudes from table 3.1, figures 3.6 and 3.7 present detailed results for a video sequence that posed for sound [nu] and [bu]:

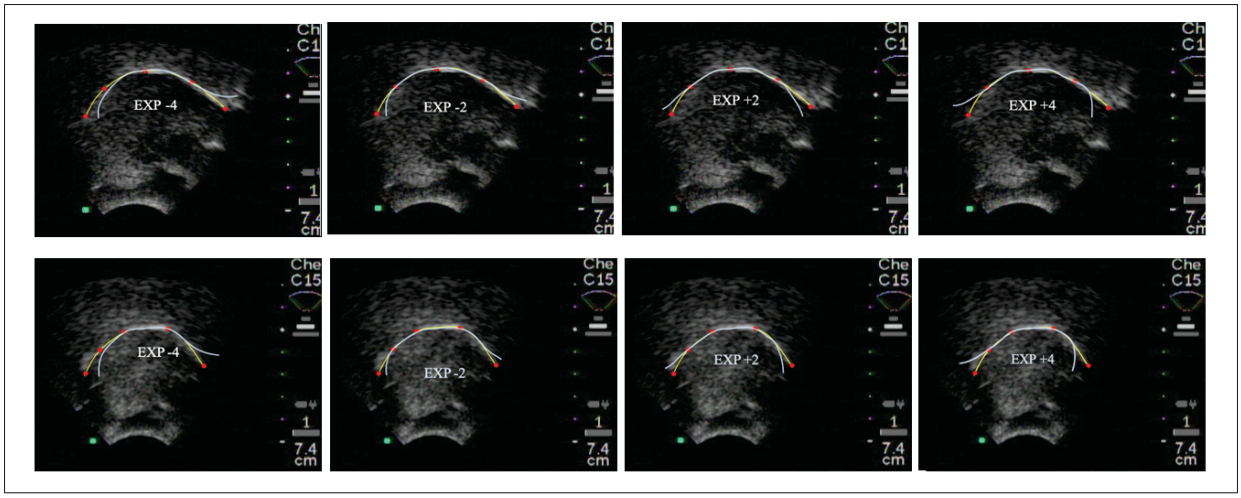

Figure 3.6 Illustration of the original contour (yellow curve) vs. the perturbed contour using exponential perturbation method (light blue curve) in a sample US image from Steinert subject. Top row images are related to sound [nu], and bottom images are related to sound [bu]

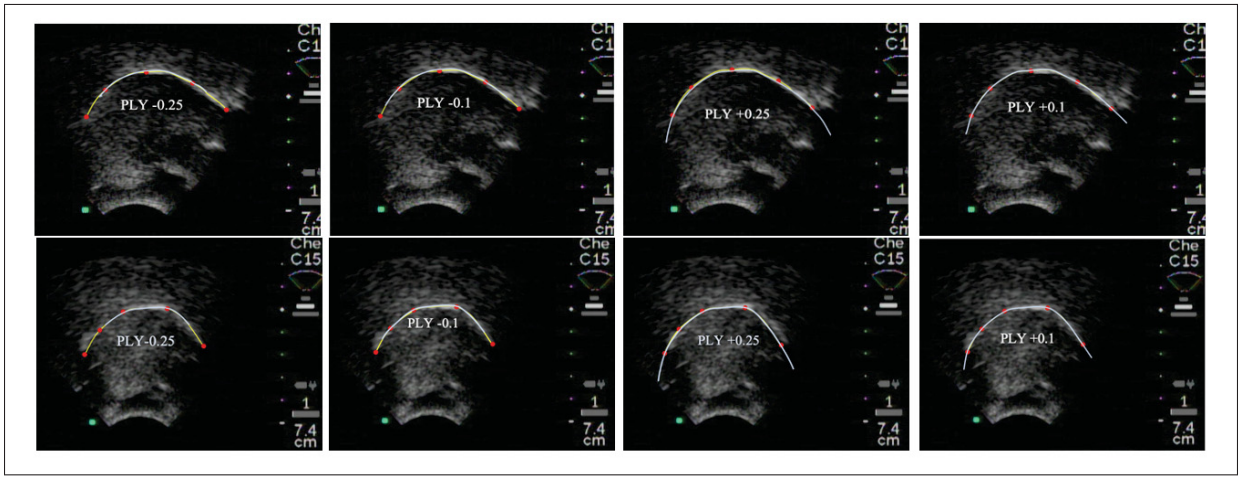

Figure 3.7 Illustration of the original contour (yellow curve) vs. the perturbed contour (light blue curve) in a sample US image from Steinert subject. Top row images are related to sound [nu], and bottom images are related to sound [bu]

These figures demonstrate how perturbation is applied to the root and tip of the original contour. As seen in the figures below, we only used perturbation magnitudes (light blue curves) which resulted in realistic tongue shapes.

3.6 Comparison with the not perturbed contour

We used an error measure to compare the perturbed tongue contour for healthy and clinical subjects. One error measure we used in our experiment is comparison of the average shape measure coefficients with the non-perturbed values. This comparison tells us about the amount of dispersion for each perturbation method from the non-perturbed values. We can analyze the behavior of shape measure when perturbation magnitude varies for both ends of the tongue. We then plotted the comparison chart for the two experiment groups. Figures 3.8 through 3.11 demonstrate the perturbation results on different models for each group of subjects:

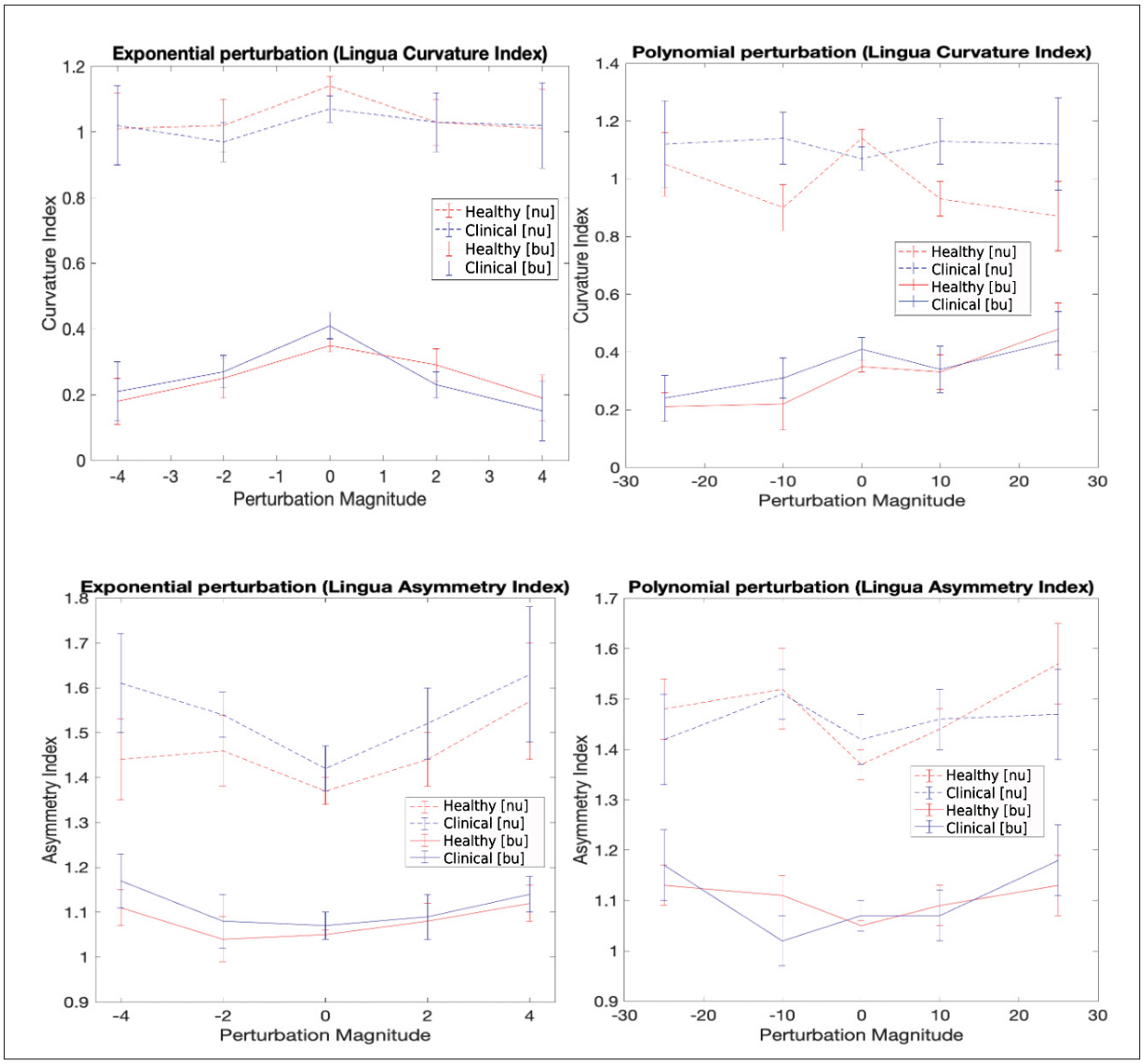

Figure 3.8 Comparison of Lingua shape model for healthy (red) and clinical (blue) subjects for sound [nu] (dashed) and [bu] (solid) lines. Vertical bars in the plot demonstrate the standard deviation for a specific perturbation magnitude level

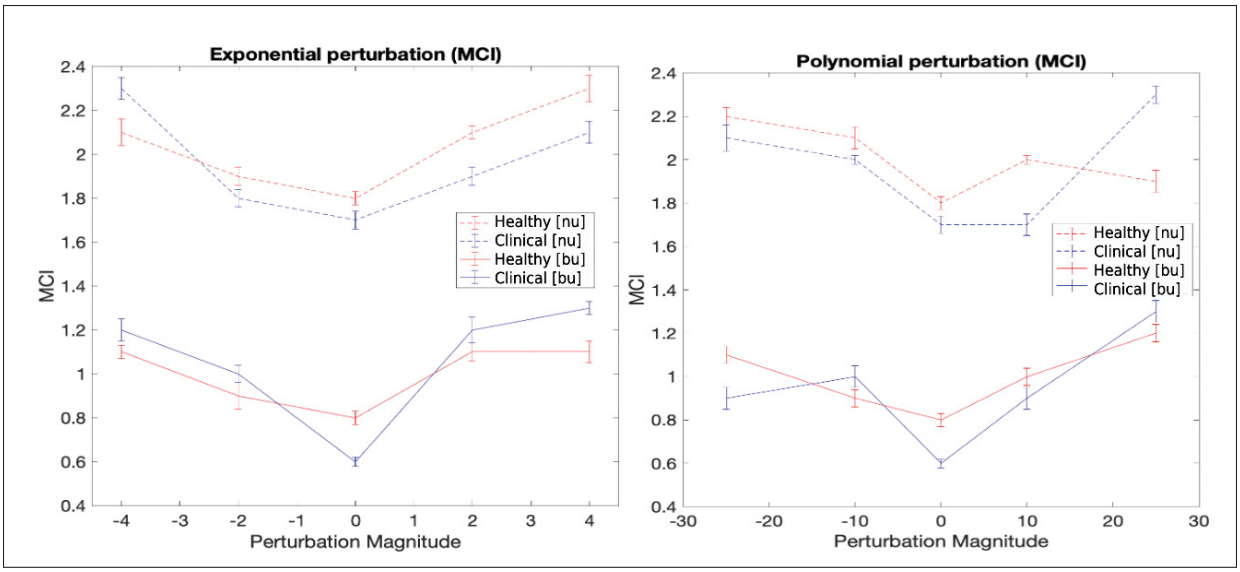

Figure 3.9 Comparison of MCI for healthy (red) and clinical (blue) subjects for sound [nu] (dashed) and [bu] (solid) lines. Vertical bars in the plot demonstrate the standard deviation for a specific perturbation magnitude level

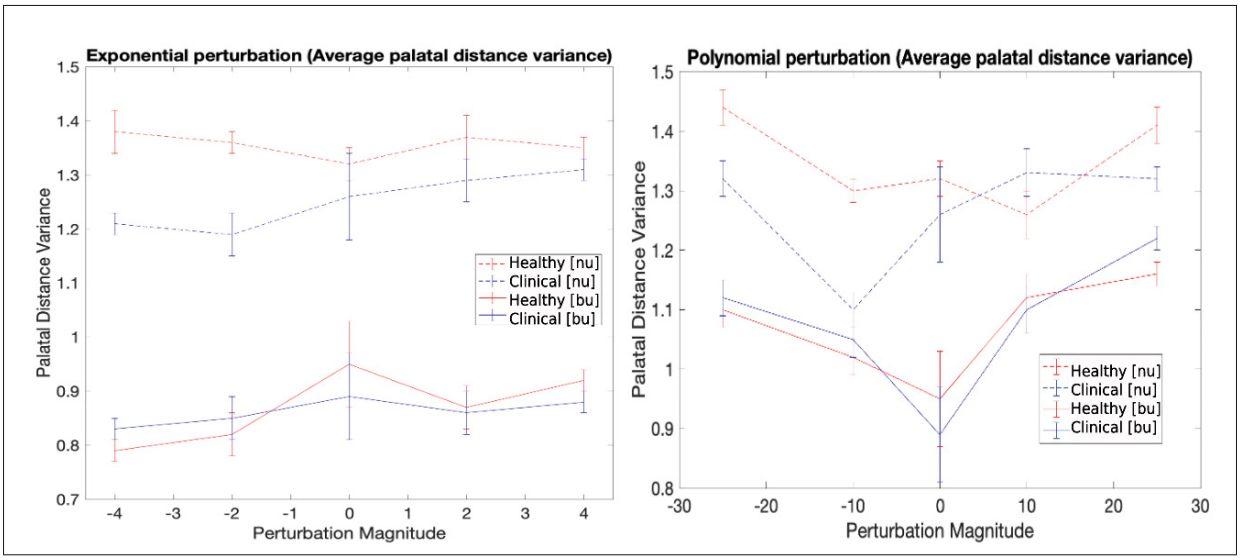

Figure 3.10 Comparison of Palatal variance index for healthy (red) and clinical (blue) subjects for sound [nu] (dashed) and [bu] (solid) lines. Vertical bars in the plot demonstrate the standard deviation for a specific perturbation magnitude level

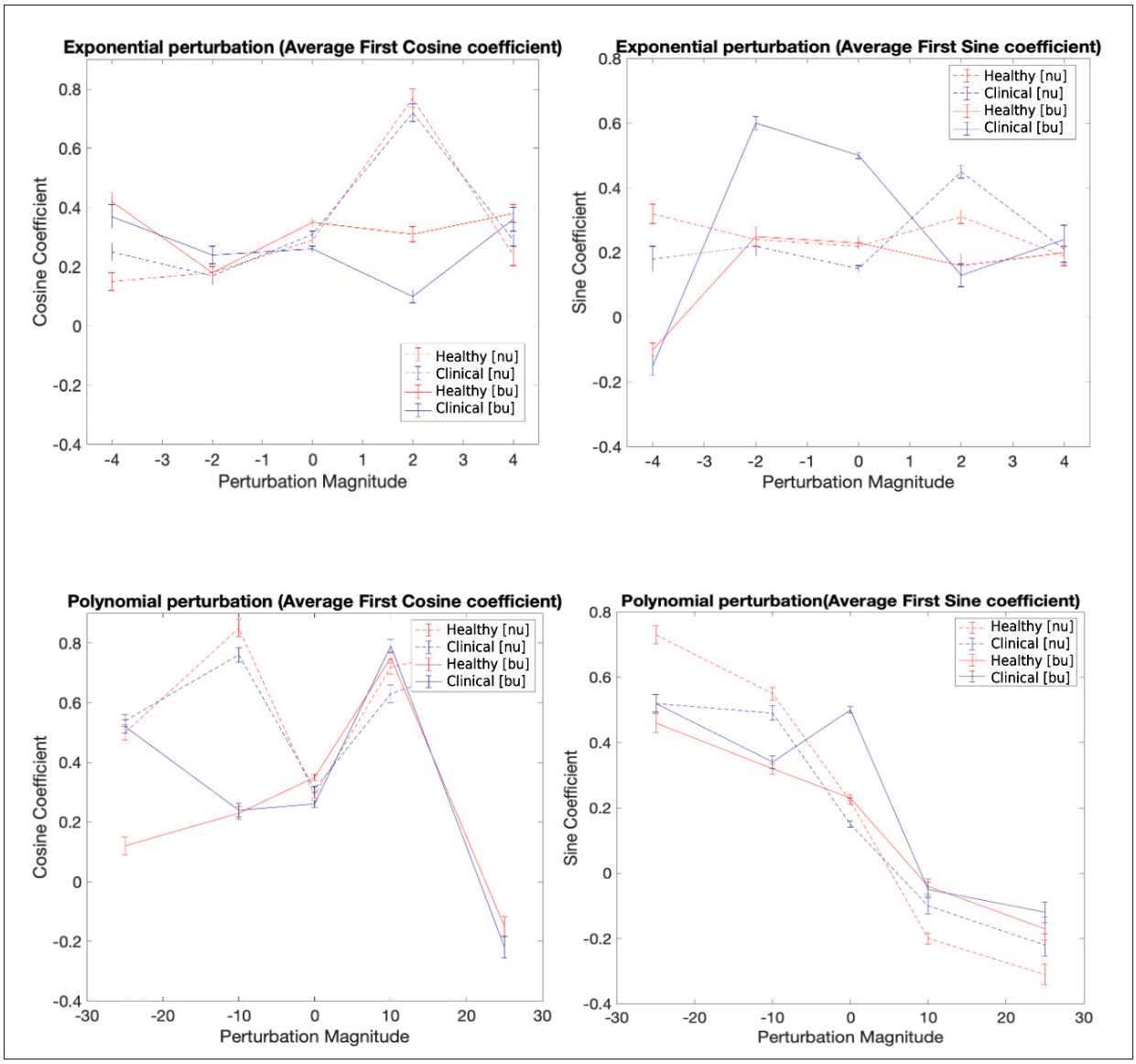

Figure 3.11 Comparison of resulting Fourier coefficients for healthy (red) and clinical (blue) subjects for sound [nu] (dashed) and [bu] (solid) lines. Vertical bars in the plot demonstrate the standard deviation for a specific perturbation magnitude level

For MCI (figure 3.9) and Lingua asymmetry index (figure 3.8), as perturbation magnitude increases, the value of the shape measure also increases compared with the zero-perturbation case in the center of the plot. In contrast, for the Lingua curvature index, as we increase the perturbation magnitude, the shape measure values decrease compared to the zero-perturbation point. This could happen because perturbation might increase the length of the triangle's base,

fitted to the tongue shape, which yields a smaller shape index. For the Fourier coefficient plot, no specific pattern was detected. That could happen for various reasons (e.g., the nature of sine and cosine coefficients and how they react to the perturbation). We observe that the palatal distance variance relatively increases as perturbation magnitude grows. One reason which can be considered is that we are calculating the variance of distances for the points that are uniformly distributed along the tongue curve, so perturbation on both ends of the tongue will affect the palatal distance variance too. Based on figures 3.8 to 3.11, the highest values for standard deviation error are close to the plot's ends, representing the extreme perturbation magnitude levels. Lower errors are observed on points close to the center of the plot (lower magnitude of the perturbation).

3.7 Correlation comparison

One limitation of direct comparisons with the zero-perturbed shape measure value, especially for shape measures like the Fourier coefficient, is that analysis does not gain information about the robustness of the shape measure. To solve this issue and determine the strength of the relationship between original and perturbed tongue shapes, we calculated the correlation based on a sequence of perturbed vs. original tongue contours. Analysis of the short articulation clips resulted in a sequence of shape measure values. The correlation coefficient between original and perturbed tongue contour shape sequences is then calculated using equation 3.2:

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{3.2}$$

where x_i is the i^{th} value in the sequence of shape measure values before perturbation, and y_i is the i^{th} value in the sequence of shape measure values after perturbation. High correlation values show that the shape descriptor is more robust to perturbation in a relative sense. Figure 3.12 summarizes the results for correlation formulation of two sounds and two experiment categories:

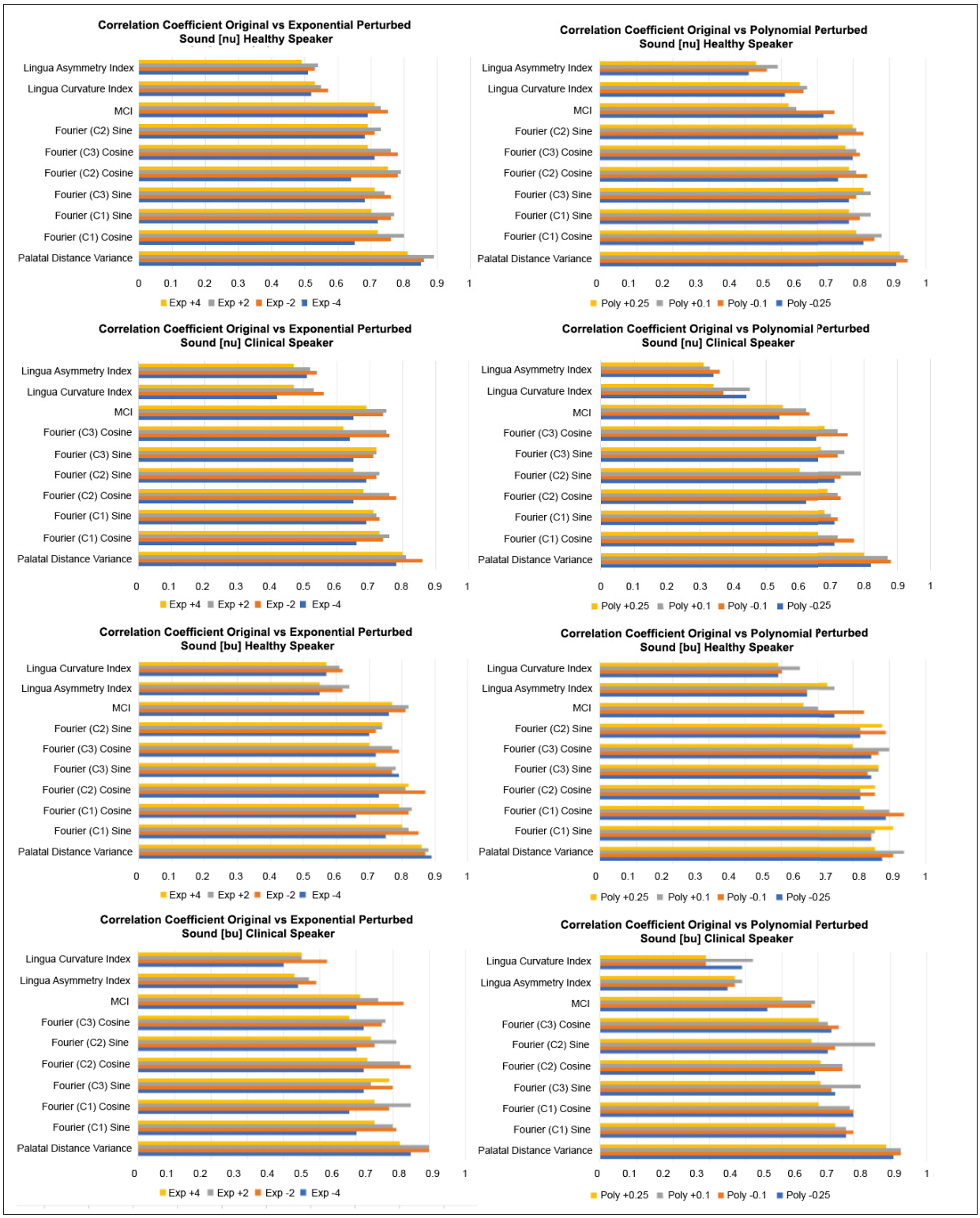

Figure 3.12 (Best viewed by zooming in on the PDF.) Comparison of correlation related to different shape measures and perturbation methods. Sorted from top to bottom based on average correlation value for different perturbation magnitude (Shape model on the bottom has the lowest average correlation, shape model on top has the highest average correlation)

Each record in the graph contains a correlation value for original vs. perturbed tongue contour in a sequence of tongue shapes extracted from US recordings. The correlation values are used to survey the robustness of each candidate shape measure. The correlation values related to the palatal distance variance have an average of 0.82 for both experiment sounds and subject categories, indicating a strong relationship between original and perturbed contour shape sequences for this measure. It suggests that the palatal distance variance can safely be used with other shape measures to select a candidate set that is robust to the noise. One reason which can be formulated for this strong correlation is that the distance of tongue to the palate is mostly affected by data points in the middle of the tongue contour, not two sides of the tongue where the perturbations have the most effect. We also observed that the Lingua curvature index and MCI are more sensitive to noise sources than other shape measures. The correlation coefficient analysis also suggests a strong positive relation between Fourier coefficients of original vs. perturbed tongue shapes, which indicates that the changes in the value of the coefficients for perturbed and original tongue contour have similar behaviour.

3.8 Linear Discriminant Analysis

In this experiment, we used the linear discriminant analysis (LDA) method to analyze the effectiveness of each shape model in classifying different tongue shapes for the two sounds [nu] and [bu]. LDA is a classification method that can be used for multivariate datasets. For this experiment, we used the default MATLAB implementation of LDA classification. This classifier searches for a group of transformed features constructed from a linear combination of initial dataset features. The resulting classification maximizes the between-class variance while minimizing the within-class variance. The optimal linear combination of features is named the discriminant function. The classifier is then constructed based on the determined discriminant function from the previous step (Dawson *et al.*, 2016). We calculated the error rate to demonstrate the capability of each shape model in classifying tongue shapes according to their corresponding sound. We used the leave-one-out cross-validation method, which is also used by Dawson *et al.* (2016) to separate the training and test dataset. In this method, the

classifier is trained with one record omitted from the original dataset. Then the same record is used to test the accuracy of the classifier. We repeated this until all samples in the dataset were once used as the test record. For healthy and clinical subjects, we calculated the model coefficient for tongue shapes perturbed with different magnitude values listed in table 3.1. We calculated the average classification rate for each subject and different perturbation magnitudes. In this experiment, we used six shape models to construct the dataset for the LDA classifiers. Table 3.2 summarizes the classification rate of sounds [nu] and [bu] for each perturbed tongue contour at a specific magnitude level related to each experiment category:

Table 3.2 Average classification rate of different shape models for healthy (H_*) and clinical (C_*) for sounds [nu] and [bu] after perturbation with two different methods and four magnitude levels (categorized as extreme and moderate levels) listed in table 3.1

Model	**Magnitude**	**Healthy**	**Clinical**
Fourier C1 (no perturbation)		0.78	0.72
Exponential perturbation	-4	0.67	0.58
	-2	0.72	0.61
	+2	0.75	0.64
	+4	0.72	0.56
Polynomial perturbation	-0.25	0.67	0.64
	-0.1	0.72	0.69
	0.1	0.75	0.67
	0.25	0.64	0.64
Fourier C1-C2-C3 (no perturbation)		0.81	0.78
Exponential perturbation	-4	0.69	0.64
	-2	0.72	0.72
	+2	0.75	0.72
	+4	0.67	0.69
Polynomial perturbation	-0.25	0.67	0.64
	-0.1	0.72	0.72

Model	Magnitude	Healthy	Clinical
	0.1	0.74	0.72
	0.25	0.67	0.75
Lingua 2D space curvature and asymmetry (no perturbation)		0.64	0.61
Exponential	-4	0.64	0.56
	-2	0.56	0.58
	+2	0.61	0.56
	+4	0.64	0.61
Polynomial	-0.25	0.61	0.56
	-0.1	0.64	0.61
	0.1	0.56	0.58
	0.25	0.58	0.56
MCI (no perturbation)		0.75	0.75
Exponential	-4	0.64	0.61
	-2	0.72	0.72
	+2	0.75	0.69
	+4	0.72	0.61
Polynomial	-0.25	0.69	0.64
	-0.1	0.72	0.72
	0.1	0.75	0.75
	0.25	0.64	0.61
Fourier_C1+Pal (no perturbation)		0.78	0.81
Exponential	-4	0.72	0.75
	-2	0.81	0.78
	+2	0.78	0.78
	+4	0.72	0.81
Polynomial	-0.25	0.64	0.72

Model	Magnitude	Healthy	Clinical
	-0.1	0.78	0.81
	0.1	0.81	0.78
	0.25	0.72	0.69
Fourier_C1_C2_C3+Pal (no perturbation)		0.75	0.78
Exponential	-4	0.75	0.72
	-2	0.72	0.75
	+2	0.78	0.75
	+4	0.72	0.72
Polynomial	-0.25	0.61	0.64
	-0.1	0.72	0.78
	0.1	0.78	0.75
	0.25	0.72	0.72

The results show that adding the palatal information to the first Fourier coefficient increases the average classification accuracy for perturbed tongue shapes. A 12% increase observed for the average accuracy score of extreme level perturbation magnitudes and a 14% growth observed for the average accuracy score of moderate level perturbation magnitudes. The first Fourier coefficient plus palatal information has the highest average classification accuracy after applying perturbation (75%), which is close to the average classification accuracy performed by the first three Fourier coefficients (70%). The Lingua asymmetry and curvature index have the lowest average accuracy scores, close to 59% after applying perturbation. The MCI model gained an average accuracy of 68% for perturbed tongue contours, which performed close to the first Fourier coefficient descriptor with an average accuracy score of 66% after applying perturbation.

3.9 Coefficient clustering

We also performed coefficient clustering to compare the effectiveness of the DFT shape measure for clustering tongue shapes according to the uttered speech sound. For this purpose, we

calculated the Fourier coefficient value related to 18 repetitions of the two sounds ([nu], [bu]) produced by healthy and clinical experiment subjects. We used the average ± one standard deviation range to create clusters for sine and cosine coefficients. Figure 3.13 demonstrates the first cosine and sine coefficients and resulting clusters:

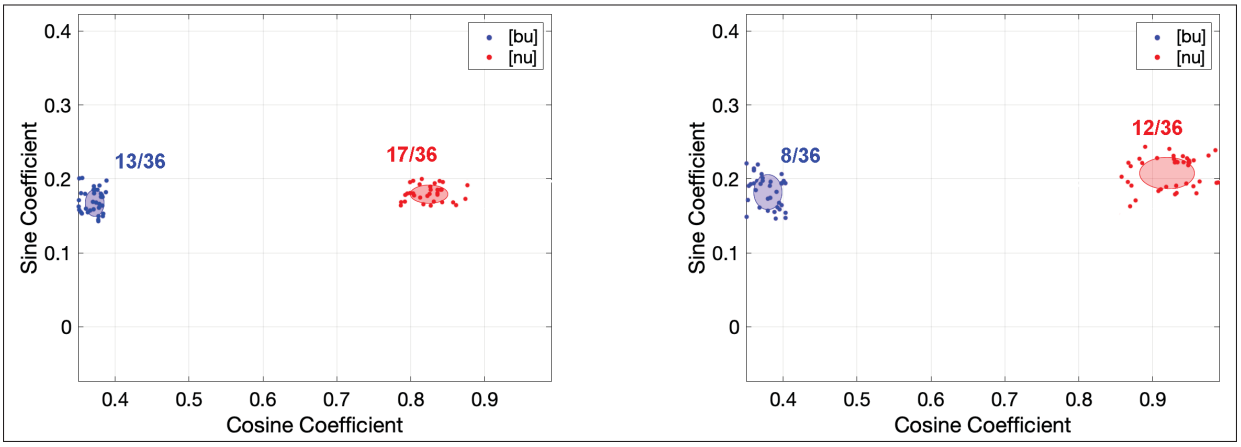

Figure 3.13 (Best viewed by zooming in on PDF.) This figure shows the Cosine and Sine coefficients calculated for ten subjects and 36 iterations for each sound using the GetContour software. Datapoints are clustered using average ± one standard deviation. Left: data points are collected from six healthy subjects. Right: data points are collected from four clinical subjects

The figure shows that cosine and sine coefficients have successfully separated tongue shapes into two clusters for both healthy and clinical subjects. The results also show that clinical subjects produce larger clusters compared to healthy subjects. This can be seen primarily in sound [nu], which created a greater spread over the x-axis compared to the healthy subjects producing the same sound.

We also repeated the same experiment with adding the palatal distance variance. We added a fourth dimension (palatal information) to the clustering plot and redid the one standard deviation clustering based on three features. Figure 3.14 shows the new clusters after considering the palatal information:

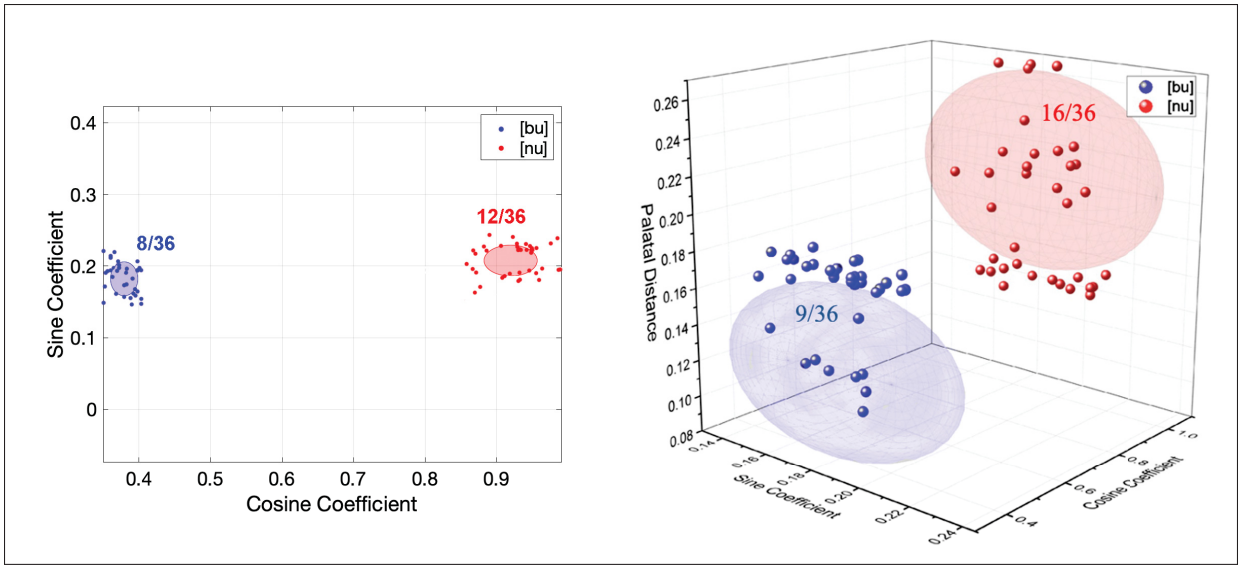

Figure 3.14 Left side: before adding palatal information for clinical subjects, Right side: after adding palatal information, four samples were included within one standard deviation for [nu] and one sample for [bu]. The number of points in each ellipsis is written in the plot

After adding the palatal information, data points inside the one standard deviation cluster became more uniform in clinical subjects, and the data spread decreased. The improvement can mainly be observed for the sound [nu]. For healthy subjects, there were no improvements between the two clustering plots.

3.10 Conclusion

Because of the jaw's shadow, lateral margins of the tongue might be missing in the US image (Preston *et al.*, 2017), or the tongue tip is usually missing in US images because of the air beneath it. In some cases, because of the US probe rotation or head movements, the image of the tongue might be rotated, or the anterior and posterior parts of the tongue might be missing in the US image. Also, the speckle noise, an inherent property of ultrasound images, might result in a noisy tongue curve (Michailovich & Tannenbaum, 2006) which can be the source of error in quantifying the tongue shape. In this chapter, candidate shape measures were experimentally assessed, and the effect of raw tongue contour perturbations on each one was evaluated. The

goal of applying a perturbation function to the tongue contours is to simulate the impact of missing parts of the tongue on the output of each model. The result of this experiment shows the existence of a strong positive correlation between original and perturbed Fourier shape descriptor and Fourier model combined with the palatal information. We also observed that using the LDA classifier, the Fourier coefficients combined with the palatal information (defined as a new feature of the dataset) could successfully improve the classification accuracy for the two sounds [nu] and [bu].

This work searched for a robust tongue shape model to quantify the tongue shape in ultrasound recordings of healthy and clinical subjects. This involves presenting a novel shape model to quantify the shape of the tongue in ultrasound recordings. The combination of the discrete Fourier transforms and tongue to hard palate distance variance, presented in this thesis, can successfully identify two types of sounds surveyed in this experiment.

CONCLUSION AND RECOMMENDATIONS

4.1 Contribution

The contribution of this thesis can be described as follows:

1. The key contribution of this thesis is an enhanced tongue shape quantification method, which is a combination of the DFT method and tongue to palate distance variance. The contour quantification method can summarize the vocal tract shape in ultrasound recordings. This shape characterization method is compared with the original Fourier transform method for healthy and clinical subjects. The results of the experiment showed that by using the DFT method in combination with the palatal distance information, the samples within one standard deviation interval increased by 33% for [nu] and 13% for [bu] related to the clinical subjects.
2. Another significant contribution of this thesis is the analysis of the robustness of various shape characterization methods to contour perturbation (polynomial and exponential) functions. Since the tongue root/tip might be missing because of the probe's misalignment or the air beneath those parts, it is necessary to consider analyzing the accuracy of the underlying quantification method to such a source of the noise. The analysis in section 3.8 demonstrates that the combination of Fourier coefficient and palatal information obtains higher classification accuracy in comparison with MCI and Lingua shape models. This analysis also showed that the average classification accuracy of the first Fourier coefficient model grows when combined with the palatal information.

4.2 Future work

Many areas can be investigated to increase the proposed quantification method's robustness and classification accuracy related to different sounds. This experiment only added the palatal distance information to the original DFT method, but other parameters can also be considered for

tongue shape or vocal tract shape analysis. For example Whalen, Kang, Magen, Fulbright & Gore (1999) studied the relation between pharynx shape and tongue positions. Moreover, we can also analyze the relationship between shape parameter variations and acoustic parameters such as formants (Brown & Anderson, 2006). With the advancements described in this section, we can improve the accuracy of shape descriptors to precisely quantify more complex tongue shapes. In this experiment we had potential limitations to access large sample data. We only surveyed two sounds ([nu] and [bu]) for the classification section of this experiment. However, classification with a larger dataset can better demonstrate the effectiveness of the novel shape model, for instance, analyzing the contrasting phonemes (e.g. [u] vs. [a] vs. [i]) might result in different classification accuracy scores. As a part of future developments, we can use more sample sounds and related tongue contours in this experiment.

BIBLIOGRAPHY

Andrade-Miranda, G. (2017). *Analyzing of the vocal fold dynamics using laryngeal videos*. (Ph.D. thesis).

Barbier, G., Perrier, P., Payan, Y., Tiede, M. K., Gerber, S., Perkell, J. S. & Ménard, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. *PLOS ONE*, 15(4), e0231484. doi: 10.1371/journal.pone.0231484. Publisher: Public Library of Science.

Bouhour, F., Bost, M. & Vial, C. (2007). [Steinert disease]. *Presse Medicale (Paris, France: 1983)*, 36(6 Pt 2), 965–971. doi: 10.1016/j.lpm.2007.01.002.

Brown, E. K. & Anderson, A. (2006). *The encyclopedia of language & linguistics*. Amsterdam; Boston: Elsevier. Retrieved from: https://www.sciencedirect.com/science/referenceworks/9780080448541.

Chen, W., Lee, N. G., Byrd, D., Narayanan, S. & Nayak, K. S. (2020). Improved real-time tagged MRI using REALTAG. *Magnetic Resonance in Medicine*, 84(2), 838–846. doi: 10.1002/mrm.28144.

Cleland, J. & Scobbie, J. M. (2020). The dorsal differentiation of velar from alveolar stops in typically developing children and children with persistent velar fronting. *Journal of Speech, Language and Hearing Research*. Retrieved from: https://strathprints.strath.ac.uk/74400/. Num Pages: 37.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, 120(1), 407–15. doi: 10.1121/1.2205133. Place: USA Publisher: Acoust. Soc. America.

Dawson, K. M., Tiede, M. K. & Whalen, D. H. (2016). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics & Phonetics*, 30(3-5), 328–344. doi: 10.3109/02699206.2015.1099164. Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/02699206.2015.1099164.

Faucher, G., Karimi, E., Ménard, L. & Laporte, C. (2019). Automatic palate delineation in ultrasound videos. pp. 422–426. Retrieved from: http://intro2psycholing.net/ICPhS/papers/ICPhS_471.pdf.

Gosztolya, G., Pintér, , Tóth, L., Grósz, T., Markó, A. & Csapó, T. (2019). *Autoencoder-Based Articulatory-to-Acoustic Mapping for Ultrasound Silent Speech Interfaces*.

Guillermic, M., Lanoy, M., Strybulevych, A. & Page, J. H. (2019). A PDMS-based broadband acoustic impedance matched material for underwater applications. Retrieved on 2020-12-08 from: /paper/A-PDMS%E2%80%90based-broadband-acoustic-impedance-matched-Guillermic-Lanoy/77074cf34a2580347eeaceb0b677032e886e799f/figure/0.

Havenhill, J. E. (2019). *Articulatory Strategies for Back Vowel Fronting in American English*. Australasian Speech Science and Technology Association Inc. Retrieved from: http://hub.hku.hk/handle/10722/269631.

Hixon, T. J. (2014). Pharyngeal-oral function and speech production. In *Pharyngeal-oral function and speech production*. Plural Publishing.

Karimi, E., Ménard, L. & Laporte, C. (2019). Fully-automated tongue detection in ultrasound images. *Computers in Biology and Medicine*, 111, 103335. doi: 10.1016/j.compbiomed.2019.103335.

Kier, W. M. & Smith, K. K. (1985). Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society*, 83(4), 307–324. doi: https://doi.org/10.1111/j.1096-3642.1985.tb01178.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1096-3642.1985.tb01178.x.

Lancia, L. & Tiede, M. (2012). A survey of methods for the analysis of the temporal evolution of speech articulator trajectories. (pp. 233–271).

Laporte, C. & Ménard, L. (2018). Multi-hypothesis tracking of the tongue surface in ultrasound video recordings of normal and impaired speech. *Medical Image Analysis*, 44, 98–114. doi: 10.1016/j.media.2017.12.003.

Lee, s. a., Wrench, A. & Sancibrian, S. (2015). How To Get Started With Ultrasound Technology for Treatment of Speech Sound Disorders. *Perspectives on Speech Science and Orofacial Disorders*, 25, 66. doi: 10.1044/ssod25.2.66.

Li, M., Kambhamettu, C. & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, 19(6-7), 545–554. doi: 10.1080/02699200500113616.

Liljencrants, J. (1971). Fourier series description of the tongue profile. *Speech Transmission Laboratory-Quarterly Progress Status Reports*, 12(4), 9–18. Publisher: Citeseer.

Lu, Y., Hsu, H., Goldstein, L. & Toutios, A. (2021). Effect of vocal tract morphology on tongue shaping for American English //. *The Journal of the Acoustical Society of America*, 150(4), A188–A188. doi: 10.1121/10.0008076. Publisher: Acoustical Society of America.

Michailovich, O. V. & Tannenbaum, A. (2006). Despeckling of Medical Ultrasound Images. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 53(1), 64–78. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3639001/.

Ménard, L., Aubin, J., Thibeault, M. & Richard, G. (2012). Measuring tongue shapes and positions with ultrasound imaging: a validation experiment using an articulatory model. *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP)*, 64(2), 64–72. doi: 10.1159/000331997.

Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A. & Whalen, D. H. (2017). Ultrasound Images of the Tongue: A Tutorial for Assessment and Remediation of Speech Sound Errors. *Journal of Visualized Experiments : JoVE*, (119). doi: 10.3791/55123.

Preston, J. L., McCabe, P., Tiede, M. & Whalen, D. H. (2019). Tongue shapes for rhotics in school-age children with and without residual speech errors. *Clinical Linguistics & Phonetics*, 33(4), 334–348. doi: 10.1080/02699206.2018.1517190. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02699206.2018.1517190.

Stolar, S. & Gick, B. (2013). An index for quantifying tongue curvature. *Canadian Acoustics - Acoustique Canadienne*, 41, 11–15.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19(6-7), 455–501. doi: 10.1080/02699200500113558. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02699200500113558.

Stone, M., Liu, X., Chen, H. & Prince, J. L. (2010). A preliminary application of principal components and cluster analysis to internal tongue deformation patterns. *Computer methods in biomechanics and biomedical engineering*, 13(4), 493–503. doi: 10.1080/10255842.2010.484809.

Tiede, M. (2020). How to "GetContours" from ultrasound imaging. Retrieved from: https://ultrafest2020.indiana.edu/abstracts/UltraFest_IX__Tiede_MasterClass.pdf.

Tuta, L., Nicolaescu, I., Mariescu-Istodor, R. & Digulescu, A. (2019). A Principal Component Analysis (PCA) based Method for Shape Extraction. *Journal of Military Technology*, 2, 17–28. doi: 10.32754/JMT.2019.2.03.

Whalen, D. H., Kang, A. M., Magen, H. S., Fulbright, R. K. & Gore, J. C. (1999). Predicting mid-sagittal pharynx shape from tongue position during vowel production. *Journal of speech, language, and hearing research: JSLHR*, 42(3), 592–603. doi: 10.1044/jslhr.4203.592.

Zharkova, N. (2013). Using Ultrasound to Quantify Tongue Shape and Movement Characteristics. *The Cleft Palate-Craniofacial Journal*, 50(1), 76–81. doi: 10.1597/11-196. Publisher: SAGE Publications.

Zharkova, N. (2018). An Ultrasound Study of the Development of Lingual Coarticulation during Childhood. *Phonetica*, 75(3), 245–271. doi: 10.1159/000485802. Publisher: Karger Publishers.

Zharkova, N., Gibbon, F. & Hardcastle, W. (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical linguistics & phonetics*, 29, 1–17. doi: 10.3109/02699206.2015.1007528.