University of South Florida
Scholar Commons

Graduate Theses and Dissertations

Graduate School

2005

Reliability of Hand Measures of Ultrasound Analysis

Sarah A. Hardin
University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the American Studies Commons

Scholar Commons Citation

Hardin, Sarah A., "Reliability of Hand Measures of Ultrasound Analysis" (2005). *Graduate Theses and Dissertations.*
https://scholarcommons.usf.edu/etd/2915

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Reliability of Hand Measures of Ultrasound Analysis

by

Sarah A. Hardin

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Communication Sciences and Disorders
College of Arts and Sciences
University of South Florida

Major Professor: Stefan A. Frisch, Ph.D.
Jean C. Krause, Ph.D.
Patricia Blake-Rahter, Ph.D., CCC-A.

Date of Approval:
July 25, 2005

Keywords: speech production, inter-rater reliability, alveolar, velar, reproducibility measures

© Copyright 2005, Sarah A. Hardin

Dedication

This thesis project is dedicated to my mentor and thesis chair, Dr. Frisch, for his contributions to psycholinguistic and linguistic research. Thank you for the opportunity to work as a research assistant in the speech science laboratory. Thank you also for the opportunity to use the work with hand measures of ultrasound analysis to create a thesis project of which I am very proud. During the course of this project you have been dependable, flexible, and at the same time, attentive to every detail. I know many future students will benefit tremendously, as I have, from having you as a mentor.

Acknowledgments

I would like to acknowledge the friends and family who supported me throughout the completion of this project. I'd like to start by thanking my parents, Denise and Gaspar Monte, and Al and Pat Hardin, for their love, support, and encouragement. I would like to thank Merete Moeller Glasbrenner and Erline Nakano for their friendship, encouragement and moral support over the past two years. I am grateful to have found friendships to last a lifetime with you.

I would like to acknowledge the motivating words of loved ones during the writing process. Thanks to my dad who said, "Just remember…if this is published, your name will be on it forever…so make it good!" I would like to thank George, my love, for your ability to turn every mountain into a molehill. Thank you for being a constant source of inspiration and support.

I would like to thank my sister Sarah Monte and aunt Gracie Monte for checking up on me regularly. Thank you for always being there when I needed you most. I would like to acknowledge three lifelong friends who have provided unconditional love and support and have kept me laughing for years, Kathryn Sheridan, Lucy Moon, and Jessica Costantino. You are the friends I have grown up with and look forward to growing old with. Whenever I take things too seriously, you are always there to laugh at me. Thanks for the reality checks and for all the good times. You mean the world to me.

I would like to acknowledge my committee member Dr. Jean C. Krause, Ph.D. for contributing her time and expertise to my project. I would like to acknowledge my committee member Dr. Patricia Blake-Rahter, Ph.D., CCC-A, for her commitment to excellence in clinical supervision and mentoring at the University of South Florida.

I would like to thank research assistants Adrienne Stearns and Sylvie Wodzinski for the use of their measurements in my thesis. Thank you Adrienne for the time you spent training and re-training me. I really enjoyed the time we spent together in the lab.

Table of Contents

List of Tables ..... **i**

List of Figures ..... **ii**

Abstract ..... **iii**

Chapter One: Introduction **1**

Imaging Techniques *2*
  Xeroradiography. 3
  Magnetic Resonance Imaging. 3
  Ultrasound Imaging. 4

Ultrasound use in Speech Research *7*

Clinical Research Using Ultrasound in Speech Pathology *9*

Ultrasound Imaging Reliability Studies *12*

Ultrasound Research Used in this Study *13*

Chapter Two: Methods **15**

Raters *16*

Stimuli *16*
  Experiment 1: Closure Angle Data. 16
  Experiment 2: Tongue Twisters # 1. 17
  Experiment 3: Tongue Twisters # 2. 18

Procedures *18*

Measurements *18*
  Experiment 1: Closure Angle. 19
  Experiment 2. 20
  Experiment 3. 21

Chapter Three: Results **23**

Experiment 1: Dorsum Angle Data *23*
  Frame Choice: Speaker 1. 23
  Frame Choice; Speaker 2. 24

Closure Points.. 26
Velar Angle.. 27

Experiment 2: Tongue Twister Data.. *33*
Frame Choice.. 33
Tongue Blade Points..... 35
Velar Closure Points... 36
Tongue Blade Angle Measures.. 36
Velar Angle Measures.... 37
Dorsum Distance.. 38
Global View. 39

Experiment 3: Tongue Twister # 2.. *41*
Frame Choice: R2i and R3.. 41
R2e and R2i. 41
R2e and R3..... 42
Tongue Blade Points..... 44

R2i and R3.. *44*

R2i and R2e.. *44*

R2e and R3.. *44*
Velar Closure Points..... 45
Tongue Blade Angle Measures. 47
Velar Angle Measures.... 48
Dorsum Distance.. 48
Global View.. 49

Chapter 4: Discussion.. **51**

Frame Choice.. 51
Closure Points.... 51
Velar Angle; Exp.1... 52
Tongue Blade Angle Measures. 52
Dorsum Distance.. 53
Summary.. 53

References... **54**

List of Tables

Table 1 Average distance between velar closure points marked by R1 and R2 . 26

Table 2 Velar angle differences for speakers 1 and 2 27

Table 3 Mean measurement difference and average 2 SD measurement variability for the dorsum angle data. 32

Table 4 Difference in location of tongue blade measurement points for tongue twister data for R2 and R3. 35

Table 5 Difference in location of dorsum measurement points for tongue twister data for R2 and R3 36

Table 6 The difference in tongue blade angle measures of R2 and R3 for alveolar and velar consonants 37

Table 7 Differences in velar angle measures derived for R2 and R3 38

Table 8 Distance between anterior and posterior tongue blade points selected by experienced and inexperienced measurers. 44

Table 9 Distances between anterior and posterior velar closure points selected by experienced and inexperienced measurer 46

Table 10 Tongue blade angle measures by experienced and inexperienced raters 47

Table 11 Velar angle measures obtained by experienced and inexperienced raters. 48

Table 12 Differences in dorsum distance measures by experienced and inexperienced raters 49

List of Figures

Figure 1 Sagital view of the tongue body with ultrasound; tongue tip to the right 6

Figure 2 Measurements applied to ultrasound image of the tongue 20

Figure 3 Difference in choice of frame to measure velar stop closure (R2 – R1) 25

Figure 4 Coarticulation patterns for word and non-word data for speaker 1, for measures by R1 and R2. 29

Figure 5 Coarticulation patterns for word and non-word data for speaker 2, for measures by R1 and R2. 30

Figure 6 Difference in choice of frame for alveolar and velar stop closure (R2 – R3) 34

Figure 7 Comparison of overall data pattern for tongue twister experiment for R2 and R3 40

Figure 8 Difference in choice of frame for measurement between R2i, R2e, and R3 43

Figure 9 Scatterplot representation of alveolar versus velar articulations for measures by R2i, R2e, and R3 50

Reliability of Hand Measures of Ultrasound Analysis

Sarah A. Hardin

ABSTRACT

As ultrasound imaging gains popularity in speech research, an important question to address is the reliability of the measures taken from these images. This study examines the reliability of hand measures of ultrasound data collected by graduate student researchers in the University of South Florida’s speech science lab. Speech production data from “Ultrasound analysis of velar fronting” (Wodzinski, 2004) and “Ultrasound study of errors in speech production” (Frisch, 2003) were used to obtain inter-rater reliability measures. This study compares the rater’s choice of video frame depicting alveolar or velar closure image, anterior and posterior points of closure, tongue blade and velar angle measurements, as well as a measurement of the tongue dorsum distance from the ultrasound probe. The measures obtained by one rater before and after experience in ultrasound analysis was gained were compared for additional information on the effect of experience on the reliability of measures. Overall, the measurements were found to be reliable between raters. Although some absolute differences in measures were found, the measures obtained from different raters led to the same quantitative description of speech articulation patterns. In addition, the measurements did become more reliable with increased rater experience.

Chapter One

Introduction

Speech articulation involves complex linguistic and cognitive processes. Frameworks describing the processing components involved in language production identify stages in the production process beginning with conceptual preparation, through grammatical, morpho-phonological, and phonetic encoding, and finally the articulation of speech.

Speech production and perception studies have used multiple imaging techniques to capture elements of overt speech for measurement. Imaging techniques that have been used to view and record speech articulation include xeroradiography, magnetic resonance imaging, and ultrasound imaging. The different imaging techniques and their application in speech perception and production research will be discussed further in this introduction.

This study examines the reliability of the hand measures of ultrasound analysis as used in speech research. As the use of hand measures of ultrasound analysis in speech research becomes more popular, knowing the reliability of these measurements becomes important. This study compares results of measurements of the video recordings of speech production derived by independent raters using the same ultrasound analysis measurement procedures and instruments. Inter-rater reliability is assessed for data from three different experiments. The raters' results will be compared for their choice of video frame to measure, their choice of point location for articulatory landmarks within the video frame, and the resulting quantitative description of articulation based on these landmarks. The results of this study will determine whether hand measurements of ultrasound analysis are reliable using the current measurement procedures in the speech production and perception laboratory at the University of South Florida.

Imaging Techniques

In recent years new insights have been gained into the processes involved in the production and perception of speech due to the use of imaging techniques. Many imaging techniques have been implemented in the measurement of speech production and physiology with varying levels of success. These techniques include xeroradiography (X-ray), magnetic resonance imaging (MRI), and ultrasound. These techniques have primarily been used to record speech production for research purposes. Imaging techniques provide the opportunity to gain a greater understanding of the complexities of the speech production process by creating an image of an entire articulatory structure rather than monitoring the position of individual points as with electromagnetic articulography (EMA) or X-ray microbeam. Although a variety of new imaging and neural recording tools are available, each has advantages and disadvantages, so the study of speech articulation and language production may be most effective using a combination of techniques rather than a single tool.

Xeroradiography. X-ray imaging works by radiating images from an x-ray source through a body part and onto a film cassette. A phosphor coating inside the cassette then glows to expose the film. This film is developed to display the x-rayed image. X-ray beams are weakened as they pass through tissues of different densities. Soft tissue absorbs less x-ray energy than bone because it is less dense. This contrasting of densities allows skeletal structures, muscular structures and organs to be identified. There are some disadvantages to using x-ray imaging to capture speech production. X-ray has not been widely used to study speech because of the danger of prolonged exposure to radiation. In addition, most of the important speech articulators are composed of soft tissue, and so do not image particularly well with x-ray.

A safer implementation of x-ray imaging has been used to track pellets attached to the articulators for speech production research. In 1994 the Speech Production Database, a collection of synchronous acoustic and flesh point kinematic data recorded with x-ray microbeam, was made publicly available. It has been used in speech production studies that examine the actions of the articulators (tongue blade, dorsum, lips, and mandible) while reading test words (Westbury, Severson, & Lindstrom, 2000). While x-ray microbeam is safer to use than x-ray, it can only track a few points on the articulators and so does not provide a complete image of the articulatory structure.

Magnetic Resonance Imaging. There are many levels of transformation that move language through the stages of conception to output. The complexity of language and language processes have drawn attention to magnetic resonance imaging (MRI) as a method of examining both the neurolinguistic and articulatory processes involved in speech production. In addition, MRI was one of the first techniques to provide 3D images of the vocal tract. However, MRI images can typically only be captured at a very low frame rate (less than one scan per second). Due to the low frame rate, the use of MRI high-resolution volumetric recordings to view speech articulation is limited to the study of speech postures and movements that are not occurring in real time (Munhall, 2000). In other studies, participants have been instructed to repeat a target syllable continuously while images are recorded from different phases of production. The images can then be pieced together to create a simulated real time video of the articulation similar to the process of stroboscopy. Technology is improving and MRI may eventually have the capability to depict speech articulation in three dimensions in real time.

Electromagnetic articulography is another imaging technique used in speech production research. EMA uses magnetic fields like MRI to track pellets attached to the articulators. EMA is like x-ray microbeam in that it has a high frame rate. However, it also only tracks specific points on the articulators so it does not provide a complete image of articulation.

Ultrasound Imaging. A technique that has been gaining increasing popularity in the field of speech research is ultrasound imaging. Ultrasound imaging has been used to study speech production since the middle of the twentieth century when its general use in the medical setting became popular (Gick, 2002). The scan rate of ultrasound imaging is significantly higher than MRI (40-60 scans per second). This increased scan rate allows tongue movement to be viewed in real time.

Ultrasound imaging works by using the reflective properties of sound waves to create an image. The ultrasound transducer creates a high frequency sound wave. As the sound wave travels through the soft tissue of the tongue it is partially reflected by changes in tissue density, and fully reflected by air. When the ultrasound transducer is placed under the chin at the base of the tongue, this phenomenon causes a white line to appear in the ultrasound image at the upper surface of the tongue. The line appears in the air space between the tongue and the palate and is used as a landmark in measurement of ultrasound images.

The tongue can be viewed with ultrasound using B-mode or M-mode images. The B-mode provides 2-D images of anatomical structures such as the hyoid bone, and movements of hyoid bone, the genioglossus, geniohyoid, and mylohyoid muscles, the mandibular symphysis, as well as the tongue surface and tip as shown in figure 1. These imaging options allow for the viewing of tongue motion during speech and swallowing and the viewing of associated anatomical landmarks. Some difficulties have been encountered using ultrasound imaging of the tongue tip and epiglottis. Shadows are created by both the sublingual and epiglottic cavities that make these structures difficult to capture, depending on probe placement (Peng, Jost-Brinkmann, Miethke, & Lin, 2000). M-mode ultrasound images show the reflections of a single scan line (a 1-D image) from the ultrasound over an interval of time, somewhat similar to the point tracking techniques of x-ray microbeam and EMA.

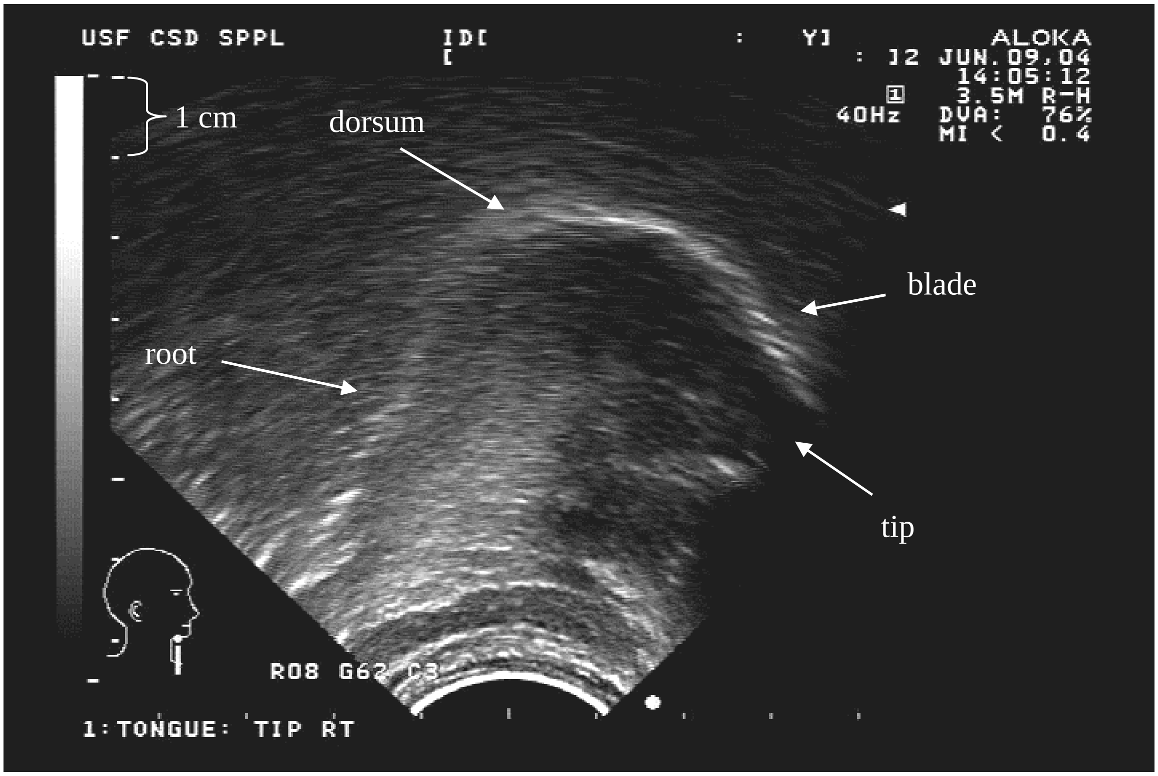


Figure 1: Sagital view of the tongue body with ultrasound; tongue tip to the right

Some research studies using ultrasound imaging to record speech production use a cushion-scanning technique (CST) to obtain more reliable measurements of speech production. CST involves the use of a transducer cushion, head stabilizer, ultrasound transducer holder and head position-recording device (Peng, et al., 2000). This system works to secure the head in a steady position during recording to ensure that the ultrasound transducer is fixed relative to the head and is not subject to variations in position or degree of transducer-skin contact. Stabilization of the head is important for recording speech production. CST is typically used in the speech perception and production lab at USF.

Due to its relatively fast scan rate and ability to image soft tissue effectively, ultrasound imaging provides more complete movement data on the tongue than any of the imaging instruments mentioned above (Stone, 1997). Also, ultrasound is safer than x-ray or MRI as there are no known hazards associated with ultrasound imaging due to its use of low power sound waves. These positive attributes, along with relatively low cost and ease of use, contribute to the growing popularity of ultrasound imaging in speech production research.

Ultrasound use in Speech Research

Ultrasound imaging is an effective way to view speech action or static postures. For example, ultrasound imaging has been used to image articulatory postures to describe speech sounds. Lundberg and Stone (1999) constructed 3D tongue surface shapes of nineteen speech sounds using multiple 2D ultrasound images. Frisch, Hardin, Nikjeh, & Stearns (2005) created a database of speech sound images using ultrasound in combination with a face video and endoscopic image.

Since ultrasound has a relatively high sampling rate it can be used to view and to record movement patterns of the tongue in real time. Some linguistic studies have focused on the use of ultrasound imaging to examine variations in the production of speech sounds in different languages. Language components that have been examined so far include the timing of articulatory events, tongue shape, and tongue movements during speech production (Gick, 2002). Current studies are examining these components across a variety of languages.

Past studies examining tongue position during speech production may have benefited from the additional information provided by ultrasound analysis. In 1967, Houde discussed the “forward looping motion” of the tongue during the production of VkV and VgV sequence sets. Houde (1967) examined x-ray images and found that the tongue body slides forward during velar closure. This motion is found for all vowel contexts, and so can not be explained by coarticulation with the vowel. A variety of researchers have tried to relate this motion to either passive or aerodynamic pressures on the tongue, as biomechanical in nature, or as created intentionally by motor control of the speaker.

A study by Perrier, Payan, Zandipour, and Perkell (2003) attempted to simulate the forward looping motion of the tongue during VCV sequence sets which included the vowels /a/, /i/, or /u/ and the consonant /k/. They claim the control of muscle movement between two sounds is based on a linear movement between muscles and targets. Therefore, the curvature of the articulatory trajectory during the looping motion rules out the theory that this motion is based on motor control of the muscles alone. Instead, they found that properties of the tongue including tongue elasticity and arrangements of the intrinsic and extrinsic muscles of the tongue contribute to passive elasticity, which creates the curvature of the looping motion in their model (Perrier, et al., 2003). These findings suggest that it is not biomechanics or passive force alone but a combination of both components that creates the forward looping motion.

The use of ultrasound imaging to view movements of the tongue would further enhance this type of speech research. The predictions of the Perrier et al. (2003) model could be compared to ultrasound video of real tongue movement in VCV sequences to see how well their simulations model the entire movement and shape of the tongue during VCV sequences. Ultrasound as an imaging tool provides objective data on speech postures and real time speech movements to support and enhance speech research. The data examined in this paper come from studies of speech articulation using CV and VCV sequences similar to those in Houde (1967) and Perrier et al. (2003).

Clinical Research Using Ultrasound in Speech Pathology

The safety and non-invasiveness of ultrasound has increased its popularity in research and clinical endeavors. Ultrasound has been used in the medical setting for many years. Its use in clinical speech language pathology seems to be a viable option. Ultrasound use could be valuable when diagnosing or treating speech impairment associated with neurogenic disorders. Some classic articulatory impairments that may be observed with ultrasound include “tongue rolling patterns” associated with Parkinson’s disease, dysarthrias, and apraxia of speech. Ultrasound may also be a viable visual feedback tool for persons learning tongue placement for speech articulation postures who are unable to rely on auditory feedback, for instance, persons within the deaf and hard of hearing populations.

One study examined tongue dorsum and laryngeal movements of a normal speaker compared to the speech movements of two people with Parkinson’s disease, one woman with senile dementia, an adult male stutterer, and an adult male with probable cranial traumatism (Keller, 1987). Tongue dorsum movements were measured for each of the participants using ultrasound imaging and voice recordings while repeating /ka/ in slow and fast repetitions. The extent and duration of ascending and descending lingual movements were compared between the normal speaker and disordered speakers. In the sample produced by the normal speaker descending lingual movements are rapid and carried out without hesitation. The ascending movement is executed with slight hesitation occurring between the lowest point (onset of regular glottal pulse oscillation visible in the audio track) and return to the highest point (preparatory movement prior to the first articulated syllable in the next set). This pattern was compared to the patterns seen in other speakers. The ultrasound recordings of the speakers with Parkinson's disease showed irregular displacement and duration of the movement pattern during slow repetitions of /ka/ compared to the normal speaker. The recording of the participant with senile dementia displayed visible signs of disturbance in lingual motor control with irregularity of movement and reduction of movement amplitude. Exaggerated initial lingual movements characterized the participant stutterer's speech. The amplitude of his lingual movement decreased with repetition. The speech of the participant affected by probable cranial traumatism showed decreases in movement amplitude during the fast repetitions (Keller, 1987).

The combination of electopalatography and ultrasound were used in a study examining the use of these feedback tools in the training of adolescents experiencing moderate to severe sensorineural hearing losses and moderately unintelligible speech (Bernhardt, 2001). The electropalatograph is a device used to monitor contact between the tongue and the palate and report points of contact between the two. The speech targets included silibant fricative place contrast (/s/ vs. /ʃ/) as well as the tense-lax high vowel contrast in (/i/ vs. /I/). The study data were trained listener transcriptions of target words before and after time spent in therapy. When applying these feedback tools together in therapy the students showed significant improvement in treatment targets as opposed to non-treatment targets. Based on these findings, it would seem that the use of ultrasound imaging has the potential to facilitate the diagnosis and treatment of a variety of speech impairments in a clinical setting.

Imaging properties of ultrasound also make it a valuable tool to observe swallowing patterns. Separate modes associated with ultrasound allow for the viewing of the sagital section of the tongue body (B-mode) and the function of specific oral components during swallowing (M-mode). This provides information on the timing and the integrity of the swallow.

Studies using this combination of sonographic techniques have assessed the duration, range of motion, and speed of tongue movement during each phase of swallowing. One study measured the swallowing of fifty-five normal persons. The swallow was divided into five phases described in the study as; phase I (shovel phase), phase II a (early transport phase), phase II b (late transport phase), phase III a (early final phase), and phase III b (late final phase) (Peng, et al., 2000). This study used a time amplitude diagram in M-mode ultrasonography that provides movement amplitude information. The M-mode image also provides a flat signal as soon as the tongue enters rest position (Peng, et al., 2000). The average duration of for all five phases of swallow was 2.43 sec. The average range of tongue motion during all phases of the swallow was 24.0 mm when viewed in the mid-sagittal plane. The speed of the swallow averaged 10.3 mm/sec.

The hyoid bone is an anatomical landmark whose movement is important for swallowing. B-mode ultrasound shows the shadow of the hyoid bone and the muscles from which it suspends. By adding the use of duplex-doppler ultrasound imaging a more accurate depiction of hyoid bone movement is provided (Sonies, Wang & Sapper, 1996). Doppler spectral patterns show movement between three phases for a normal swallow: hyoid elevates from rest, hyoid bone moves anteriorly to reach maximum displacement, and the hyoid bone moves back to resting position. Doppler spectral patterns were consistent amongst the six volunteers with a normal swallow that participated in this study. Sonies, et al. (1996) concluded that the phases of swallow could be determined by tracking the hyoid bone movement in a normal swallow. Quantitative information obtained on the phases of swallow may be valuable in the assessment of swallowing disorders. There is also the possibility that ultrasound can be used as a feedback tool to train swallowing exercises in therapy. In summary, ultrasound has the potential to be a valuable tool used by speech-language pathologists for biofeedback, diagnosis, and treatment of a variety of speech and swallowing disorders.

Ultrasound Imaging Reliability Studies

Every good research and clinical tool requires data to support its use. Although reliability data has not been established for the use of ultrasound imaging in speech research, reproducibility measures have been reported for other applications of ultrasound using Bland and Altman’s repeatability method. Bland and Altman use a correlation to compare two different measurement techniques and assess the repeatability of a method by comparing repeated measurements using one single method on a series of subjects. Bland and Altman have also developed a measure of reproducibility. It compares the values for different measurements using the same method. If measures are repeatable, the mean difference of different measurements for the same data should be zero. The coefficient of repeatability is 2 times the standard deviation of the measures for a single rater. The 2 SD window provides an estimate of the measurement variability within a single rater. The difference between the measures of different raters is compared to the coefficient of repeatability to determine whether the inter-rater measurement differences are small compared to the variability observed within the measurements of one rater.

One study used ultrasound imaging to evaluate the size of the ventricular system in children. A number of studies have provided evidence that ventricular dilation in children may be indicative of future learning disorders, autism and mental retardation (Iova, Garmshov, Androuchtchenko, Koberidse, Berg & Garmashov, 2004). The variation in measurement of the size of each part of the ventricular system within raters was reported using the means and standard deviations according to each of the age groups. Inter-rater reliability was evaluated against these means using the Bland and Altman method. The best inter-rater reliability was seen for the third ventricle (Iova et al., 2004).

Ultrasound Research Used in this Study

Studies of speech errors and timing of speech production have identified distinct phases of planning and control in the preliminary stages of language production (Munhall, 2001). Current research in the USF speech perception and production laboratory uses ultrasound-imaging techniques to examine errors in speech production in order to better understand these processes (Frisch, 2005). These studies examine articulatory characteristics of gestures in both normal and errorful speech production.

Information derived from these studies will provide an articulatory model of normal speakers from which to compare different forms of disordered speech.

In one part of these studies, the normal production of velar stops was examined. Wodzinski (2004) measured the velar closure location of the tongue preceding a Standard American English vowel in a CVC or CV syllable shape or embedded between two vowels in a VCV syllable shape. Results of the study showed with quantitative measures that the position of palatal contact during production of velars is dependent upon the frontness of the following vowel. Findings of this study help to determine the amount of coarticulatory variation that can occur in the normal production of a velar stop phoneme.

The second experiment used to derive reliability measures is a speech error study examining tongue position during the production of alveolar and velar phonemes paired with a vowel in CVC or CV syllable shapes. The third experiment used for reliability measurement in this study consists of the same type of data. However, it is a practice data set that was created to train research assistants on measurement procedures.

This study was designed to evaluate inter-rater reliability of hand measures of ultrasound analysis. Reliability measures are needed to support the use of ultrasound analysis for measuring speech production. This study will determine whether current measurement procedures for ultrasound analysis utilized in our speech science laboratory are sufficient. The results of this study will provide information on possible improvements to the current measurement protocol. Typically, ultrasound analysis is used in the medical setting for measurement of broad structures. In this study we will attempt to capture precise movements of the tongue and describe relatively small variations in speech production of alveolar and velar phonemes using ultrasound analysis. Since use of ultrasound analysis for speech production research is relatively new, reports of the reliability of its use for this purpose are limited.

In the current study each of the raters measured two separate data sets, and reliability measures were obtained. In addition, for the practice data set, intra-rater reliability was examined for the same rater measuring the data over two periods that were several months apart. In general, across all the studies, the measurements made by experienced raters are found to be reliable.

Chapter Two

Methods

Raters

The three raters (R1, R2 and R3) were adult female graduate students enrolled in the Master's degree program for speech language pathology at the University of South Florida. They worked in the Communication Sciences and Disorders department's speech science lab as research assistants. The participants did or will complete master's thesis projects in the area of speech science. R2 is the author of this thesis.

Stimuli

Experiment 1: Closure Angle Data. The data for this experiment consisted of measurements used to assess the location of velar closure along the palate for velar stop articulation with a variety of vowels. The closure angle data was measured by R1 and R2 for this study. The closure angle data was originally measured by R1 in a study that examined coarticulation between velar onset stops and the vowel that followed (Wodzinski, 2004). The recordings consisted of ultrasound images of the production of CV and CVC real words and VCV nonsense words read by two female volunteers ages 25-35, who were native speakers of Standard American English.

The real word stimulus set consisted of CVC and CV words containing an initial velar stop consonant (k or g) followed by a vowel. The real words ended in either a bilabial or labiodental coda or had no coda at all. Labial coda consonants were used in the final position of words so as not to interfere with tongue movement of the onset and vowel portion of the stimuli. The stimulus words were read six times in the context of the carrier phrase, “Say a ____ again.” Examples of real words used in this set were “Kim” and “go.”

The nonsense stimulus word set consisted of twenty-nine VCV stimuli containing a velar stop consonant (k or g) in the intervocalic position. The initial and final vowels were the same, and the vowel that was used varied across stimuli. Each VCV nonsense word was read six times in the context of the carrier phrase “Say ___ again.” Examples of the VCV nonsense words used in this set were /iki/ and /ugu/.

Experiment 2: Tongue Twisters # 1. The data for this experiment consisted of measurements of the production of tongue twisters by R2 and R3. The stimuli for this experiment were taken from a larger project that is currently in progress “Ultrasound study of errors in speech production” (Frisch, 2003). The study has two parts, a baseline portion and a tongue twister portion. The baseline portion of the study was designed to provide normal examples of alveolar and velar stop consonants. The tongue twister portion of the study was designed to elicit speech errors between onset alveolar and velar stop consonants.

The recordings consisted of ultrasound images of the production of tongue twisters using CV and CVC real words and CV nonsense syllables read by a volunteer within the Communication Sciences and Disorders department at the University of South Florida. The stimuli consisted of sets of four words, which were repeated six times consecutively.

The initial consonant in the CV and CVC real word and CV nonsense word stimuli consisted of either the alveolar phoneme /d, t/ or the velar phoneme /g, k/. The initial consonant phoneme alternated between alveolar /d, t/ and velar phonemes /g, k/ within the set to create a tongue twister. The vowel consisted of either /a/ or /ae/. The coda consonants used in the words (if any), were labial. Two examples of baseline stimuli are /ta tae tae ta/ and /gae ga ga gae/. Two examples of real word tongue twister stimuli sets are “top cap cop tab” and “dam gob gap damp.” Two examples of CV nonsense word tongue twister stimuli sets are /ta kae ka tae/ and /gae da dae ga/.

Experiment 3: Tongue Twisters # 2. The data for this experiment consisted of tongue twister recordings, similar to experiment 2, measured by R2 and R3. The stimuli measured consisted of ultrasound images of speech production recordings of tongue twisters read by USF faculty member Stefan Frisch. This recording is used in the USF speech perception and production lab to familiarize research assistants with the data and measurement procedures used in the tongue twister study.

Procedures

The procedures used for audio and video recording and hand measurement of ultrasound images of speech production were the same for all raters across the three experiments. The audio and video recordings created of the talkers were measured using the same computer, programs, and settings. For a detailed description of the placement of talkers and the procedures for audio and video recording see Frisch (2003) and Wodzinski (2004).

Measurements

All experiment data sets were measured using the programs Adobe Premiere 6.0 and Adobe Photoshop 7.0. The video recording of the tongue was viewed in Adobe Premiere 6.0. The video of tongue movement was observed frame-by-frame, until the frame closest to the midpoint of consonant closure was determined. The frame was then imported into Photoshop for measurement. Cues utilized to determine the exact closure location included direction of tongue movement preceding and following closure, flattening of the tongue against the alveolar ridge or palate, and the bright line of the tongue edge that appears when the tongue surface is motionless during closure. The audio waveform was also used to identify the appropriate frame containing stop closures and releases (Wodzinski, 2004).

Experiment 1: Closure Angle. In this experiment the rater's goal was to quantify the location of closure of the tongue dorsum against the palate. First, the rater (R1 or R2) chose the frame displaying velar closure. Within this frame, the most anterior and posterior points of the closure constriction were marked. The location of these points was used to determine the midpoint of the closure against the palate. The closure angle was derived by measuring the angle from the location of the center of the ultrasound probe (at the base of the video image) to the midpoint. The recordings were measured by R1 during the 2003-2004 school year for the thesis titled "Ultrasound Analysis of Velar Fronting". The recordings were measured by R2 in March of 2005 for this thesis. Inter-rater reliability between R1 and R2 was examined for choice of closure frame, identification of anterior and posterior closure points in the image, and resulting velar closure angle. An example of dorsum closure angle measurement is shown in figure 2.

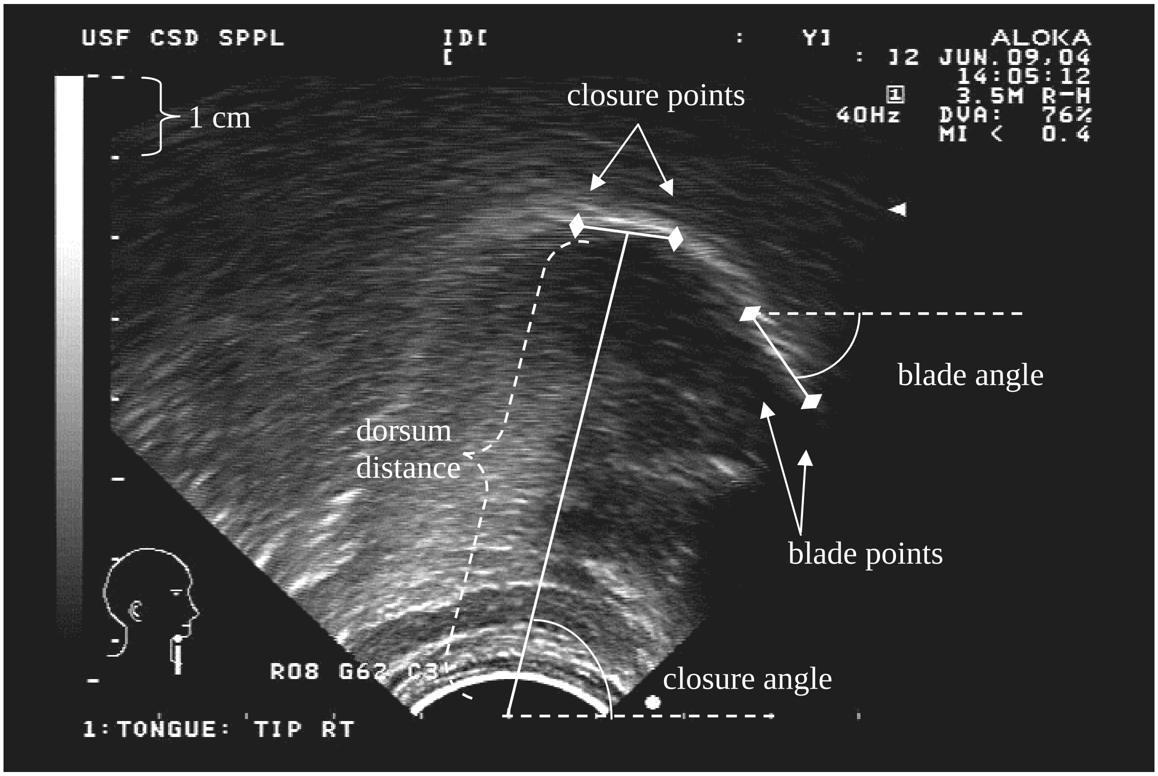


Figure 2: Measurements applied to ultrasound image of the tongue

Experiment 2. In the tongue twister experiment, R2 and R3 measured velar consonants and alveolar consonants. For velar consonants, velar closure location and closure angle measurement were derived in the same way as in experiment one. For alveolar consonants the actual closure of the tongue tip against the alveolar ridge is often not visible in an ultrasound image because the ultrasound beam from the probe is reflected by air under the tongue tip. Consequently, the measure used to assess alveolar closure is the angle of the tongue blade, computed from two points. The first point chosen was the most anterior portion of the visible tongue blade, and the second was a point about one centimeter posterior to the first along the tongue blade. Based on these two points the amount of elevation or declination of the tongue tip from the horizontal plane determines the blade angle. The tongue blade angle was also measured for velar stop consonants. An example is shown in figure 2.

Following the same procedure, the measure of dorsum distance was derived to assess degree of elevation of the tongue dorsum during production of alveolar and velar stop consonants. The dorsum distance was determined by tracing a line from the center of the ultrasound probe to the typical location of velar closure that was determined from baseline measures. The angle of closure was derived by averaging the closure points for each vowel context when following a velar consonant (g or k) in a CVC syllable set within the baseline data. An example is shown in figure 2.

Experiment 3. In the second tongue twister experiment raters derived all measurements in the same way as experiment two. This data set was measured by R3 in August of 2004 by which time R3 was already an experienced measurer of ultrasound images of consonants. This data set was measured by R2 in August of 2004 as a training data set when R2 was still inexperienced with ultrasound analysis (R2i). The same data set was measured a second time by R2 in March of 2005 once R2 had gained considerable experience with ultrasound image analysis (R2e). Experience as a variable will be assessed in this study by comparing R2’s measures before and after gaining experience (R2i vs. R2e), and by comparing the two sets of measures from R2 with the measures of another experienced researcher (R3).

The measurements from all experiments will be compared to determine inter-rater reliability. This project will compare choice of video frame of closure, points chosen to measure closure for velar consonants, points chosen to measure the tongue blade for velar and alveolar consonants, and angle measurements derived from those points. The results will be described quantitatively by examining the mean, SD, and the range of differences in measures. The results will be analyzed statistically by a t-test for angle measures derived by two raters. In addition, correlation analysis and the Bland and Altman reproducibility measure will determine if any significant difference exists in the measure derived by raters when measuring speech production data using ultrasound-imaging techniques. For experiment 3, results derived by a single rater at different levels of experience will be examined. The results obtained by this rater before and after experience has been gained will be compared using the same statistical measures. The results will also be compared to another experienced rater to assess experience as a factor affecting reliability of hand measures of ultrasound data.

Chapter Three

Results

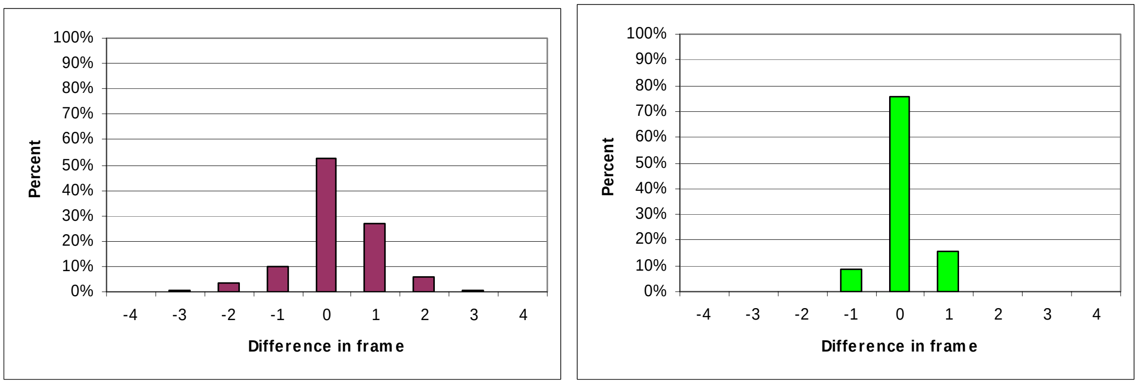
This thesis assesses reliability of hand measures of ultrasound analysis of consonant closure in speech production in three experiments. It assesses the choice of the frame from the video image, location of the anterior and posterior points used to measure alveolar and velar articulation, and the resulting tongue blade and velar angle. Experience with ultrasound measurement is also assessed as a factor impacting reliability in experiment 3.

Experiment 1: Dorsum Angle Data

Frame Choice: Speaker 1. The first experiment compares closure angle measures in CV and CVC real word data and VCV non-word data measured by two raters. The first step in the analysis procedure is to select the video frame that best captures the midpoint of closure. Consequently, frame selection was compared between raters to determine how consistently the same video frame was chosen. Differences in frame choice selected by R1 and R2 are shown in Figure 3. The differences for speaker 1 data are shown in the top row, with word data on the left and non-word data on the right. In this figure, a positive difference means that R2 selected a frame later than R1. For the CV and CVC real word data, raters chose the same frame approximately 52 percent. When the choice of frame was not the same R2 tended to choose a video image depicting alveolar and velar closure one frame later than R1. 26 percent of the time a video image was chosen one frame later and 10 percent of the time closure was marked one frame earlier. Thus, for words, R1 and

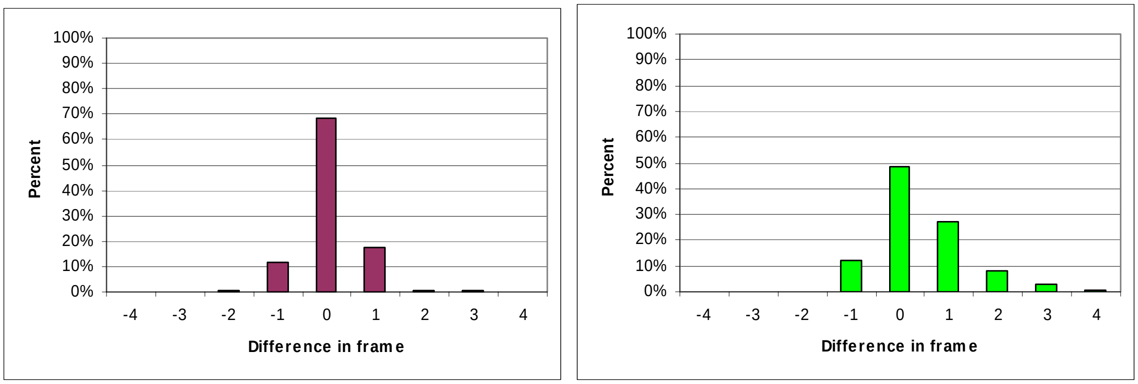
R2 were within 1 frame 88 percent of the time. Frame selection was also compared for non-word VCV data. In this data set the same frame was chosen 75 percent of the time. Again there was a slight tendency for R2 to select a frame one frame later at 16 percent versus 9 percent of the time when R2 chose a closure frame one frame earlier.

Frame Choice; Speaker 2. The same comparisons were made between raters from the productions of a second reader, shown in the bottom row of Figure 2. For CV and CVC real word data, raters chose the same frame approximately 68 percent of the time. Again, R2 demonstrated the tendency of choosing a video frame depicting closure one frame later than R1. A frame choice of one frame later occurred approximately 18 percent of the time, as opposed to one frame earlier 12 percent of the time. Frame selection was also compared for VCV non-word data. In this data set the same frame was chosen 49 percent of the time. Again, the next highest frame selection occurred one frame later 27 percent of the time. Frame choice occurred one frame earlier 12 percent of the time.



Speaker 1 word data

Speaker 1 non-word data



Speaker 2 word data

Speaker 2 non-word data

Figure 3: Difference in choice of frame to measure velar stop closure (R2 – R1)

Closure Points. The average difference of anterior and posterior points of velar closure for speakers 1 and 2 are depicted in Table 1. Distances are shown separately for anterior and posterior tongue blade points for CVC and CV real word data and VCV non-word data for speakers 1 and 2.

Table 1: Average distance between velar closure points marked by R1 and R2.

	Speaker 1		*Speaker 2*	
	Real Word	*Non Word*	*Real Word*	*Non Word*
Anterior	3.2 mm	3.1 mm	3.4 mm	2.9 mm
Posterior	6.2 mm	5.2 mm	3.4 mm	3.3 mm

Overall, the distance between points is small to medium. The average distance between velar closure points was smaller for the non-word data. More variation was seen between the posterior velar closure points chosen for both real and non-word data for speaker 1 but not for speaker 2.

To assess the reliability of this data, we can compare the measurement differences between raters to the variability within a rater using the Bland & Altman measure. Within a rater, the average distance between measurement points for repetitions of the same stimulus was 2.6 mm with a SD of 1.8. So the limit of reproducibility by the Bland and Altman measure is 6.1 mm. The measures for the anterior points for speaker 1 fall within this limit, but the measures for posterior points are near to the limit. The average differences for both real and non-word data were closer for speaker 2. For speaker 2, the difference in measurements falls within the interval of reproducibility by the Bland and Altman measure

Velar Angle. The differences in dorsum angle measures including the average angle, the mean angle difference, a paired t-test and correlation are given in table 2 fpr R1 and R2.

Table 2: Velar angle differences for speakers 1 and 2.

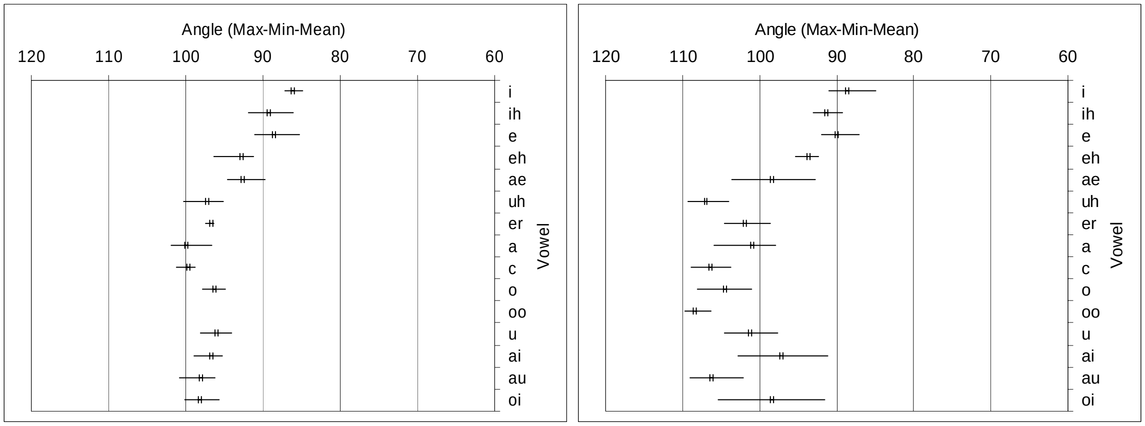
Speaker 1		
	Word	*Non-word*
Mean Angle	R1 = 83.1°, R2 = 81.5°	R1 = 80.7°, R2 = 80°
Mean Angle Diff	1.6°	0.7°
t-test	t(119) = 10.3, p<0.01	t(171) = 4.1, p < 0.01
r =	0.93	0.95

Speaker 2		
	Word	*Non-word*
Mean Angle	R 1 = 85.0° , R2 = 85.3°	R1 = 90.7°, R2 = 90.9 °
Mean Angle Diff	0.3°	0.3°
t-test	t(188) = 1.88 , n.s.	t(172) = 1.46 , n.s.
r =	0.92	0.88

The difference in angle measures for speaker 1 was statistically significant, the resulting analysis of how coarticulation between velar and the vowel affected the closure location is the same whether you use the data from R1 or R2. The agreement in coarticulation pattern observed by R1 and R2 for speaker 1 can be seen statistically in the high correlation between raters. For the word data r = 0.93 and for the non-word data r = 0.95.

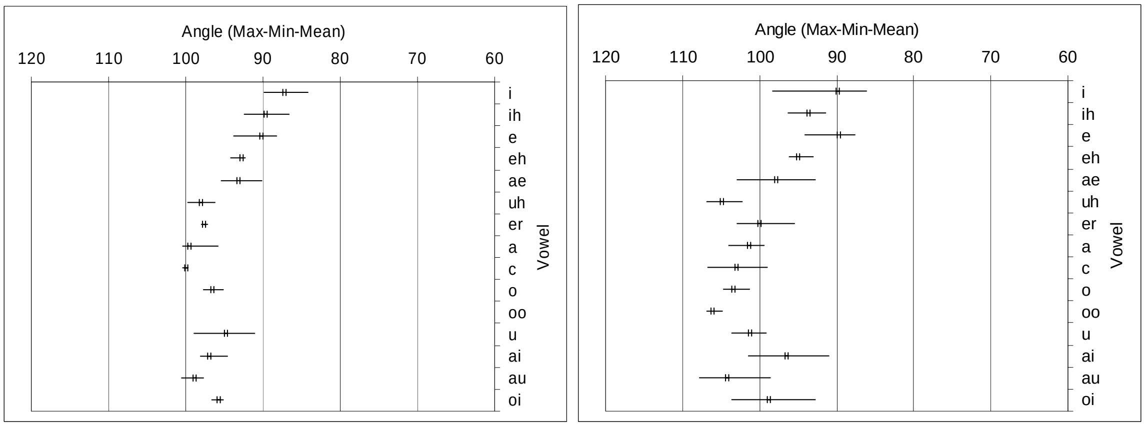
In other words, R1 and R2 agreed in overall pattern for how dorsum angle related to the following vowel, as shown in Figure 4. Figure 4 shows the mean angle and range of angle measures for speaker 1 for each vowel. The top row of Figure 3 shows the measures for R1 of the data for speaker 1. The bottom row of Figure 3 shows the measures for R2 of the data for speaker 1. The word data are shown in the left column, and the non-word data are shown in the right column. Overall, the patterns are very similar.

Figure 5 provides the same measures of word and non-word data for speaker 2. No significant difference was seen between R1 and R2 for dorsum angle measures of real word or non-word data for speaker 2. As with speaker 1, the resulting analysis of coarticulation between velar and the vowel that followed is the same whether you use the data from R1 or R2. In other words, R1 and R2 also agreed in overall pattern for how dorsum angle related to the following vowel for speaker 2. The correlation between measures for speaker 2 was also high. The correlation between raters for word data is r = 0.92 and correlation between raters for the non-word data is r = 0.88.



R1 measures of words, speaker 1

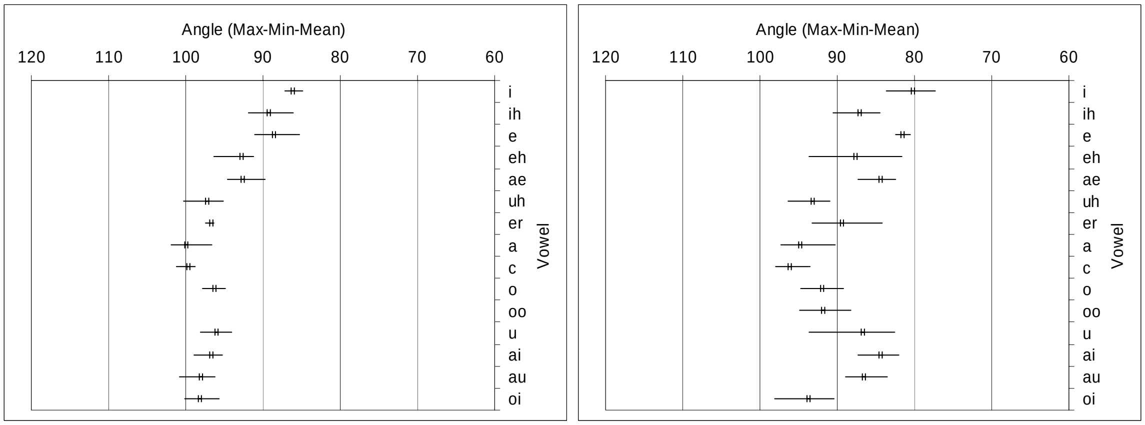
R1 measures of non-words, speaker 1



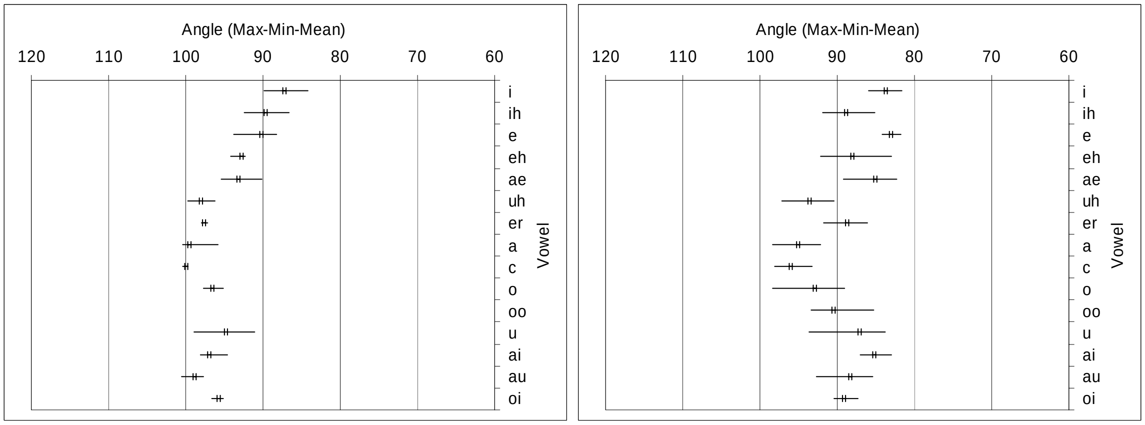
R2 measures of words, speaker 1

R2 measures of non-words, speaker 1

Figure 4: Coarticulation patterns for word and non-word data for speaker 1, for measures by R1 and R2.


R1 measures of words, speaker 2

R1 measures of non-words, speaker 2


R2 measures of words, speaker 2

R2 measures of non-words, speaker 2

Figure 5: Coarticulation patterns for word and non-word data for speaker 2, for measures by R1 and R2.

The Bland and Altman measure of reproducibility was used to determine the reproducibility of dorsum angle measures dependent upon the vowel. Table 3 shows the mean difference in angle between raters for word and non-word data for speakers 1 and 2. Table 3 also shows the limit of the reproducibility measure, 2 SD for the measure of angle for each vowel, averaged between R1 and R2’s measures. The majority of the measures proved to be reproducible, falling within two standard deviations. In a few cases, the difference fell outside of the range considered to be reliable for reproducibility, and are marked by gray shading in Table 3. Overall, the velar angle measures for each vowel are considered to be reproducible using this comparison as there is no consistent pattern to the lack of reproducibility. The angle differences that are outside of the 2 SD limit are indicated by gray shading.

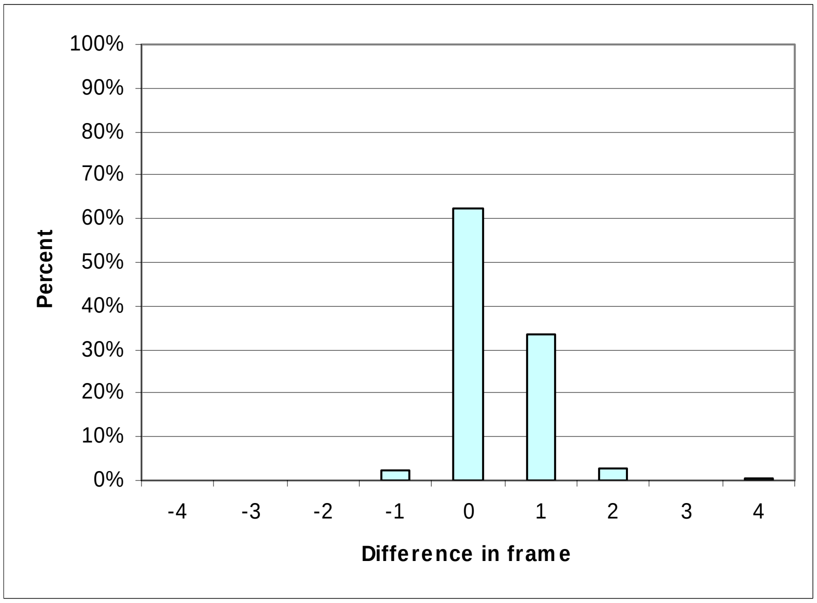
Table 3: Mean measurement difference and average 2 SD measurement variability for the dorsum angle data.

	Speaker 1				*Speaker 2*			
	Word		*Non-word*		*Word*		*Non-word*	
Vowel	*Angle diff*	*2 SD avg*	*Angle diff*	*2 SD avg*	*Angle diff*	*2 SD avg*	*Angle diff*	*2 SD avg*
i	0.6º	2.7	-1.2º	5.2	1.1º	2.9	3.4º	3.2
ih	0.5º	5.8	-2.2º	2.7	0.1º	3.7	0.8º	4.4
e	1.6º	3.8	0.3º	3.4	0.3º	3.5	1.6º	1.6
eh	1.0º	2.1	-1.4º	2.5	0.0º	2.7	0.6º	8.4
ae	1.7º	5.0	0.6º	7.1	-0.5º	3.3	0.6º	3.7
uh	0.5º	4.6	2.1º	2.9	-0.9º	2.5	-0.3º	4.0
er	2.3º	2.3	2.0º	3.7	-0.9º	0.7	0.5º	5.2
a	2.6º	5.2	-0.3º	4.5	0.4º	2.6	-0.2º	3.8
c	3.7º	3.3	3.4º	4.5	-0.2º	1.2	0.1º	2.9
o	1.8º	7.2	1.1º	2.9	-0.3º	1.9	-1.1º	4.5
oo			2.3º	1.5			0.5º	4.4
u	1.2º	2.8	0.1º	3.6	1.2º	4.1	0.8º	7.0
ai	1.2º	3.3	0.7º	7.1	-0.2º	2.9	0.7º	3.1
au	1.7º	3.5	2.1º	4.6	2.4º	2.3	2.7º	3.7
oi	2.7º	2.7	-0.4º	10.3	-0.8º	2.7	1.3º	3.4

Experiment 2: Tongue Twister Data

Frame Choice. In experiment two, alveolar and velar measurements for CV non-word and CVC real word tongue twister data obtained by two raters (R2 and R3) were compared. The tongue twister data measurements were made on a word initial alveolar consonant /t/ or /d/ or velar consonant /k/ or /g/. For experiment 2, the data for alveolar and velar consonants are analyzed separately. As with the data from experiment 1, the first choice in the measurement procedure is the choice of frame to measure. Differences in the choice of frame for alveolar and velar consonants are shown in Figure 6. Raters chose the same frame depicting alveolar closure approximately 62 percent of the time in the CV non-word and CVC word data sets. R2 chose a closure frame one frame later than R3 about 33 percent of the time. Only 2 percent of the time a video frame depicting closure was chosen one frame earlier. For velar consonants, raters chose the same frame depicting velar closure approximately 54 percent of the time in CV non-word and CVC word data. R2 demonstrated a tendency towards choosing a closure frame one frame later than R3. 39 percent of the time a video image was chosen one frame later. Again, 2 percent of the time a video frame depicting closure was chosen one frame earlier. The pattern was the same for both alveolar and velar consonants.

Alveolar consonants



Velar consonants

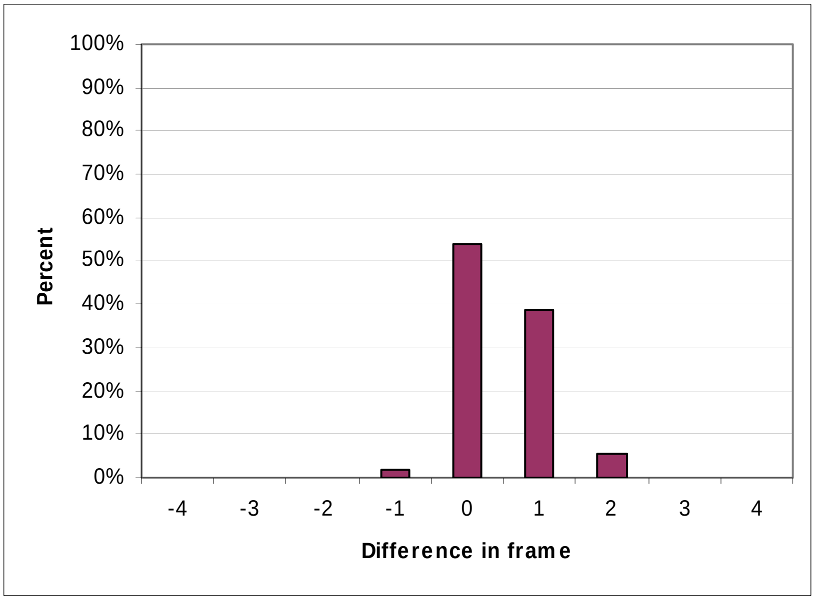


Figure 6: Difference in choice of frame for alveolar and velar stop closure (R2 – R3)

Tongue Blade Points. A comparison of tongue blade measurement points chosen by R2 and R3 is given in table 4. A comparison was made between both the anterior (tongue blade front) and posterior (tongue blade back) closure points chosen by the two raters, for both alveolar and velar consonants. On average, the distances between measurement points were small. There was generally a greater difference for the posterior point than the anterior point. Applying the Bland and Altman measure of reproducibility within a rater to the blade data point gives an average limit of reproducibility of 12.1 mm. So these measurement differences fall within the Bland and Altman limit for the most part.

Table 4: Difference in location of tongue blade measurement points for tongue twister data for R2 and R3.

Alveolar	*Anterior*	*Posterior*
Mean	2.3 mm	4.4 mm
SD	1.6 mm	3 mm
Min	0 mm	0 mm
Max	8.5 mm	13.4 mm
Velar	*Anterior*	*Posterior*
Mean	2.9 mm	3.8 mm
SD	2.5 mm	2.7 mm
Min	0 mm	0 mm
Max	13.9 mm	16.4 mm

Velar Closure Points. Anterior and posterior points chosen by R2 and R3 were also compared for velar closures. Velar closure was only measured for velar productions. The descriptive statistics are shown in Table 5.

Table 5: Difference in location of dorsum measurement points for tongue twister data for R2 and R3.

	Anterior	*Posterior*
Mean	3.5 mm	2.4 mm
SD	1.9 mm	1.5 mm
Min	0 mm	0 mm
Max	9.6 mm	8.3 mm

Tongue Blade Angle Measures. A comparison of the tongue blade angle measures that resulted from the points selected was made between R2 and R3. Summary statistics are provided in Table 6 below. Overall the tongue blade angle means and standard deviations were comparable for both alveolars and velars. The difference in rater measures for the alveolars is statistically significant. The velar blade angle measurements proved to be more reliable between the raters and are not statistically significant.

Table 6: The difference in tongue blade angle measures of R2 and R3 for alveolar and velar consonants

	Alveolar	*Velar*
Mean Angle	R2 =23.6°, R3 = 23.9°	R2 = 45.7°, R3 = 45.6°
Mean Angle Diff.	0.4°	-0.1°
t-test	t(268) =2.1, p< .05	t(266) = 0.5, n.s.
r=	.94	.83

According to Bland and Altman's reproducibility measure these angle measures are reproducible. Two standard deviations from the mean for alveolar tongue blade measures is 15.1°. The average angle difference for alveolars is .04° so this is a reproducible measure by Bland and Altman's standards. Two standard deviations of the mean for velars are 14.7°. The average difference for velars is –0.1° so this is considered reproducible by Bland and Altman's measure.

Velar Angle Measures. Velar angle measures derived by R2 and R3 were also compared. Summary statistics are provided in Table 7 below. The Bland and Altman method was applied to velar angle measures. Two standard deviations of the mean within raters is 4.9 °.

Table 7: Differences in velar angle measures derived for R2 and R3

	Velar Angle
Mean Angle	R2 = 85.5°, R3 = 86.4°
Mean Angle Diff.	-0.01°
t-test	t(272) = 12.0, p < .001
r=	.89

Dorsum Distance. The productions were measured for the degree of dorsum raising by measuring the distance from the ultrasound probe to the tongue dorsum along the angle where dorsal closure is typically observed for the vowel for the speaker. Table 8 provides descriptive statistics and statistical tests for the difference in dorsum distance measures between R2 and R3 for both alveolar and velar consonants. Overall, the differences in distance are small, and somewhat smaller for velars than alveolars. There is a statistically significant difference between the measures of R2 and R3, however the correlation between the two raters is high, especially for the alveolars. By the Bland and Altman method, the limit of reliability for alveolars would be 6.4 mm, and for velars would be 2.6 mm, so both measures are considered reliable.

Table 8: Dorsum distance measures for R2 and R3.

	Alveolar	*Velar*
Mean Dist.	R2 = 46.0 mm, R3 = 46.8 mm	R2 = 51.0 mm, R3 = 51.3 mm
Mean Diff.	-13 mm	-5 mm
t-test	t(269) = 20.3, p < .001	t(255) = 9.0, p < .001
r =	0.95	0.78

Global View. As in experiment 1, the measures by the different raters can be compared as to how they provide an overall picture of the data from the experiment. Figure 7 shows the measures of the productions from speech error experiment two for velar and alveolar consonants for the two raters (R2 and R3). The data for alveolar consonants is shown in the top row. The data for velar consonants is shown in the bottom row. The measures by R2 are shown in the left column. The data for R3 are shown in the right column. Each production is quantified by the tongue blade angle and the dorsum distance for the consonant. These two measures are the most relevant to the articulation of alveolar and velar stops respectively. Figure 6 shows that the majority of the alveolar and velar stop productions occupy distinct regions of the graph. For a few tokens, an error was produced and the measures of the articulation show an articulation that was typical of the opposite category. In comparing the different raters, the graphs for the two raters look very similar to one another. Thus, it appears that the measurement procedures used to quantify articulation in the tongue twister experiment provide the same overall data for different raters.

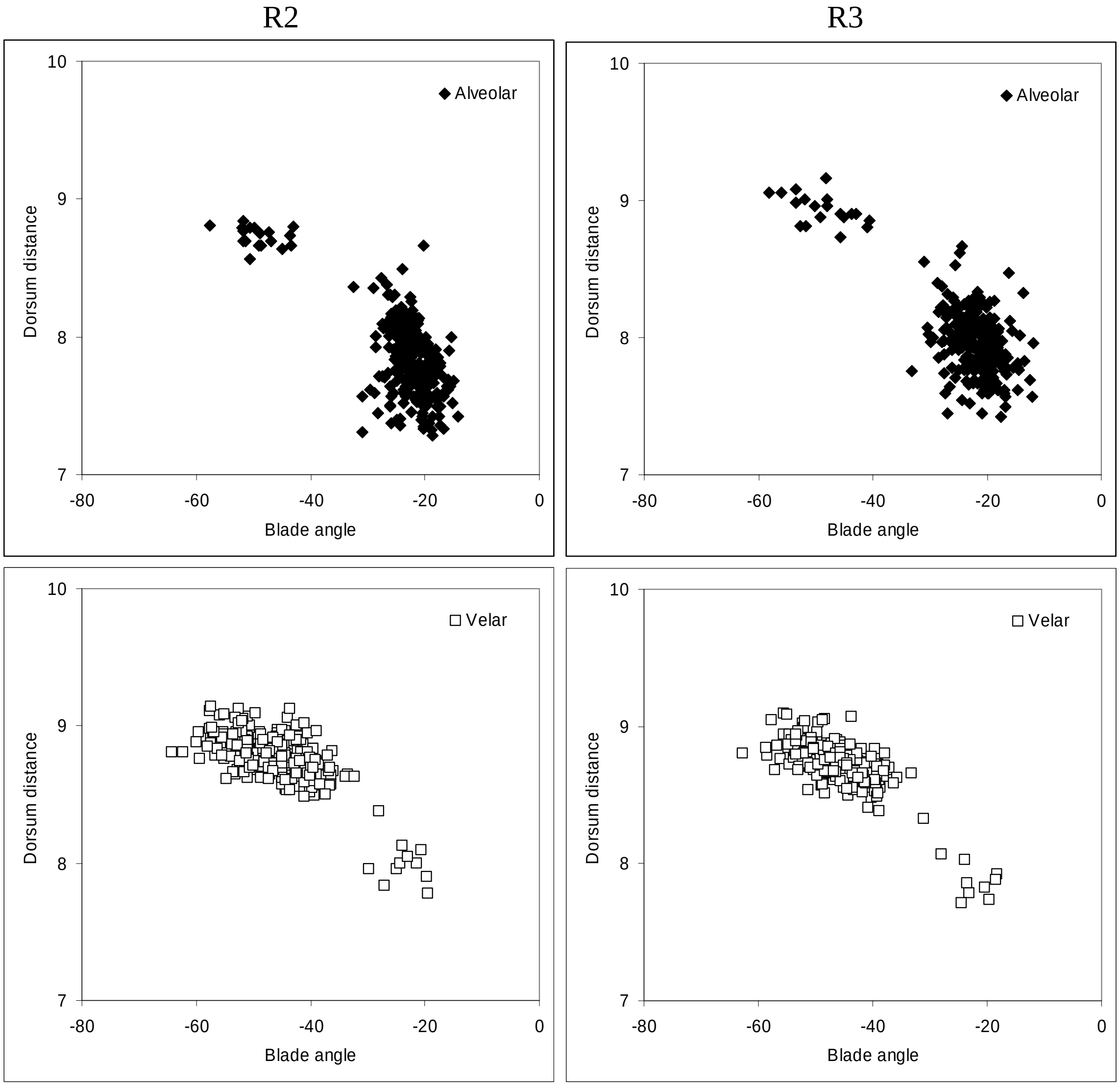


Figure 7: Comparison of overall data pattern for tongue twister experiment for R2 and R3

Experiment 3: Tongue Twister # 2

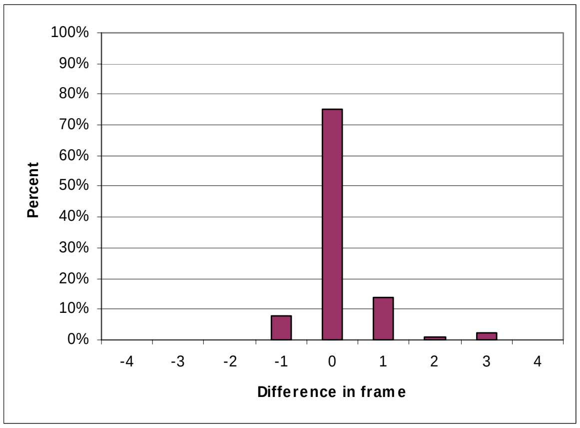
Ultrasound analysis measures of tongue twister baseline data were compared between R2 and R3. Tongue twister baseline data measures were also compared between R2’s measures when inexperienced in ultrasound analysis (R2i) and R2’s measures seven months later upon gaining considerable experience in measurement technique (R2e). Both experienced and inexperienced R2 measures were compared with the measures derived by R3, an experienced measurer, to assess reliability. The tongue twister baseline data consisted of CV syllables beginning with either the alveolar phoneme /t/ or /d/ or the velar phoneme /g/ or /k/.

Frame Choice: R2i and R3. The frame choice was not divided by alveolar and velar measures for this section since no large difference was observed in experiment two. Differences in frame choice for the second set of tongue twister data are shown in Figure 8 below. In comparing the video image frames chosen for alveolar and velar closure location for baseline data between R2i and R3, exact frame agreement was seen 75 percent of the time. R2i chose a closure image one frame later than R3 approximately 12 percent of the time.

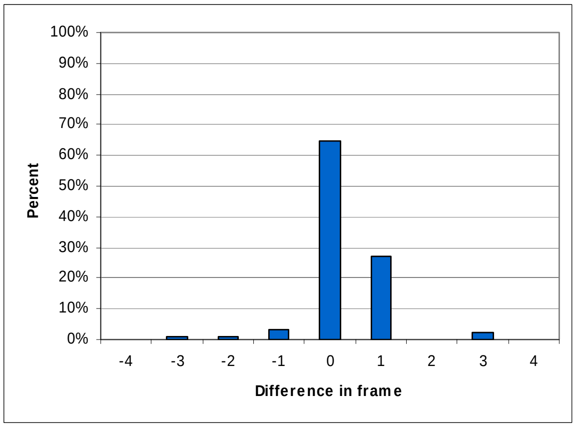
R2e and R2i. In comparing the video image frames chosen for alveolar and velar closure location for baseline data between R2e and R2i; exact frame agreement was seen 65 percent of the time. Upon gaining experience R2e chose the frame depicting closure as one frame later than the R2ie approximately 27 percent of the time. In both instances experience increased R2’s tendency to choose the closure location as one frame later than when inexperienced.

R2e and R3.In comparing the video image frames chosen for alveolar and velar closure location for baseline data between R2e and R3; exact frame agreement was seen 56 percent of the time. Upon gaining R2e chose the frame depicting closure as one frame later than R3 approximately 35 percent of the time.

R2i vs. R3



R2i vs. R2e



R2e vs. R3

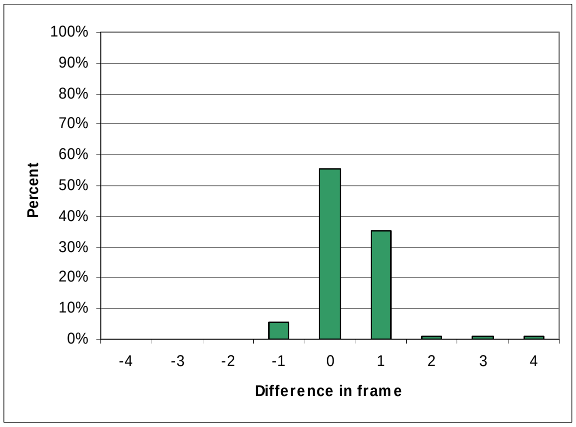


Figure 8: Difference in choice of frame for measurement between R2i, R2e, and R3

Tongue Blade Points. The analysis of differences in choice of tongue blade closure point was not broken up for alveolar and velars because no large differences were noted in experiment two. Summary statistics of anterior and posterior closure points are provided in Table 8 below. A comparison was made between the anterior (tongue blade front) closure point chosen by two separate raters and the posterior (tongue blade back) closure point chosen by two separate raters.

Table 8: Distance between anterior and posterior tongue blade points selected by experienced and inexperienced measurers.

Anterior			
	R2i and R3	*R2i and R2e*	*R2e and R3*
Mean	1.0 mm	0.8 mm	0.4 mm
SD	0.7 mm	0.5 mm	0.3 mm
Min.	0 mm	0 mm	0 mm
Max.	3.0 mm	1.4 mm	1.9 mm
Posterior			
	R2i and R3	*R2i and R2e*	*R2e and R3*
Mean	0.8 mm	0.7 mm	0.3 mm
SD	0.6 mm	0.4 mm	0.3 mm
Min.	0 mm	0 mm	0 mm
Max.	2.3 mm	1.4 mm	1.6 mm

In comparing point locations chosen between R2i and R3 and R2e and R3; the average mean distance between the anterior tongue blade points (tongue blade front) became closer with experience (from 1.0 mm to 0.4 mm.). The average distance between the posterior tongue blade points (tongue blade back) also became closer with experience (from 0.8 mm to 0.3 mm). The range narrowed with experience for both anterior and posterior closure points. The greatest consistency in choice of point location was seen between the R2e and R3 measures. Results indicate that experience did impact choice of points of tongue blade closure.

Velar Closure Points. A comparison was made between the anterior (velar front) closure point chosen by two separate raters and the posterior (velar back) closure point chosen by two separate raters. One factor by which rater's measures are compared is experience. Table 9 below shows the mean, standard deviation and range of distances between closure points selected by raters.

Table 9: Distances between anterior and posterior velar closure points selected by experienced and inexperienced measurers.

Anterior			
	R2i and R3	*R2i and R2e*	*R2e and R3*
Mean	0.3 mm	0.5 mm	0.5 mm
SD	0.3 mm	0.3 mm	0.3 mm
Min	0 mm	0 mm	0 mm
Max	1.7 mm	1.9 mm	1.9 mm
Posterior			
	R2i and R3	*R2i and R2e*	*R2e and R3*
Mean	0.7 mm	0.3 mm	0.3 mm
SD	0.4 mm	0.3 mm	0.3 mm
Min	0 mm	0 mm	0.1 mm
Max	2.3 mm	1.1 mm	1.9 mm

In comparing point locations chosen between R2i and R3 and R2e and R3; the average distance between the anterior velar points (velar front) did not become closer with experience (from 0.3 mm to 0.5mm.). The average distance between the posterior velar point (velar back) locations improved with experience (from 0.7 mm to 0.3 mm). Overall the differences were small.

Tongue Blade Angle Measures. The tongue blade angle measures were not divided for alveolar vs. velar because no large differences in reliability were observed between the two in experiment two. The tongue blade measure comparisons shown in Table 10 below are divided by, R2i versus R3 measures, R2i versus R2e measures and R2e versus R3 measures. Overall, the mean difference in angle measure and standard deviation became closer with experience. A significant difference in tongue blade angle was seen in the comparisons, involving the inexperienced rater R2i. The comparisons of the experienced measurers show no significant difference. The highest correlation was also seen between the experienced measures. Blade angle measures to became more reliable with experience in this experiment.

Table 10: Tongue blade angle measures by experienced and inexperienced raters

	R2i and R3	*R2i and R2e*	*R2e and R3*
Mean Angle	R2i = -31° R3 = -21°	R2i = -32° R3 = -21°	R2e = -32° R3 = -30°
Mean Angle Diff	9.5 °	10.9 °	1.4 °
t(87)	9.0, p < .001	8.8, p < .001	1.6, n.s.
r =	0.58	0.53	0.86

Velar Angle Measures. The velar angle measures were also compared amongst R2i versus R3, R2i versus R2e, and R2e and R3 in Table 11 below. The mean and standard deviation of differences in velar angle measures became smaller with experience. The range of differences in the angle measures became closer. In both comparisons with an inexperienced measurer the velar angle measures were significantly different. By the time R2 had gained experience and was compared with another experienced measurer the differences were no longer significant. A significant increase in correlation of velar angle measures is seen with experience.

Table 11: Velar angle measures obtained by experienced and inexperienced raters.

	R2i and R3	*R2i and R2e*	*R2e and R3*
Mean Angle	R2i = 79.3° R3 = 81.0°	R2i = 79.4° R3 = 80.9°	R2e = 80.9° R3 = 81.0°
Mean Angle Diff	1.6 °	1.5 °	0.1 °
t(43)	5.4, p < .001	4.5, p < .001	0.5, n.s.
r =	0.80	0.75	0.90

Dorsum Distance. Dorsum distance measures were obtained for both alveolars and velars. The results are displayed in Table 12. Differences in dorsum distance measures were extremely small. Correlation was high for all three comparisons.

Table 12: Differences in dorsum distance measures by experienced and inexperienced raters

	R2i and R3	*R2i and R2e*	*R2e and R3*
Mean Diff	-0.1 mm	0.1 mm	0 mm
t(87)	6.9, p < .001	5.9, p < .001	1.8, n.s.
r =	0.98	0.96	0.97

Global View. As in the previous experiments, the measures by the different raters can be compared as to how they provide an overall picture of the data from the experiment. Figure 6 shows the measures of the productions from speech error experiment two for velar and alveolar consonants for the three raters (R2i, R2e, R3). Each production is quantified by the tongue blade angle and the dorsum distance for the consonant. These two measures are the most relevant to the articulation of alveolar and velar stops respectively. Figure 9 shows that the alveolar and velar stop productions occupy distinct regions of the graph. In comparing the different raters, the graphs for the two experienced raters look much more similar to one another, and different from the inexperienced rater, especially for the blade angle for velars.

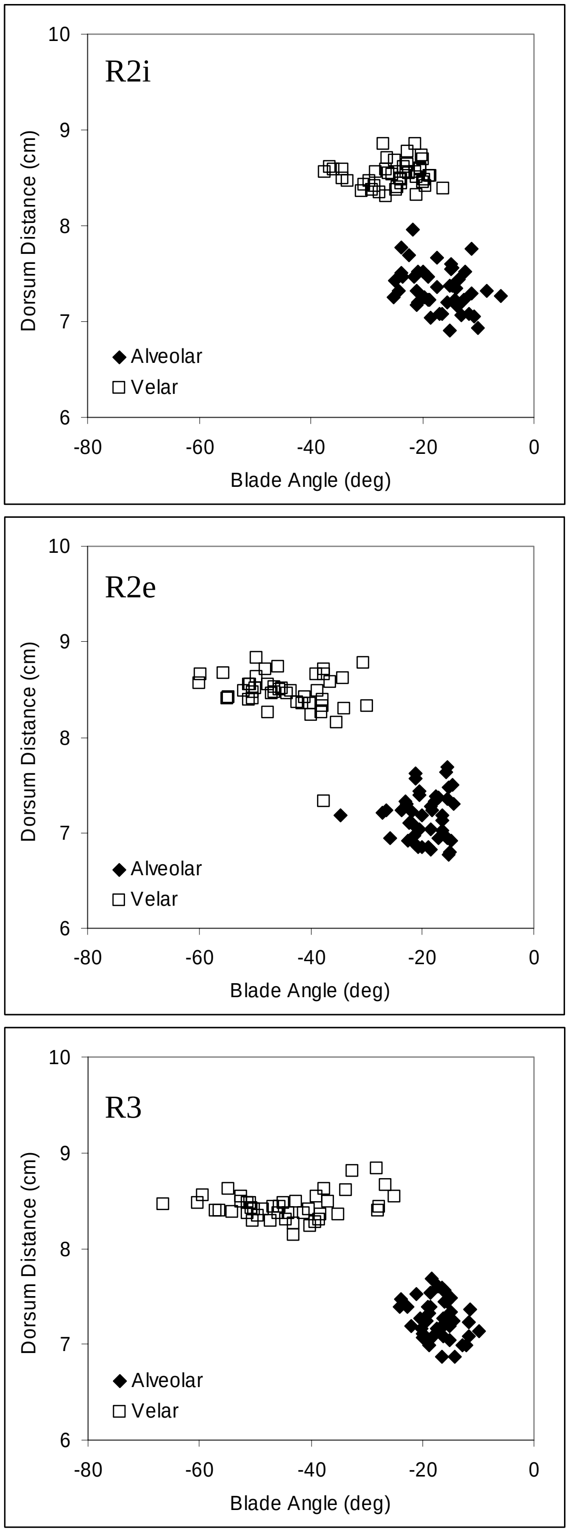


Figure 9: Scatterplot representation of alveolar versus velar articulations for measures by R2i, R2e, and R3

Chapter 4

Discussion

This thesis examined the reliability of hand measures of ultrasound analysis. The measurements of three raters were compared based upon their choice of video frame depicting closure, anterior and posterior tongue blade and dorsum closure points, blade and velar angle, and dorsum distance. The variable of experience was introduced in experiment three to assess the impact of experience on reliability.

Frame Choice. The raters' choice of video frame was fairly consistent throughout the three experiments. The raters chose the same frame depicting closure the majority of the time. The next highest percentage depicted one rater (R2) choosing a closure image that was either one frame behind or one frame ahead of the compared rater (R1 or R3). The closeness in frame choice indicates that the raters were measuring the same closure image consistently but sometimes at slightly different points during the closure. R2 demonstrated a tendency towards choosing a video image frame one frame later than the other raters comparatively.

Closure Points. In the Closure Angle Data experiment the raters' average choice of anterior closure point was closer than their average choice of posterior closure point for both speakers one and two. For both speakers, closure points were closer for the VCV non-word data than the CV or CVC real word data.

In the Tongue Twister Study, the raters' average anterior point was closer than their average posterior point chosen for the tongue blade measures. For the velar angle measures the opposite was true, however, the differences were minute. Overall, no large differences were seen between alveolar and velar measures, suggesting that measurements for these two phonemes are equally reliable.

In Tongue Twisters #2 experiment the raters' choice of anterior and posterior closure points for tongue blade measures became closer on average with experience. For velar measures posterior closure points became closer as anterior points remained consistent across the three comparisons (R2i, R2e, R3).

Velar Angle; Exp.1. A significant difference was seen in dorsum angle measures for speaker 1. However, despite the significant difference, the patterns of articulation were the same for both raters. Although slight variability was seen in the range of angles derived for each vowel, the average angle measurements were similar for both raters. For speaker two there were no significant differences in angle measurements. Patterns of articulation were also observed to be the same. The majority of the angle measurements for vowels are considered to be reproducible by Bland and Altman's measure of reproducibility. A few vowels fell just outside the range of reproducibility, however, these vowels appeared not to be reproducible in only one context. They were reproducible in other contexts presented. The contexts in which they appeared to be reproducible or not reproducible were inconsistent.

For experiments 2 and 3 the difference in velar angle measure was small between raters. In experiment 3 the difference in average velar angle measure became closer with experience. Correlation between velar angle measures also increased with experience.

Tongue Blade Angle Measures. The average tongue blade measure was very close between raters in experiment two. Tongue blade angle averages grew significantly closer in experiment three with experience. Difference in average angle measures decreased from a 9.5° to a 1.4° difference. Correlation increased from r = 0.58 to r = 0.86. The measure of tongue blade angle appears to be the measure most affected by the added variable of experience.

Dorsum Distance. The average dorsum distance obtained in experiments two and three revealed extremely small differences in average distance (0.1 mm to –0.1 mm). The dorsum distance measure appears to be very reliable for either a velar or alveolar production. Experience did not appear to impact dorsum distance measures obtained by the raters' as the measure was very reliable even for the inexperienced rater.

Summary. Overall, hand measures of ultrasound analysis do appear to be a reliable way to quantify the articulation of alveolar and velar stop consonants. Choice of frame to analyze and the measurement points for articulatory landmarks of the stops were reliable for three different raters, and data from four different talkers. While a larger study with more raters and more speakers would be desirable, these preliminary findings suggest that ultrasound analysis is a good method for studying lingual articulation.

References

Akgul, Y.S., Kambhamettu, C. & Stone M. (1998) Extraction and tracking of the tongue surface from ultrasound image sequences. *Computer Vision and Pattern Recognition*

Astheimer, J.P., Pilkington, W.C., Waag, R.C. (2004). "Reduction of variance in statistical estimates of spectra for use in correction of ultrasonic aberration." Paper presented at the 147th meeting of Acoustical Society of America IN *Journal of Acoustical Society of America 115 (5).*

Bernhardt, B., Gick, B., Bacsfalvi, P. & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics and Phonetics, 17 (3).*

Bernhardt, (power point presentation) "Evaluating ultrasound as a visual feedback tool in speech therapy" University of British Columbia.

Bland & Altman (2003). Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol, 22, (*85-93).

Frisch, S.A., Hardin, S.A., Nikjeh, D.A. & Stearns, A.M. (2005). The University of South Florida audiovisual phoneme database, v. 1.0. Journal of the Acoustical Society of America 117: 2574.

Frisch, S. A. (2003). Ultrasound study of errors in speech production. Proposal R03 DC006164, funded by NIH-NIDCD Small Grants Program.

Frisch, S. A. (2005). The phonetics of speech errors. Paper presented at the Third workshop on Ultrasound Applications in Linguistic Research, University of Arizona.

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *The Journal of International Phonetic Association, 32 (2),* 113-121.

Hutchinson, TP. (1997). Improving rater agreement studies. *AJR Am J Roentgenol, 168(5).*

Iova et al. (2004). Evaluation of the ventricular system in children using transcranial ultrasound: reference values for routine diagnostics. *Ultrasound in Medicine and Biology, 30 (6),* (745-751).

Lindblom, B. (1996). Role of articulation in speech perception: clues from production. *Acoustical society of America, 99(3).*

Lindblom et al. (2002). The trough effect: implications for motor speech programming. *Phonetica*, *59 (4).*

Lofqvist, A. & Gracco, V.L. (1994). Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context. *Phonetica, 51*, (52-67).

Lundberg & Stone (1999). Three-dimensional tongue surface reconstruction: practical considerations for ultrasound data. *Journal of Acoustical Society of America.106 (5*), (2858-2867).

Munhall, K.O. (2001). Functional imaging during speech production. *Acta Psychologica*, *107* (95-117).

Patton, N. & Aslam, T. (2005). Statistical analysis of agreement in measurement comparison studies. *Eye*, *19(3).*

Peng et al. (2000). Ultrasonographic measurement of tongue movement during swallowing. *Journal of Ultrasound Medicine*, *19 (1).*

Perrier et al. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *The Journal of the Acoustical Society of America*, *114 (3)* (1582-1599).

Pouplier, M. (2004) The role of coda consonants in triggering speech errors: An ultrasound study. *Journal of Acoustical Society of America*, 115 (5).

Rankin, G. & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clinical Rehabilitation, 12 (3).*

Stone, M. (1997). Laboratory techniques for investigating speech articulation. In W.J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 11-32). Cambridge, Massachusetts: Blackwell Publishers.

Sonies, B., Wang, C. & Sapper, D. (1996). Evaluation of normal and abnormal hyoid bone movement during swallowing by use of ultrasound duplex-doppler imaging. *Ultrasound in Medicine and Biology*, *22 (9).*

Valentin, L. & Bergelin, I. (2002). Intra- and interobserver reproducibility of ultrasound measurements of cervical length and width in the second and third trimesters of pregnancy. *Ultrasound Obstet Gynecol, 20,* (256-262).

Wodzinski, S. M. (2004). Ultrasound analysis of velar fronting. Unpublished Master's Thesis, University of South Florida, Tampa, FL.