

3-31-2006

Production and Perception of Place of Articulation Errors

Adrienne M. Stearns
University of South Florida

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

Scholar Commons Citation

Stearns, Adrienne M., "Production and Perception of Place of Articulation Errors" (2006). *Graduate Theses and Dissertations*.
<http://scholarcommons.usf.edu/etd/3891>

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.

Production and Perception of Place of Articulation Errors

by

Adrienne M. Stearns

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.
Catherine L. Rogers, Ph.D.
Jean C. Krause, Ph.D.

Date of Approval:
March 31, 2006

Keywords: ultrasound, articulatory gestures, slips of the tongue,
consonants, tongue twister

© Copyright 2006, Adrienne M. Stearns

Dedication

This thesis project is dedicated to Tom Sievers, for his unwavering support and constant encouragement. Without his love, steadfastness, and occasional capering, I would not have had the confidence or strength to finish this journey.

Acknowledgements

I would like to thank friends and family that have supported me and given me encouragement during this project. First of all, I would like to express how much I have enjoyed working with my thesis chair, Dr. Stefan Frisch. His support and patience helped me complete this thesis, which I am very proud of, and provided me with an opportunity to learn more about ultrasound research and contribute to the field. Dr. Frisch was always dependable and flexible, and I benefited greatly from having him as a mentor.

I would like to thank my committee member Dr. Catherine Rogers for showing me how interesting speech science is. I would also like to thank my committee member Dr. Jean Krause for her encouragement and support during this project. Thanks to both of you for your insightful comments and for your contributions to a lively and entertaining defense.

I would like to thank my friends who have tried to keep me on an even keel during the last two years. Danielle Melowic and Shannon Weekley have always cheered me on while providing a good laugh when it was needed. It has been a joy to work with Sylvie Wodzinski in the lab and I have come to count her as a true friend.

Finally, I would like to thank my family for their ever-present encouragement. My parents, Robert and Freda Stearns, and my aunt and uncle, Pat and Paul Stearns, always told me that I could do anything I wanted to. Their support bolstered my determination to complete this project.

TABLE OF CONTENTS

List of Tables	iii
List of Figures.....	iv
Abstract.....	vi
Chapter One: Introduction	1
Patterns Observed in Research on Naturally Occurring Speech Errors.....	2
Patterns Observed in Research on Experimentally Induced Speech Errors	6
Speech Error Assessment Utilizing Instrumentation	8
Ultrasound and Speech Error Research	10
Purpose.....	13
Chapter Two: Ultrasound Error Elicitation.....	14
Stimuli.....	14
Participants.....	15
Procedure.	16
Measurement.....	17
Results.....	20
P1	25
P2	30
P3	34
P4	38
Discussion.....	41
Chapter Three: Perception of Place of Articulation Errors.....	44
Stimuli.....	44

Procedure.....	45
Participants.....	45
Results.....	45
Identification.....	46
Reaction Times.....	49
Discussion.....	54
Chapter Four: Conclusions/Discussion.....	55
Conclusions.....	55
Discussion.....	56
Clinical Uses.....	58
References.....	61
Appendix A: Perception Results for Each Stimulus.....	64

LIST OF TABLES

Table 1	Tongue twisters read by participants	15
Table 2	Stimulus classification system.....	24
Table 3	Category classification of P1's productions	27
Table 4	Category classification of P2's productions	32
Table 5	Category classification of P3's productions	35
Table 6	Category classification of P4's productions	39
Table 7	Number of stimulus presentations by category and percentage of error selection by listeners.....	47
Table 8	Reaction times and z-scores for perception of alveolar production targets.....	51
Table 9	Reaction times and z-scores for perception of velar production targets.....	51

LIST OF FIGURES

Figure 1	Sagittal view of the tongue body with ultrasound; tongue tip to the right.....	12
Figure 2	Cushion-scanning technique within head stabilizing frame	17
Figure 3	Measurements applied to ultrasound image of the tongue	18
Figure 4	Ultrasound images of alveolar (top) and velar (bottom) productions with mean and 2 SD ranges depicted for tongue dorsum and tongue blade.....	22
Figure 5	Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for P1	26
Figure 6	Articulations for P1 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance	29
Figure 7	Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for P2	31
Figure 8	Articulations for P2 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance	33
Figure 9	Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for P3	34
Figure 10	Articulations for P3 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance	37
Figure 11	Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for P4	38
Figure 12	Articulations for P4 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance	40
Figure 13	Distributions of perceptual judgements.....	49

Figure 14 Reaction time to stimuli on the basis of production target and 3-way error categorization.....53

Production and Perception of Place of Articulation Errors

Adrienne M. Stearns

ABSTRACT

Speech errors have been utilized since the beginning of the last century to learn more about how speech is produced, both physically and cognitively. Collection of speech errors has progressed from writing down naturally occurring speech errors to recording experimentally induced speech errors to current studies, which are using instrumentation to record acoustic and kinematic information about experimentally induced speech errors. One type of instrumentation being used in articulatory research is ultrasound. Ultrasound is gaining popularity for use by those interested in learning how speech is physically produced because of its portability and noninvasiveness. Ultrasound of the tongue during speech provides visual access to the articulatory movements of the tongue.

This study utilizes ultrasound recordings of speech errors in two ways. In Experiment 1, ultrasound images of participants' tongues were recorded while they read tongue twisters designed to elicit speech errors. The tongue twisters were CVC words or CV syllables with onset velar or alveolar stops. Within the ultrasound video, the angle of the tongue blade and elevation of the tongue dorsum were measured during the onset stop closure. Measurements of tongue twisters were compared to baseline production

measures to examine the ways in which erroneous productions differ from normal productions. It was found that an error could create normal productions of the other category (i.e., categorical errors) or abnormal productions that fell outside the normal categories (i.e., gradient errors).

Consonant productions extracted from ultrasound video were presented auditory only to naïve listeners in Experiment 2. Listeners heard a variety of normal, gradient error, and categorical error productions. Participants were asked to judge what they heard as the onset sound. Overwhelmingly, the participants heard normal productions as well as gradient error productions as the target sound. Categorical error productions were judged to be different from the target (e.g., velar for alveolar). The only effect of erroneous production appears to be a slight increase in reaction time to respond with a choice of percept, which may suggest that error tokens are abnormal in some way not measured in this study.

CHAPTER ONE: INTRODUCTION

Speech errors have been reported by many people in many ways. Sigmund Freud in 1901 termed speech errors as “faulty actions.” The colloquial label for Freud’s faulty actions is “Freudian slip.” Freud attributed the speech error to an error in human action, speech, or memory caused by the unconscious mind. These errors often appear to the observer as being bizarre and nonsensical. He concluded that articulation errors reveal a “hidden” cognitive process of which even the speaker is unaware. Reverend William Archibald Spooner was reported to regularly make speech errors by transposing the first consonant or consonant cluster of words. Very few speech errors were verifiably recorded as being produced by him, and most of those have been determined to be purposeful, such as “You’ve missed my history lectures → You’ve hissed my mystery lectures” (MacKay, 1970). However, the term “spoonerism” became popular in describing accidental speech errors indicative of a faulty phonological process where the initial consonants of words are transposed.

Speech errors provide insight into the speech production process by bringing to light the range of possibilities that exist for articulatory actions. Research into slips of the tongue has concluded that language processing makes use of phonological constituents as units (e.g., features, segments, onsets, rimes). Speech errors can involve the rearrangement of a single segment, syllable onset, syllable rime, or whole syllable,

suggesting that each of these constituents exist as a unit in language structure and processing (Fromkin, 1971). At the same time, speech errors are more likely to occur between segments that occupy the same position within the word, syllable, and prosodic structure, suggesting that speech is generated by using word “templates” into which segments are fitted (Shattuck-Hufnagel, 1992). Stemberger and Trieman (1986) and Dell (1990) found that frequently occurring words and segments are more protected from speech errors, which they explain using an activation/competition model of speech production. By learning more about how speech errors are produced, and perceived, more can be learned about the processes involved in production and perception of normal speech.

Patterns Observed in Research on Naturally Occurring Speech Errors

Fromkin gathered data on naturally occurring speech errors by logging errors noticed while in conversation with others or described by others to her (Fromkin, 1971). She analyzed errors linguistically and found that speech errors follow a pattern of rules. She concluded that most speech errors occur at the level of the phonological segment and that only those segments that are near to each other and in parallel structural positions (e.g., word final and word final or word initial and word initial) are involved in errors. In other words, prosodic constituents of words (e.g., onset, rime, final consonant) are involved in errors rather than random selections of phonemes throughout the word(s) in error. Grammatically non-permissible phones or phonetic sequences did not occur in the errors in her data. It was also determined that errors may be semantic in nature and that

the similarity of the phonological form of words appears to play a role in word substitutions.

Garrett (1975, 1976) and Shattuck (1975) collected a corpus of approximately 3,400 speech errors that occurred in spontaneous speech (MIT corpus). The errors collected were either heard directly by Garrett or Shattuck, or reported by friends. Garrett stated that the errors in his corpus revealed four generalizations: (1) interacting elements are similar to each other, (2) interacting environments are also similar, (3) phonological well-formedness is preserved, and (4) speech segments in an error interact with both word and phrasal stress. In the phoneme errors in Garrett's corpus, consonants and vowels did not substitute for each other in the creation of an error. That is, a consonant exchanged with other consonants and a vowel exchanged with other vowels, but consonants did not exchange for vowels or vice versa. Garrett stated that two nonphonological processing levels for word forms exist that affect the production of speech errors. One is the "real time" constraint, which is affected by adjacency of elements and the other is a "functional" constraint, which is sensitive to the structural placement of the target word in the sentence. Therefore, generation of a speech error is constrained not only by similar elements in similar environments, but also by syntactic well-formedness.

Shattuck-Hufnagel and Klatt (1979) concluded that most phoneme errors occur as a result of mis-selection between two similar phonemes competing for a single prosodic location in an utterance. They studied an expanded version of the MIT corpus containing nearly 6,000 errors. They found that phoneme substitutions and exchanges accounted for approximately 35% of the total errors. The remaining errors consisted either of single

phoneme errors that lacked an identifiable target segment or errors involving longer sequences of segments, morphemes, words, or phrases. Shattuck-Hufnagel and Klatt state that the planning process used to generate speech follows three steps: (1) planning segments or phonemes, (2) sequencing structurally-defined slots, and (3) the integration mechanism, which consists of a “scan-copier to insert segments into the slots, a bookkeeper to check off or delete segments that have been copied, and an output error monitor.” Malfunctions in any part of this process could account for errors in selecting the incorrect segment, incorrect placement of the segment, or overuse or underuse of the segment if the integration mechanism fails.

Dell and Reich (1981) endeavored to methodologically create a corpus of naturally occurring slips of the tongue. They instructed students who had not yet taken coursework related to theories of speech errors to collect examples of slips of the tongue heard personally during a 1-month timeframe. The students were instructed not to record speech errors that they were unsure of hearing correctly or that they had not heard themselves. Over the course of two years, five data collection periods were initiated, resulting in a corpus of approximately 4,000 English examples. The analysis of the corpus was restricted to word level errors and sound errors that involved only the initial consonant due to the fact that these types of errors are easier to detect and therefore, more likely to be accurately collected by the students. In the analysis of initial consonant errors, they found a large percentage of speech errors created real words instead of nonsense items. Using two-word pairs, anticipation, perseveration, and transposition errors were analyzed. Anticipation errors are cases where a segment is produced before its intended position. In anticipation errors, they found that the erroneous word, which

occurs first, was more likely to be a real word than a nonsense word. Perseveration errors are cases where a segment is produced again after being produced in its intended position. In perseveration errors, they also found that the erroneous word, which occurs second, was more likely to be a real word than a nonsense word. Transposition errors are cases where two segments exchange position. In transposition errors, they found that the first word was more likely to be a real word than nonsense, but the second word was not. This suggests that the production of the first word in the transposition is the actual error, and the erroneous production of the second word occurs because of the remaining unused intended phoneme from the production of the first word. Given the strong tendency for error productions to create real words rather than nonsense, Dell and Reich concluded that the phonological forms in the mental lexicon influence the outcome of a phonological error.

Stemberger and Treiman (1986) found that the second consonant (C2) in word-initial consonant clusters is more likely to be produced in error than the first consonant (C1) of the cluster. Their experiments show that C2 is more often deleted than C1 during a speech error. They also demonstrate that when a consonant is added to a word during a speech error, a C2 is added more often than a C1. As a result, they determined that C1 and C2 are distinct positions within the template of the syllable onset in single syllable words. They propose that C1 positions are used for both initial singletons and the initial consonant in a cluster whereas C2 positions are used only in clusters. Stemberger and Treiman state that activation discrepancies between C1 and C2 during speech production are the cause of the difference in error frequencies. C1, being the first sound in the word, activates many other words beginning with the same letter when it is heard by the

listener. On the other hand, C2 only serves to narrow the field and is not as important in distinguishing the word from other options. The extensive use of C1 positions across a larger variety of words also leads to greater activation levels in the lexicon for C1. This increased activation level protects C1 from errors relative to C2.

Patterns Observed in Research on Experimentally Induced Speech Errors

Baars, Motley, and MacKay (1975) developed experimental procedures to induce speech errors in their subjects rather than relying on naturally occurring errors. The authors sought to create a functional separation of the input of the target from the articulation of the corresponding error outcome. In this way, they were able to generate spoonerism-type errors in their subjects. By presenting lexical and nonsense word pairs to naïve listeners, the authors sought to discover if lexical or nonsense spoonerisms were predominant. The results of their experiments demonstrated that spoonerisms generally occur in favor of creating lexical outcomes, suggesting that subjects edit their output before producing it. However, if the subject is primed to read nonsense words, even if the spoonerism would create lexical outcomes, the words are interpreted as nonsense, and the subject does not edit their output to create lexical items.

It has also been determined that speech errors generally occur in content words, rather than function words. Dell (1990) concluded that this is a result of usage frequency, which is related to increased activation. Function words are used far more frequently than content words in speech. Dell proposes that extreme familiarity with function words as a result of their high frequency rather than their status in a closed class is why they are rarely involved in speech errors. He also addresses the concept that function words have

a special status in language production. He states that function words can only be considered “special” if they are treated as whole units and not as segments that fit within a frame. Functional words may be stored as whole units (i.e., segments already associated with a frame) due the high frequency with which they are used in speech. Therefore, it is far less likely that function words will be produced incorrectly since they are automatically inserted into place as whole units in utterances.

Shattuck-Hufnagel (1992) determined that sounds that appear in the same prosodic position in the word are more likely to interact in speech errors as a result of their similar placement. Her study examined the role of shared word onset position and shared stress in influencing error rates. She concluded that pairs of word-onset consonants interact in segmental speech errors more often than pairs of consonants in two different word positions. Also, more errors occur between pairs of consonants that share stress than between pairs that do not share stress. Therefore, word onset position must be considered one of the prosodic constituents of phonological encoding for speech production, distinct from a syllable onset position.

Meijer (1997) conducted three experiments to determine the interaction of syllable position and stress in a word and the production of speech errors. When matching for word syllable and stress, speech errors between initial consonants were produced twice as often as speech errors between word onsets and word-internal segments, confirming previous research that word onset is especially vulnerable to speech errors. This finding was further upheld by results from his second experiment in which word-internal structures were targeted versus word-initial structures. Meijer presented a sentence containing two disyllabic target words and two interfering words. Two

conditions were presented: (1) where a word internal phoneme could interact to create a lexical item (e.g., *nephew* + *copy* = *coffee*) and (2) where a word onset phoneme could interact with a word internal phoneme to create a lexical item (e.g., *feather* + *copy* = *coffee*). Even when word onsets were interacting with word-internal phonemes, it was the word-onset that displaced word-internal phonemes rather than the other way around, which shows that word onsets are particularly volatile in creating speech errors. Finally, the idea that vowels cannot interact with consonants in speech errors was shown to be a problem of opportunity rather than phonotactic constraints. In a speech error where substituting a vowel instead of a consonant would produce a word, interactions did occur and their frequency of occurrence seemed comparable with the more common interactions between consonants (e.g., *candy* → *candle*).

Speech Error Assessment Utilizing Instrumentation

All previous data discussed were obtained using transcription of speech errors either in a laboratory setting or as they occurred by chance in normal conversation. However, by relying on auditory perception translated into individual segments by transcription, bias may have been introduced into the data collection process. IPA is, by definition, a categorical system of transcribing individual phoneme sounds. This system does not provide a straightforward means to encode half-productions or undershooting of the target phoneme, nor does it allow for articulatory combinations that occur when a speaker mistakenly attempts to produce two phonemes at the same time. In addition, the ability of listeners to accurately perceive these types of productions is unclear. As a result of lack of confidence in the reliability of transcribed speech errors to accurately

reflect the speech production phenomena, interest developed in technological approaches to recording and analyzing speech errors.

Mowrey and MacKay (1990) conducted one of the first experiments to use articulatory instrumentation to collect data on speech errors. Their experiment used electromyographical (EMG) data to examine muscle fiber activity in speech errors. Their results show that errors involving abnormal muscular activation do occur. These errors occur at a gestural level that is smaller than the phoneme unit and they found these errors were not detectable in their own perception. Since the errors are not detectable without instrumentation, the results cast doubt on the conclusions drawn from transcribed speech errors that rely solely on human perception.

Speech errors can be produced by undershooting the target articulation or by overshooting the target articulation (Boucher, 1994). Boucher examined x-rays of articulators in motion while speakers uttered phrases. Auditorally, the speech errors sounded distinctly like a completely different phoneme than the target. However, the x-ray data show that the errors were a result of either not articulating enough or moving the articulators too far in an attempt to produce the target phoneme. He concluded that restricting the experimenter's encoding of speech errors to an alphabetic transcription may obscure important aspects of the nature of the speech error itself.

Acoustic analysis of speech errors involving /s/ and /z/ by Frisch and Wright (2002) concluded that errors of devoicing /z/ occur acoustically/articulatorily more often than errors of voicing /s/. However, auditory perception appears to be more sensitive to voicing errors of /s/, which then is comprehended as /z/. Therefore, in a transcription of speech errors, it would appear that /s/ is in error far more often than /z/, even though the

opposite is actually the case. Frisch and Wright also reported speech errors that were recorded that were not phonologically well-formed, which empirically contradicts previous claims that all errors occur within phonologically well-formed segments. They concluded that their findings were incompatible with models of speech production that involve selecting, organizing, and integrating phonetic segmental units, as suggested by Shattuck-Hufnagel (1979).

Pouplier (2003) used an electromagnetic midsagittal articulometer (EMMA) to investigate speech errors. Participants had magnetic coils that received electromagnetic signals attached to specific points along the midline of their tongues. The signals from the coils were then tracked to determine the movement of the tongue during articulation. She found that erroneous gestural movements are produced at different rates. She also found that it is more common for gestures from one phoneme to intrude upon another rather than for the gestures of the target phoneme to be reduced. Pouplier presented auditory stimuli gathered from the EMMA study to participants to obtain their perceptual judgements of the initial consonants. Contrasting /t/ and /k/ in a perceptual experiment, intruding gestures of the tongue dorsum during production of /t/ has a systematic perceptual effect (i.e., hearing /k/) whereas tongue tip intruding gestures during production of /k/ did not significantly affect identification.

Ultrasound and Speech Error Research

Ultrasound imaging has been gaining popularity as a method to study articulatory gestures during speech production. Ultrasound studies completed in the 1960s and 1970s were completed using large, cumbersome equipment that only created 1-dimensional

images (Gick, 2002). Since that time, technology has progressed and current ultrasound workstations have the capability to produce 2-dimensional images with far greater temporal resolution using real-time B-scan (Stone, 2005). Real-time B-scan is a mode wherein individual 1-dimensional images are combined to reconstruct a 2-dimensional grayscale image in the shape of a 90-120 degree wedge. In this mode, modern ultrasound machines can obtain 30 scans per second or more, which is considered fast enough to accurately capture most speech movements.

Ultrasound images are produced by placing a transducer, or probe, against the subject's skin while the transducer emits a sound wave between 2 and 10 MHz. This sound wave travels through soft tissue (like the tongue) and is reflected by air back to the transducer (Figure 1). The image that is created reveals where soft tissue stops and air begins. For instance, when the probe is placed under the chin, the sound waves travel through the tongue but are reflected back from the top of the tongue where it meets air in the oral cavity. This reflection is shown as a white "line" following the curve of the tongue and is used as a landmark in measurement of ultrasound images. The upper surface of the tongue is the black edge at the underside of the white line where the air reflection begins.

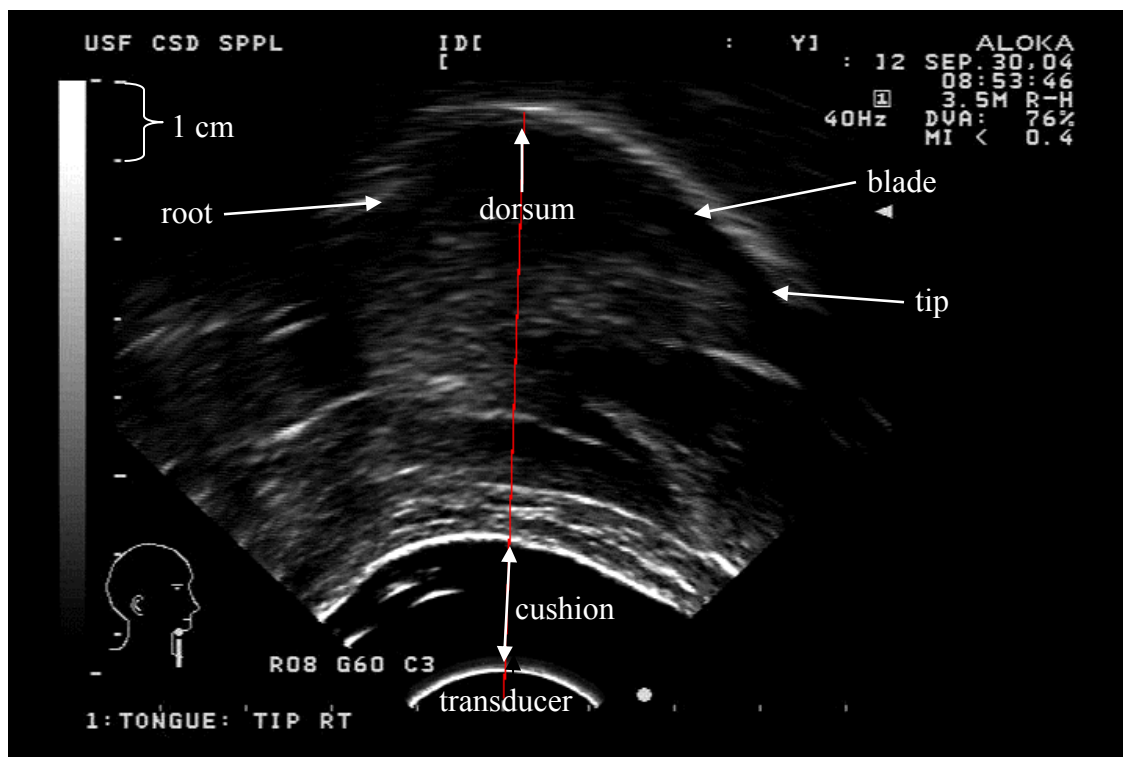


Figure 1 Sagittal view of the tongue body with ultrasound; tongue tip to the right

Due to the placement of the ultrasound probe under the speaker's chin, shadows are cast in the image by the hyoid bone, mandible, and sublingual cavity (Gick, 2002). The shadows have the potential to obscure the tongue tip and tongue root, depending on the placement of the probe and the subject's anatomy. Securing the probe in a fixed position relative to the participant's chin improves reliability of the ultrasound images produced. To further improve the reliability of the ultrasound probe positioning in relation to the articulators, a cushion placed over the probe allows for jaw movements during speaking without dislocating the probe (Peng, Jost-Brinkmann, & Miethke, 1996).

Purpose

Clearly, differences exist between instrumental speech error data and perceptual speech error data. Perceptual speech error data recorded by transcription have been used to conclude that all errors are phonologically well-formed and that errors are categorical in nature. Instrumental speech error results demonstrate that errors can violate phonotactic constraints and are not always categorical in nature (Mowrey & MacKay, 1990; Pouplier & Goldstein, 2005). The distinction between how errors are articulated and how they are perceived is explored in this thesis in two experiments. Experiment 1 examines the articulation of tongue twisters using ultrasound imaging of the tongue. Experiment 2 uses productions from Experiment 1 in a perception task where naïve listeners are asked to identify productions with a variety of articulatory characteristics.

CHAPTER TWO: ULTRASOUND ERROR ELICITATION

Experiment 1 used an ultrasound workstation to record images of articulators during spoken tongue twisters. To reduce demands on short-term memory, the participants were provided a written list of tongue twisters to read. The tongue twisters contrasted alveolars /t, d/ with velars /k, g/ in the initial position of words and nonwords. Based upon research that has shown that repetition of tongue twisters tends to result in speech errors, the participants repeated each tongue twister six times (Shattuck-Hufnagel, 1992). To create more “natural” phonological speech errors, the speakers were instructed to speak at a normal rate to reduce the articulatory/motor programming load (Wilshire, 1999). It was expected that the speakers would produce both “categorical errors” where a normal phoneme from the wrong category is produced, as well as “gradient errors” that involve an abnormal or incomplete articulatory gesture.

Stimuli

The tongue twisters were comprised of 22 phrases containing four words or nonsense syllables each (Table 1). In the first eight tongue twisters, all four CV syllables in each twister began with the same phoneme (/t/, /d/, /k/, or /g/) and were used to obtain baseline measures of articulation for each speaker. The second set of six tongue twisters consisted of CVC words wherein the word initial phonemes alternated in place of

articulation only (i.e., /t/ and /k/ or /d/ and /g/) and contrasted only between tongue tip and tongue dorsum raising gestures. The CVC words contained the target consonants as onsets, low vowels (/æ/ and /ɑ/), and labial final consonants. The final eight tongue twisters were CV syllables (nonwords) with the same alternating initial phonemes and the same low vowels. Low vowels were used in all the tongue twisters to reduce interference during articulation of stop consonants. Labial coda consonants were used in the CVCs to avoid introducing additional tongue gestures that might obscure clear observation of the target initial consonants (Wodzinski, 2004; Pouplier & Goldstein, 2005). Each tongue twister or tongue twister-like utterance was repeated six times, following the speech error elicitation procedure of Shattuck-Hufnagel (1992).

Table 1 Tongue twisters read by participants

Baseline data	Words	Nonsense syllables
/tɑ tæ tæ tɑ/	top cap cop tab	/tɑ kæ kɑ tæ/
/gæ gɑ gɑ gæ/	tap cab cob top	/gæ dɑ dæ gɑ/
/kɑ kæ kæ kɑ/	dam gob gap damp	/kɑ tæ tɑ kæ/
/dæ dɑ dɑ dæ/	cap top tab cop	/dæ gɑ gæ dɑ/
/kæ kɑ kɑ kæ/	cab tap top cob	/kæ tɑ tæ kɑ/
/gɑ gæ gæ gɑ/	gob dam damp gap	/dɑ gæ gɑ dæ/
/tæ tɑ tɑ tæ/		/tæ kɑ kæ tɑ/
/dɑ dæ dæ dɑ/		/gɑ dæ dɑ gæ/

Participants. Four (4) volunteers from an undergraduate speech and hearing sciences course participated in the experiment. They received extra credit for their participation. Each participant had completed a phonetics course to ensure that they could read stimuli written in International Phonetic Alphabet (IPA) font. All were

females between 19 and 45 years of age. All participants were monolingual native speakers of American English who reported no past speech or hearing disorders.

Procedure. Lingual articulation was imaged using an Aloka SSD-1000 ultrasound machine with a 3.5 MHz convex probe. The participant was seated in a wheel-less straight-backed chair with her back against the back of the chair and feet flat on the floor. The probe was held stationary using the cushion-scanning technique (Peng, et al., 1995) and was placed along the inferior midline of the participant's mandible pointing upward toward the tongue (Figure 2). The participant sat with her head protruded forward slightly to provide the best image of the tongue body. The participant's head was stabilized using a frame with contact points at the forehead, left and right sides of the head, and the back of the head, similar to the HATS system (Stone & Davis, 1995). Due to the interconnectedness of the support structures of the stabilization system, any head movement caused the ultrasound probe to move concurrently, maintaining a fixed relationship between the head and the probe.



Figure 2 Cushion-scanning technique within head stabilizing frame

Measurement

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 (following Wodzinski, 2004).

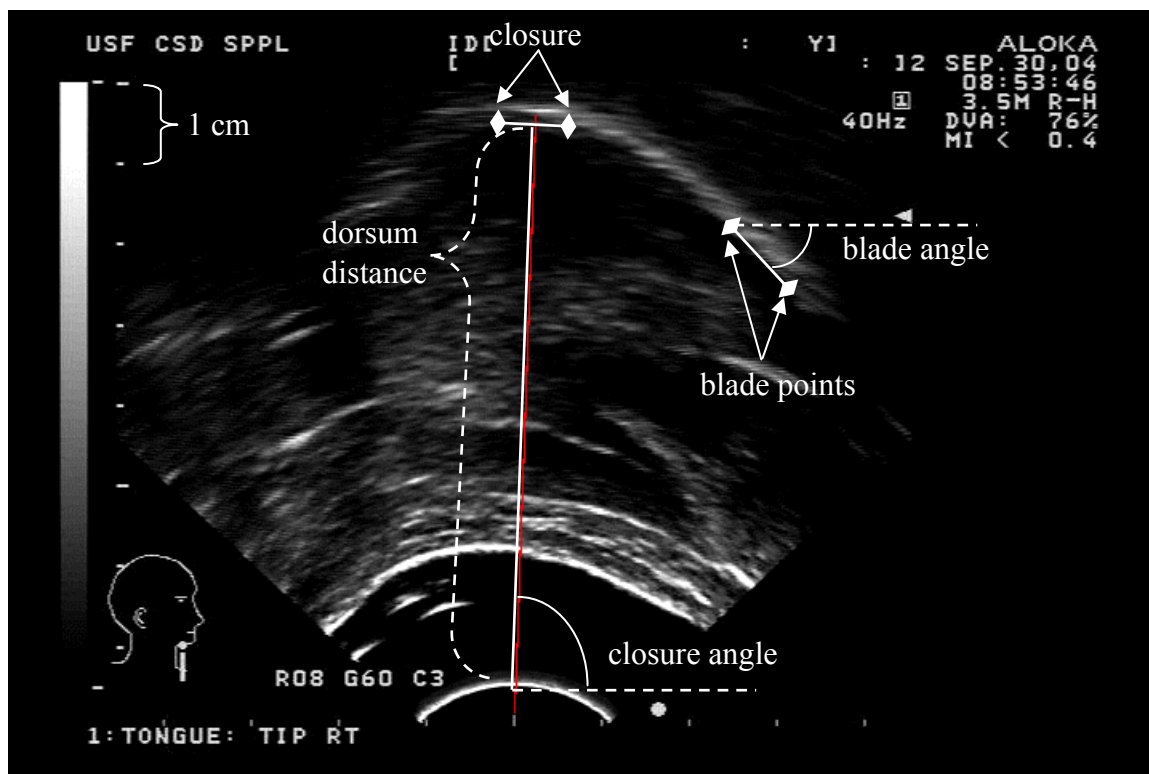


Figure 3 Measurements applied to ultrasound image of the tongue

Closure location for velar productions was measured by identifying the most anterior and posterior points of velar closure. Closure location was then quantified by computing the angle from the horizontal plane at the base of the ultrasound image to the midpoint between the anterior and posterior closure points. The closure angle for velar productions from baseline twisters was averaged by vowel (i.e., /ka/ and /ga/ vs. /kæ/ and /gæ/). Since /æ/ is a front vowel and /a/ is a back vowel, the velar closure position changed accordingly due to coarticulation. The average closure angle provided data about the general direction of closure that was typical for the participant for that vowel. The tongue dorsum elevation was measured along the average closure angle for all

productions (velar and alveolar), depending upon which vowel was contained in the utterance.

Tongue blade angle was measured for both alveolar and velar productions. In velar productions, the tongue blade declines sharply away from the palatal closure location (Figure 3). 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 angle of elevation of the tongue blade is used to assess the alveolar closure gesture. The angle is computed from two points that are measured by the angle of declination from a horizontal. The first point chosen was the most anterior portion of the visible tongue blade/tip, 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.

Studies of the reliability of measurements from ultrasound have found that measures are reliable (Stone, Sonies, Shawler, Weiss, & Nadel, 1983). However, Stone's study was based on measurement procedures that were somewhat different from those used in this study. Reliability of ultrasound measures used in the University of South Florida Speech Perception and Production Lab was investigated by Hardin (2005). 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 impact of experience measuring was also assessed by comparing measurements from one rater who progressed from inexperienced to experienced over the course of her research. Hardin confirmed that overall reliability

of ultrasound measures was acceptable, and the measures from different raters led to similar results.

Results

Using comparisons with the baseline measurements from each participant, regular (or normal) articulations were differentiated from irregular (or abnormal) articulations within the target items (tongue twisters) based on the distribution of tongue blade and tongue dorsum measures. The mean and standard deviation (SD) of measures from the baseline stimuli were calculated to provide reference values for stimulus classification. Productions were categorized by determining whether a measurement fell within or outside of the range of 2 SD from the mean of the baseline measurements for tongue blade angle and tongue dorsum elevation for both velar and alveolar target sounds. Stimuli were classified to reflect how individual measurements from the stimulus were related to the baseline measures. Words or syllables that contained articulations determined to be within 2 SD of the mean of the baseline measures were categorized as normal articulations (following Pouplier, 2003).

To determine if an articulation error was gradient or categorical, it was compared not only to the mean for its own category but also to the mean for the other category (Figure 4). For instance, if a production of /t/ or /d/ was articulated with the tongue blade far lower than the mean for that sound (i.e., more than 2 SD below the mean for alveolars), it was also compared to see if the tongue blade was so low as to be within 2 SD of the mean of the velar category. Likewise, if a velar consonant articulation was determined to have a much higher tongue blade than normal, then it was compared to the

alveolar category to determine if it fell within 2 SD of the mean for alveolar productions. If the measurement fell outside of 2 SD from the mean for its own category, but not within 2 SD of the other category, then it was considered to be a gradient error because it was produced in the space between the two categories. However, if a measurement fell outside of 2 SD from the mean of its own category and within 2 SD of the mean of the other category, that production was considered to be a categorical error for the tongue blade. The same process was used to determine error type for tongue dorsum measurements in alveolar and velar productions.

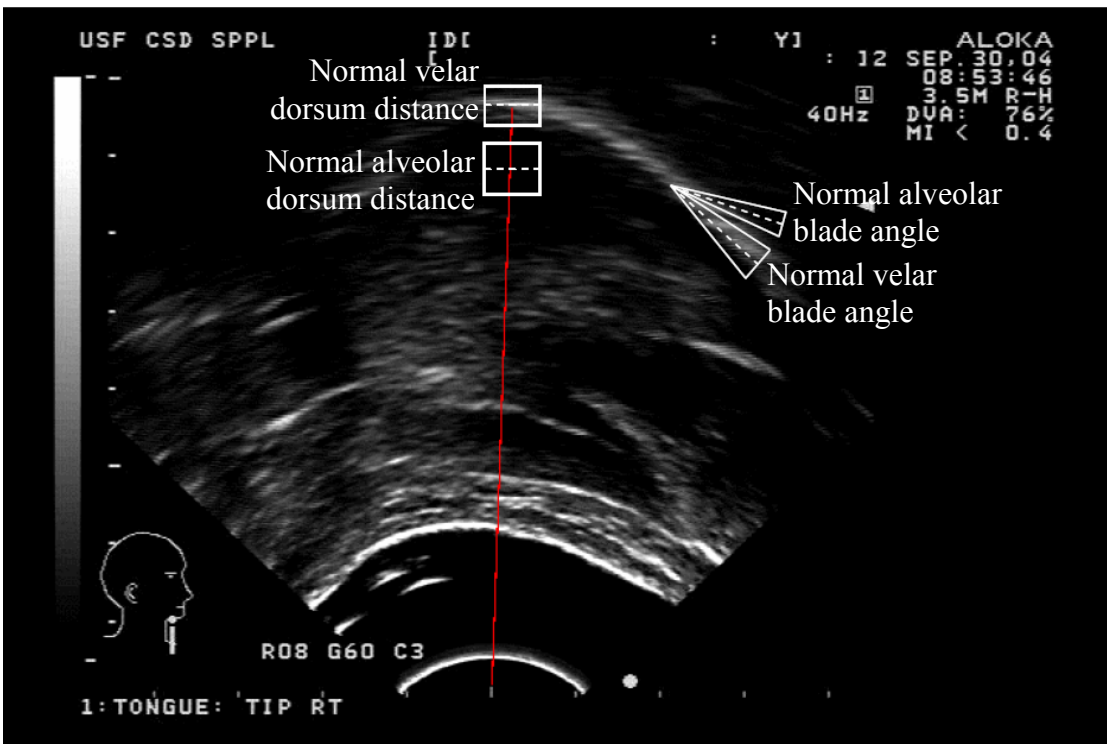
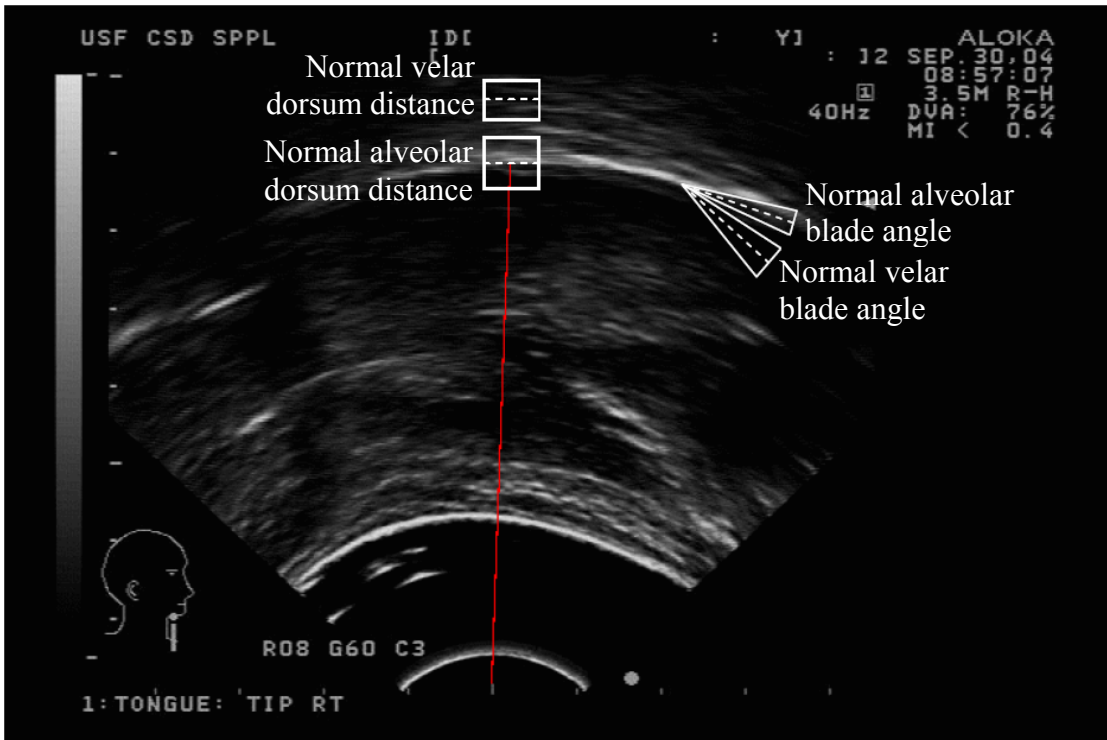


Figure 4 Ultrasound images of alveolar (top) and velar (bottom) productions with mean and 2 SD ranges depicted for tongue dorsum and tongue blade

Classification codes were assigned to articulatory measures to reflect the position of the tongue blade and tongue dorsum in relation to normal articulation, gradient error, or categorical error (Table 2). “B” was used to designate the tongue blade and “D” was used to designate the tongue dorsum. “N” was used to classify a normal articulation, “G” was used to classify a gradient error, and “C” was used to denote a categorical error. Therefore, “GBCD” refers to a stimulus that was measured to have a gradient error in the tongue blade angle and a categorical error in the tongue dorsum elevation. The measurements for each participant were categorized to determine the number of “normal”, “gradient error”, and “categorical error” productions.

Table 2 Stimulus classification system

Classification	Definition
NBND	Normal Articulation <ul style="list-style-type: none"> • Within 2 SD of baseline measurements for target category
NBGD	Gradient error in tongue dorsum distance <ul style="list-style-type: none"> • Dorsum distance outside mean + 2 SD ranges for both categories
NBCD	Categorical error in dorsum distance <ul style="list-style-type: none"> • Dorsum distance outside 2 SD of baseline measures for target category • Dorsum distance is within 2 SD of baseline measures for other category
GBND	Gradient error in blade angle <ul style="list-style-type: none"> • Blade angle outside mean + 2 SD ranges for both categories
CBND	Categorical error in blade angle <ul style="list-style-type: none"> • Blade angle outside 2 SD of baseline measures for category • Blade angle is within 2 SD of baseline measure for other category
GBGD	Gradient error in both blade angle and dorsum distance <ul style="list-style-type: none"> • Measures for both blade angle and dorsum distance fall outside mean + 2 SD for target category and other category
GBCD	Categorical error in dorsum distance Gradient error in blade angle <ul style="list-style-type: none"> • Dorsum distance outside of 2 SD of target category and within 2 SD of other category • Blade angle outside mean + 2 SD ranges for both categories
CBGD	Categorical error in blade angle Gradient error in dorsum distance <ul style="list-style-type: none"> • Blade angle outside of 2 SD of target category and within 2 SD of other category • Dorsum distance outside mean + 2 SD ranges for both categories
CBCD	Categorical error in blade angle and dorsum distance <ul style="list-style-type: none"> • Blade angle and dorsum distance within 2 SD of other category

Each participant was measured independently. Due to differences in head size and relative placement of the probe from articulators between participants, the measures for one participant are not directly comparable to another. However, the overall patterns of normal productions are quite similar, while the error data present a continuity of error values from virtually error free to error ridden. Data were analyzed using a within-

subjects approach, using each participant's baseline measures to examine the productions from the word and syllable tongue twisters for that participant.

P1. Figure 5 shows the range of production results for P1 and provides a visual comparison of their location in articulatory space. Within both the blade angle and dorsum distance panels, the productions are divided between alveolar and velar, which are subdivided by vowel and then again by baseline and alternating twisters. The boxes represent the 2 SD surrounding the mean measurements for blade angle in degrees and dorsum distance in centimeters. The extent to which productions were measured to be outside of the 2 SD ranges, either above or below, are marked by the range lines extending above and/or below the boxes. Where range lines are not present either above or below a 2 SD box, no production was measured to be outside of 2 SD for that target. The means are located in the center of the 2 SD boxes. The means and 2 SD ranges were calculated for blade angle and dorsum distance measures for both baseline productions and alternating twisters by onset consonant and vowel. The mean blade angles and dorsum distances for velar productions appear to be slightly lower for the alternating twisters than for the baseline measures for P1. While the 2 SD boxes are roughly the same size, overall they are marginally lower for the twisters. The mean and range for the alveolars appear similar in all contexts, indicating that P1's productions were measured to occur in roughly the same place in articulatory space.

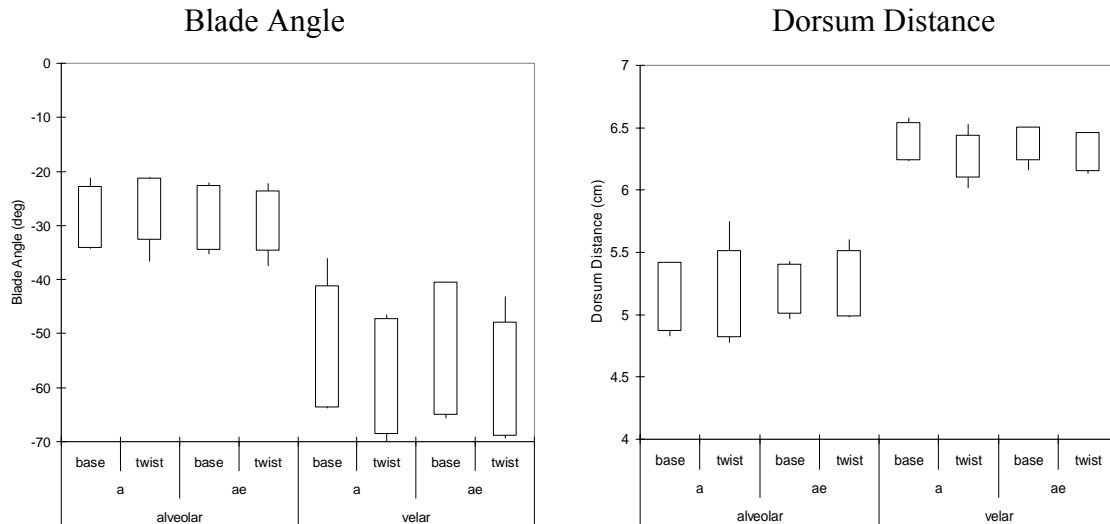


Figure 5 Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for P1

Table 3 shows the categorization of P1's 536 measurable productions. Table 3 shows that, while P1 had only two categories of errors (GBND and NBGD), they accounted for 13% of her total productions. Eight of the total 71 productions classified as gradient errors in either blade or dorsum gestures were produced during baseline measures, which is consistent with other participants. The majority of her error gestures consisted of NBGD in the nonword and word twisters where a velar target was produced with a slightly lower dorsum elevation than expected (as seen in Figure 5). However, this difference may be due to normal coarticulation and not truly erroneous production.

Table 3 Category classification of P1's productions

Blade Angle	Dorsum Distance	Baseline	Nonwords	Words	Total
NB	ND	185	159	121	465
	GD	5	37	21	63
	CD				0
GB	ND	3	1	4	8
	GD				0
	CD				0
CB	ND				0
	GD				0
	CD				0
Total		193	197	146	536

In Figure 6, the two measurements of each consonant production for P1 are displayed in scatter plots. The x-axis is the angle of declination of the tongue blade and y-axis is the distance of the tongue dorsum from the probe in the typical direction of closure for that vowel. The dotted lines represent the means of dorsum distance and blade angle while the boxes represent 2 SD surrounding the mean in which 95% of the articulations should fall. The box in the upper left corner of the figure represents the “normal” location for velar productions. The box in the lower right corner of the figure represents the “normal” location for alveolar productions. The figures divide the twister productions into baseline (first row), alveolar target items in alternating twisters (second row), and velar target items in alternating twisters (third row). The left column is /a/ vowel stimuli and the right column is /æ/ vowel stimuli. P1 had very small SD ranges around the mean tongue blade angle and dorsum distance, compared to the other participants. Her productions of baseline twisters and alternating twisters were very tightly packed into their target production boxes, with no categorical errors measured. As

discussed previously, P1 produced a small number of gradient errors that can be seen as those items outside of the 2 SD boxes, particularly in the two velar target panels.

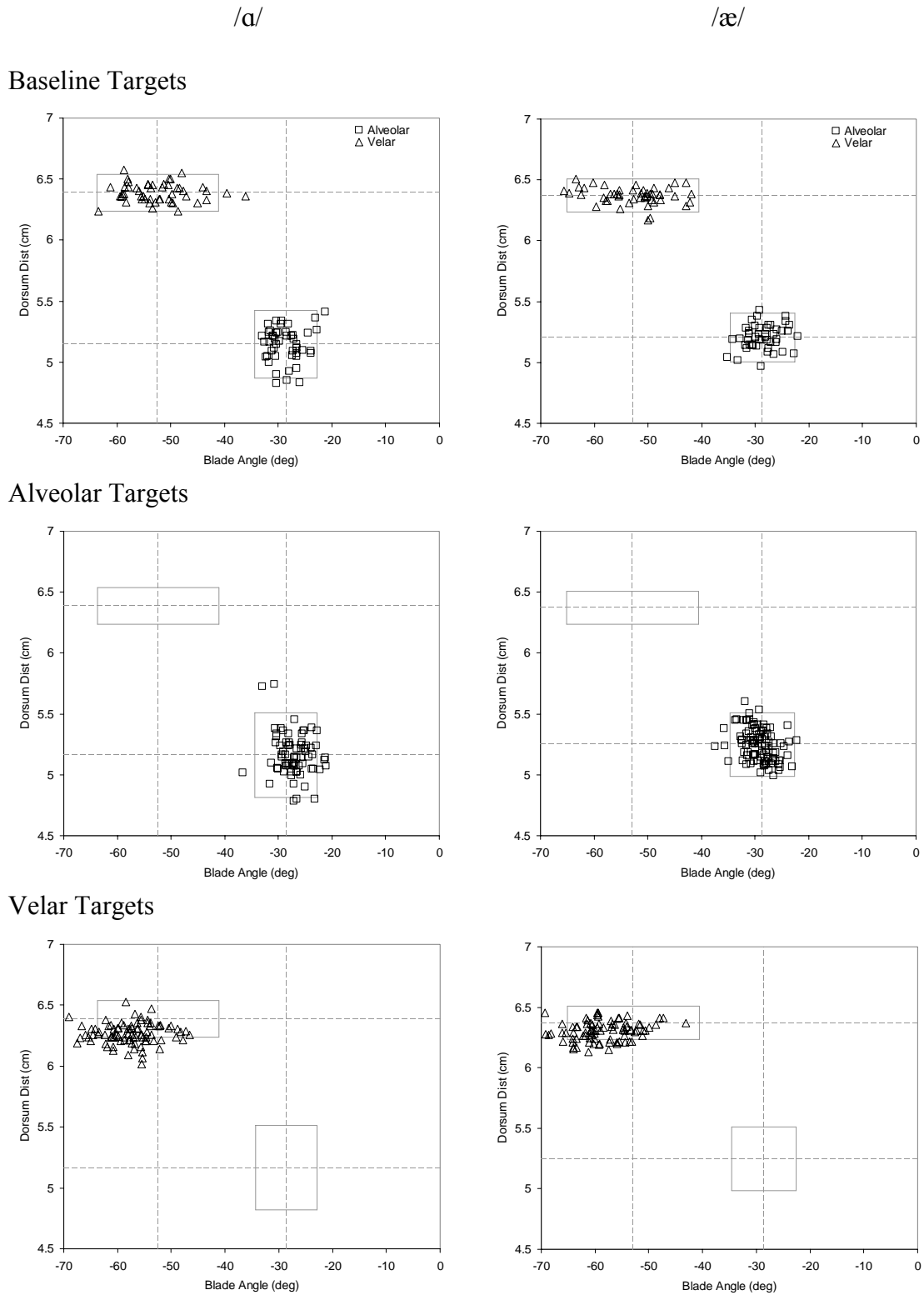


Figure 6 Articulations for P1 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance

P2. When determining the mean articulation measurements of the baseline twisters and comparing them to measurements generated from articulations within alternating twisters for *P2*, it was discovered that coarticulation had a pronounced effect on tongue dorsum elevation for alveolars for at least this participant and perhaps others as well (most noticeably *P3*). A visual comparison of the 2 SD boxes for alveolar productions in the dorsum distance panel of Figure 7 clearly shows that the means between baselines and alternating twisters are substantially different. This discrepancy was discovered while categorizing the productions for *P2*. Therefore, baseline measures for tongue dorsum elevation during alveolar productions in alternating twisters were obtained for *P2* by using the mean and standard deviation from the alternating twister productions, excluding measurements that fell within the velar category for dorsum distance (i.e., categorical errors).

While the sizes of the 2 SD boxes vary between baseline and twister productions within the velar and alveolar blade angles and the velar dorsum distance, the midpoints within each set are comparable. This suggests that, unlike the alveolar dorsum distance, these productions had more variability but were centered on the same average target, as would be expected when errors are elicited. The very long range lines that are present in Figure 7 for alternating twisters reflect errorful productions. Those lines indicate that some production(s) fell in the articulatory space in which the other category was normally produced. For instance, the bottom of the line for the blade angle of alveolar alternating twisters with the vowel /æ/ appears to fall just below the midpoint for the box representing the 2 SD range for velar alternating twisters with the vowel /æ/, suggesting

at least one categorical error. At the same time, the top of the line for the dorsum distance of alveolar targets in alternating twisters with the vowel /a/ appears to be slightly higher than the midline for velar target productions in alternating twisters with the vowel /a/, which also suggests at least one categorical error.

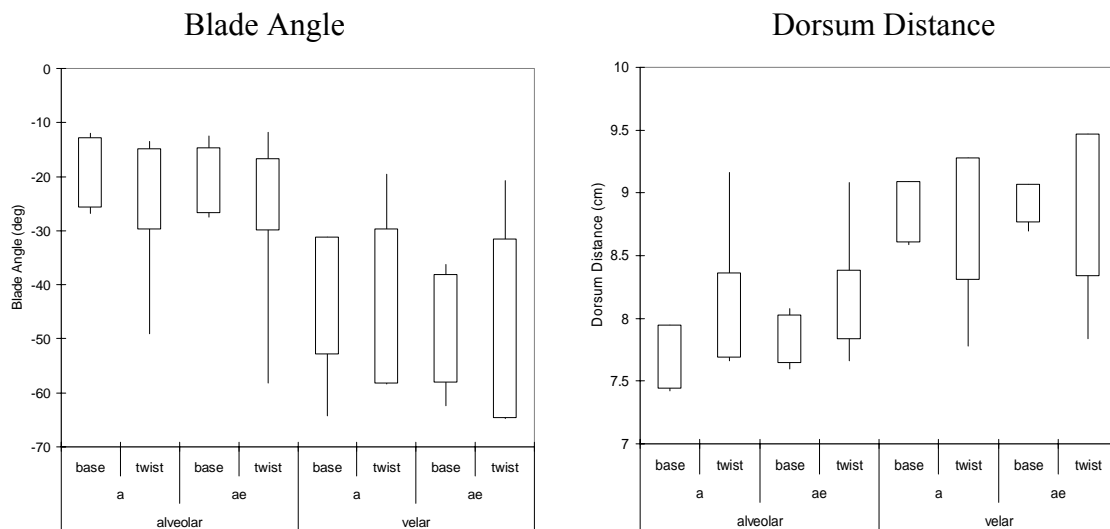


Figure 7 Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for P2

Table 4 shows that P2 produced a total of 541 measurable articulations. She produced 11 gestures within baseline measures that fell outside of the 2 SD cutoff for normal articulations, and they either involved a gradient blade angle error with normal dorsum distance (GBND) or a normal blade angle with gradient dorsum distance error (NBGD). No baseline production was measured to be a categorical error. Normal gestures comprised 83% of P2's productions. The word twisters generated 33 articulation errors (23% of all word productions) while the nonword twisters generated 41 errors (20% of all nonword productions). Unlike the word twisters, the nonword

twisters generated errors in all but one category, CBGD, which had no errors from any set of twisters.

Table 4 Category classification of P2's productions

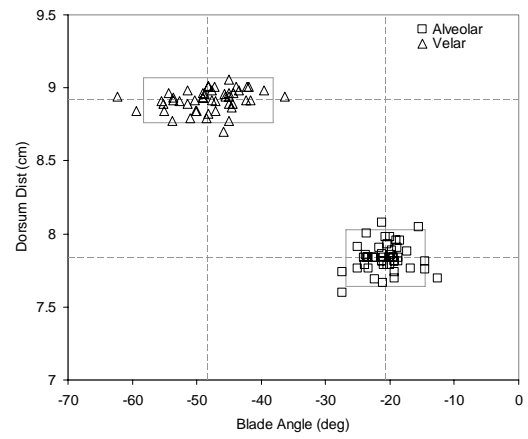
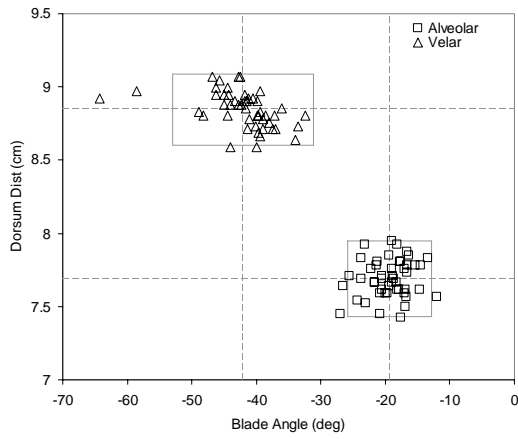
Blade Angle	Dorsum Distance	Baseline	Nonwords	Words	Total
NB	ND	181	165	110	456
	GD	6	7	8	21
	CD		2		2
GB	ND	5	8	15	28
	GD		3	1	4
	CD		1	1	2
CB	ND		1		1
	GD				0
	CD		19	8	27
Total		192	206	143	541

Figure 8 shows P2's articulations in scatter plots. In baseline productions, the majority of her productions fall neatly within the 2 SD boxes. For both /a/ and /æ/ vowels, there were two velar productions articulated with substantially lower blade angle than normal. However, since the focus of this paper is on articulations that are closer to the other category rather than farther away, these productions are not considered to be erroneous. When comparing P2's productions of alveolar-initial twisters to the mean and 2 SD range, P2 made several articulation errors involving productions that fell into the velar box (categorical errors) and in the space between the velar box and the alveolar box (gradient errors). A handful of gestural errors made while trying to approximate a velar target are evidenced in the alveolar boxes in the bottom row of figures. The categorical errors are within the outlines of the box while the gradient errors fall in the space between the two boxes.

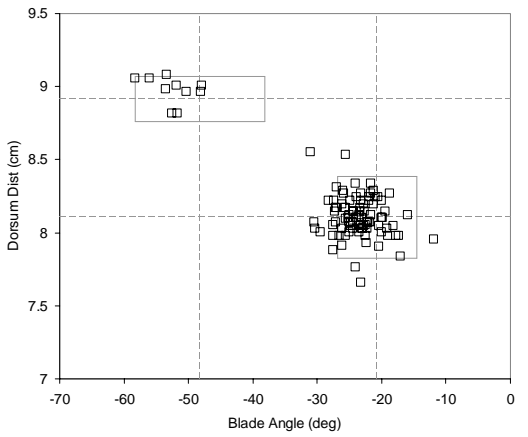
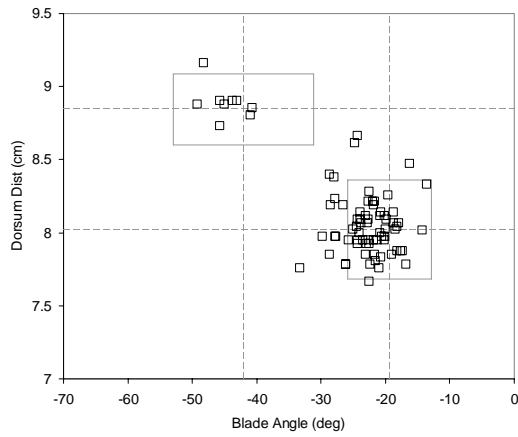
/a/

/æ/

Baseline Targets



Alveolar Targets



Velar Targets

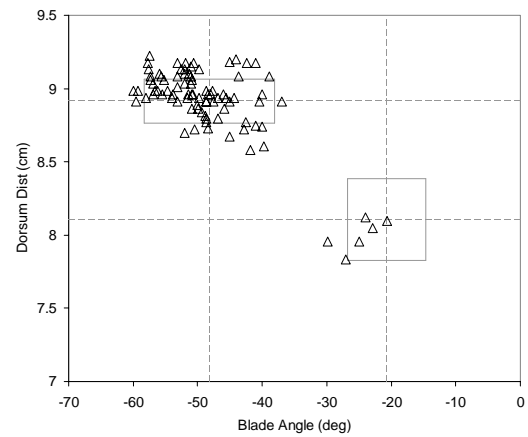
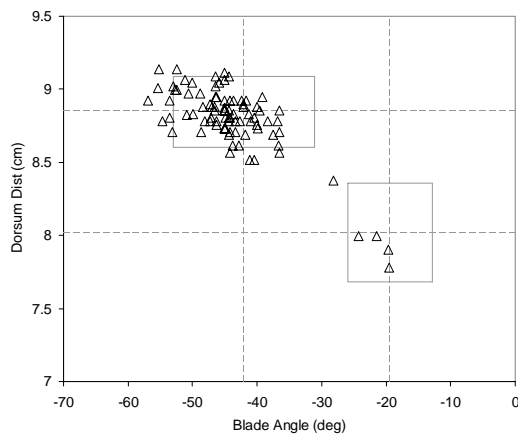


Figure 8 Articulations for P2 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance

P3. The midpoints of the 2 SD boxes for dorsum distance for alveolar target productions of baseline twisters and alternating twisters appear to be substantially different for P3, which is similar to the pattern for P2 (Figure 9). As a result, error categorization for alveolar blade angle for P3 used the same procedure as for P2. As with P2, the ends of the range lines for each set of alternating twisters fall within the range of the 2 SD boxes for the other category (i.e., alveolar vs. velar), suggesting categorical errors in all alternating twister sets.

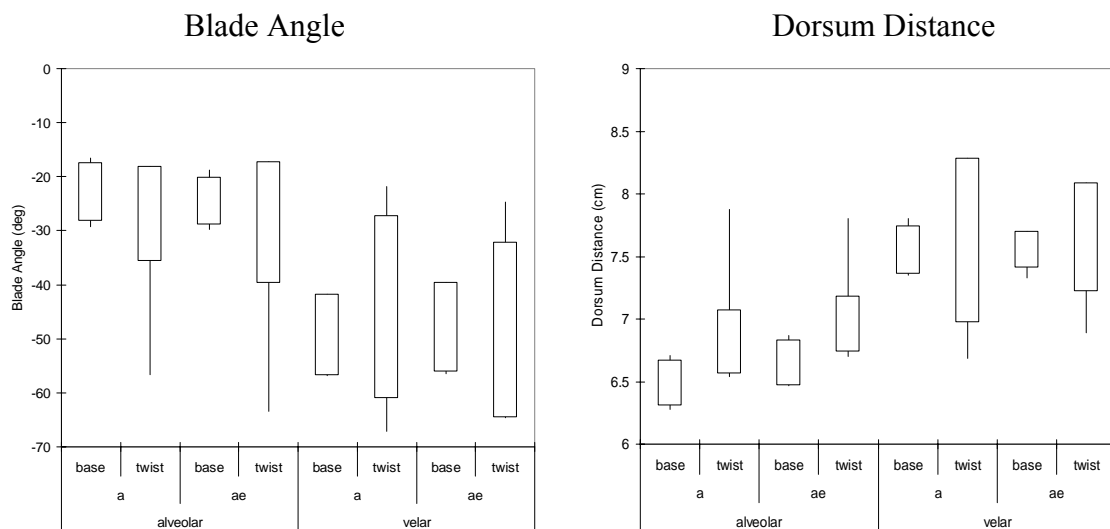


Figure 9 Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for P3

P3 had very small SD ranges around the mean tongue blade angle and dorsum distance for her baseline measures. However, she did make several categorical and gradient errors. Table 5 shows that P3 produced a total of 589 measurable articulations. Eight of the total 125 errors were produced during baseline measures. P3 produced errors

in all but one category, NBCD. The largest error category is GBND, with 52 of the 125 total errors. Of the 117 gesture errors produced in the nonword or word twisters, the percentage of errors produced in nonword twisters and word twisters was comparable at 29% each. Categorical errors in both tongue blade and tongue dorsum comprised 34 of the 117 errors, which is 29%. P3 produced 68 errors while reading the nonword twisters and 49 while reading the word twisters.

Table 5 Category classification of P3's productions

Blade Angle	Dorsum Distance	Baseline	Nonwords	Words	Total
NB	ND	180	163	121	464
	GD	5	2		7
	CD				0
GB	ND	3	34	28	65
	GD		3	3	6
	CD		8	1	9
CB	ND		2	1	3
	GD		1		1
	CD		18	16	34
Total		188	231	170	589

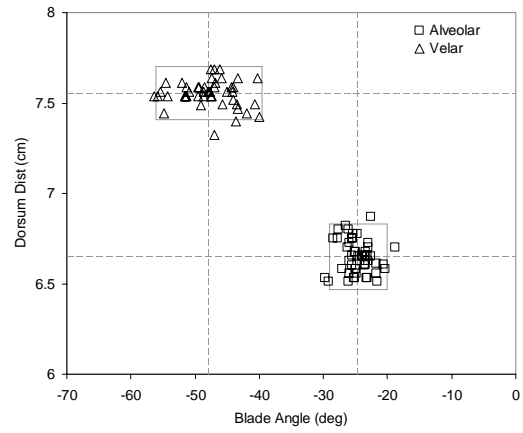
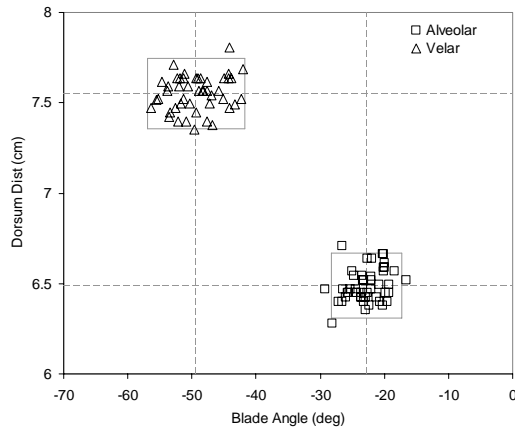
Figure 10 shows that a large number of P3's velar articulations in alternating twisters had a higher dorsum distance than the mean generated from the baseline measures. This pattern was matched by the categorical error productions of alveolar targets in velar articulatory space. The consistency of the elevated dorsum distance for velars during alternating twisters is mysterious. Since velar stops involve raising the tongue dorsum to the palate, the baseline dorsum distance for closure should be the physical upper limit of tongue raising. Three possibilities exist to explain this finding.

1. Coarticulation may have caused a shift in closure location between baseline and alternating twister productions so that the angle used to measure dorsum distance is not the same as the direction of closure.
2. A shift in the participant's head position relative to the ultrasound probe between the baseline and alternating twister portions of the experiment may have occurred. While the participants were placed in an apparatus to restrict head movement, a small upward shift in head position is not impossible.
3. The participant may have hyperarticulated the alternating twisters in an effort to emphasize the different productions. This may have involved lowering the jaw onto the probe thereby displacing the probe slightly, or raising the dorsum more fully to enhance the contrast.

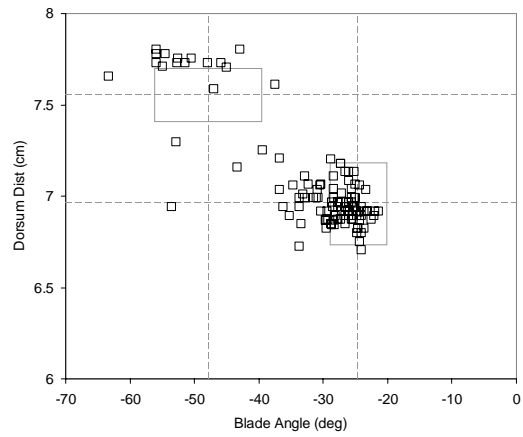
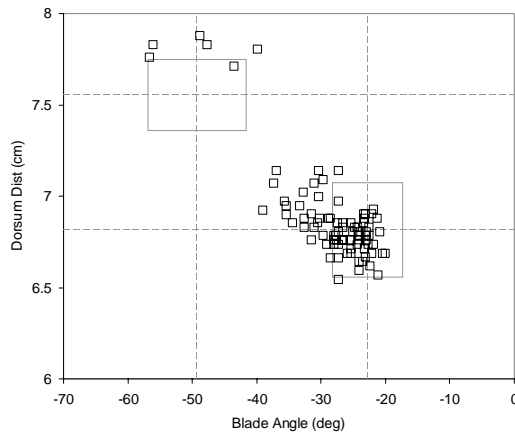
/a/

/æ/

Baseline Targets



Alveolar Targets



Velar Targets

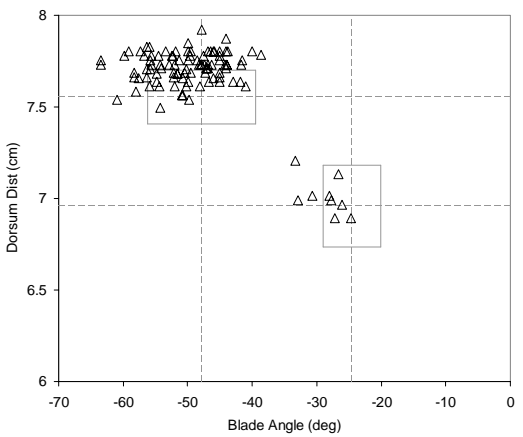
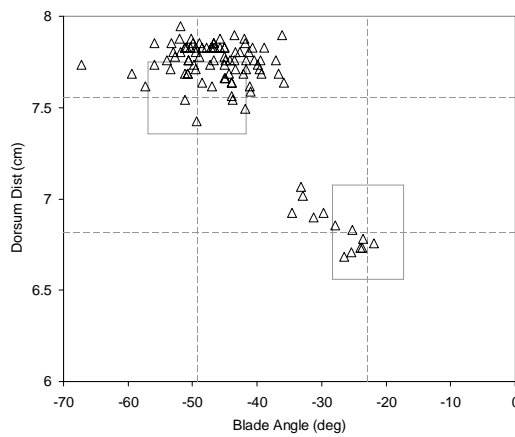


Figure 10 Articulations for P3 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance

P4. *P4* had a pattern of production somewhat unlike those for the other participants. For this participant, blade angle measures overlapped substantially between alveolar and velar targets, especially in the baseline productions. In the dorsum distance measures for alveolar targets, there is some indication that the mean was shifted higher than those of the baseline measures due to coarticulation (as for *P2* and *P3*). For the dorsum distance measure for velars, there is some indication of a shift upward in the alternating twisters, as observed in *P3*.

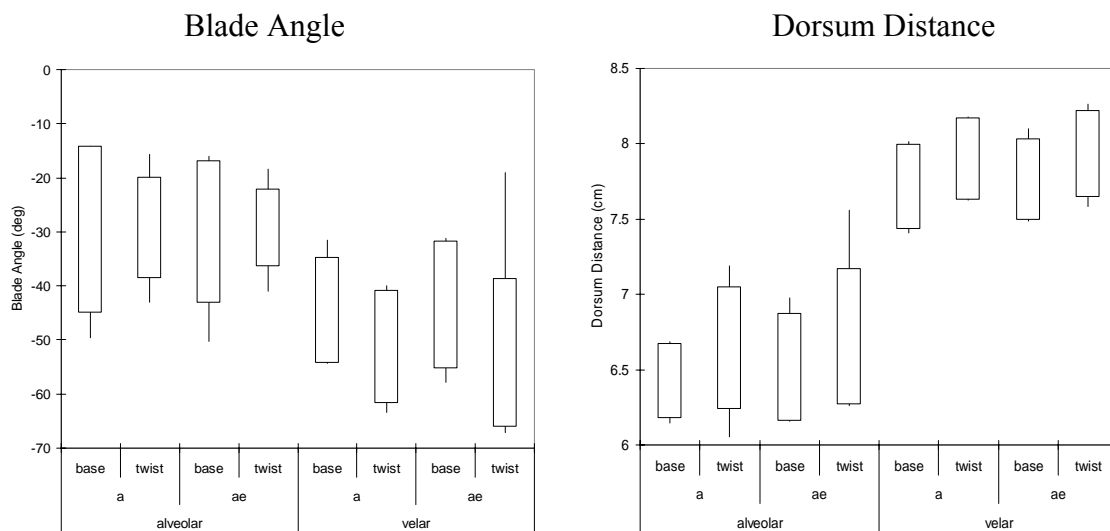


Figure 11 Standard deviations surrounding means and ranges of productions for blade angle and dorsum distance for *P4*

P4 produced a total of 527 measurable articulations (Table 6). Due to the overlap between blade angle categories for alveolar and velar productions, no productions can be classified as gradient errors for blade angle. Considering only dorsum distance measures, she made 6 errors during baseline productions, which is comparable to other participants.

P4 made a total of 8 dorsum distance errors either in the nonword or word twisters, only one of which was categorical (classified as NBCD).

Table 6 Category classification of P4's productions

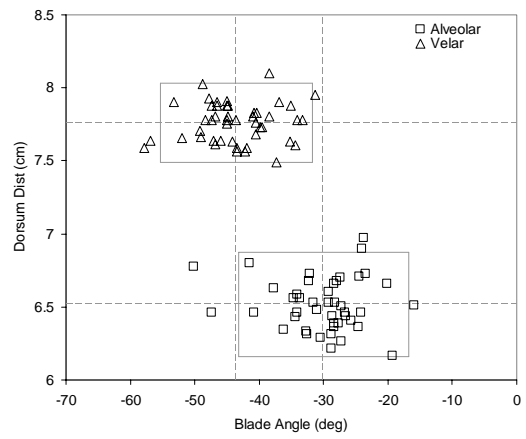
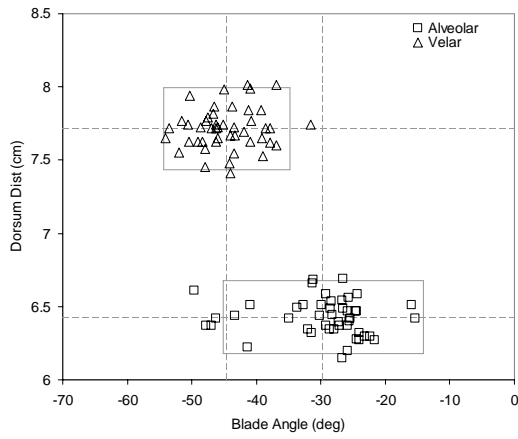
Blade Angle	Dorsum Distance	Baseline	Nonwords	Words	Total
NB	ND	170	189	145	504
	GD	6	5	2	13
	CD		1		1
GB	ND				0
	GD				0
	CD				0
CB	ND	8		1	9
	GD				0
	CD				0
Total		184	195	148	527

Figure 12 shows the scatter plots of P4's measures. P4 produced one categorical error in blade angle in the velar + /æ/ twisters and none in the other three twister categories. This error, classified as CBND in Table 6, is located in the bottom right panel of Figure 12. One of P4's productions with an alveolar target and /æ/ vowel falls within the 2 SD box for velars; however, it is categorized as NBCD due to the overlap in blade angle between the velar and alveolar categories. It is likely that the large 2 SD boxes for dorsum elevation for alveolars are camouflaging gradient errors in dorsum raising, but there is no straightforward means to assess how much of the variability is due to coarticulation with the velars and how much is due to gestural intrusion in a gradient speech error.

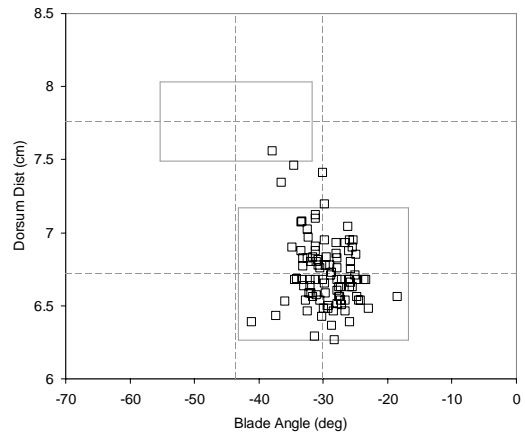
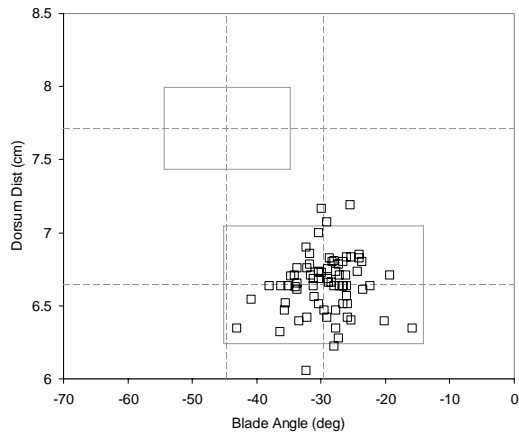
/a/

/æ/

Baseline Targets



Alveolar Targets



Velar Targets

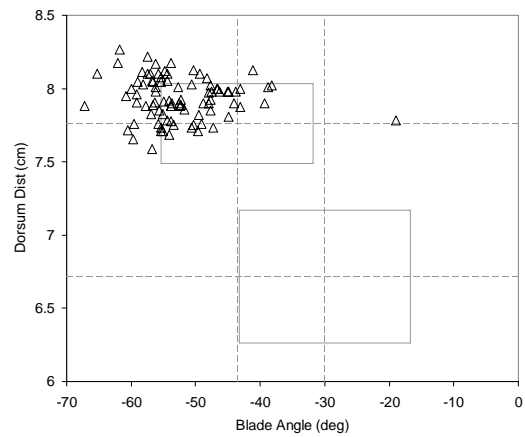
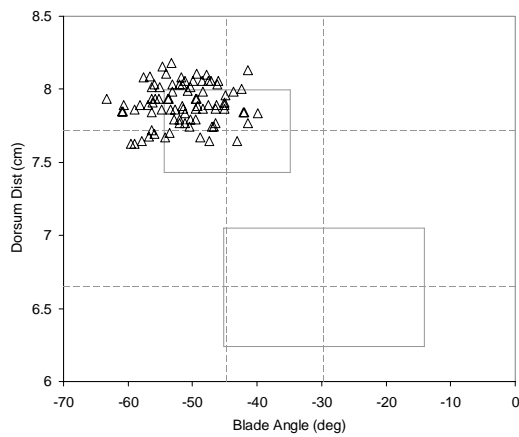


Figure 12 Articulations for P4 delineated by vowel and baseline, alveolar, and velar productions by blade angle and dorsum distance

Discussion

The productions for P1 were basically error free. A small amount of increased variability in the alternating twister productions was observed, which may have been gradient errors, or may have only been due to coarticulation. The productions measured for P2 and P3 are quite similar to one another. They differ from P1 in that both participants generated numerous speech errors while reading the tongue twisters. The overlapping mean and standard deviation boxes for blade angle for P4 make it more difficult to directly compare those measurements to other participants. This discussion will therefore focus on the data from P2 and P3 as both P2 and P3 have obvious loci of productions for alveolar and velar targets and both produced a number of errors.

P2 and P3 had clear categorical errors where both the blade angle and dorsum distance values were normal for a production of the other category (i.e., a normal production of the wrong segment). A small number of productions were measured to be categorical in only one dimension. P2 had categorical errors in dorsum distance for alveolar targets combined with a normal blade angle or a gradient error in blade angle. P2 had categorical blade angle errors for alveolar targets combined with both normal dorsum distance and gradient error dorsum distance.

P2 and P3 had clear gradient errors, where one dimension was intermediate between a normal production for its target category and a normal production for the other category. In most cases these were closer to the normal values, but the large number of instances where the productions are in space clearly distant from the cloud of normal productions suggests that at least some are gradient errors and not just instances of coarticulation. This replicates findings by Mowrey and MacKay (1990), Frisch and

Wright (2002), and Pouplier (2003). A small number of productions are gradient errors in both dimensions and are found in the space directly between the 2 SD boxes, which might be viewed as gestural blends where gestures for both segments compete to be articulated.

Pouplier and Goldstein (2005) concluded that the majority of their production errors consisted of “intrusion errors.” They stated that articulatory gestures of one phoneme (e.g., /k/) were more likely to intrude upon another articulatory gesture (e.g., /t/) than for the articulatory gesture for a phoneme to be reduced. In particular, many errors produced in their experiment exhibited the intrusion of tongue dorsum raising on /t/, resulting in a simultaneous co-production of two gestures. The data in this experiment do not show a large number of co-productions (which would appear above the normal alveolar box and to the right of the normal velar box in Figures 6, 8, 10, and 12).

For P2, a few alveolar productions do seem to fit this pattern; however, categorical errors with a fully raised dorsum and lowered tongue tip were much more common. Also, some of P2’s velar productions seem to have a small reduction in dorsum raising. The reduction errors seem to be about as frequent as categorical errors in the velars where the tongue dorsum is not raised and the tongue blade is raised. For P3, it appears some small increase in dorsum raising frequently occurs for an alveolar target, which is correlated with a small decrease in blade angle. This may be comparable to the dorsum intrusion error of Pouplier and Goldstein (2005). Categorical errors in alveolar targets were also quite common for P3. For velar targets, categorical errors were most common. A small number of velar targets may have a reduction of dorsum raising, especially for /a/.

Overall, these data do show some similarity to the findings of Pouplier and Goldstein (2005). However, these data contain fewer gradient errors and more categorical errors than observed by Pouplier and Goldstein. This is probably due in part to a difference in the production task that was used: Pouplier and Goldstein collected a large number of repetitions of two-word phrases, while this study used fewer repetitions of four-word phrases. In addition, this study used a lower rate of production. It would not be surprising that a larger number of faster repetitions would lead to more gradient articulations than were found with the tongue twister task of this study, due to the increased motor planning demand produced by attempting to speak a large of number of repetitions as quickly as possible.

Many errors classified as gradient fell just outside of the 2 SD boxes and so may just be increased variability in articulation due to the more strenuous nature of the tongue twister task. However, a few cases cannot be wholly dismissed as variability, but also are not categorical errors. These cases show that gradient errors observed by Pouplier and Goldstein (2005), as well as by Frisch and Wright (2002) and Mowrey and MacKay (1990), occur in this data as well. These cases are rare in comparison to the occurrences of categorical errors, which were the most common error type for P2 and P3. The higher rate of categorical errors probably reflects the more language-like nature of the tongue twister task, which suggests that most of the errors observed in naturally occurring speech error corpora are categorical errors.

CHAPTER THREE: PERCEPTION OF PLACE OF ARTICULATION ERRORS

The second experiment is based on the instrumental speech error data from Experiment 1 that revealed variations in articulation. Words and nonwords containing misarticulated phoneme gestures from these data were presented to naïve listeners to obtain their perceptual judgements of the phoneme produced. Listener judgements of the phoneme produced was compared to actual measures from instrumental analysis using ultrasound. Based on research by Mowrey and MacKay (1990) and Pouplier and Goldstein (2005), it was expected that the listeners would only perceive the production as different from the target when the misarticulated gestures were categorically different. When presented with “correct” articulations or gradient errors, the listeners were expected to perceive the target phoneme. Reaction times for responses were also collected to see if reaction times would reflect gradience in the productions that might not be seen in the identification data.

Stimuli

The stimuli were taken from the previous experiment involving ultrasound imaging of participants articulating tongue twisters. The stimuli presented to participants were compiled from examples of regular and irregular articulations uttered by P2 from Experiment 1. Several potential stimulus items were not used due to partial productions

of the word or syllable involving the initial consonant only or other factors interfering with clear auditory perception of the target sound (e.g., laughing). Otherwise, all available irregular articulations were used. For each irregular articulation that was used, a normal articulation of the same syllable in the same twister from a different repetition was also included, where possible.

Procedure. The experiment took place in a sound-controlled room where each participant sat in one of four carrels. Each carrel housed an LCD monitor with a mouse and headphones. The stimuli were presented using TDT System 3 stimulus presentation and recording equipment and were randomized using ECOS/Win experiment software. The stimuli were played over headphones at approximately 65 dB. The participant was presented with four options (i.e., /t/, /k/, /d/, or /g/) for each stimulus item presented. The response options were presented on the computer monitor and the participants used the mouse to click on their choices. When the participant responded, the reaction time delay between the onset of the stimulus and the response was also recorded by the computer. They were instructed to listen carefully, as each stimulus was only played once.

Participants. Forty-six (46) students in an undergraduate speech anatomy course participated in the experiment. They received extra credit for their participation. Subjects were between 19 and 45 years of age. All participants were monolingual native speakers of American English who reported no past speech or hearing disorders.

Results

It was expected that responses matching the target phoneme would occur for normal articulations and gradient errors. It was expected that categorical errors would be

more likely to be perceived as different than the target phoneme. Reaction times were expected to demonstrate slower reactions to gradient errors than to normal productions or categorical errors.

Identification. Table 7 shows the percentage of responses that reflected the perception of a place of articulation error from the target, classified by production category. The table shows the average over all presentations for stimuli for that category to all participants. Stimuli that were judged to be in a different voicing category than the target (i.e., /d/ for /k/ or /t/) were eliminated from the analysis. Appendix A contains a complete list of target items and percentage of place of articulation error judgements by item. The analysis reveals that perception was nearly invariant across all participants. Participant perception of the stimuli was uniform in all but a few cases. All but two of the 21 stimuli in category CBCD (categorical error) were judged to be 100% in the other category than the production target. Of the two stimuli that were judged to be less than 100% in the other category, only two people judged one stimulus to sound the same as the target and only one person judged the other stimulus to be the same as the target. One of the four stimuli for category GBGD (gradient in blade angle, gradient in dorsum distance) was a velar target, while the remaining three were alveolar targets. The velar target was judged by all participants to be different than the target and all the alveolar targets were judged to be the same as their target category, which explains the exact 25% overall percentage of error selection. Examining analysis of the measurements for this stimulus, it was determined that the velar stimulus fell just outside the 2 SD box for the alveolar category. In fact, all GBGD stimuli fell roughly in this area and were heard as alveolar productions. Table 7 also reveals that many items measured to be gradient

errors were not perceived to be categorically different from their target by any participant.

Table 7 Number of stimulus presentations by category and percentage of error selection by listeners

Category	Number of Stimuli	Percentage of Error Selection
NBND	54	0.7%
NBGD	9	0.2%
NBCD	2	0.0%
GBND	22	0.6%
GBGD	4	25.0%
GBCD	0	--
CBND	1	0.0%
CBGD	0	--
CBCD	21	99.7%

Figure 13 illustrates how the stimuli were perceived by the participants as a function of the articulatory measures. As in Experiment 1, the boxes on the figure outline 2 SD from the means for each target sound, which are shown as dotted lines. The diagonal lines through the centers of the panels indicate the midpoint between alveolar and velar articulations. Perceptions of the productions are perfectly predicted by this dividing line. Articulations on the alveolar side were heard as alveolars and articulations on the velar side were heard as velars. Note that this perceptual boundary is fairly far in articulatory space from the region of normal productions for these sounds. This suggests that there can be considerable variability in production without changing the identity of the percept, which is an issue that will be addressed more fully in the discussion.

The few stimuli of the “normal” or “gradient error” categories that were perceived by some listeners to be different from the target had no distinguishing characteristics

from other stimuli in the same categories. Those items designated in Figure 13 as “mixed” are items that more than one person judged differently from the majority of the participants and are shown on the figure as red circles. Items that just one person responded to as different from the target category are not marked as mixed as these may have been accidental incorrect responses rather than genuine differences in perception. All of the mixed items are found in the alveolar targets. In the upper left panel, which shows alveolar targets for the vowel /a/, the mixed item was measured to be a categorical error production (CBCD) but was judged by two participants to be a target alveolar production. Three mixed items are in the upper right panel that shows alveolar targets for the vowel /æ/. Two were measured to be normal productions (NBND) and one was measured to have a gradient error of the blade angle (GBND). The two NBND stimulus items were judged to be categorically different by five and seven participants respectively, while the GBND stimulus was judged to be different by four participants. Of the 53 stimulus items within the NBND category (“normal”), an additional five items were judged by only one person to be categorically different from the target, which may be attributable to an accidental incorrect selection of consonant choice by the participant during the experiment process.

In category NBGD (gradient error in dorsum distance), one participant judged one item out of nine to be categorically different. Two of 22 items in the GBND category (gradient error in blade angle) were judged by one participant to be categorically different. In the category GBGD, four stimulus items were presented to listeners. Three of the items were perceived to be the target 100% of the time. One of the GBGD

stimulus items was judged to be categorically different by 100% of the participants. The GBGD item judged to be categorically different was a velar target that was measured to be close to the alveolar box, shown on Figure 13 in the lower left panel.

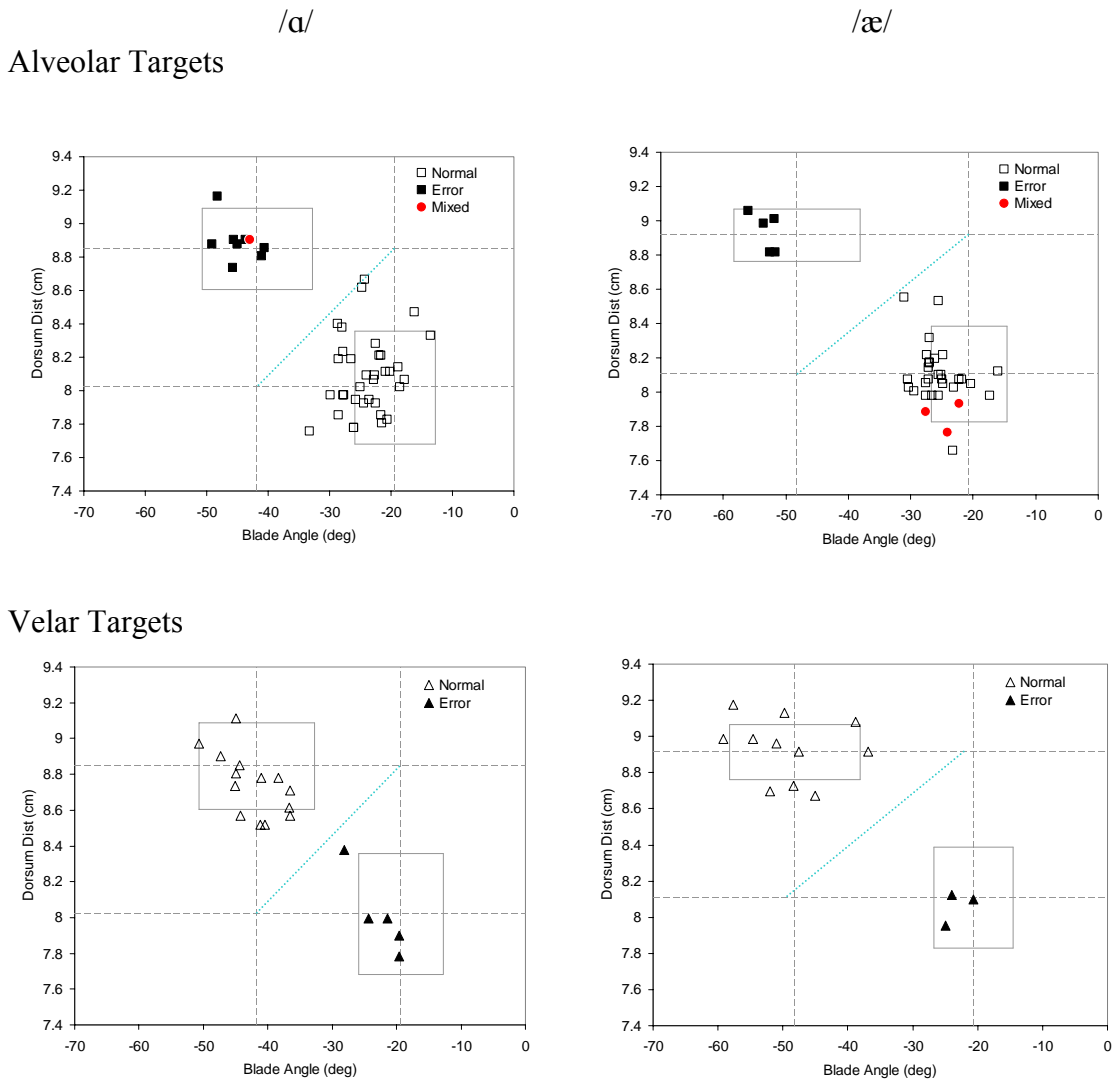


Figure 13 Distributions of perceptual judgements

Reaction Times. Reaction times for identification of the onset phoneme were gathered during the course of the experiment. The participants were given no particular

instructions to answer quickly; therefore, reaction times between participants varied greatly. Also, the experiment used a mouse for judgement selection rather than a “button box,” which has the potential to introduce additional variability in the lag time between selection of choice and logging of an answer. However, some evidence of patterns in reaction time was observed despite these limitations.

Due the variation in reaction times for each participant, z-scores were calculated to standardize the results and allow statistical comparisons across participants. Z-scores are used to compare the relative standings of items from distributions with different means and/or different standard deviations. The z-score for an item indicates how far and in what direction that item deviates from its distribution’s mean and is expressed in units of its distribution’s standard deviation. Therefore, the z-scores for each participant reveal how quickly or slowly they responded to a category of stimuli, based upon their individual mean and standard deviation. Z-scores that are near zero indicate a reaction time that was near average whereas z-scores that are positive correspond to a longer response time than average and z-scores that are negative correspond to a shorter response time than average. In an effort to control for outliers, z-scores that were more than 3 SD from the average were discarded before averaging across subjects and stimuli.

The reaction times are broken down by alveolar targets and velar targets in Tables 8 and 9, respectively. Within each table, the number of stimuli is given, and the reaction times are further broken down by the average reaction time, SD of reaction time, average z-score, and SD of z-score. Tables 8 and 9 show that the pattern in actual reaction times matches the pattern in z-scores, suggesting that the participants had similar overall patterns of reaction time even though they had mean reaction times that differed

by 500-1000 ms in some cases. These tables also show that some of the error categories had very few stimulus items, so the reaction times for these categories are potentially not representative.

Table 8 Reaction times and z-scores for perception of alveolar production targets

Category	Number of Stimuli	Average RT (ms)	SD RT (ms)	Average Z-score	SD Z-Score
NBND	38	1966	512	-0.25	0.93
NBGD	2	2078	575	0.02	0.79
NBCD	2	2152	533	0.27	0.85
GBND	21	2034	531	-0.05	0.91
GBGD	1	2043	517	-0.03	0.83
CBGD	3	1962	496	-0.23	0.76
CBCD	15	2139	534	0.24	1.22

Table 9 Reaction times and z-scores for perception of velar production targets

Category	Number of Stimuli	Average RT (ms)	SD RT (ms)	Average Z-score	SD Z-Score
NBND	16	2085	538	0.10	0.90
NBGD	7	2131	579	0.18	0.95
GBND	1	1957	480	-0.20	1.29
GBGD	1	1942	608	-0.32	1.50
CBCD	7	1993	569	-0.20	1.01

Given that several of the gradient error production categories have very few stimulus items in them, these data were aggregated for statistical analysis. The gradient error categories were combined into a single group, resulting in three production error groups: Normal, Gradient Error, and Categorical Error. Analysis of variance (ANOVA) was performed with two independent variables: Target (Alveolar, Velar) and Error Category (Normal, Gradient Error, and Categorical Error). The dependent variable was the z-score for the reaction time. ANOVA was performed for z-scores averaged over

items within subjects (analysis by subjects) and also for z-scored averaged over subjects within items (analysis by items). For the main effect of target, there was no significant difference [by subjects $F(1,270) = 0.166, p > .05$; by items $F(1,108) = 0.046, p > .05$]. In other words, on average the alveolar targets were responded to at the same rate as velar targets. For the main effect of error category, results were mixed [by subjects $F(2,270) = 6.031, p < .01$; by items $F(2,108) = 1.568, p > .05$]. A post-hoc Tukey HSD test for the analysis by subjects found that responses to the normal productions were faster than responses to the gradient errors and categorical errors, but the gradient errors and categorical errors were responded to at the same rate. The interaction of target and error category was significant [by subjects $F(2,270) = 80.875, p < .01$; by items $F(2,108) = 15.042, p < .01$]. The rate of response to different error types was different depending on whether the target was alveolar or velar.

This interaction can clearly be seen in Figure 14, which shows the average reaction time grouped by target and error category as in the ANOVA. Figure 14 shows that for alveolar targets, normal productions were responded to more quickly than categorical error productions. On the other hand, for velar targets, normal productions were responded to more slowly than categorical error productions. This interaction can actually be explained if the data are considered in terms of the listeners' percept rather than the production target. A categorical error for an alveolar target is produced in the velar articulatory space, resulting in a velar percept. Similarly, a categorical error for a velar target is produced in the alveolar articulatory space, resulting in an alveolar percept. Comparing the reaction time for normally produced alveolars (first column in Figure 14) versus alveolars that are the result of errors (last column in Figure 14), the responses to

the normal alveolars are 27 ms faster than the responses to the velar errors produced as alveolars. There is an analogous pattern for the velars. Comparing the reaction time for normally produced velars (fourth column in Figure 14) versus velars that are the result of errors (third column in Figure 14), the responses to the normal velars are 54 ms faster than the responses to the alveolar errors produced as velars. With this perceptual pattern in mind, a second trend also emerges. Overall reaction times to alveolar percepts (first, second, and last columns) are on average 115 ms faster than reaction times to velar percepts (third, fourth, and fifth columns). There is also some evidence that reaction times to gradient errors are slower than reaction times to normal items. This is seen most clearly for the alveolars, but the small trend for the velars is also in the same direction.

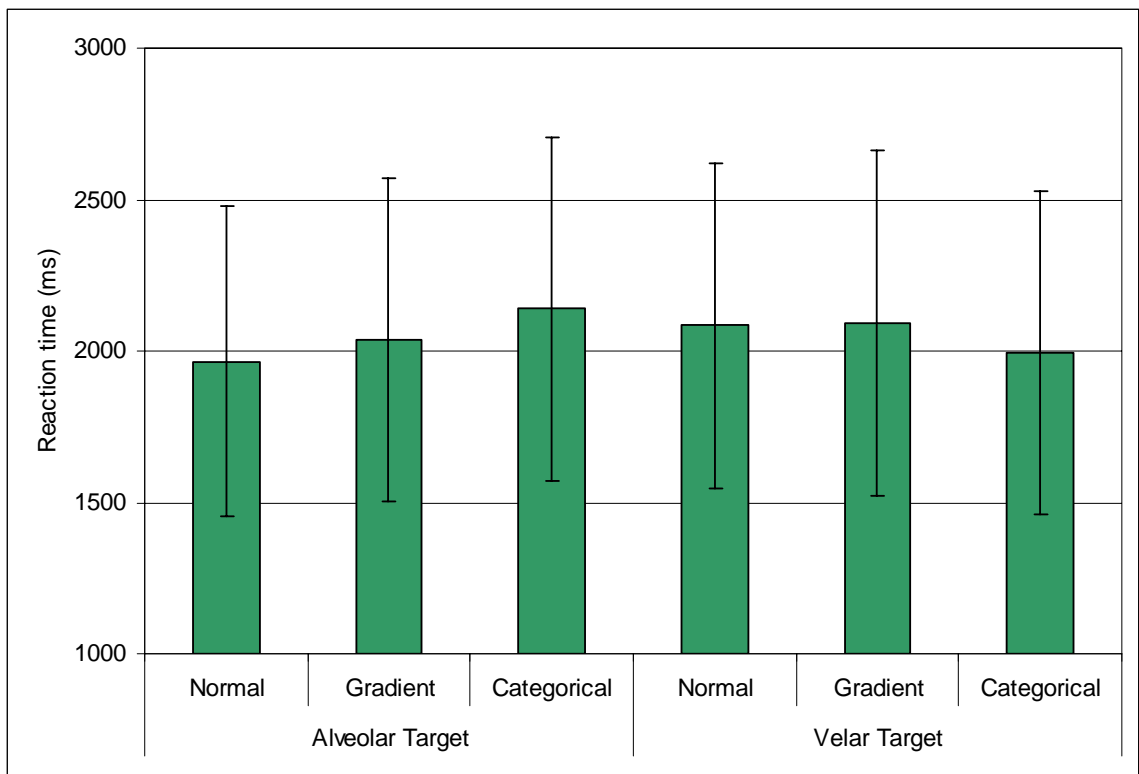


Figure 14 Reaction time to stimuli on the basis of production target and 3-way error categorization

Discussion

Overall, the results of this perception experiment follow a very simple pattern: stimuli that were measured to be within velar articulatory space were heard as velar productions and stimuli that were measured to be within alveolar articulatory space were heard as alveolar productions. Only two stimuli were measured to be very near to the “line” between velar space and alveolar space, but both fall just on the alveolar side of the articulatory space. Both of these productions were heard as alveolars by the participants. The reaction time analysis found suggestive evidence that error productions are not exactly the same as normal productions. This finding should be investigated further using more sensitive procedures.

Overall, the findings are compatible with the claims of Mowrey and MacKay (1990) and Frisch and Wright (2002) that gradient errors are perceptually heard as normal. The results are also comparable to the study of Pouplier and Goldstein (2005), who found gradient errors were generally not perceived as misproductions by naïve listeners. Pouplier and Goldstein’s claim that dorsum intrusions on alveolars are more perceptible than tongue tip intrusion on velars could not be tested due the small number of intrusions in the present data.

CHAPTER FOUR: CONCLUSIONS/DISCUSSION

Conclusions

Experiment 1 used ultrasound to capture articulation in action in tongue twisters and quantify where it happened in articulatory space. Analysis of the measurements revealed that participants made both categorical errors of articulation and gradient errors of articulation. Some of the errors made by the speakers were produced as a completely different phoneme. On the other hand, some of the errors made by the speakers were only somewhat different from the target, but were different enough to fall outside of the normal articulatory space for the target.

Experiment 2 used data collected during Experiment 1 and presented tokens of normal productions, gradient errors, and categorical errors to naïve listeners. As one would expect, listeners perceived normal and categorical error productions as the target or the other of the target, respectively. Interestingly though, perception of gradient errors appeared to depend on where in articulatory space the token was produced. Gradient errors produced on the alveolar side of the articulatory space were perceived to be alveolars and gradient errors produced on the velar side of the velar articulatory space were heard as velars. Analysis of reaction times from this experiment revealed that listeners were slower to respond to gradient and categorical error productions compared

to normal productions. On average, responses to alveolar percepts were faster than responses to velar percepts, obscuring the error pattern.

Discussion

Transcription of speech errors has provided a great deal of linguistic information about how speech is produced. However, for speech errors to be perceived by listeners in running speech, they necessarily needed to be obvious. Speech errors are also constrained by transcribing them into IPA, which imposes a label upon what was heard. Experiment 1 revealed that errors occur that are not within the normal range for either the target category or the other category. These findings are similar to those by other researchers who have used instrumentation to quantify speech errors (Mowrey & MacKay, 1990; Frisch & Wright, 2002; Pouplier & Goldstein, 2005). To determine how gradient errors are perceived, they were included as stimulus items along with normal productions and categorical errors in Experiment 2. The perception of all the gradient errors depended upon where in articulatory space they were produced. If the gradient error was produced on the alveolar side of the space, then it was perceived to be an alveolar; however, if it was produced on the velar side of the space, then it was perceived as a velar. This indicates that transcription of speech errors is not accurate enough to record small articulatory errors made by speakers.

One possible explanation for the way gradient errors are perceived is found in quantal theory (Stevens, 1972). The thesis of the quantal theory is that certain, relatively large changes in articulator position will cause little change in the acoustic signal, while other, relatively small changes in articulator placement will cause large changes in the

acoustic signal. The extent of the acoustic change appears to be related to the particular region of the vocal tract where the articulation is located. In certain critical regions, a slight adjustment of articulatory placement will cause a large acoustic change in sound. Therefore, the boundary line between articulatory spaces may be the critical region that causes the production to be perceived as either on target or in error. The relatively large area between the “box” of normal articulation for velars and alveolars and the boundary line between the two categories in perception would allow for a great deal of variability of articulation that would still result in the listener perceiving the target sound. Quantal theory also proposes that, in general, languages use phonemes that are articulated in regions that are large enough to withstand individual variations of production. These large articulatory spaces are thought to account for relatively steady state productions while the articulators are moving through the area. The findings of this study support these ideas that relatively large variability in production did not affect identification.

While quantal theory may provide a convenient explanation for the correlation of gradient error perception and articulatory space, the reaction time data are not as clear. The possibility exists that the measurements from the ultrasound recordings do not provide a complete picture of the productions. The difference in reaction times for normal and error productions (especially the categorical errors) suggests that the listeners were affected by some cue that was unaccounted for by articulatory measures of blade angle and dorsum distance during the stop closure. Further research into reaction times for perception of speech errors needs to be conducted to determine if some as yet unmeasured part of the utterance interferes with perception of errors.

Clinical Uses

It has been demonstrated in this thesis and elsewhere that articulatory gestures produced slightly differently than “normal” can produce sounds that are heard as the target by the listener (Boucher, 1994; Frisch & Wright, 2002; Mowrey & MacKay, 1990; Pouplier, 2003; Pouplier & Goldstein, 2005). The perception of articulatory gestures that are more than slightly different as categorical errors implies that the gesture is perceived as that category closest to which it is produced. Ultrasound could provide a visual clue as to where an individual is producing a particular phoneme. The use of ultrasound in a clinical setting with clients who have articulation errors could provide an accessible visual tool in remediating articulation impairments. Many a speech-language pathologist can relate anecdotes about trying to explain to a client, young or old, hearing-impaired or normal hearing, how to move the tongue to produce the target sound. Providing a real time ultrasound image to the client of their own tongue and how it moves could provide a valuable asset to speech-language pathologists. While the clinician instructs the client where and how to move the tongue, the client can, at the same time, watch to see if they are moving their tongue to the appropriate location. The clinician can point to the screen and show the client the general area where they need to put their tongue in order to approximate the target sound. The results of this thesis show that the speaker can produce a perceptually correct sound even if the articulators are not in exactly the same location each time, or in the ideal position desired by the clinician. Therefore, guiding the client to an approximation of the target sound may be enough to help them produce a perceptually accurate phoneme.

Research into the usability of ultrasound in field applications has demonstrated that a reliable ultrasound image can be produced if the subject rests his/her head against a stable surface, such as a headrest against a wall (Gick, Bird, & Wilson, 2005).

Combining the use of a headrest with a fixed ultrasound probe, such as on a microphone stand, further enhances the reliability of the ultrasound image. Finally, in order for a clinician to utilize these methods in combination with therapy, placing the ultrasound monitor where the client could see it easily would allow for interaction between articulation and visual feedback.

While Bernhardt, Gick, Bacsfalvi, and Adler-Bock (2005) point out that portable ultrasound machines may not generate the consistency of data collection that stationary machines with fixed transducers provide, they do allow treatment to be provided in locations convenient for the client. Their research has shown that use of ultrasound as visual feedback can improve the client's understanding of what they are being taught by the clinician. Bernhardt, et al., suggest that the clinician demonstrate the use of the ultrasound probe for the client not only to show how to position it properly (if a fixed transducer is not used) but also to provide a visual ultrasound image of the desired target production. A sagittal view of the tongue from the transducer provides a descriptive image of velar and alveolar productions, vowel productions, approximants /r/ and /l/, and sibilants and affricates. Rotating the transducer to create a coronal image can provide another image of sibilants and affricates that shows grooving of the tongue.

Descriptive research has been conducted that uses ultrasound to examine the shape of the tongue during the articulation of vowels and consonants. This research has generated a great deal of information regarding how the tongue postures and moves to

create sounds. Clinical research on the application of ultrasound as a tool for abnormal speech production and on the use of ultrasound to provide articulatory feedback is just beginning.

REFERENCES

- Baars, B., Motley, M., & MacKay, D. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382-391.
- Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics*, 19(6/7), 605-617.
- Boucher, V.J. (1994). Alphabet-related biases in psycholinguistic enquiries: considerations for direct theories of speech production and perception. *Journal of Phonetics*, 22(1), 1-18.
- Dell, G.S., & Reich, P. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611-629.
- Dell, G.S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5(4), 313-349.
- Frisch, S.A., & Wright, R. (2002). The phonetics of phonological speech errors: an acoustic analysis of slips of the tongue. *Journal of Phonetics*, 30, 139-162.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47(1), 27-52.
- Freud, S. (1901). The psychopathology of everyday life. *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Vol. 6.
- Garrett, M. (1975). The analysis of sentence production. In G. Bower (ed.). *The Psychology of Learning and Motivation*. New York: Academic Press. 133-177.
- Garrett, M. (1976). Syntactic processes in sentence production. In E.C.T. Walker & Wales (Ed.). *New Approaches to Language Mechanism*. Amsterdam: North Holland.
- Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association*, 32(2), 113-121.
- Gick, B., Bird, S., & Wilson, I. (2005). Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics & Phonetics*, 19(6/7), 503-514.

- Hardin, S.A. (2005). *Reliability of Hand Measures of Ultrasound Analysis*. Unpublished Master's Thesis, University of South Florida, Tampa, FL.
- MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8, 323-350.
- Meijer, P. (1997). What speech errors can tell us about word-form generation: The roles of constraint and opportunity. *Journal of Psycholinguistic Research*, 26(1), 141-158.
- Mowrey, R., & MacKay, I. (1990). Phonological primitives: electromyographic speech error evidence. *Journal of the Acoustical Society of America*, 88(3), 1299-1312.
- Peng, C.L., Jost-Brinkmann, P.G., & Miethke, R.R. (1996). The cushion-scanning technique: A method of dynamic tongue sonography and its comparison with the transducer-skin coupling scanning technique during swallowing. *Journal of Academic Radiology*, 3, 239-244.
- Poupier, M. (2003). *Units of Phonological Encoding: Empirical Evidence*. Ph.D. dissertation, Yale University.
- Poupier, M. & Goldstein, L. (2005). Asymmetries in the Perception of Speech Production Errors. *Journal of Phonetics*, 33, 47-75.
- Shattuck, S. (1975). *Speech Errors and Sentence Production*. Unpublished doctoral dissertation. Department of Psychology, Massachusetts Institute of Technology, Cambridge, MA.
- Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18, 41-55.
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42, 213-259.
- Stemberger, J., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, 25, 163-180.
- Stevens, K.N. (1972). The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data. In E.E. David, Jr. and P.B. Denes (Eds.). *Human Communication: A Unified View*. New York: McGraw-Hill, 51-66.
- Stone, M. (2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19(6/7), 455-501.

- Stone, M. & Davis, E. (1995). A head transducer support system for making ultrasound images of tongue/jaw movement. *Journal of Acoustical Society of America*, 98, 3107-3112.
- Stone, M., Sonies, B., Shawler, T., Weiss, G., & Nadel, L. (1983). Analysis of real-time ultrasound images of tongue configuration using grid-digitizing system. *Journal of Phonetics*, 11, 207-218.
- Wilshire, C.E. (1999). The “tongue twister” paradigm as a technique for studying phonological encoding. *Language and Speech*, 42(1), 57-82.
- Wodzinski, S.M. (2004). *Ultrasound Analysis of Velar Fronting*. Unpublished Master’s Thesis, University of South Florida, Tampa, FL.

APPENDIX A: PERCEPTION RESULTS FOR EACH STIMULUS

Code	Stimulus	Count	Percentage
NBND	n1kae1r2	46	0.0%
	n1ta2	46	0.0%
	n1ta4	46	0.0%
	n1ta6r2	46	0.0%
	n3kae3	46	0.0%
	n3kae4	46	0.0%
	n3ta1	46	0.0%
	n3ta3	45	0.0%
	n3tae4	46	0.0%
	n3tae5	46	0.0%
	n3tae7	46	0.0%
	n4da1	45	0.0%
	n4da2	45	2.2%
	n4dae2	46	0.0%
	n4dae3	46	0.0%
	n5ta2	46	0.0%
	n6da1	46	0.0%
	n6da2	46	0.0%
	n6da4	46	2.2%
	n6dae5	46	2.2%
	n6ga1	46	0.0%
	n6ga2	46	2.2%
	n7kae2	46	0.0%
	n7ta2	46	0.0%
	n7ta3	46	0.0%
	n8da1	46	0.0%
	n8da5	46	0.0%
	n8dae3	46	0.0%
	n8ga3	46	0.0%
	w1ka5	46	0.0%
	w1ta3	46	0.0%
	w1ta4	46	0.0%
	w2ta2	46	0.0%
	w3dae1r4	46	0.0%
	w3dae1r5	46	0.0%
	w3dae2r2	46	0.0%
	w3dae2r3	46	0.0%
	w3gae3	46	0.0%
	w4ka1	46	0.0%
	w4ka3	46	0.0%
w4kae1	46	0.0%	

APPENDIX A (Continued)

Code	Stimulus	Count	Percentage
NBND	w4kae2	46	0.0%
	w4ta2	46	0.0%
	w4ta4	46	0.0%
	w4tae2r2	46	10.9%
	w4tae3	46	0.0%
	w4tae5	46	0.0%
	w5ka2	46	0.0%
	w5ka3	46	0.0%
	w5ta1	46	0.0%
	w5tae3	46	0.0%
	w5tae4	46	15.2%
	w6dae1r2	46	0.0%
	w6ga5	46	2.2%
	NBGD	n1kae1r1	46
n3kae1		46	0.0%
n7ta4		46	0.0%
w3dae2r5		46	0.0%
w3gae2		46	0.0%
w4ka2		46	0.0%
w4ka5		46	0.0%
w5ka1		46	0.0%
w5ka4		46	2.2%
NBCD	n3ta2	46	0.0%
	n8da3	46	0.0%
GBND	n3kae5	46	0.0%
	n3ta7	46	0.0%
	n3tae3	46	0.0%
	n3tae6	46	2.2%
	n4da5	46	0.0%
	n5ta4	46	0.0%
	n5tae5	46	0.0%
	w2ta6	46	0.0%
	w3dae1r1	46	0.0%
	w3dae1r2	46	0.0%
	w3dae1r3	46	0.0%
	w3dae2r1	46	0.0%
	w3dae2r4	46	0.0%
	w4ta1	46	0.0%
	w4ta3	46	0.0%
	w4ta5	46	0.0%
w4ta6	46	0.0%	

APPENDIX A (Continued)

Code	Stimulus	Count	Percentage
GBND	w4tae6	46	0.0%
	w5ta5	46	2.2%
	w5tae1	46	0.0%
	w5tae2	46	8.7%
	w6dae1r3	46	0.0%
GBGD	n1ta1r2	46	0.0%
	n1ta3	46	0.0%
	n6ga6	46	100.0%
	w1ta1	46	0.0%
CBND	n1ta1r1	46	0.0%
CBCD	n1ta5	45	100.0%
	n1ta6r1	46	100.0%
	n4da4	45	100.0%
	n4dae4	44	100.0%
	n4dae6	46	100.0%
	n4ga4	46	100.0%
	n6da3	46	100.0%
	n6da5	46	100.0%
	n6da6	46	97.8%
	n6dae3	46	100.0%
	n7kae6	46	100.0%
	n7ta1	45	100.0%
	n8da2r1	45	95.6%
	n8dae4	44	100.0%
	n8ga4	46	100.0%
	w1ka6	46	100.0%
	w2ta3	46	100.0%
	w4kae3	46	100.0%
	w4kae4	46	100.0%
	w4tae1	46	100.0%
	w4tae2	eliminated – heard as place and voice error	
w6ga6	46	100.0%	