2016-06-01

The Association Between Articulator Movement and Formant Trajectories in Diphthongs

Katherine Morris McKell

Brigham Young University

The Association Between Articulator Movement and

Formant Trajectories in Diphthongs

Katherine Morris McKell

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Christopher Dromey, Chair
Kristine Tanner
Ray Merrill

Department of Communication Disorders

Brigham Young University

June 2016

ABSTRACT

The Association Between Articulator Movement and
Formant Trajectories in Diphthongs

Katherine Morris McKell
Department of Communication Disorders, BYU
Master of Science

The current study examined the association between formant trajectories and tongue and lip movements in the American English diphthongs /aɪ/, /aʊ/, and /ɔɪ/. Seventeen native speakers of American English had electromagnetic sensors placed on their tongues and lips to record movement data along with corresponding acoustic data during productions of the diphthongs in isolation. F1 and F2 trajectories were extracted from the middle 50% of the diphthongs and compared with time-aligned kinematic data from tongue and lip movements. The movement and formant tracks were converted to z-scores and plotted together on a common time scale. Absolute difference scores between kinematic variables and acoustic variables were summed along each track to reflect the association between the movement and acoustic records. Results show that tongue movement has the closest association with changes in F1 and F2 for the diphthong /aɪ/. Lip movement has the closest association with changes in F1 and F2 for the diphthong /aʊ/. Results for the diphthong /ɔɪ/ suggest tongue advancement has the closest association with changes in F2, while neither lip movement nor tongue movement have a clearly defined association with changes in F1. These results suggest that for diphthongs with the lip rounding feature, lip movement may have a greater influence on F1 and F2 than previously considered. Researchers who use formant data to make inferences about tongue movement and vowel space may benefit from considering the possible influence of lip movements on vocal tract resonance.

Keywords: speech science, formant, diphthong, speech acoustics, kinematic, articulation

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF THESIS STRUCTURE

The structure of this document includes elements required by the university and department, but is also modeled after peer-reviewed articles in the field of Communication Disorders. Appendix A contains an annotated bibliography, and Appendix B contains the participant consent form.

Introduction

Researchers who examine speech articulation have available to them several different methodologies. The two most common involve acoustic or kinematic approaches. Acoustic analysis is widely used because it is relatively inexpensive and is also noninvasive (Kent, Weismer, Kent, Vorperian, & Duffy, 1999). High-quality microphones and recording devices are easily available, and because no apparatus makes contact with the tongue or other articulators during recording, speech is unimpeded and is therefore more likely to be natural than when invasive procedures are used.

Acoustic information is useful for examining speech because of the association between acoustic measures and the movements of the articulators, particularly the tongue. It has been established that the frequency of the first formant (F1) in a vowel is influenced by the height of the tongue, and the frequency of the second formant (F2) by tongue advancement (Ferrand, 2007; Weismer, Martin, Kent, & Kent, 1992). Analyzing F1 and F2 allows researchers to indirectly infer patterns of vertical and horizontal movement of the tongue during speech. Additionally, F1 and F2 of the corner vowels in English (/i/, /u/, /ɑ/, and /æ/) can be used to calculate the vowel space area of a speaker (Tjaden & Wilding, 2004), which reflects the articulatory acoustic working space, and therefore how distinct the vowel contrasts might be.

Formant-based analyses of speech have many applications. For example, recent research has shown that analyzing formants from disordered speech is a robust enough method to reveal systematic changes in articulation following treatment. Sapir, Ramig, Spielman, and Fox (2010) found that measurements of vowel formant centralization not only differentiated dysarthric speakers from healthy speakers, but also differentiated pre- and post-treatment speech. The researchers found that measurements of vowel space area and vowel formant centralization were

significantly different in dysarthric speakers who had received one month of Lee Silverman Voice Treatment versus dysarthric speakers who had received no treatment. Roy, Nissen, Dromey, and Sapir (2009) found a significant increase in the vowel space of speakers with muscle tension dysphonia following single treatment sessions involving laryngeal reposturing maneuvers and/or manual circumlaryngeal treatment. Measurements using the vowel space area and the vowel articulation index showed an increase in both following treatment, which suggested an improvement in articulatory movements.

While acoustic analysis is clearly a useful research tool, it has several inherent limitations. First, there is some ambiguity regarding the relative contributions of different articulators to the acoustic features of a speech sound. The vocal tract consists of multiple resonating cavities (the pharynx, the oral cavity, and the nasal cavity) that can influence the acoustic signal. These pharynx and oral cavity change their shape depending on how the articulators move. A change in a formant results from the movements of several articulators (e.g., the lips and the tongue) or from several parts of an articulator (e.g., the tongue blade and the base of the tongue), making it difficult to isolate the contribution of a specific articulator. Additionally, speech sounds are not uttered in isolation; they are coarticulated. As a result, an articulatory gesture may be influenced by the anticipation or perseveration of features from a neighboring speech sound.

A second potential limitation of acoustic analysis is associated with the quantal theory of speech (Stevens, 1989), which states that a change in the position of the articulators does not have a uniform, one-to-one correspondence with changes in the acoustic signal. This theory is based in the discovery that a movement from point A to point B by the tongue may produce a relatively small change in the acoustic signal, whereas a movement from point B to point C of

the same distance may result in a relatively large acoustic change. In other words, the magnitude of change in the acoustic signal varies depending on where a sound is made in the vocal tract.

A third limitation of acoustic analysis is that idiosyncratic differences in articulation are known to occur. The motor equivalence principle (Hughes & Abbs, 1976) suggests that different speakers can accomplish the production of perceptually similar sounds in different ways. For example, it is well documented that in rhotic varieties of North American English, speakers exhibit variation in how they produce the phoneme /r/ (Mielke, Baker, & Archangeli, 2016). Some produce a variation of the *bunched /r/,* where the tongue tip is flat and the back of the tongue is raised. Others produce a variation of the *retroflex /r/,* where the tongue tip is raised toward the alveolar ridge and the back of the tongue is flat. While the tongue position of these two articulatory configurations is quite different, the two productions are perceptually indistinguishable (Twist, Baker, Mielke, & Archangeli, 2007). Additionally, when examining the acoustic data, researchers have found no consistent difference between the first three formants of the bunched /r/ versus the retroflex /r/ (Zhou et al., 2008).

To address the limitations of acoustic methods mentioned above, it would be useful to determine how closely acoustic information reflects articulator movement by comparing acoustic variables with more direct measures of movement—in other words, kinematics. Kinematic methods track the movement of the articulators directly. Cinefluorography, magnetic resonance imaging tagging, ultrasound systems, and electromagnetic articulography have all been used for this purpose (Mefferd & Green, 2010; Moll, 1960; Stone et al., 2007; Stone, Langguth, Woo, Chen, Prince, 2014).

A small but growing body of research has used both acoustic and kinematic analysis to examine speech. Many of these studies have used kinematic data and acoustic data in

complementary roles, assuming that F1 reflects tongue height and F2 reflects tongue advancement. However, a few recent studies (e.g., Dromey, Jang, & Hollis, 2013) have begun to directly address the relationship between kinematic and acoustic data. For the most part, acoustic measurements have been shown to reflect kinematic measurements of movement fairly accurately. Lee (2014) found that F1 and F2 in the diphthong /aɪ/ were highly predictive of kinematic measurements. Not only did F1 correlate with vertical tongue movement and F2 with horizontal tongue movement; but, unexpectedly, the movement of both formants correlated with both tongue height and tongue advancement.

Other research has painted a slightly more complex picture of the relationship between acoustics and kinematics. Mefferd and Green (2010) found that in the production of the vowel transition in /ia/ by typical speakers at different rates and loudness levels, kinematic measures of changes in tongue displacement correlated strongly with changes in acoustic vowel distance; however, changes in tongue movement spatiotemporal variability did not correlate with changes in formant variability. Yunusova et al. (2012) examined three vowel transitions in speakers with ALS and typical speakers, comparing the association between acoustic measures of F2 slope and F2 range with kinematic measures of tongue movement speed and displacement. The researchers found a moderately strong correlation between F2 slope and tongue movement speed. However, the association between F2 range and tongue displacement was much weaker. Dromey et al. (2013) examined the diphthongs /ɔɪ/, /aɪ/, /aʊ/ and /eɪ/ and found that while F1 and F2 were often predictive of tongue movement, there were several exceptions, particularly in F1. F1 change for the diphthong /aɪ/ was highly predictive of kinematic measurements across speakers, and /aʊ/ and /eɪ/ had moderately strong correlations between acoustic and kinematic measures. However, the size of the tongue displacement did not correlate with the strength of the acoustic–kinematic

relationship across speakers. Additionally, in a large majority of speakers, F1 movement for /ɔɪ/ did not correlate with vertical tongue movement as anticipated.

These studies have begun to reveal nuances in the relationship between formants and tongue movement. It would be useful to further expand on these findings by addressing some of the limitations of these studies. None of the studies mentioned above have reported on articulators other than the tongue. Dromey et al. (2013) and Yunusova et al. (2012) each tracked one fleshpoint on the tongue. Mefferd and Green (2010) and Lee (2014) each tracked several fleshpoints on the tongue. None of the studies tracked movement of the lips. In the current study, sensors were placed on the tongue and the upper and lower lip of participants. The purpose of these placements was to gather data about how the movement of articulators other than the tongue may contribute to changes in formant histories. Dromey et al. (2013) suggested that protrusion of the lips, which lengthens the vocal tract and tends to lower all formant frequencies, may have influenced the strength of the acoustic–kinematic relationship in diphthongs, particularly for /ɔɪ/, which was produced in the context of the word *boy*. The bilabial feature of /b/ may have been preserved through the /ɔɪ/ diphthong. Dromey et al. noted that the diphthong /aɪ/, which had a stronger acoustic–kinematic relationship, was produced between alveolar sounds. Examining data from sensors on the upper and lower lips of speakers will allow us to examine the contribution of these articulators to changes in F1 and F2 in diphthongs.

Another limitation addressed in this study was the possibility of context effects. Of the four acoustic–kinematic studies discussed above, Dromey et al. (2013) sampled the most comprehensive list of sounds, recording four English diphthongs. Yunusova (2012) examined three vowel transitions, and Mefferd and Green (2010) and Lee (2014) each considered just one vowel. The vowels in these studies were sampled in a variety of contexts. Each study used

stimuli that embedded the target vowels in sentences. Lee (2014) chose to use an hVd context to minimize coarticulation effects. This study showed the clearest and cleanest relationship between acoustic and kinematic measurements. In the current study, we chose to examine diphthongs in isolation in order to eliminate the effects of coarticulation.

In the current study, we compare formants from productions of the American English diphthongs /aɪ/, /aʊ/, and /ɔɪ/ with kinematic data from the tongue and lips in order to examine how closely acoustic data reflect the measured movements.

Method

Participants

Twenty individuals with normal speech (as judged by the experimenters), 10 men and 10 women, took part in the study. Their ages ranged from 20 to 34, and the median age was 25. All were native speakers of American English with no identifiable regional accent. Data from three of the speakers were not included in the study due to formant tracking errors in PRAAT. The remaining 17 speakers whose data are reported here included 9 males and 8 females. Before data collection, participants signed a consent form, which had previously been approved by the Brigham Young University Institutional Review Board. After data collection, each participant received compensation of $10.

Stimuli

Participants read four sets of stimuli (see Table 1). List A comprised a set of individual diphthongs: /ɔɪ/, /eɪ/, /aʊ/, /aɪ/, and /oʊ/. Participants were given an example of how to pronounce the diphthong (e.g., *oy as in "boy"*) but produced the diphthongs in isolation when reading the list. List B included each of the diphthongs embedded in an hVd context. List C included each diphthong in an rVl context. List D included a diphthong-loaded sentence: *The boy gave a shout*

at the sight of a cake, you know. These sets of stimuli added one vowel, /oʊ/, to the four vowels

examined in Dromey et al. (2013). Participants read each set of stimuli five times through.

Table 1

Stimuli

List A	List B	List C	List D
oy as in "boy"	I say hoyed again	roil	The boy gave a shout
ay as in "day"	I say hayed again	rail	at the sight of a cake,
ow as in "cow"	I say how'd again	rowel as in "vowel"	you know.
i as in "tie"	I said hide again	rile	
o as in "hoe"	I say hoed again	role	

The order of presentation of the lists was randomized for each participant. For the

purposes of the present study, we report on the isolation context only. We also chose to report

only on /ɔɪ/, /aʊ/, and /aɪ/, since these diphthongs involve more movement between the onset and

offset vowels.

Equipment

During each speech task, the acoustic signal was recorded into a Dell Optiplex 990

computer via an AKG C2000B microphone that was positioned approximately 30 cm from the

speaker's mouth. The acoustic signal passed through a Focusrite Scarlett 2i2 preamplifier. The

kinematic signal was tracked using an NDI Wave system. Both the acoustic and kinematic

signals were recorded with NDI WaveFront software. The microphone signal was sampled at

22050 Hz and the kinematic signal at 400 Hz.

Procedure

Each participant sat in an Acoustic Systems sound-attenuating booth on a chair,

approximately 90 cm from the stimuli, which were printed in black, 36-point font on white

paper. Using latex gloves, tongue depressors, and PeriAcryl®90 cyanoacrylate tissue adhesive,

the experimenters glued 5 electromagnetic sensors as follows: (TM) on the superior surface of

the tongue, approximately 3 cm posterior to the tip, at midline; (TT) on the superior surface of the tongue, 1 cm posterior to the tip, at midline; (J) on the lower incisors, at midline; (UL) on the vermillion border of the lower lip at midline; (LL) on the vermillion border of the upper lip at midline. Silver DriAid tongue drying pads were used to prepare the tongue for the tissue adhesive. Rather than being placed directly on the lower incisors, the J sensor was glued to a small patch (approximately 5 x 10 mm) of Stomahesive that had been placed over the teeth to prevent possible damage to the enamel. A reference sensor was attached to the bridge of an eyeglass frame (without lenses) that participants wore during speech tasks. The data collected in this study were part of a larger research project. Participants spoke continuously for at least 20 minutes to adapt to the presence of the sensors before the data included in this study were collected.

Data Analysis

The NDI Wave system generated time-aligned output files for audio and kinematic data. These records were imported into a custom MATLAB application, which was used to segment the target diphthongs from the audio recording. Diphthongs were visually segmented from the microphone waveform display, and segmentation points were confirmed using audio playback. MATLAB exported each diphthong segment as a wav file for audio and as a text file for the sensor movements. All kinematic signals were low-pass filtered at 10 Hz to remove noise.

The isolated diphthong audio recordings were analyzed with PRAAT (version 5.4.17) acoustic analysis software to extract the F1 and F2 histories during the diphthongs. The display was adjusted to show 5 formants with a window length of 25 ms and a dynamic range of 30 dB. A default ceiling frequency of 5500 Hz was used for all audio recordings for the women, and a 5000 Hz default ceiling for the men. However, adjustments were made to the ceiling to correct

for formant tracking errors for individual speakers. During analysis, there were several instances of discontinuity in the formant tracking that required manual correction. The formant records were exported from PRAAT with formant values recorded at 1 ms intervals. The text files generated by PRAAT were re-imported into the MATLAB application, where they were time-normalized along with the kinematic record using a linear Fourier interpolation algorithm to ensure equivalent sampling intervals for all data.

Finally, a file was exported from the MATLAB application that contained the F1 and F2 histories, along with time-aligned records for each sensor. This record included only the middle 50% of each diphthong in order to exclude formant tracking errors, which were common in the onset and offset of the diphthong. Each middle 50% of the diphthong contained 500 data points for the kinematic track and for the acoustic track, regardless of the actual segment duration in ms. This was possible because of the Fourier time normalization process that equalized the record lengths.

In order to focus on tongue and lip movements, the number of variables was reduced by deriving just four metrics from three of the five sensors. Since mid and front tongue share a great deal of movement, we determined to focus on just the x and y movements of TT, the tongue tip sensor. To measure lip protrusion, we examined the x movement of LL, the lower lip sensor. To measure lip aperture, we computed the Euclidian distance between LL and UL, the lower and upper lip sensors, respectively. The variables for our analysis of the connection between movement and acoustics were as follows (see Table 2):

Table 2

Variables

Independent Variables	Dependent Variables
1. Tongue advancement (tongue tip sensor, x)	1. F1 track during diphthong transition
2. Tongue height (tongue tip sensor, y)	2. F2 track during diphthong transition
3. Lip protrusion (lower lip sensor, x)	
4. Lip aperture (distance between lower lip & upper lip sensors)	

Correlation and regression models were initially considered but subsequently rejected for evaluation of the contribution of individual articulator movements to the diphthong acoustics. These statistical methods assume independence of samples and normal distribution of the data. They lack validity for time-series data because such datasets artificially inflate the metrics for correlation and regression models. Therefore, a novel method for analyzing the data was devised.

The kinematic and acoustic records of the diphthong transitions were converted into z-scores with a custom MATLAB application. This eliminated the difference in units (mm and Hz) so that the kinematic and acoustic data could be plotted on the same axes. The absolute z-score difference between the kinematic data track and the acoustic data track of each diphthong was computed at each of the 500 points for each variable, and these differences were summed to generate an index that reflected the association between the movement and acoustic variables. A lower value reflected a closer match between the movement and acoustic records. It was anticipated that relationships between independent and dependent variables would be either direct or inverse, given that numerous studies have shown F1 to be inversely related to tongue height and F2 to be directly related to tongue advancement. Thus, direct and inverse relationships between each kinematic variable and F1 and F2 were computed.

In a preliminary analysis, Figure 1 and Figure 2 illustrate how formant tracks and kinematic tracks can be plotted on the same graph after the data are z-transformed. In Figure 1, F1 is plotted along with the positive values for all kinematic variables. There appears to be a direct relationship between F1 and lip aperture (*lip ap*) and possible inverse relationships between tongue x, tongue y, and lip protrusion (*lip x*), respectively, and F1.
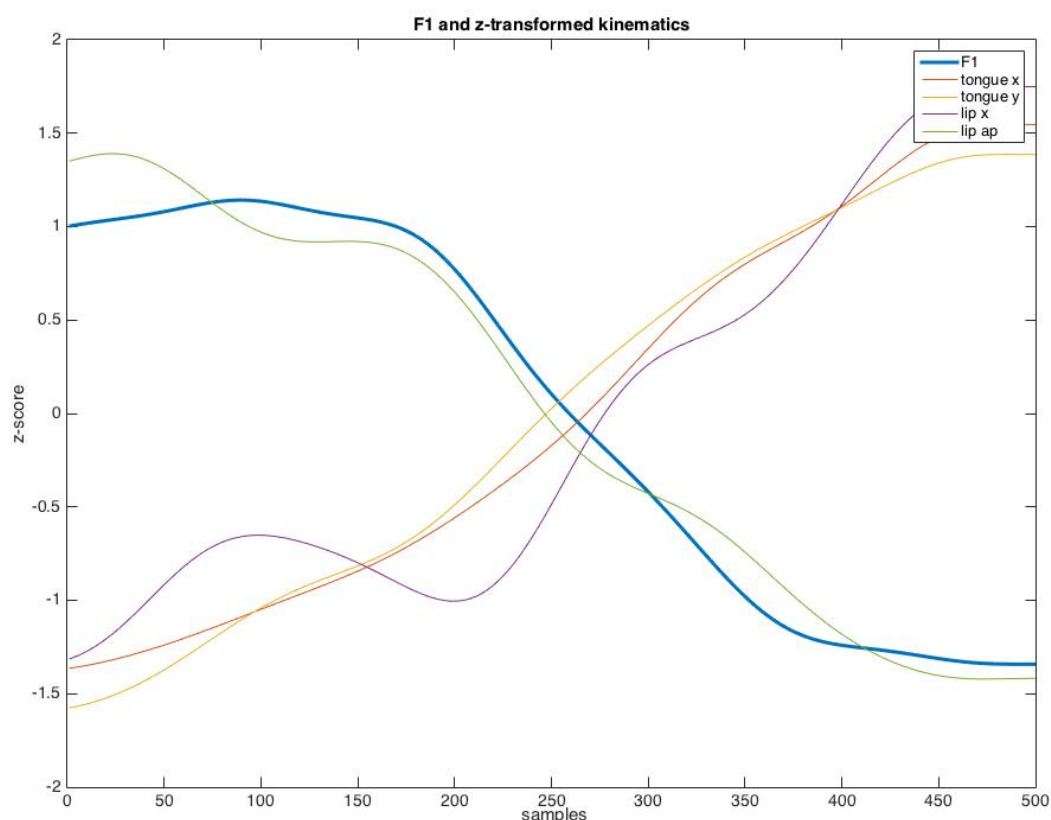
Figure 1. F1 and z-transformed kinematic variables of the first /aɪ/ token for subject F1. The F1 track is plotted together with positive values of each kinematic variable.

The same variables can be plotted again with inverted (negative) values for tongue x, tongue y, and lip protrusion, along with the positive values for lip aperture. When laid over the F1 track, the plot in Figure 2 reveals how closely each kinematic variable corresponds to changes in F1 movement during the middle 50% of the diphthong, across the 500 samples.
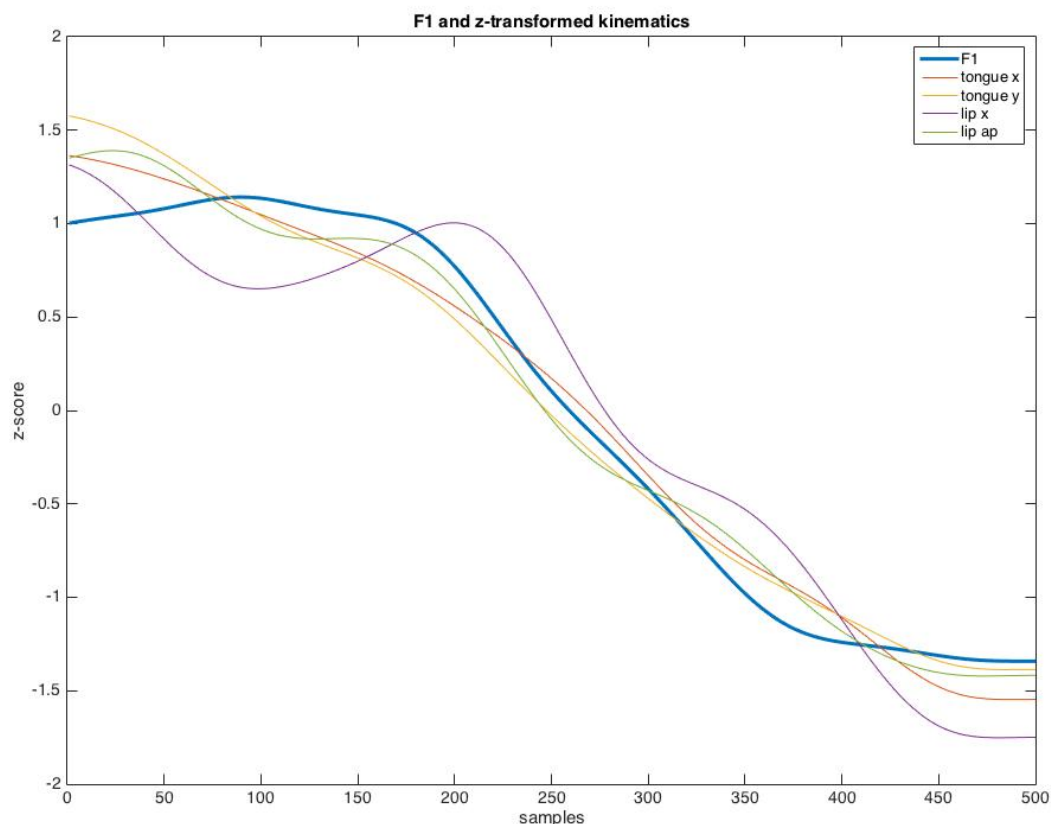
Figure 2. F1 and z-transformed kinematic variables of the first /aɪ/ token for subject F1. The F1 track is plotted together with negative values for all tongue x, tongue y, and lip x variables. Lip aperture (*lip ap*) is plotted with positive values.

The z-score difference sums for the 5 repetitions of each diphthong were averaged for individual speakers, both for the inverse and direct relationship between the acoustic and kinematic signals. Plots of the means and 95% confidence intervals were graphed using SPSS software to compare the relationships between the F1 and F2 transitions and movement patterns in the four articulatory metrics. The closer the match between an F1 or F2 track and a kinematic track, the closer the z-score difference sum was to zero. Therefore, a value closer to zero suggested that the specific kinematic variable was a better predictor of the change in F1 or F2 than a mean further from zero. Figure 3 shows an example of the z-score sums for each

kinematic variable compared with F1 and F2 data for the diphthong /aʊ/ from a single speaker. Each of the 4 kinematic variables was plotted twice—once showing a potential direct relationship, and once showing a possible inverse relationship—against F1 and F2, for a possibility of 16 relationships.
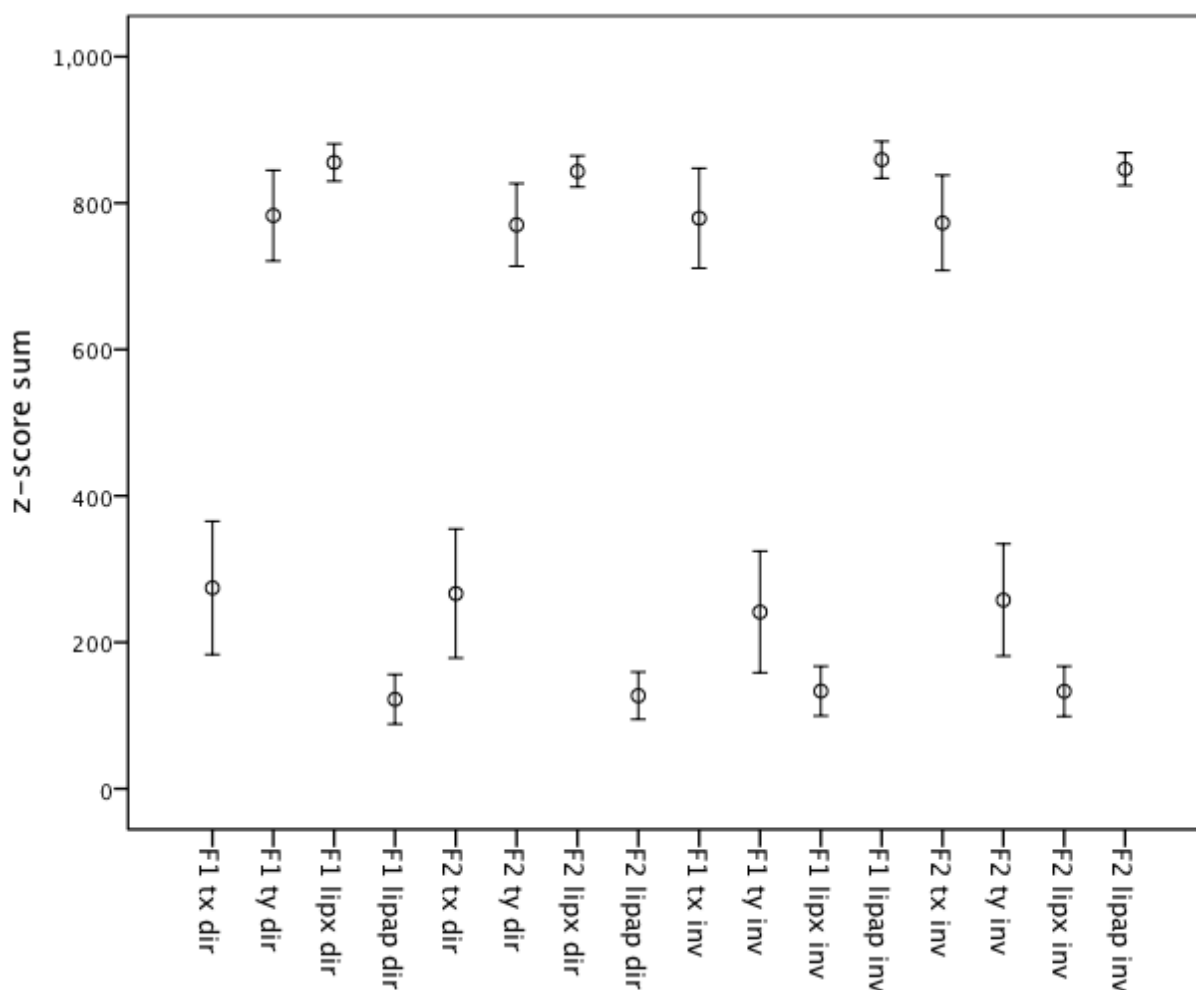
Figure 3. Mean (and 95% confidence intervals) of z-score difference sums for direct and inverse kinematic variables. Variables closer to zero contribute to changes in F1 and F2 during the /aʊ/ transition.

The kinematic variables (used as predictors of the diphthong acoustics) for the three diphthongs naturally separated out into more predictive or less predictive variables. This was

anticipated for the following reason: If a kinematic/acoustic variable pair with a direct association had a lower z-score difference sum (reflecting a stronger contribution of the movement to the acoustics), then the inverse relationship for the same variable pair would be weaker (a higher difference sum). For the z-score sum graphs included in the next section, we show only the more predictive variables (i.e. the lower section of the plot in Figure 3) for each of the three diphthongs examined.

Results

Because of the exploratory nature of the analyses used in the current study, inferential statistics were not applied. Instead, the observed patterns and differences will be presented descriptively.

/aɪ/ Diphthong

Results for /aɪ/ show that as the tongue advanced, F1 decreased and F2 increased, as expected (see Figure 4), based on previous studies. As tongue height increased, F1 decreased and F2 increased. Examination of the lip measures showed that as lip protrusion increased, F1 decreased and F2 increased. As lip aperture increased, F1 increased and F2 decreased. The difference sums for tongue advancement and tongue height were closer to zero than the values for lip protrusion and lip aperture for both F1 and F2. This suggests that tongue advancement and height were more predictive of changes in F1 and F2 than lip protrusion and aperture for the diphthong /aɪ/.

Figure 4. Mean (and 95% confidence intervals) of z-score difference sums for all speakers for kinematic variables contributing to changes in F1 and F2 during the /aɪ/ transition.

A look at the formant and kinematic tracks of the first /aɪ/ token for each speaker plotted together (see Figure 5) suggests that, aside from several outliers, data patterned fairly uniformly across speakers.

Figure 5. Formant and kinematic tracks from all speakers for /aɪ/. The first token of the

diphthong /aɪ/ produced by each speaker is plotted to allow a visual evaluation of interspeaker

variability.

/aʊ/ Diphthong

Results for /aʊ/ show that as tongue advancement increased, F1 and F2 increased (see

Figure 6). Conversely, as tongue height increased, F1 and F2 both decreased. As lip protrusion

increased, F1 and F2 decreased. As lip aperture increased, F1 and F2 increased. The respective

means of lip protrusion and lip aperture were closer to zero than the respective means of tongue

advancement and tongue height for both F1 and F2, suggesting that lip movement was more

predictive of changes in F1 and F2 for the diphthong /aʊ/.
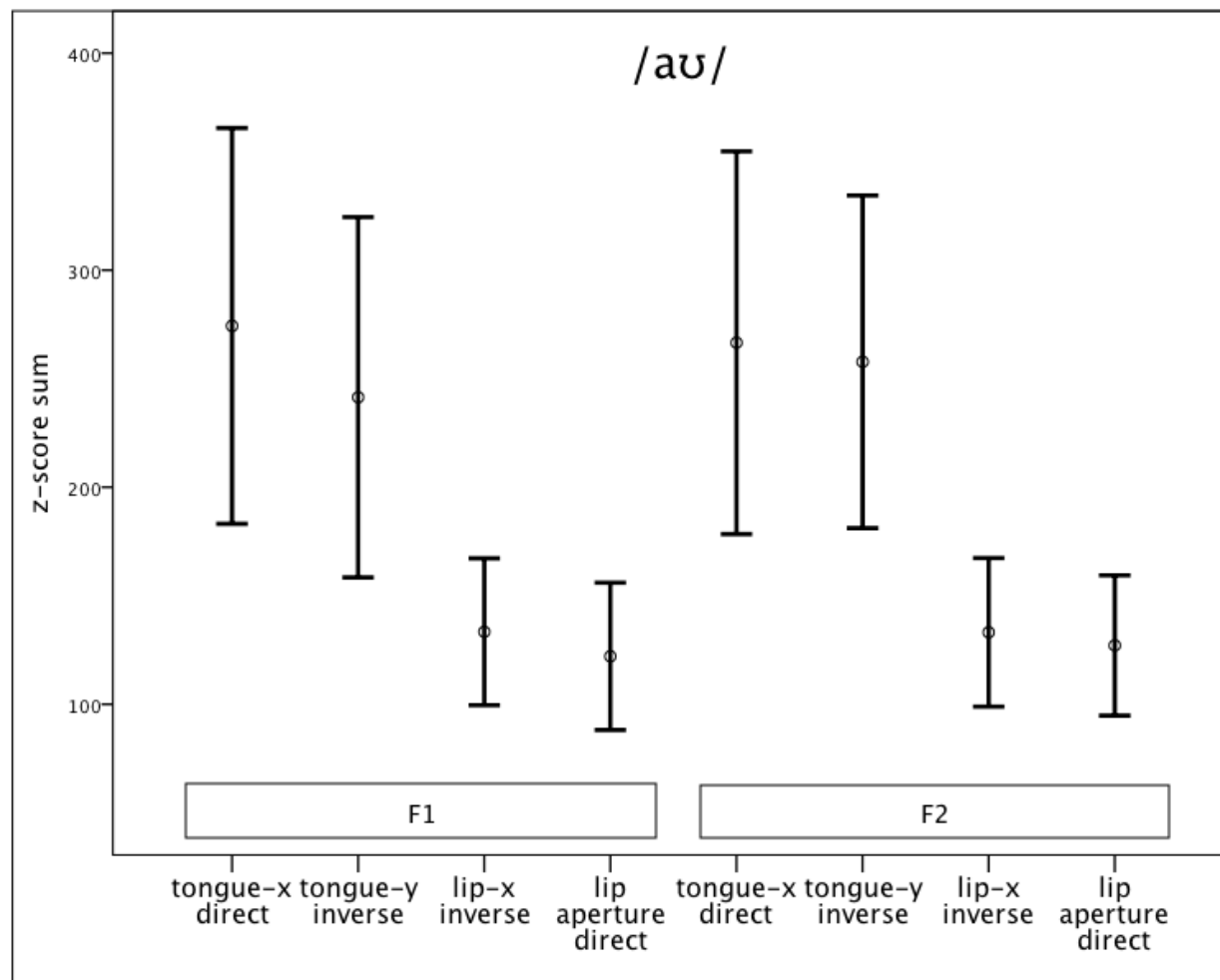
Figure 6. Mean (and 95% confidence intervals) of z-score difference sums for all speakers for kinematic variables contributing to changes in F1 and F2 during the /aʊ/ transition.

Plotting individual tokens for all 17 speakers together (see Figure 7) reveals that the data were somewhat less uniform in the diphthong /aʊ/ than in the diphthong /aɪ/. There appears to be more variation among the speakers.
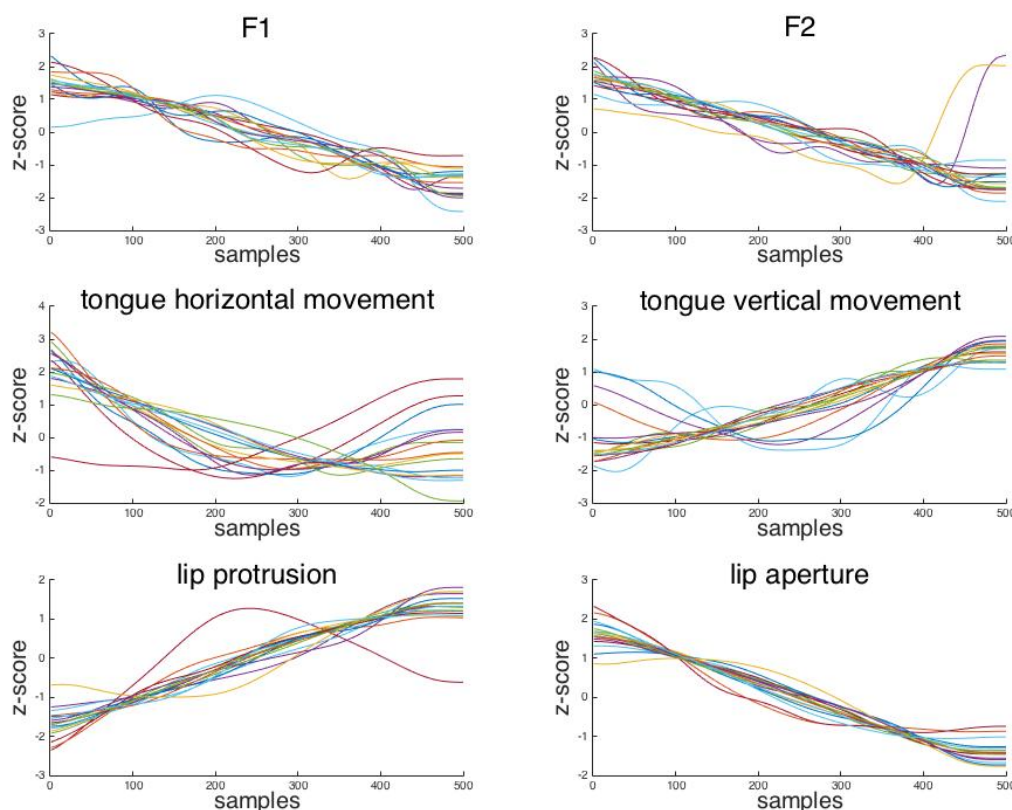
Figure 7. Formant and kinematic tracks from all speakers for /aʊ/. The first token of the diphthong /aʊ/ produced by each speaker is plotted to allow a visual evaluation of interspeaker variability.

/ɔɪ/ Diphthong

Results for the /ɔɪ/ diphthong show that all of the movement variables were less predictive of changes in F1 than changes in F2 (see Figure 8). As tongue advancement increased, the F2 frequency increased. As tongue height increased, F2 also increased. As lip protrusion increased, F2 decreased. However, as lip aperture increased, F2 increased. The mean of the z-score difference sums for the combination of tongue advancement and F2 was much closer to zero than the means for the three other movement variables for F2, suggesting that tongue advancement was a stronger contributor to changes in F2 than any of the other variables.
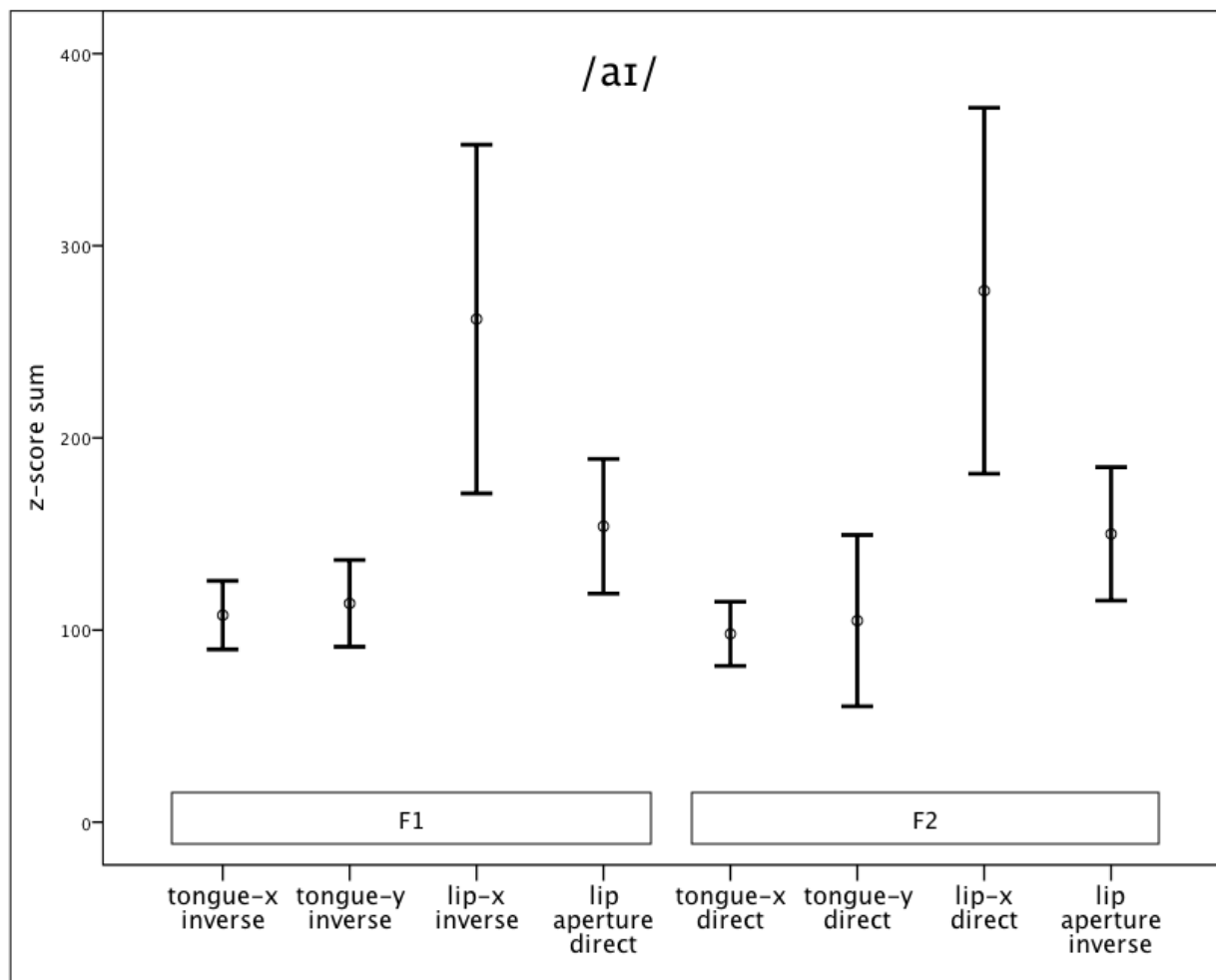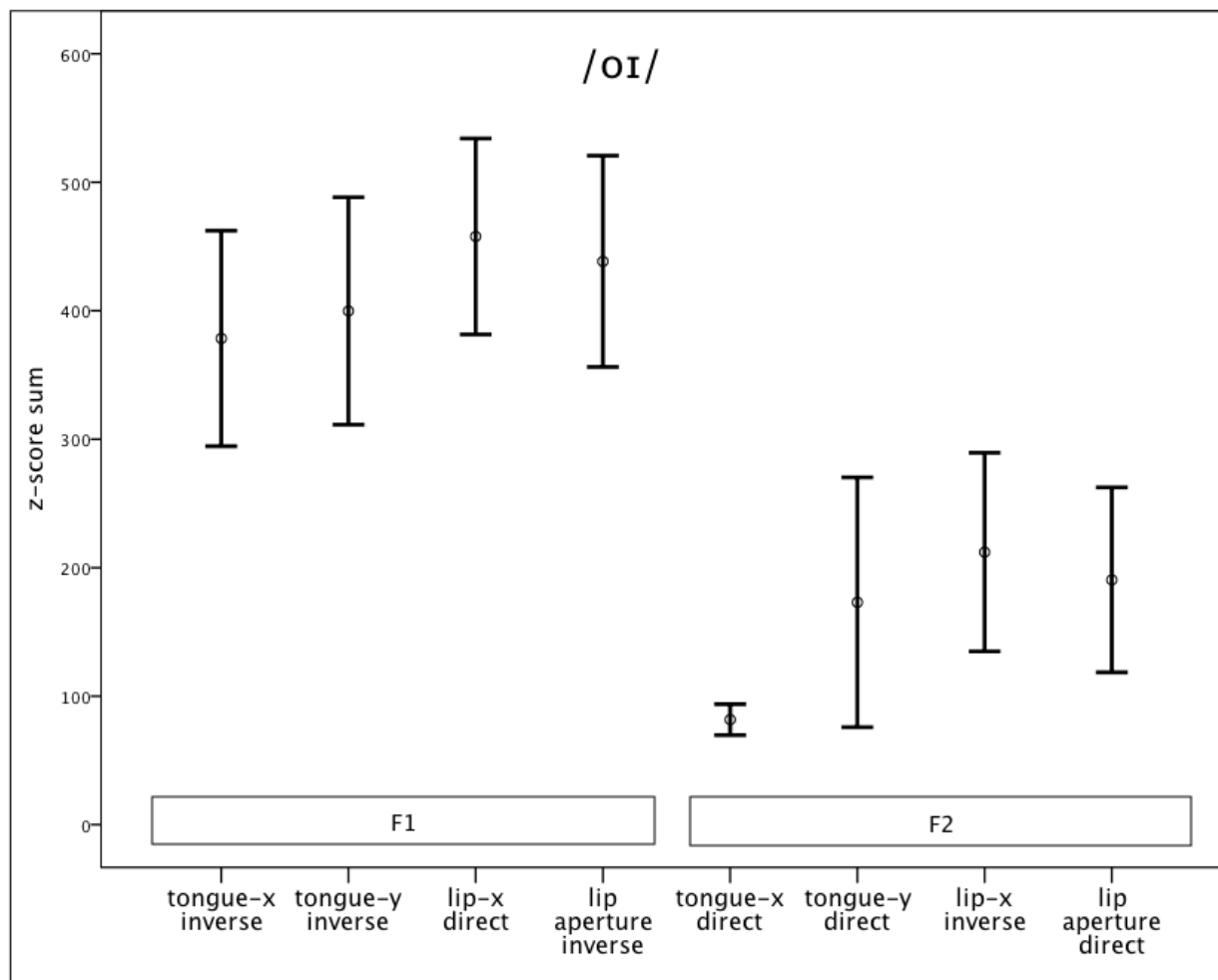
Figure 8. Mean (and 95% confidence intervals) of z-score difference sums for all speakers for kinematic variables contributing to changes in F1 and F2 during the /ɔɪ/ transition.

As mentioned, all movement variables seemed to be more weakly associated with changes in F1 than with changes in F2. The respective means of tongue advancement and tongue height were slightly closer to zero than the means of lip protrusion and lip aperture, suggesting tongue movements may be slightly more predictive of changes in F1. However, the error margins of all four movement variables overlap, suggesting the slight differences may not be meaningful.

Plots comparing the first /ɔɪ/ token of each speaker (Figure 9) show noticeable variation in F1 formant tracks across speakers. Variation in F1 is greater in /ɔɪ/ than either /aɪ/ or /aʊ/.
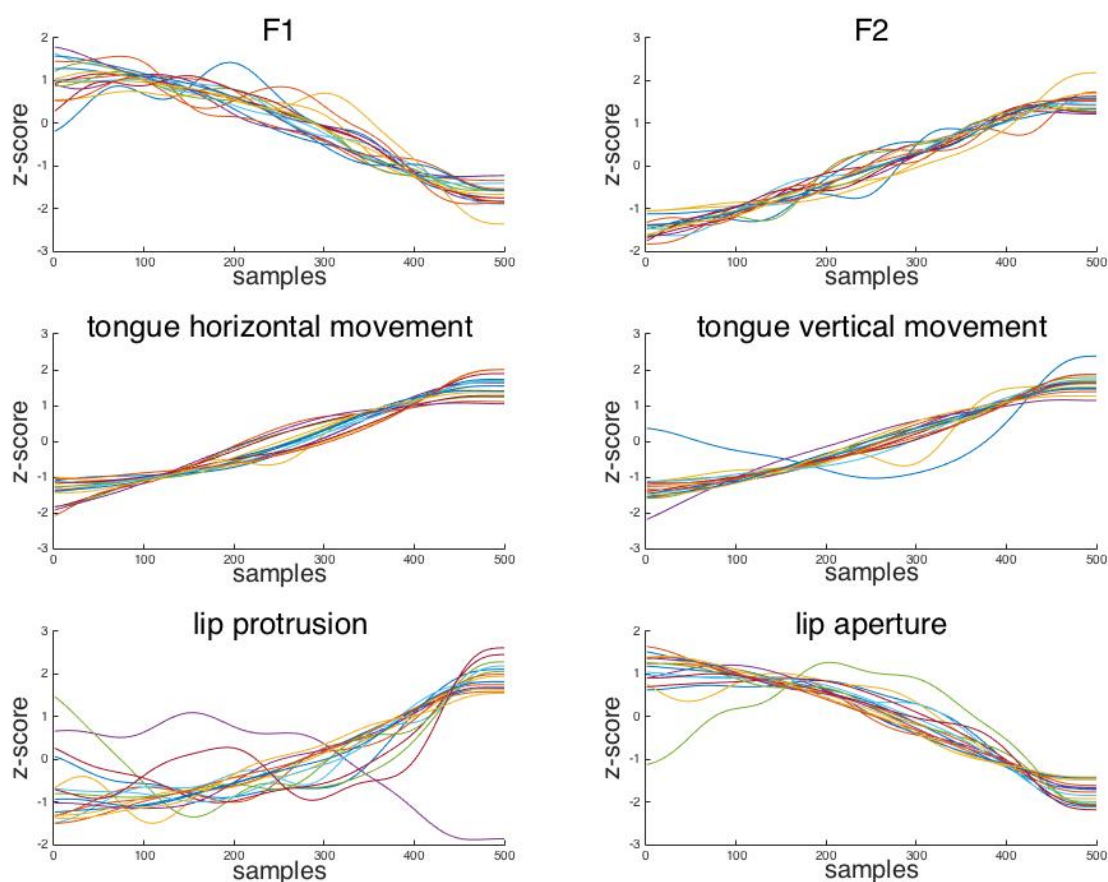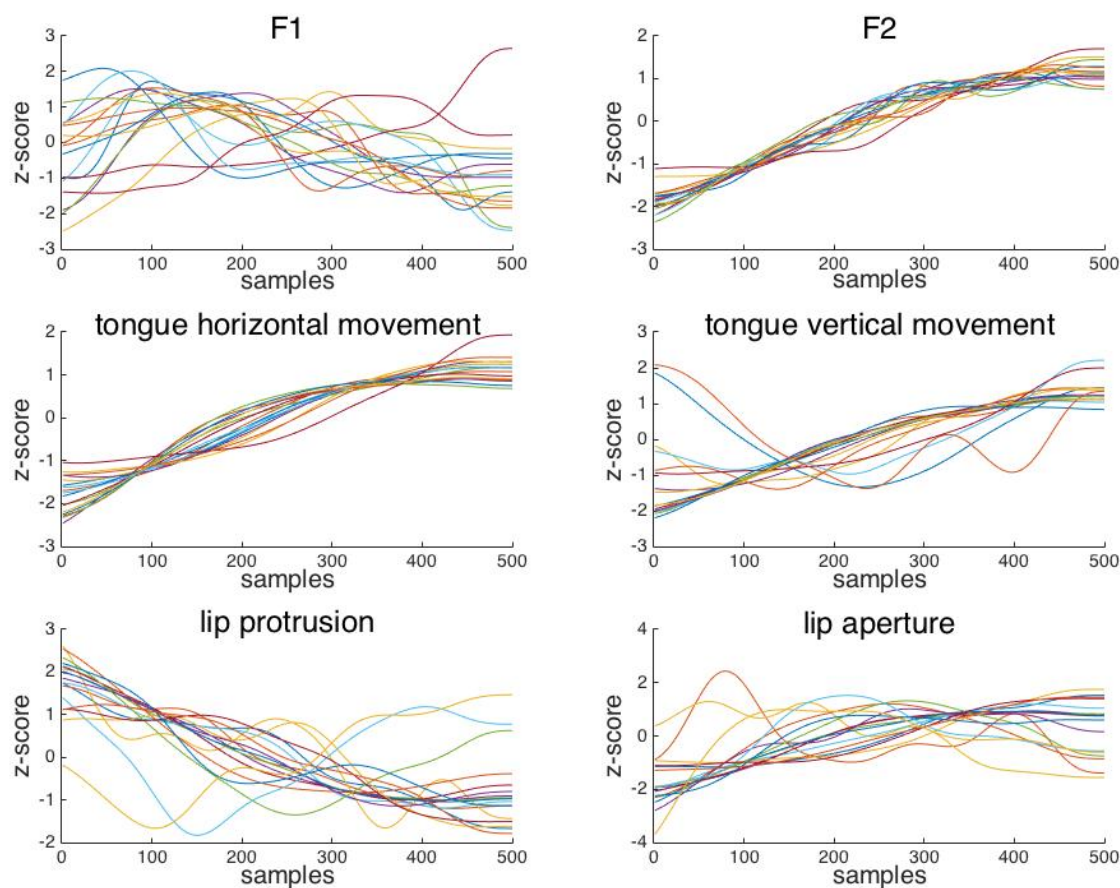
Figure 9. Formant and kinematic tracks from all speakers for /ɔɪ/. The first token of the diphthong /ɔɪ/ produced by each speaker is plotted to allow a visual evaluation of interspeaker variability.

An informal comparison within speakers showed greater variation in F1 tracks in /ɔɪ/ than was seen in other diphthongs for a number of speakers. Figure 10 shows a representative example from a single speaker of a comparison between F1 and the other acoustic and kinematic variables. For this speaker, the F1 tokens varied much more widely than any of the other variables.
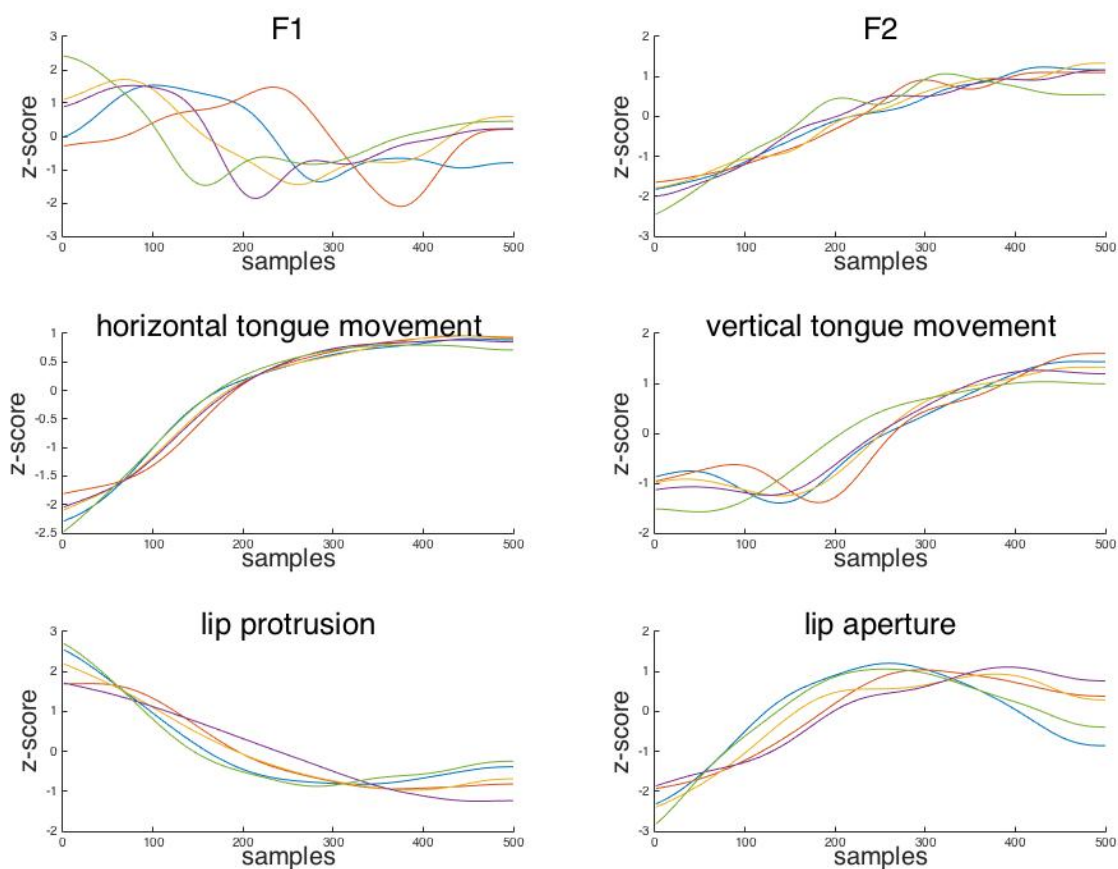
Figure 10. Formant and kinematic tracks of /ɔɪ/ tokens for subject F9. All 5 tokens of /ɔɪ/ for subject F9 were plotted for F1, F2, and each kinematic variable.

Another plot from the same speaker (see Figure 11) shows a comparison of F1 tracks between /aɪ/, /aʊ/, and /ɔɪ/ in one speaker. The formant tracks appear to be more variable for /ɔɪ/ than for /aɪ/ or /aʊ/.
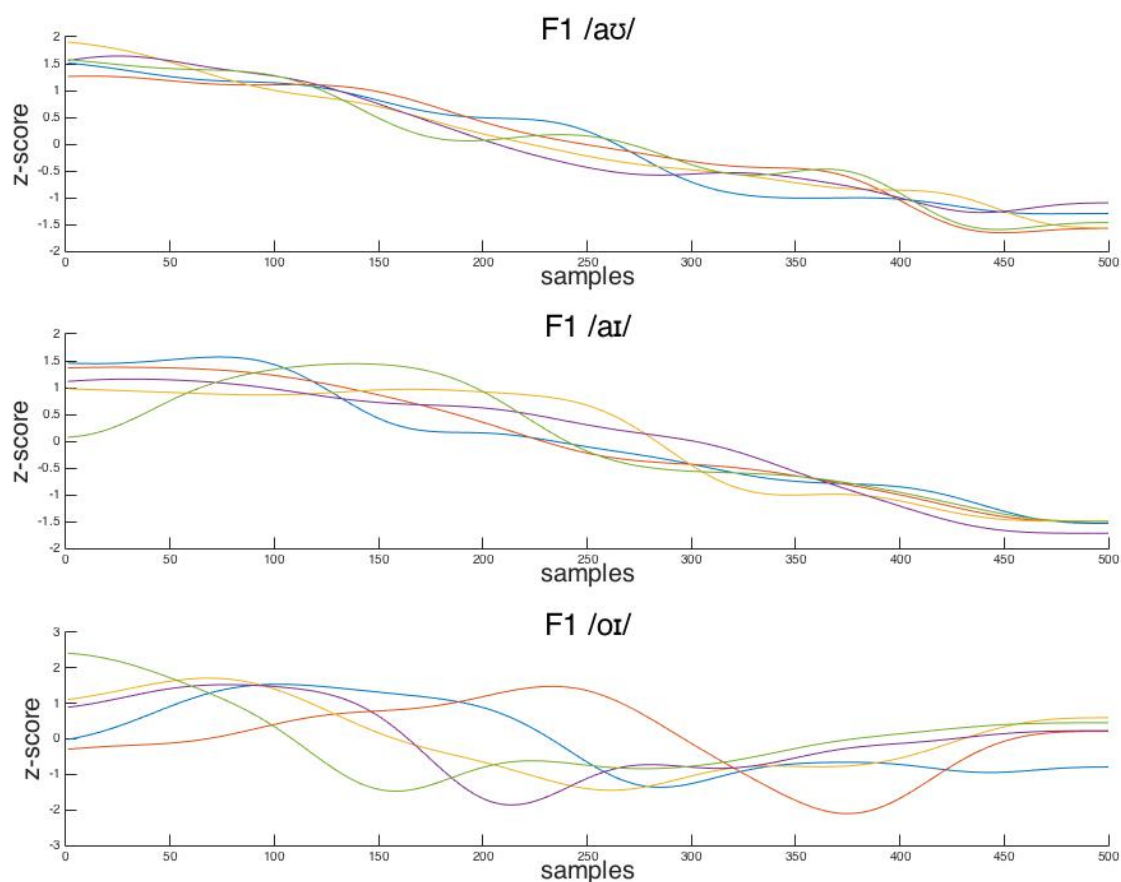
Figure 11. F1 tracks for all tokens of /aɪ/, /aʊ/, and /ɔɪ/ for subject F9. The variability between F1 tracks for all tokens of the 3 diphthongs is shown.

Discussion

The purpose of this study was to examine how well movements of the tongue and lips predicted changes in F1 and F2 during the English diphthongs /aɪ/, /aʊ/, and /ɔɪ/. The results suggest that the association between tongue movement and changes in F1 and F2 may be more complex than generally assumed, in part because of the contribution of the lips.

In the diphthong /aɪ/, the typical associations between tongue movement and changes in F1 and F2 were reflected in the data. As tongue height increased, F1 decreased; and as the tongue advanced, F2 increased. These relationships between tongue movement and changes in

formants have been observed numerous times in previous research and were expected (Kent et al., 1999). In directly comparing acoustic and kinematic data, Dromey et al. (2013) found that changes in F2 had a strong positive correlation with tongue advancement in the diphthong /aɪ/, and F1 had a strong negative correlation with tongue height. Of the four diphthongs examined in Dromey et al. (2013), /aɪ/ exhibited the strongest relationship of any of the diphthongs between changes in F1 and tongue height.

A less anticipated finding in the current study was that increases in tongue height and tongue advancement were both associated with the predicted decrease in F1 and increase in F2. This result is consistent with recent work by Lee (2014), who found that F1 and F2 both correlated with tongue height and tongue advancement in the diphthong /aɪ/. The author provided a caution for this finding, however, that is relevant to the current study. Lee pointed out that F1 and F2 are inherently correlated in the diphthong /aɪ/. In this diphthong, the tongue moves from a position lower and further back to a position higher and more front, meaning the tongue both rises and advances forward throughout the diphthong. Because the tongue is rising and advancing forward from /a/ to /ɪ/, we would expect both F1 and F2 to change during the diphthong. A diphthong in which the tongue was not making the same movement from low to high and back to front may not demonstrate the same strength of association between tongue movements and changes in F1 and F2.

Another finding for the diphthong /aɪ/ was that lip aperture seemed to make a stronger contribution to changes in F1 and F2 than lip protrusion. When considering the features of the vowels that make up /aɪ/, this finding was not unexpected. Neither /a/ nor /ɪ/ have the feature of lip rounding. However, they do differ in tongue height. The vowel /a/ is low, while the vowel /ɪ/ is high. Thus, the jaw is more open at the beginning of the transition and more closed at the end.

The lips are coupled with the jaw, and so one would expect the lips to also move from a more open to a more closed position from the beginning to the end of the transition. Dromey, et al. (2013) noted that "the generally accepted view is that F1 is strongly influenced by the height of the tongue and jaw and F2 is to a large extent linked to tongue advancement in the mouth" (p. 316). The results in the current study suggest that lip aperture (which is inherently coupled with jaw height) has a similar influence on both the F2 and the F1 frequency.

In the diphthongs /ɔɪ/ and /aʊ/, the relationships between tongue movement and changes in F1 and F2 were less straightforward than were seen in /aɪ/. In /aʊ/, as tongue advancement increased, both F1 and F2 increased. As tongue height increased, both F1 and F2 decreased. The association between tongue advancement and an increase in F2 was expected, as was the association between tongue height and a decrease in F1. Previous research by Dromey et al. (2013) also found a modest negative association between tongue height and changes in F1 in the diphthong /aʊ/. As in the current study, the association between F1 and tongue height was weaker than it was in the diphthong /aɪ/. However, Dromey et al. (2013) found that the positive correlation between F2 and tongue advancement was stronger in /aʊ/ than in /aɪ/, something we did not find in the current data.

The associations between tongue advancement and an increase in F1 and between tongue height and a decrease in F2 were not expected. However, tongue movements overall were less closely associated with changes in F1 and F2 in /aʊ/ than lip movements were. The changes in lip protrusion and lip aperture more closely matched the changes in formants throughout the diphthong. As the lips protruded, F1 and F2 both decreased. As lip aperture increased, both F1 and F2 increased. It is not entirely unexpected that lip movements in /aʊ/ would have a greater influence on the changes in formants than lip movements in /aɪ/, because /aʊ/ has greater lip

protrusion during the diphthong than /aɪ/ does. As mentioned, lip rounding lengthens the vocal tract and thus tends to cause all formants to decrease in frequency (Kent et al., 1999). What is notable is that the lip movements in /aʊ/ were overall more predictive of formant changes than the tongue movements were. This suggests that for /aʊ/, and perhaps for other vowels with lip rounding, tongue movements may not be the most prominent contributor to changes in F1 and F2.

In the diphthong /ɔɪ/, the association between movement variables and changes in F1 and F2 was even more surprising than it was in /aʊ/. Tongue advancement was the variable most closely associated with F2, which was expected. Dromey et al. (2013) found the positive correlation between tongue advancement and F2 was stronger in /ɔɪ/ than in any of the other diphthongs examined. What was somewhat surprising was that in the current study, F1 did not seem to have a close association with any of the tongue or lip movement variables. This finding was unexpected but not without precedent. Dromey et al. (2013) also found movement variables to be poorly associated with changes in F1 for the diphthong /ɔɪ/. Only one of the 20 speakers in the study showed a strong negative correlation between F1 and tongue height, while the remaining 19 speakers were divided between weak negative and positive correlations. The authors suggested coarticulation effects might have partly accounted for this result. The diphthong /ɔɪ/ was produce in the word *boy*, and the bilabial feature of the word-initial /b/ may have been preserved through the diphthong. The preserved lip-rounding feature might have strengthened the association between the lip movements and the changes in the formants while weakening the association between the tongue movements and the changes in the formants.

The findings in the current study, however, show that when /ɔɪ/ is produced in isolation, the same result occurs. When /ɔɪ/ was spoken in isolation, the association between vertical

tongue movement and changes in F1 was weak, and there was a great deal of variability in the data across participants and even within participants. This suggests that, as Dromey et al. (2013) hypothesized, lip movements influence changes in F1 for the diphthong /ɔɪ/. However, preserved lip-rounding from the phoneme /b/ may not completely account for the influence of the lips. Lip-rounding of the /ɔ/ vowel, even when not coarticulated with /b/, seems to have a similar effect on changes in F1 during the diphthong.

One potential explanation for why the results for /ɔɪ/ seem to be particularly complex is that the tongue and lips may be having opposing acoustic influences throughout the diphthong. In /ɔɪ/, the tongue moves from low to high through the transition, which would typically correspond with a decrease in F1. However, also through the transition, the lips are moving from a protruding posture to a more neutral posture, which movement would typically result in an increase in all frequencies as the length of the vocal tract decreases. It may be that F1 is sensitive both to vertical tongue movement and to this overall change in frequencies caused by the lips. Because F1 is simultaneously influenced by vertical tongue movement that would cause the formant frequency to decrease and lip retraction that would cause the formant frequency to increase, neither vertical tongue movement nor lip retraction show a close association to changes in F1. It could be speculated that F1 is in an acoustic "tug-of-war" that neither lip movement nor tongue movement appear to be winning during the diphthong.

Limitations of the Current Study and Directions for Future Research

One limitation of the current study is that because descriptive rather than inferential statistics were used, we cannot conclude that an association between an independent and dependent variable is statistically significant, nor describe in conventional terms what the strength of the relationship is.

Another limitation comes from the complexity of the speech mechanism. Individual morphology of the anatomical structures for speech varies from speaker to speaker and may affect an individual's productions. Additionally, structures other than the tongue and lips can influence the resonance of the vocal tract. Since we measured only one point on the tongue and one on each lip, our data do not account for acoustic changes that may be attributed to, for example, the length of an individual speaker's vocal tract or the movement of an individual speaker's larynx.

While using the middle 50% of the diphthong reduced formant tracking errors considerably, there may have been instances where formant tracking errors extended within the middle 50% that could have affected the data. Additionally, several diphthong tokens that were difficult to analyze with PRAAT's formant tracking algorithm required manual correction.

For individual speakers, only 5 tokens of each diphthong were used in the analyses. In some instances, there was a great deal of variation among the 5 tokens within a speaker, and in some cases, there may have been an outlier among the 5 tokens. Either of these scenarios would affect the average of the diphthong tracks for that speaker and would not best represent how the speaker typically produced the diphthong. To further examine the effects of lip protrusion on changes in F1 and F2, future research may benefit from a larger group of participants with more tokens from each speaker. In the current study, we only had one diphthong that lacked the lip-rounding feature and two diphthongs that had the lip rounding feature. Another recommendation for further research would be to include more vowels with and without the lip-rounding feature for a broader analysis of the effects of lip protrusion on formant changes.

Speakers in the current study lived in the same region. However, formal data was not collected from speakers to establish what variety of English each spoke or what regions of the

United States they had lived in previously. While this would not have influenced data that compared tongue and lip movement with formant movement within a speaker, it may have influenced overall results when tracking trends in tongue and lip movements. Further research may benefit from accounting or controlling for differences in varieties of English spoken by participants.

Conclusion

The current study demonstrates that speech articulators make different relative contributions to F1 and F2 in different English diphthongs. For the diphthong /aɪ/, which lacks lip rounding, tongue movement may be the best predictor of changes in F1 and F2. However, for the diphthong /aʊ/, which has significant lip rounding, lip movement may be a better predictor of changes in F1 and F2 than tongue movement. For the diphthong /ɔɪ/, tongue movement and lip movement that act in opposition with each other may muddy the acoustic waters and make it challenging to discover the association between articulator movements and formant changes. These data have important implications for researchers who use acoustic methods to make inferences about articulator movement during vowels. The current study used a novel method of data analysis. Future research comparing acoustic and kinematic data may benefit from using this approach to analyze similar sets of data.

References

Dromey, C., Jang, G. O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication, 55*, 315–328. doi:10.1016/j.specom.2012.09.001

Ferrand, C. T. (2007). *Speech science: An integrated approach to theory and clinical practice* (2nd ed.). Boston, MA: Allyn & Bacon.

Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica, 33,* 199–221.

Kent, R. D., Weismer, G., Kent, J. F., Vorperian H. K., & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress and potential. *Journal of Communication Disorders*, *32,* 141–186.

Lee, J. (2014). Relationship between the first two formant frequencies and tongue positional changes in production of production of /aɪ/. *Journal of the Acoustical Society of America, 135,* 2294. doi:10.1121/1.4877541

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 53,* 1206–1219. doi:10.1044/1092-4388(2010/09-0083)

Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /r/. *Language, 92*, 101–140. doi: 10.1353/lan.2016.0019

Moll, K. L. (1960). Cinefluorgraphic techniques on speech research. *Journal of Speech and Hearing Research, 3,* 227–241.

Roy, N., Nissen, S. L., Dromey, C., & Sapir, S. (2009). Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of Communication Disorders, 42,* 124–135. doi: 10.1016/j.jcomdis.2008.10.001

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C.. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. [Research Note]. *Journal of Speech, Language, and Hearing Research, 53,* 114–125. doi: 10.1044/1092-4388(2009/08-0184)

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3–45.

Stone, M., Langguth, J. M., Woo, J., Chen, H., & Prince, J. L. (2014). Tongue motion patterns in post-glossectomy and typical speakers: A principal components analysis. *Journal of Speech, Language, and Hearing Research, 57,* 707–717. doi: 10.1044/1092-4388(2013/13-0085)

Stone, M., Stock, G., Bunin, K., Kumar, K., Epstein, M., Kambhamettu, C., … Prince, J. (2007). Comparison of speech production in upright and supine positions. *Journal of the Acoustical Society of America, 122*, 532–541.

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47,* 766–783. doi:10.1044/1092-4388(2004/058)

Twist, A., Baker, A., Mielke, J., & Archangeli D. (2007). Are covert /ɹ/ allophones really indistinguishable? *University of Pennsylvania Working Papers in Linguistics, 13*, Article 16.

Weismer, G., Martin, R., Kent, R. D., & Kent, J. F. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *Journal of the Acoustical Society of America, 91,* 1085–1098.

Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica, 64*, 94–102. doi: 10.1159/000336890

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of retroflex and bunched American English /r/. *Journal of the Acoustical Society of America, 123*, 4466–4481. doi: 10.1121/1.2902168

APPENDIX A: ANNOTATED BIBLIOGRAPHY

Dromey, C., Jang, G. O., & Hollis, K. (2013). Assessing correlations between lingual
movements and formants. *Speech Communication, 55*, 315–328.
doi:10.1016/j.specom.2012.09.001

Objective: The authors explored the relationship between formants and tongue movements. The
purpose was to compare changes in F1 and F2 with magnetically tracked lingual movements to
see how closely the acoustic measures reflected the kinematic measures during diphthongs.
Method: Twenty participants repeated the sentence *The boy gave a shout at the sight of the cake*
three times. The acoustic signal was recorded via a microphone. A sensor was placed 1 cm
posterior to the tip of each participant's tongue at midline and recorded via a magnetic tracking
instrument. *Results:* Data analysis revealed that for F1, /aɪ/, /aʊ/ and /eɪ/ exhibited a negative
correlation between F1 movement and vertical tongue movement. The diphthong /aɪ/ showed the
strongest correlation, while /ɔɪ/ exhibited the weakest correlation. For /ɔɪ/, only one speaker had
a strong negative correlation between F1 and vertical tongue movement. For F2, all four
diphthongs exhibited a positive correlation between F2 and anteroposterior tongue movement for
the majority of speakers. *Conclusion:* Associations between formants and lingual movements
were variable across diphthongs. The relationship between formants and lingual movement may
be more complex that generally assumed. *Relevance to the current work:* The authors speculated
that coarticulation may have influenced the relationship between acoustic and kinematic data.

Ferrand, C. T. (2007). *Speech science: An integrated approach to theory and clinical practice*
(2nd ed.). Boston, MA: Allyn & Bacon.

Relevance to the current work: In discussing vocal tract resonance, the author explains that
vowels are characterized acoustically by the first three formants. Formant frequencies are related
to the size and shape of the oral and pharyngeal cavities. The author states that the frequency of
F1 is related to the volume of the pharyngeal cavity, while the length of the oral cavity influences
F2. The frequency of F1 is also influenced by how tight the vocal tract constriction is. As the
tongue moves forward and back and high and low, the volumes of the oral cavity and the
pharyngeal cavity change, which changes the frequencies of F1 and F2. As tongue height
increases, the pharyngeal cavity becomes larger, which decreases the frequency of F1. As the
tongue advances in the vocal tract, the oral cavity becomes smaller, which increases the
frequency of F2. This understanding of how tongue movement influences changes in the
frequencies of F1 and F2 is relevant to the current study, because our experiment examines this
relationship.

Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of
speech: Implications for the operation of motor equivalence. *Phonetica, 33,* 199–221.

Objective: This study examined how speakers accomplished articulatory goals and what role
motor equivalence played in that process. *Method:* Six native female speakers of American
English produced [hæbæb, hibib, hɛbɛb] in the carrier phrase *That's a _____ again*. A strain
gauge transduction system was used to track movements of the upper lip, lower lip, and jaw,
while audio was recorded with a microphone. Subjects repeated each sentence ten times at a

normal speaking rate and ten times at a rate that was faster than normal. *Results:* Data showed that the overall vertical opening of the lips had a small variation across repetitions. However, the displacement contributions of the lower lip and jaw had considerable variation. When the jaw had a relatively large contribution, the displacement contribution of the lower lip was relatively small, and vice versa. The jaw and lower lip seemed to be quite sensitive to changes in the movement of the other and were the primary contributors in the vertical opening of the lips. The upper lip contributed minimally to the vertical opening, although in several cases, it compensated for extremely reduced contributions of the jaw and lower lip. The degree of motor equivalence varied across speakers. Rate of speech did not appear to alter the degree of displacement of any of the articulators examined. *Conclusion*: Motor equivalence appears to be a principle that can be observed in the speech mechanism. *Relevance to current work:* Different speakers accomplish articulatory tasks in different ways. The current study took this into account.

Kent, R. D., Weismer, G., Kent, J. F., Vorperian H. K., & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress and potential. *Journal of Communication Disorders*, *32,* 141–186.

Objective: This study summarizes applications of acoustic analysis to dysarthric speech and proposes acoustic analyses that may be used to assess disordered speech and voice. *Relevance to the current work:* The paper states that as a general rule, F1 frequency has an inverse relationship with tongue height and F2 frequency has a direct relationship with tongue advancement. Additionally, lip rounding decreases the frequency of all formants. All formant frequencies are influenced by the length of the speaker's vocal tract. In the current study, we examine whether tongue movement is the primary contributor to changes in formant frequencies.

Lee, J. (2014). Relationship between the first two formant frequencies and tongue positional changes in production of production of /aɪ/. *Journal of the Acoustical Society of America, 135,* 2294. doi:10.1121/1.4877541

Objective: The purpose of the study was to examine the relationship between F1 and F2 trajectories and changes in tongue height and advancement during the diphthong /aɪ/. *Method:* Ten native speakers of American English participated in the study. They each produced three repetitions of the sentence *I say hide again.* Kinematic information was recorded via three coils attached to each speaker's tongue tip, body, and dorsum. Only data from the tongue body was analyzed for this study. Audio was recorded simultaneously with the kinematic data. *Results:* F1 decreased as tongue height increased and as the tongue advanced, and F2 increased at tongue advanced and as tongue height increased. F1 and F2 has similar relationships with both x and y movements. These correlations were significant within speakers and across speakers. *Conclusion:* F1 and F2 showed the expected relationship to tongue advancement and tongue height. *Relevance to the current work:* Changes in F1 and F2 were both associated with tongue height and tongue advancement.

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 53,* 1206–1219. doi:10.1044/1092-4388(2010/09-0083)

Objective: The authors studied the relationship between tongue kinematic and speech acoustic changes in response to adjustments to speaking rate and loudness. *Method:* Ten adults with typical speech produced the sentence *Tomorrow Mia may buy you toys again* in four conditions: typical, fast, slow, and loud speech. Articulatory movements were recorded with three-dimensional electromagnetic articulography that tracked four sensors on the tongue. Data for the vowels /ia/ were analyzed. *Results:* The authors compared acoustic and kinematic data to measure phonetic specification and phonetic variability. Kinematic data showed that slow speech elicited significantly larger lingual displacements than the other conditions. Loud speech showed larger lingual displacement than typical and fast speech, and fast speech displacements were significantly smaller than typical speech. Acoustically, vowel distances were significantly larger during slow speech than during the other speaking conditions. Vowel distances were also larger during loud speech than fast speech. The spatiotemporal index (STI) was calculated to determine kinematic and acoustic phonetic variability. In the kinematic domain, loud speech was significantly less variable than slow speech. Loud speech was less variable than typical and fast speech, and typical speech was less variable than slow and fast speech. In the acoustic domain, there were no significant differences in acoustic variability among the different speech conditions. *Conclusion:* Changes in tongue displacement correlated with changes in acoustic vowel difference; however, changes in tongue movement STI variability had no relationship with changes in formant variability. *Relevance to the current work:* Acoustic changes may not always reflect kinematic changes. The authors call for a greater understanding of how changes in the kinematic domain relate to changes in the acoustic domain.

Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /r/. *Language, 92*, 101–140. doi: 10.1353/lan.2016.0019

Objective: This study explored the tongue positions used by native speakers of American English to produce the phoneme /r/. The purpose was to establish phonological patterns for the different tongue positions used to make /r/. *Method:* Fifteen females and 12 males were imaged by ultrasound while saying monosyllabic words in the carrier phrase *Please say _____ again.* Stimulus words were chosen to elicit the /r/ before or after three vowels, in word-initial and word-final positions. Others were separated from the word edge by a consonant. *Results:* Two subjects produced retroflex /r/, 16 produced bunched /r/, and 9 varied between the bunched /r/ and the retroflex /r/. For the 9 speakers who varied, retroflexed /r/ was significantly more frequent before vowels than after vowels. The vowels /ɑ/ and /o/ conditioned more retroflexion than the vowel /i/. Retroflexion was less common after coronal and labial consonants. *Conclusion:* The /r/ allophony patterns, like dark and light /l/, seem to be influenced by the articulation of other sounds. However, unlike dark and light /l/, the allophony patterns of /r/ are relatively complex and do not appear to be shared by communities of speakers. This may be due to the fact that bunched and retroflex /r/ are perceptually indistinguishable. *Relevance to the current work:* Native speakers of American English produce /r/ using different tongue positions.

Moll, K. L. (1960). Cinefluorgraphic techniques on speech research. *Journal of Speech and Hearing Research, 3,* 227–241.

Objective: The purpose of the study was to examine how cinefluorography might be used to study positions and movements of the articulators during speech. The author gives a description of the equipment, how it can be used to image the speech articulators, and provides an example study. *Method:* Two young adult females produced sustained vowels and six disyllables while cinefluorographic pictures were taken of the speech articulators. A frame-by-frame tracing method was used to measure the movements of the articulators during speech. The data were plotted to show the magnitude over time for velopharyngeal contact, velum-pharynx distance, tongue-alveolus distance, and incisal opening. The findings have implications regarding the physiology of speech. For example, opening of the velopharyngeal port always preceded the onset of phonation by several frames. *Conclusion:* Cinefluorography is a useful and promising method for studying speech articulation. *Relevance to the current work*: Cinefluorography is one method that has been productively used to examine the kinematics of speech.

Roy, N., Nissen, S. L., Dromey, C., & Sapir, S. (2009). Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of Communication Disorders. 42,* 124–135. doi: 10.1016/j.jcomdis.2008.10.001

Objective: This paper investigated how vowel articulation (measured by acoustic vowel space) changed in individuals with muscle tension dysphonia before and after receiving manual circumlaryngeal treatment. *Method:* The recordings used in this study were taken from an archive of speech samples of speakers with voice disorders. The samples used were 111 women with muscle tension dysphonia who showed improvement following manual laryngeal reposturing and/or circumlarngeal massage. Speech samples were the second and third sentences from The Rainbow Passage. The recordings for each speaker were used to make acoustic measures from formant data of four extracted vowels: /i, æ, ɑ, u/. Quadrilateral vowel space area (QVSA) and vowel articulation index (VAI) were calculated for each speaker. *Results:* Both QVSA and VAI increased significantly from the pre-treatment samples to the post-treatment samples. *Conclusion:* Manual circumlaryngeal therapy appears to improve articulatory acoustics. *Relevance to the current work:* Acoustic measures based on vowel formants are used in research to make inferences about lingual movement in speakers.

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. [Research Note]. *Journal of Speech, Language, and Hearing Research, 53,* 114–125. doi: 10.1044/1092-4388(2009/08-0184)

Objective: The authors compared two acoustic metrics, the vowel space area and the formant centralization ratio, to determine which was more effective in differentiating healthy speech from dysarthric speech. *Method:* Participants included 14 healthy speakers in addition to 38 speakers with Parkinson's disease and dysarthria. 19 of the dysarthric speakers had received Lee Silverman Voice Treatment. The three sentence stimuli were *The blue spot is on the key*, *The potato stew is in the pot*, and *Buy Bobby a puppy*. The vowel /i/, /u/, and /ɑ/ were extracted, and

formant measurements of the vowels were used to calculate VSA and FCR for pre-treatment and post-treatment samples. *Results:* For pre-treatment samples, the VSA did not show a significant difference between dysarthric speakers and healthy controls, while the FCR measure differentiated significantly between the two groups. The FCR data did not reveal an effect of gender; whereas, the VSA did show a gender effect. The FCR measure and the VSA measure both showed differences between pre- and post-treatment dysarthric speech samples, but the FCR data showed a more robust effect. *Conclusion:* The FCR more effectively differentiated the speakers with dysarthria from the healthy controls than did the VSA. The FCR was sensitive enough to differentiate treatment effects, but it was insensitive to gender effects. *Relevance to the current work:* In this study, the authors used acoustic measurements garnered from vowel formants to differentiate pre- and post-treatment dysarthric speech in addition to dysarthric and healthy speech.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3–45.

Relevance to the current work: The author describes the quantal nature of speech. He explains that as sounds are articulated, the acoustic parameter is more sensitive to changes in articulation in some ranges of movement than others. He suggests that this phenomenon is a factor in shaping phonology in language. Boundaries between phonemes may reflect boundaries between areas with articulatory-acoustic sensitivity. This may help explain the inventory of distinctive features in language. The quantal nature of speech as explained by the author is an important consideration when analyzing and interpreting the data in the current study.

Stone, M., Langguth, J. M., Woo, J., Chen, H., Prince, J. L., 2014. Tongue motion patterns in post-glossectomy and typical speakers: A principal components analysis. *Journal of Speech, Language, and Hearing Research, 57,* 707–717. doi: 10.1044/1092-4388(2013/13-0085)

Objective: The authors examined how patients who had received a glossectomy treatment moved their tongues during the /s/ phoneme relative to healthy controls. *Method:* Three glossectomy patients and 10 typical speakers participated in the study. The glossectomy patients had previously had tumors extracted from the lateral portion of the tongue, allowing the tongue tip to remain intact. Patients said the phrase *a geese* while MRI data and speech recordings were collected. *Results:* The authors used the velocity field to quantify the direction and velocity of tissue points between the MRI time-frames. Principal Components Analyses were performed on the time-frames of the following MRI slices: midsagittal, tumor/small motion, and nontumor/large motion. It was hypothesized that glossectomy patients would have smaller motion on the tumor side of the tongue, larger motion on the nontumor side, and would be more likely to use a laminal /s/ than an apical /s/. Analysis of speaker differences in movement between the tumor side of the tongue showed greater differences in motion pattern between patients and controls than other slices did. Analysis of data from the nontumor side of the tongue did not show larger motions as hypothesized. Only three of the controls and one of the patients used a laminal /s/, while all the other participants used an apical /s/. *Conclusion:* The tumor side of the tongue of glossectomy patients differed in its movement from typical speakers, while the nontumor side of the tongue of glossectomy patients exhibited no differences in movement from typical speakers. Both glossectomy patients and controls showed variability within each group

regarding tongue position during /s/. The majority of glossectomy patients and healthy controls used an apical /s/. *Relevance to the current work:* This study used magnetic resonance imaging to collect information about speech articulation in typical and disordered speakers.

Stone, M., Stock, G., Bunin, K., Kumar, K., Epstein, M., Kambhamettu, C., … Prince, J. (2007). Comparison of speech production in upright and supine positions. *Journal of the Acoustical Society of America, 122*, 532–541.

Objective: This study explored the effects of gravity on tongue position in speech. The purpose was to determine what interaction gravity has with tasks and speakers. *Method:* Seven males and six females repeated the words *bang, golly,* and *dash* while an ultrasound machine collected midsagittal images of each speaker's tongue. The speech tasks were completed while the speakers were in either an upright position or a supine position. Acoustic data were also recorded. Range of motion (ROM) and tongue contour were calculated using ultrasound image sequences for each speaker's upright and supine repetitions. Formants were extracted at vowel onsets or offsets for the purpose of collecting consonant information. Formants were also extracted at the onset and midpoint of /l/ and at the midpoint of the vowels /u/, /a/, and /i/. *Results:* Only 13 of the 168 comparisons for acoustic measures were significant, a number consistent with chance. This indicates that the effects of gravity on acoustics were minimal. For 39 speaker and word comparisons for ROM, 27 had averages larger in upright condition. The remaining 12 were larger in the supine condition. RMS differences were calculated between each pair of upright and supine tongue contours. There was a significant effect for speaker but not phoneme. Differences in speaker and phoneme were also calculated for the pharyngeal zone, which showed no differences in phonemes but significant differences for 10 speakers. Data showed intra-speaker and inter-speaker variability in which direction participants moved their pharyngeal tongue in the supine position as compared to the upright position. *Conclusion:* Speakers are variable in their tongue displacement strategies for speaking in a supine position, particularly in the posterior tongue. Individuals may preserve tongue position more at the constriction location of a phoneme than at other areas of the vocal tract while in the supine position. *Relevance to the current work:* This study used acoustic data and ultrasound imaging in complementary roles to examine tongue movements during speech.

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47,* 766–783. doi:10.1044/1092-4388(2004/058)

Objective: This study explored the effects of rate and loudness change on vocal tract acoustics for speakers with dysarthria and healthy controls. *Method:* Participants included 15 speakers with dysarthria secondary to multiple sclerosis (MS), 12 speakers with dysarthria secondary to Parkinson's disease (PD), and 15 typical speakers. Participants read a passage loaded with the target vowels /i/, /ɑ/, /æ/, and /u/ and the consonants /s/, /ʃ/, /t/, and /k/ in habitual, loud, and slow speaking conditions. Vowel space area was calculated for each condition. F2 transitions characteristics were extracted for the diphthongs /ɑɪ/ and /eɪ/. First-moment difference measures were used to calculate working space for fricative and stop consonants. Ten listeners rated intelligibility of each of the speakers while listening to an extract of the reading passage. *Results:* Vowel acoustic working space was significantly larger in the slow condition for healthy speakers

and speakers with MS. Vowel working space was not significantly different across conditions for speakers with PD. However, vowel working space was smaller in speakers with PD for every condition relative to typical speakers. First-moment difference measures for fricatives were analyzed and exhibited a smaller difference for speakers with PD versus typical speakers. Analysis for first-moment differences measure for stops showed that difference measures were larger in the loud condition versus the habitual condition. They were largest in the loud condition. There were smaller difference measures for the speakers with than the control group. For the diphthong /ɑɪ/, speakers with PD had significantly shallower F2 slopes than healthy controls. For /eɪ/, F2 slopes were steeper for the loud condition versus the slow condition and the habitual condition versus the slow condition. They were also steeper for the control group relative to the MS group. Loud speech was rated significantly higher for intelligibility than habitual speech. For the speakers with PD, intelligibility was higher in the loud condition relative to the habitual condition. Changes in vocal tract acoustic output did not correlate with intelligibility ratings. *Conclusion:* Those treating dysarthria may want to consider different strategies, such as reducing rate or increasing loudness, depending on what population they are working with and which phonemes are affected by the dysarthria. *Relevance to the current work:* This study used vowel working space calculated from quadrilateral corner vowels.

Twist, A., Baker, A., Mielke, J., & Archangeli D. (2007). Are covert /ɹ/ allophones really indistinguishable? *University of Pennsylvania Working Papers in Linguistics, 13*, Article 16.

Objective: To determine whether listeners can distinguish between bunched and retroflex /r/. *Method:* Fourteen native speakers of American English and 11 native speakers of Mandarin participated in the study. Participants listened to a set of monosyllabic words with /r/ in different phonetic contexts. They also listened to /r/ segments extracted from the monosyllabic words. For each discrimination task, they listened to a set of four sounds and were required to choose whether the second or third sound differed from the others. *Results:* For the whole-word stimuli, responses were not affected by word position or language of the listener. For the segment stimuli, responses were not affected by articulation of the phoneme or by language of the listener. *Conclusion:* Listeners do not appear to systematically perceive a difference between bunched and retroflex /r/. *Relevance to the current work:* Very different tongue positions may produce perceptually identical sounds.

Weismer, G., Martin, R., Kent, R. D., & Kent, J. F. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *Journal of the Acoustical Society of America, 91,* 1085–1098.

Objective: The purpose of this study was to describe acoustic characteristics of vowels from speakers with ALS. *Method:* Subjects were 15 healthy male controls and 25 male speakers with ALS. Stimuli were 12 words taken from a single-word speech intelligibility test. Words were spoken at a comfortable rate and loudness and recorded. Formant data were extracted by tracing the midpoint of F1 and F2 from the initial to final glottal pulse of the vowel nucleus. *Results:* Speakers with ALS tended to have longer transition durations, larger transition extents, shallower transition slopes, more centralized vocalic gestures at the onset of the transitional segment, longer durations of vocalic nuclei, and greater variability between speakers as compared to

typical speakers. Qualitative analysis showed speakers with ALS tended to have more movement at the onset of the F1 trajectory and occasionally at the onset of the F2 trajectory. *Conclusion:* Acoustic features of speakers of ALS are different from the acoustic features of typical speakers. *Relevance to the current work:* Analyzing formant trajectories is productive for measuring differences between typical and disordered speech.

Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica, 64*, 94–102. doi: 10.1159/000336890

Objective: This study examined the relationship between acoustic and kinematic measures and intelligibility in dysarthric and typical speech. *Method:* Participants included 31 male and female speakers diagnosed with various ALS subtypes. They were divided into two subgroups based on their speaking rate (AN = normal rate; AS = slowed rate). A group of healthy controls also participated in the study. Speakers repeated the sentences *Say doily again* and *I love Seattle in the spring*. Speakers with ALS read them at their normal rate, and healthy controls read them at their normal and half of their normal speaking rate (CN = normal rate; CS = slowed rate). Tongue movement was recorded using an electromagnetic system that tracked a magnet on the tongue blade. Acoustic data were also recorded. Target sounds for analysis were /dɔ/, /ɔɪ/, and /jæ/. *Results:* F2 slope was significantly shallower in the CS group and the AS group for all three sound transitions. The CS group differed from the AN group in /jæ/ and the CN group in /ɔɪ/. The CS group's F2 range was expanded compared to the CN group. F2 range was most reduced in speakers with ALS. The duration of the sound transition was slower in the CS group compared to the CN group and both groups of speakers with ALS. The AS group had slower durations than the AN group. For the average speed measure, the CS group was slower than the CN group. Likewise, the AS group had slower average speed than the AN group. The association between F2 slope and speed in speakers of ALS was moderately strong. Speaking rate was associated with duration, F2 slope, and movement speed. Speech intelligibility was associated with F2 slope; however, kinematic measures were not associated with intelligibility. In examining the association between acoustics and kinematics, it was found that movement speed had a significant effect on F2 slope, even after controlling for duration. Tongue displacement, however, was found to be weakly associated with F2 range, particularly when controlling for duration. Only F2 slope in /jæ/ seemed to relate to tongue displacement. *Conclusion:* Examining F2 slope in speakers with ALS is productive for obtaining information about tongue movements and may have applications for tracking disease progression and speech intelligibility. *Relevance to the current work:* Acoustic and kinematic speech data were related for certain measures but not for others.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of retroflex and bunched American English /r/. *Journal of the Acoustical Society of America, 123* , 4466–4481. doi: 10.1121/1.2902168

Objective: The authors investigated F4 and F5 differences between two American speakers with different productions of /r/ to determine whether they pattern consistently. *Method:* Participants included two speakers of American English. One produced a bunched /r/, and the other produced

a retroflex /r/. MRI data and acoustic data were recorded while the participants produced a set of utterances containing /r/ in various contexts. *Results:* F4 and F5 differed in spacing between the two speakers. Simple tube modeling suggested F3, F4, and F5 are the first, second, and third resonances of the back cavity. For the bunched /r/, the resonances could be explained by modeling the back cavity as a quarter-wavelength tube. For the retroflex /r/, the resonances could be explained by modeling the back cavity as a half-wavelength tube. However, data also showed F4 and F5 were influenced by the front cavity for the bunched /r/, perhaps due to higher coupling between the front and back cavities in this tongue position. *Conclusion:* F4 and F5 patterned differently between the two subjects and could be explained by the length of the cavity posterior to the constriction. These differences have potential for distinguishing the two /r/ productions in acoustic analysis. *Relevance to the current work:* The two variations of the American English /r/ phoneme have different tongue productions but are perceptually identical and acoustically similar for F1–F3.

Consent to be a Research Subject

Introduction

This research study is being conducted by Christopher Dromey, a professor in the department of Communication Disorders at Brigham Young University to determine how movements of the tongue and lips change under several conditions (voicing, whispering, silently mouthing the words). You were invited to participate because you are a native speaker of English and have no history of speech, language, or hearing disorders**.**

Procedures

If you agree to participate in this research study, the following will occur:

- you will be seated in a sound-treated recording booth in room 106 of the John Taylor Building
- six small sensor coils will be attached with dental adhesive to your tongue, teeth, and lips and one to the frame of eyeglasses (no corrective lenses) that you will wear
- while you speak, the researchers will record the movements of these articulators and audio record your speech
- you will read sentences from a sheet in front of you under several conditions: normal speech, whispering, and silent mouthing of the words
- the total time commitment will be less than 60 minutes

Risks/Discomforts

You may feel uncomfortable having the sensors attached with dental glue inside your mouth. These may cause you to mildly lisp on some sounds at first. For several hours after the study you may be able to feel a slight residue on your tongue, which will disappear within a day. This technology has been widely used at other research centers and no problems for the research subjects have been reported.

Benefits

There will be no direct benefits to you. It is hoped, however, that through your participation researchers may learn about the way speech articulator movements may change under different voicing conditions. This may expand our understanding of the way the brain controls speech movements in healthy individuals and could lead to further work that would help people with speech disorders.

Confidentiality

The research data will be kept in a locked laboratory on a password protected computer and only the researcher will have access to the data. At the conclusion of the study, all identifying information will be removed and the data will be kept in the researcher's locked office. Arbitrary participant codes, but no names, will be used on the computer files or paper records for this project in order to maintain confidentiality. In presentations at conferences and in publications based on this work, only group data will be reported.

Compensation

You will receive $10 cash for your participation; compensation will not be prorated. For BYU students, no extra credit is available.

Participation

Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your class status, grade, or standing with the university.

Questions about the Research

If you have questions regarding this study, you may contact Christopher Dromey at (801) 422-6461 or dromey@byu.edu for further information.

Questions about Your Rights as Research Participants

If you have questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Name (Printed): _____ Signature _____ Date: _____