2018-04-01

The Association Between Articulator Movement and Formant Histories in Diphthongs Across Speaking Contexts

Janae Valyn Christensen
Brigham Young University

The Association Between Articulator Movement and Formant

Histories in Diphthongs Across Speaking Contexts

Janae Valyn Christensen

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

Master of Science

Christopher D. Dromey, Chair
Kristine Marie Tanner
Shawn L. Nissen

Department of Communication Disorders

Brigham Young University

ABSTRACT

The Association Between Articulator Movement and Formant
Histories in Diphthongs Across Speaking Contexts

Janae Valyn Christensen
Department of Communication Disorders, BYU
Master of Science

This study examined the effect of context on the association between formant trajectories and tongue and lip kinematics in the American English diphthongs /aɪ/ and /aʊ/. Seventeen native speakers of American English had electromagnetic sensors placed on their tongue and lips to record kinematic signals that were time-aligned with the corresponding acoustic recording. Speakers produced the diphthongs in isolation, in a single word rVl context, in a phrase hVd context, and in a sentence context. Kinematic data and the F1 and F2 trajectories were extracted from the middle 50% of each diphthong production. To allow direct comparison of signals with different units of measurement, all data were converted to z-scores. The z-score records were plotted together on common axes. For each tracked sensor from each diphthong production, an absolute difference between the kinematic and acoustic variables was calculated. Average z-score difference sums were calculated for each speaker's /aɪ/ and /aʊ/ production in each context, and this measure was called the Acoustic Kinematic Disparity Index (AKDI). A repeated measures ANOVA was used to test for main context effects on the AKDI, with concurrent contrasts to test for differences between the baseline (isolated diphthong) condition and the more complex phonetic contexts. The results revealed that context has a significant impact on acoustic and kinematic relationships. The sentence context resulted in the highest number of significantly different AKDI values when compared to the isolated condition, the single word rVl context resulted in the second highest number, and the phrase level hVd context resulted in the least differences. These findings suggest, therefore, that more complex phonetic contexts have a greater effect on the acoustic and kinematic relationship. These findings imply that caution is warranted in relying on acoustics to draw inferences about articulator movements in complex phonetic contexts. These results further indicate that the investigation of sounds produced in one context does not necessarily allow a straightforward generalization to other contexts.

Keywords: formants, diphthong, acoustic, kinematic, articulation, phonetic context

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF THESIS STRUCTURE

This thesis, *The Association Between Articulator Movement and Formant Histories in Diphthongs Across Speaking Contexts,* adheres to the journal-formatted structure required by Brigham Young University Graduate Studies and the Communication Disorders Department. It is also modeled after Communication Disorders related peer-reviewed articles. An annotated bibliography is included in Appendix A.

Introduction

For decades, researchers have relied on acoustic measures to quantify aspects of speech because this approach is affordable and convenient. It provides objective metrics that are reflective of the actions of the speech subsystems (Kent, Weismer, Kent, Vorperian, & Duffy, 1999). Acoustic data have often been used as an index of vocal tract movement during speech. The first formant has been shown to be inversely related to tongue height. Likewise, the second formant increases with tongue advancement (Ferrand, 2015). Although many studies have relied on formant data to draw inferences about articulatory activity, the relationship between movement and acoustics is complex and sometimes nonlinear (Stevens, 1989).

Roy, Nissen, Dromey, and Sapir (2009) used acoustic measures of vowels to study the impact of manual circumlaryngeal voice treatment on articulation in individuals with muscle tension dysphonia. Changes in the acoustic vowel space after treatment suggested that muscle tension dysphonia affects more than the larynx alone, and that improving voice quality was associated with an enlarged vowel space. Laaksonen, Rieger, Happonen, Harris, and Seikaly (2010) demonstrated the applicability of acoustic evaluation of vowel space as they completed a longitudinal study of participants undergoing radial forearm free flap reconstruction of the tongue. By acoustic means they were able to objectively examine the effect of altered anatomy in the oral cavity on a speaker's ability to produce vowels.

While acoustic analyses are well established as a research methodology, there are a number of limitations. At a basic level, the first two formants are representative of tongue height and tongue advancement, respectively. However, the vocal tract is composed of multiple resonating cavities that attenuate or enhance formants. Because these cavities are interconnected, it can be difficult to estimate their relative contributions to vowel acoustics. Additionally,

acoustic measures cannot isolate the individual contributions of specific articulators to the acoustics, since the combined actions of several structures will contribute to the sound that is produced.

Another limitation of acoustic analysis derives from the motor equivalence principle, whereby individual speakers produce perceptibly identical sounds through different articulatory means. Individuals can vary in the way each articulator contributes to the production of a sound, yet still achieve the desired output. An example is the way that native speakers of American English can produce /r/ with two different tongue configurations. These speakers use either a *bunched* /r/ or a *retroflex* /r/. Although the two forms of /r/ are produced in different ways, they are perceived as the same phoneme by listeners.

An additional limitation of acoustic analysis is associated with the quantal theory of speech (Stevens, 1989). This theory explains that articulatory movements do not have a consistent and linear one-to-one relationship with acoustic measures. Researchers have determined that movement of the tongue from point A to point B does not produce the same degree of acoustic change as tongue movement from point B to point C, when points C and A were equidistance from point B. This shows that the magnitude of acoustic change is dependent on the specific location within the vocal tract where the movement occurs.

Unlike acoustic variables, kinematic measures directly quantify articulatory movements. Unfortunately, kinematic methods are more invasive and more expensive than acoustic recordings. Kinematic measures identify exactly where in space the articulators are. If kinematic measures and acoustic recordings of the same speech events are compared, acoustic measures can be validated by means of the kinematics; in other words, the kinematics will reveal how much we can rely on approximating articulatory movements via acoustics. Diphthongs are a

suitable research target for combined acoustic and kinematic studies because diphthongs involve movement. By tracking diphthong transitions via both acoustic and kinematic means, signals may be compared to determine how well the acoustic changes reflect the movements.

A study by Dromey, Jang, and Hollis (2013) examined this correlation between acoustics and kinematics of diphthongs the context of the sentence, "The boy gave a shout at the sight of the cake." The researchers examined the association of the first and second formants with the kinematic measures of vertical and horizontal tongue movement during diphthong production. A single point on the tongue was tracked. The purpose of the study was to determine the accuracy with which acoustic measures represent lingual movements. As the researchers analyzed the data, they determined that the widespread assumption of straightforward links between tongue movements and formant frequencies may not reflect the complexity of diphthong acoustics and their connection to kinematics. They suggested that this lack of direct relation may be especially true for diphthongs in natural speech, although they only analyzed diphthongs in a single sentence.

McKell (2016) also examined the relationship between formants and kinematic measures of articulation in diphthong production. Unlike Dromey, et al. (2013), McKell measured both lingual and labial movements. McKell studied the diphthongs /aɪ/, /aʊ/, and /ɔɪ/. By adding lip sensors, articulators that had not been tracked in previous acoustic versus kinematic diphthong analyses, McKell was able to determine that the articulators (labial or lingual) contributing most to F1 and F2 changes varied by diphthong. The study revealed that tongue movement may be the best predictor of F1 and F2 changes for the diphthong /aɪ/, likely because the diphthong lacks lip rounding. McKell determined lip movement may be the best predictor of F1 and F2 changes in the diphthong /aʊ/, likely due to the substantial lip rounding of the offglide vowel. McKell found

that /ɔɪ/ presented challenges in determining the relationship between the articulators and formant changes because the actions of the lips and tongue during this diphthong acted in opposite directions on F1.

These combined acoustic and kinematic studies have contributed useful information to the literature on speech production. However, the conclusions of Dromey et al. (2013) and McKell (2016) were limited in their ecological validity, or the extent to which the findings could be generalized to everyday speech. McKell only analyzed diphthongs produced in isolation and Dromey et al. (2013) analyzed diphthongs in only one context. Sounds produced in isolation in a highly-controlled laboratory environment are not necessarily representative of typical, everyday communication. While these studies have improved our understanding of the relationship between acoustics and kinematics, their ecological validity has been restricted due to the limited range of speaking contexts.

Because diphthongs analyzed in the study by McKell (2016) were produced in isolation, it remains unclear whether or how the relationship between acoustics and kinematics changes across phonetic contexts. The range of phonetic contexts includes structures such as consonant-vowel-consonant, *hide*, or phrase level, *I say hide again*. Rather than phonemes being spoken in isolation, everyday speech is influenced by coarticulation and includes substantial allophonic variation. Allophonic variations are alternative productions of a phoneme which do not change the meaning of the word. For example, whether the /s/ pronounced in *sue* has a rounded lip shape or the /s/ pronounced in *see* has a retracted lip shape, the /s/ phoneme remains the same despite the alteration in production. Allophonic variations would not be observed in isolated sounds; therefore, it would be valuable to examine diphthongs across a range of contexts in the present study.

Coarticulation is the principle of individual speech sounds being influenced by surrounding sounds. A phoneme will not be the same in isolation as it will be when it has neighboring sounds. Coarticulation exists in two main forms: anticipatory and perseverative. Anticipatory coarticulation occurs when a current speech sound is influenced by a characteristic of an upcoming speech sound. Consider, for example, the /k/ sound in *cool* versus *keep*. In *cool*, while producing the /k/ phoneme the lips assume a rounded position in preparation for the rounded vowel that follows. In the word *keep*, while producing the /k/ phoneme the lips assume a retracted position, in preparation for the non-rounded vowel that follows. Perseverative coarticulation occurs when a speech sound is influenced by a characteristic of a previously occurring sound. This concept can be demonstrated by the words *top* and *mop*. The vowels of these words become nasalized or non-nasalized in consequence of the preceding consonant. The /ɑ/ in *top* is non-nasal since the preceding sound is non-nasal. The /ɑ/ in the word *mop* is nasal due to the nasality of the preceding /m/ carrying over and affecting the vowel.

Coarticulatory effects can extend beyond directly neighboring sounds. Magen (1997) studied the boundaries of coarticulation. This researcher determined that coarticulatory effects can extend beyond the boundaries of the syllable. Magen's work is relevant to the current study because he considered the effect of coarticulation of individual syllables and phrases. When diphthongs are analyzed only in isolation, no coarticulation is accounted for. If diphthongs were analyzed in the production of a consonant-vowel-consonant format, only two phonemes (the first and last) would have the potential to have coarticulatory effects on the diphthong. In a phrase however, there are many phonemes that have the ability to spread their coarticulatory effects beyond their syllable, thus affecting the diphthong production in a variety of ways. Whether in consonant-vowel-consonant form or phrase form, coarticulation will most likely affect the

overall production of diphthongs. It is unknown, however, whether or not the coarticulation will affect the relationship between the acoustics and kinematics of those productions.

Farnetani and Faber (1992) demonstrated the importance of context as they analyzed the coordination of the tongue and the jaw across isolated words versus connected speech. The researchers studied an individual's production of the Italian vowels /i/ and /a/ across three utterance types: vowel-consonant-vowel (VCV) pseudowords in isolation; consonant-vowel-consonant-vowel (CVCV) words in isolation; and consonant-vowel-consonant-vowel (CVCV) words in pre-final phrase positions in complex natural sentences. Farnetani and Faber determined, as they expected, that the tongue and jaw can function "largely independent of each other" (pg. 209). The tongue and jaw can perform either synergistically or independently. When moving synergistically, the jaw performs the gross movements to deliver the tongue to the proper position for the tongue to achieve precise shaping of the vocal tract. When performing independently, articulators can move in a vertical or horizontal plane, or even in opposite directions, at the same time. The researchers explained that during connected speech, individuals attempt to conserve time and energy by partially centralizing articulator movements. Centralization means that the articulators are not reaching maximal displacement as they would if the sound were being produced in isolation. Since the speaker's articulator movements are partially centralized, the speaker has other articulators adjust their movement patterns to create the target sound. The need to analyze actions of the tongue and lips during diphthong production across contexts is supported by Farnetani and Faber's claim that articulatory roles adjust during connected speech to compensate for centralization.

As the change in articulatory roles is not observed with sounds produced in isolation, comparing articulator movements across an array of contexts is important. Not only does the role

of compensatory relations of the tongue and lips need to be analyzed, but also how such relations are reflected in the acoustic output. Analysis is warranted as the speaker may cause one articulator to attempt to "pick up the slack" of another articulator to maintain the desired general vocal tract shape and thus affect the acoustic output.

Nittrouer (1991) analyzed the effect of suprasegmental factors, such as rate and stress, on the relative timing of articulator movements, specifically the tongue and jaw. She determined that spatiotemporal relationships among articulatory movements were affected not only by phonetic context but also by suprasegmental factors. This means that the contextual effects of connected speech, such as rate and stress, impacted the articulatory relations. This suggests that the relations of the tongue and lips during diphthong production vary during connected speech versus diphthongs produced in isolation, as suprasegmental factors do not apply to isolated diphthongs. Diphthongs produced in isolation, as researchers have previously analyzed, do not represent the phonetic structure of everyday diphthong production (connected speech), or factors such as rate and stress.

Luce and Charles-Luce (1985) also recognized that articulatory features may vary across phonetic contexts. Although vowel and closure duration served as acoustic correlates to determine a voiced versus voiceless stop in specific contexts, it had not been determined if this finding was still applicable across various phonetic and sentence contexts. They determined that vowel duration presented the most consistent temporal cue of voicing across changes in sentence position, local phonetic environment, place of articulation, and inherent vowel duration.

These studies support the hypothesis that the relationships between acoustics and kinematics may change as the context varies. In the current study, we compare the acoustic and kinematic relationships in the American English diphthongs /aɪ/ and /aʊ/, across the contexts of

isolation, single words, and phrases. The investigation will use data collected during McKell's (2016) study. The aim is to learn how closely the acoustic-kinematic relationships from isolated diphthongs represent the acoustic-kinematic relationships of the same diphthongs in other speaking conditions. The ecological validity of isolated diphthong analyses will be enhanced if the findings from this context are comparable to the data from diphthongs produced in sentences.

Method

Participants

The data analyzed for this study were selected from a larger archival dataset. Twenty individuals, 10 men and 10 women, participated in the study. Their ages ranged from 20 to 34, and the median age was 25. Due to formant tracking error in Praat (version 6.0.36) acoustic analysis software, data from three of the speakers were not included in the study. The remaining 17 speakers included 9 males and 8 females. All participants were native speakers of American English and had no specific regional accent, as determined by the researchers. Participants signed a consent form, which was approved by the Brigham Young University Institutional Review Board, before participating in data collection and received a $10 compensation for their time.

Stimuli

Participants read four sets of stimuli, with five repetitions per set. The sets were completed in a randomized order for each participant. The stimuli are displayed in Table 1. List A was composed of the isolated diphthongs /ɔɪ/, /eɪ/, /aʊ/, /aɪ/, and /oʊ/. Examples of the diphthongs' ideal sounds were given in the form of short words such as *"boy"* for *oy*. However, during data collection for List A, the diphthongs were produced in isolation. List B was composed of diphthongs in hVd context. List C was composed of diphthongs in rVl context.

List D was the diphthong-rich sentence: *The boy gave a shout at the sight of the cake, you know.*
Only the diphthongs /aɪ/ and /aʊ/ were analyzed in this study.

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 the
ow as in "cow"	I say how'd again	rowel as in "vowel"	cake, you know.
i as in "tie"	I say hide again	rile	
o as in "hoe"	I say hoed again	role	

Equipment

An AKG C2000B microphone was placed 30 cm from the speaker's mouth. The acoustic
signal passed through a Focusrite Scarlett 2i2 preamplifier and was recorded to a Dell Optiplex
990 computer. The microphone signal sampling rate was set to 22050 Hz. The kinematic signals
were recorded with an NDI Wave system, which sampled them at 400 Hz. Both the acoustic and
kinematic signals were recorded by the NDI WaveFront software.

Procedure

Participants sat in a chair in an Acoustic Systems sound-attenuating booth 90 cm from the
printed stimuli. These were printed in black ink on white paper, in 36 pt font. The researchers
wore latex gloves and used tongue depressors as they applied the five electromagnetic sensors
with PeriAcryl®90 cyanoacrylate tissue adhesive. Silver DriDaid tongue drying pads were used
to prepare the tongue for the adhesive. The locations of the electromagnetic sensors were as
follows: the tongue mid sensor (T-MID) was placed at midline on the superior surface of the
tongue, approximately 3 cm posterior to the tongue tip; the tongue tip (T-TIP) sensor was also

placed on the superior surface of the tongue at midline, 1 cm posterior to the tip; the jaw sensor (JAW) was placed on the midline of the lower incisors, on a 5 x 10 mm patch of Stomahesive to prevent potential enamel damage from the adhesive; the upper lip (UL) sensor was placed at midline on the vermillion border of the upper lip; the lower lip (LL) sensor was placed at midline on the vermillion border of the lower lip. Participants wore an eyeglass frame with a reference sensor placed on the bridge. The speakers read aloud for 20 minutes to adapt to the presence of the electromagnetic sensors before producing the stimuli.

Data Analysis

For the diphthongs /aɪ/ and /aʊ/, the NDI wave system generated time-aligned output files for audio and kinematic data, which were imported into a custom MATLAB application, where the diphthongs were segmented from the longer recording. Diphthongs were segmented visually on the basis of both acoustic and kinematic waveforms and by using audio playback for confirmation. MATLAB exported audio segments as a wav file and also exported the sensor data as a text file.

Praat (version 6.0.36) acoustic analysis software was used to extract F1 and F2 histories at 1 ms intervals during the diphthong recordings. The default ceiling frequency was set to 5500 Hz for women and 5000 Hz for men. Text files from PRATT were reimported into MATLAB. In MATLAB, the kinematic recordings were time-normalized with a Fourier linear interpolation algorithm. The purpose of the time-normalization was to ensure equivalent sampling intervals and durations for all productions in the study.

Because formant frequencies and articulator movements have different units of measurement, MATLAB converted the strings of numbers into z-scores to create records in units that could be directly compared. Diphthongs are composed of three segments: onglide, transition,

and offglide. As the movement of the transition was the desired research target, the middle 50% of the diphthong was analyzed. A sample of 500 data points was taken for each diphthong (the middle 50% of the diphthong). The plots of acoustics and kinematics, each composed of 500 data points, were compared for each diphthong in each context. A z-score difference sum was calculated for each diphthong segment. This was done by summing the absolute difference between each acoustic and kinematic z-score for all 500 sample points.

The study focused on tongue and lip movements. The specific articulatory movements included x and y movement of the following sensors: upper lip, lower lip, tongue tip, and tongue middle. The lower lip horizontal movement was used to reflect lip protrusion, and thus vocal tract lengthening. The Euclidian distance between the lower lip and upper lip sensors was computed as an index of lip aperture (LIP).

The relationship between acoustics and kinematic movements was reflected in the z-score difference sum. A low sum indicated that the patterns of the kinematics and acoustics were similar to each other. A high sum indicated large differences between the acoustics and kinematics. An average absolute z-score difference of each speaker's 5 repetitions was computed for each condition. The resulting index is referred to as the acoustic-kinematic disparity index (AKDI). The AKDI is a measure without units as it is derived from z-scores.

Results

Acoustic and kinematic relations were compared across the phonetic context conditions using SPSS (version 25). A repeated measures ANOVA was used to test for main effects in the AKDI, with concurrent contrasts to test for differences between the baseline (isolated diphthong) condition and the more complex phonetic contexts. Thus, context was the within-subjects factor and gender was initially used as a between-subjects factor to determine if the patterns for men

and women differed. Since initial testing revealed that there were no main or interaction effects for gender, this factor was removed, and the statistics reported below are for men and women combined. The ANOVA main effect RM ANOVA results are reported in Table 2 and the contrast results are reported in Table 3. In instances where the Mauchly test revealed violations of the sphericity assumption of the ANOVA, Huynh-Feldt adjusted (non-integer) degrees of freedom were used.

/aɪ/ and F1 Association

Descriptive statistics of the /aɪ/ and F1 relationships are shown in Table 4 and are further illustrated in Figure 1. There was a significant main effect of context on the T-MID X and F1 association. Contrast analyses revealed that the AKDI of the single word rVl and sentence contexts was significantly higher than in the isolated condition. There was also a significant effect of context on the T-MID Y and F1 association. Contrast analysis revealed that the AKDI of the rVl context was significantly higher than in the isolated condition.

A significant effect of context on the T-TIP X and F1 association was found. Contrast analyses revealed that the AKDI of the phrase hVd, single word rVl, and sentence contexts was significantly higher than for the diphthong in the isolated condition. There was a significant effect of context on the T-TIP Y and F1 association. Contrast analyses revealed that the AKDI of the rVl and sentence condition was significantly higher than in the isolated condition.

There was a significant effect of context on the LL X and F1 association. Contrast analysis revealed that the AKDI of the sentence condition was significantly higher than in the isolated condition. There was a significant effect of context on the LL Y and F1 association. Contrast analyses revealed that the AKDI was significantly higher in the single word rVl and sentence contexts than in the isolated condition.

There was also significant effect of context on the lip aperture and F1 association. Contrast analyses revealed that the AKDI was significantly higher in the single word rVl and sentence contexts than in the isolated condition.

Table 2

ANOVA Main Effects

	Main Effect RM ANOVA Results			
Sensor	*df*	*F*	*p*	η_ρ^2
/aɪ/-F1				
T-MID X	1.860, 29.767	7.975	0.002	0.333
T-MID Y	1.608, 25.728	8.510	0.003	0.347
T-TIP X	2.350, 37.594	25.896	<0.001	0.618
T-TIP Y	1.778, 28,447	20.774	<0.001	0.565
LL X	2.189, 35.028	6.713	0.003	0.296
LL Y	2.509, 40.140	9.974	<0.001	0.384
LIP	2.272, 36.344	11.933	<.001	0.427
/aɪ/-F2				
T-MID X	1.957, 31.315	14.378	<0.001	0.473
T-MID Y	1.676, 26.821	8.682	0.002	0.352
T-TIP X	2.317, 37.965	36.059	<0.001	0.693
T-TIP Y	2.389, 38.219	20.867	<0.001	0.566
LL X	3, 48	6.937	0.001	0.302
LL Y	3, 48	10.874	<0.001	0.405
LIP	1.796, 28.733	10.193	0.001	0.389
/aʊ/-F1				
T-MID X	3, 48	5.742	0.002	0.264
T-MID Y	2.356, 37.699	6.319	0.003	0.283
T-TIP X	3, 48	1.676	0.185	0.095
T-TIP Y	3, 48	16.510	<0.001	0.508
LL X	1.960, 31.353	8.027	0.002	0.334
LL Y	1.934, 30.951	8.627	0.001	0.350
LIP	1.807, 28.904	9.921	0.001	0.383
/aʊ/-F2				
T-MID X	2.410, 38.560	13.415	<0.001	0.456
T-MID Y	3, 48	6.176	0.001	0.279
T-TIP X	3, 48	5.318	0.003	0.249
T-TIP Y	3, 48	8.033	<0.001	0.334
LL X	3, 48	2.000	0.127	0.111
LL Y	2.038, 32.609	3.022	0.062	0.159
LIP	1.893, 30.293	2.925	0.072	0.155

Table 3

Contrast Results

	hVd Contrast			rVl Contrast			Sentence Contrast		
	F	p	η_ρ^2	F	p	η_ρ^2	F	p	η_ρ^2
/aɪ/-F1									
T-MIDX	3.968	0.064	0.199	20.734	<0.001***	0.564	19.897	<0.001***	0.554
T-MIDY	0.232	0.637	0.014	9.580	0.007***	0.375	3.960	0.064	0.198
T-TIPX	5.366	0.034*	0.251	18.452	0.001***	0.536	52.747	<0.001***	0.767
T-TIPY	2.076	0.169	0.115	30.197	<0.001***	0.654	5.177	0.037*	0.244
LLX	1.957	0.181	0.109	2.382	0.142	0.130	4.594	0.048*	0.223
LLY	0.307	0.587	0.019	7.077	0.017**	0.307	11.057	0.004***	0.409
LIP	0.822	0.378	0.049	12.032	0.003***	0.429	11.998	0.003***	0.429
/aɪ/-F2									
T-MIDX	7.104	0.017**	0.307	3.924	0.065	0.197	21.818	<0.001***	0.577
T-MIDY	1.234	0.283	0.072	9.914	0.006***	0.383	0.138	0.715	0.009
T-TIPX	7.59	0.014**	0.322	4.864	0.042*	0.233	55.636	<0.001***	0.777
T-TIPY	2.231	0.155	0.122	43.724	<0.001***	0.732	7.425	0.015**	0.317
LLX	1.127	0.304	0.066	4.203	0.057	0.208	5.248	0.036*	0.247
LLY	0.34	0.568	0.021	12.015	0.003*	0.429	14.189	0.002 *	0.47
LIP	0.214	0.65	0.013	10.403	0.005*	0.394	11.007	0.004***	0.408
/aʊ/-F1									
T-MIDX	4.401	0.052	0.216	6.297	0.023**	0.282	1.371	0.259	0.079
T-MIDY	5.852	0.028*	0.268	2.802	0.114	0.149	5.794	0.029*	0.266
T-TIPX	2.484	0.135	0.134	0.783	0.389	0.047	0.371	0.551	0.023
T-TIPY	6.368	0.023**	0.285	19.914	<0.001***	0.554	2.949	<0.001***	0.156
LLX	0.04	0.843	0.003	4.716	0.045*	0.228	10.424	0.005***	0.394
LLY	0.495	0.492	0.03	6.072	0.025*	0.275	9.090	0.008***	0.362
LIP	0.144	0.71	0.009	6.476	0.022**	0.288	11.268	0.004***	0.413
/aʊ/-F2									
T-MIDX	2.741	0.117	0.146	8.723	0.009***	0.353	4.593	0.048*	0.223
T-MIDY	0.001	0.977	0.000	0.121	0.733	0.007	14.862	0.001***	0.482
T-TIPX	3.785	0.07	0.191	1.294	0.272	0.075	4.749	0.045*	0.229
T-TIPY	0.004	0.949	0.000	4.964	0.041*	0.237	17.894	0.001***	0.528
LLX	4.594	0.048	0.223	4.254	0.056	0.210	2.457	0.137	0.133
LLY	7.539	0.014	0.320	5.250	0.036	0.247	1.484	0.241	0.085
LIP	5.719	0.029	0.263	4.213	0.036	0.208	0.079	0.782	0.005

Note. *p<.05, **p<.025, ***p<.001

All within-subject contrasts had the degrees of freedom 1,16.

Table 4

/aɪ/ and F1 Descriptive Statistics

Sensor	Isolated		hVd		rVl		Sentence	
	M	SD	M	SD	M	SD	M	SD
T-MID X	107.8	34.7	240.2	259.5	253.5	130.9	406.4	254.1
T-MID Y	113.9	43.9	108.8	23.3	248.9	174.0	160.1	78.0
T-TIP X	112.1	37.0	205.2	156.6	246.8	124.2	549.9	233.3
T-TIP Y	131.2	46.5	115.3	32.6	349.9	163.0	189.5	100.1
LL X	261.9	176.5	201.6	103.5	361.3	166.7	406.1	208.1
LL Y	159.8	84.5	146.6	76.9	263.6	140.6	327.0	189.3
LIP	154.0	68.2	140.0	59.4	280.9	130.5	316.9	190.9
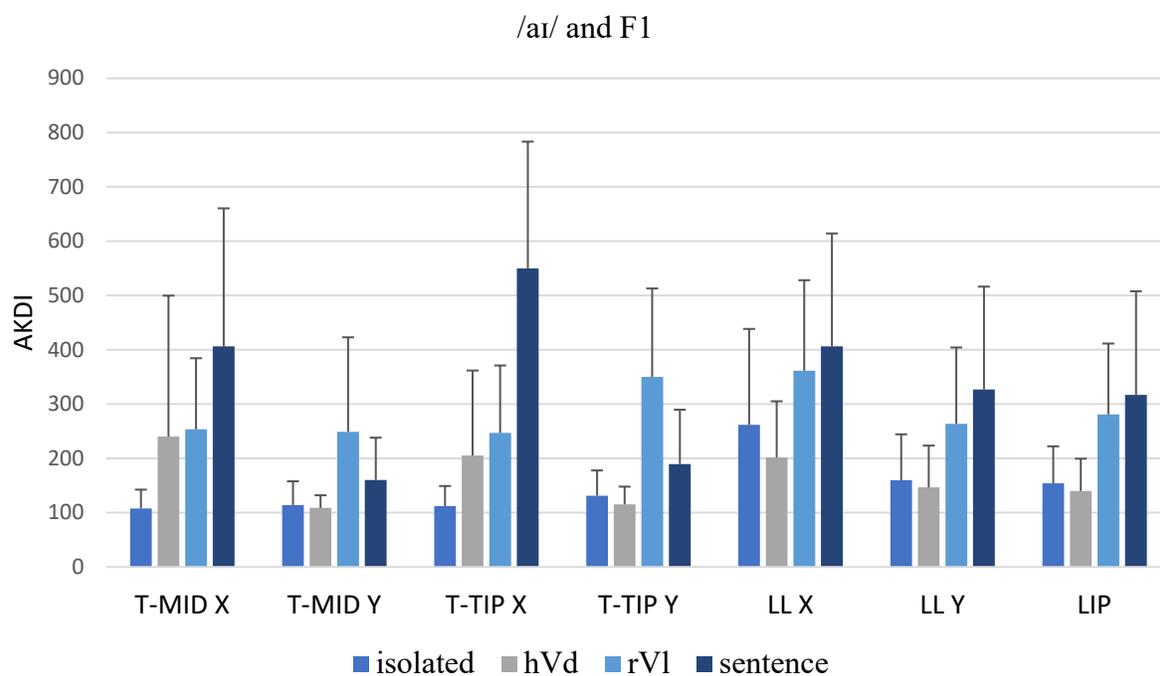
Figure 1. Mean AKDI and standard deviation of each /aɪ/ and F1 relationship. A small AKDI represents a strong acoustic-kinematic relationship and a large AKDI represents a weak acoustic-kinematic relationship.

/aɪ/ and F2 Association

Descriptive statistics of the /aɪ/ and F2 relationships are shown in Table 5 and are further illustrated in Figure 2. There was a significant effect of context on the T-MID X and F2 association. Contrast analyses revealed that the AKDI of the single word rVl and sentence contexts was significantly higher than in the isolated condition. There was also a significant effect of context on the T-MID Y and F2 association. Contrast analysis revealed that the AKDI of the single word rVl condition was significantly higher than the AKDI of the isolated condition.

A significant effect of context on the T-TIP X and F2 association was found. Contrast analyses revealed that the AKDI of each condition (phrase hVd, single word rVl, and sentence) was significantly higher than in the isolated context. The relationship between T-TIP Y and F2 was also significantly affected by context. Contrast analyses revealed that the AKDI of the single word rVl and sentence contexts was significantly higher than in the isolated context.

There was a significant effect of context on the LL X and F2 association. Contrast analysis revealed that the AKDI of the sentence context was significantly higher than in the isolated context. A significant effect of context on the LL Y and F2 association was found. Contrast analyses revealed that the AKDI of the single word rVl and sentence contexts was significantly higher than in the isolated context.

There was a significant effect of context on the lip aperture and F2 association. Contrast analyses revealed that the AKDI of the single word rVl and sentence contexts was significantly higher than the AKDI of the isolated context.
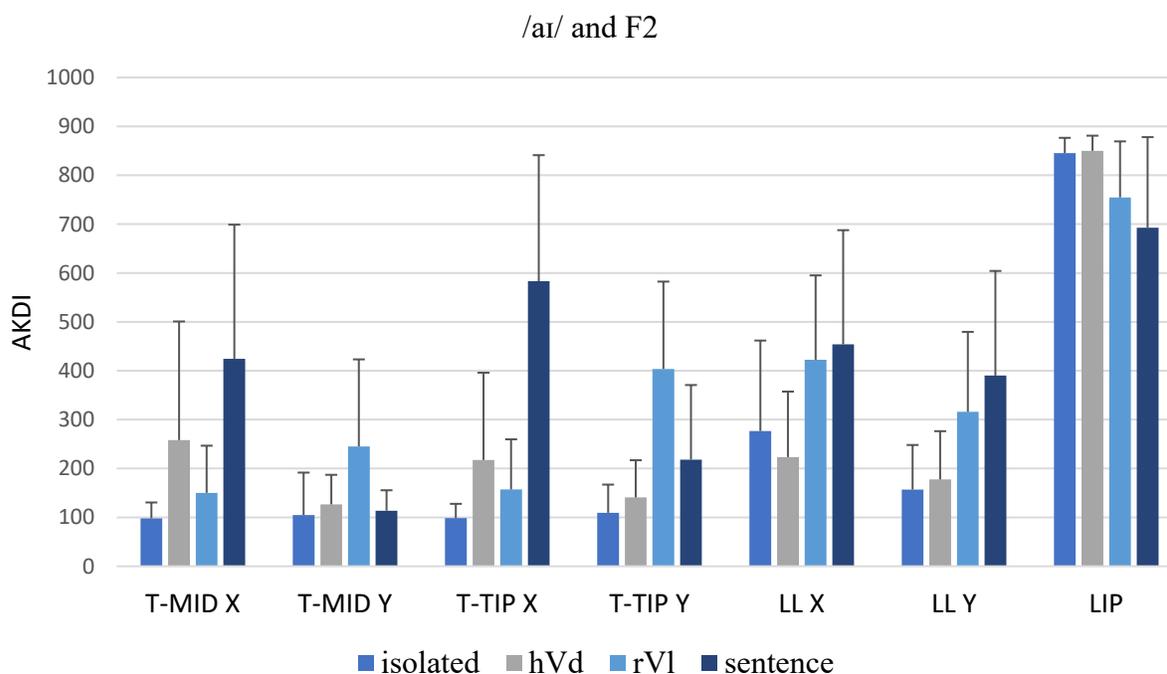
Table 5

/aɪ/ and F2 Descriptive Statistics

Sensor	Isolated		hVd		rVl		Sentence	
	M	SD	M	SD	M	SD	M	SD
T-MID X	98.0	32.5	258.1	242.7	150.2	96.5	424.4	274.5
T-MID Y	104.9	86.7	126.6	60.5	245.1	178.2	113.7	41.8
T-TIP X	98.5	29.2	217.3	178.6	157.2	102.5	583.1	258.0
T-TIP Y	109.4	57.7	140.8	76.2	403.7	178.7	218.2	152.6
LL X	276.6	185.2	223.3	134.1	422.5	172.6	454.1	233.5
LL Y	156.9	91.2	177.6	98.7	316.0	163.6	390.2	214.0
LIP	845.3	31.2	850.0	31.0	754.6	114.7	692.7	185.4

Figure 2. Mean AKDI and standard deviation of each /aɪ/ and F2 relationship.

/aʊ/ and F1 Association

Descriptive statistics of the /aʊ/ and F1 relationships are shown in Table 6 and are further illustrated in Figure 3. There was a significant effect of context on the T-MID X and F1 association. Contrast analysis revealed that the AKDI of the single word rVl context was significantly higher than in the isolated condition. There was also a significant effect of context on the T-MID Y and F1 association. Contrast analyses revealed that the AKDI of the phrase hVd was significantly lower than in the isolated condition. The AKDI of the sentence context was significantly higher than in the isolated context.

There was no significant effect of context on the T-TIP X and F1 association. There was a significant effect of context on the T-TIP Y and F1 association. Contrast analyses revealed that the AKDI of the phrase hVd context was significantly lower than in the isolated condition. The AKDI of the single word rVl context was significantly higher than in the isolated condition.

There was a significant effect of context on the LL X and F1 association. Contrast analyses revealed that the AKDI of the single word rVl and the sentence contexts was significantly higher than in the isolated condition. A significant effect of context on the LL Y and F1 association was found. Contrast analyses revealed that the AKDI of the single word rVl and sentence contexts was significantly higher than in the isolated context.

There was a significant effect of context on the lip aperture and F2 association. Contrast analyses revealed a higher AKDI for the single word rVl and sentence contexts compared to the isolated diphthong condition.

Table 6

/aʊ/ and F1 Descriptive Statistics

Sensor	Isolated		hVd		rVl		Sentence	
	M	SD	M	SD	M	SD	M	SD
T-MID X	274.3	177.4	192.7	100.5	407.0	201.9	357.9	187.7
T-MID Y	241.5	161.5	136.3	68.0	369.7	212.3	368.6	230.4
T-TIP X	363.7	230.4	291.2	173.0	425.2	172.7	409.5	183.3
T-TIP Y	261.9	195.1	130.2	41.2	541.8	227.2	387.3	197.2
LL X	133.4	65.8	137.3	54.4	193.4	83.5	262.1	153.8
LL Y	125.1	66.0	113.0	36.2	188.2	84.8	239.6	139.6
LIP	122.1	66.0	115.4	40.9	185.4	78.5	257.5	150.0

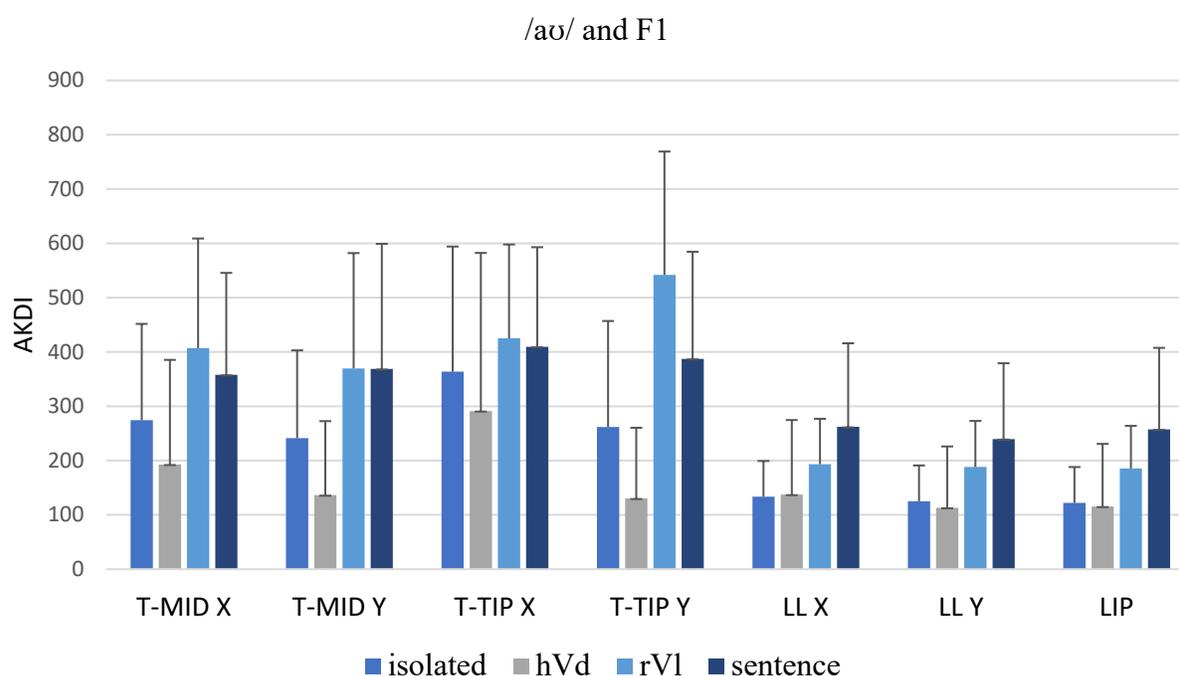*Figure 3.* Mean AKDI and standard deviation of each /aʊ/ and F1 relationship.

/aʊ/ and F2 Association

Descriptive statistics of the /aʊ/ and F2 relationships are shown in Table 7 and are further illustrated in Figure 4. There was a significant effect of context on the T-MID X and F2 association. Contrast analyses revealed that the AKDI in the single word rVl context was

significantly higher than in the isolated context. The AKDI in the sentence context was significantly lower than in the isolated context. There was a significant effect of context on the T-MID Y and F2 association. Contrast analysis revealed that the AKDI in the sentence context was significantly higher than in the isolated condition.

A significant effect of context on the T-TIP X and F2 association was observed. Contrast analysis revealed that the AKDI of the sentence context was significantly lower than in the isolated context. There was also a significant effect of context on the T-TIP Y and F2 association. Contrast analysis revealed that the AKDI of the single word rVl and sentence contexts was significantly higher than in the isolated context. There was no significant effect of context on the LL X and F2 association, the LL Y and F2 association, or the lip aperture and F2 association.

Table 7

/aʊ/ *and F2 Descriptive Statistics*

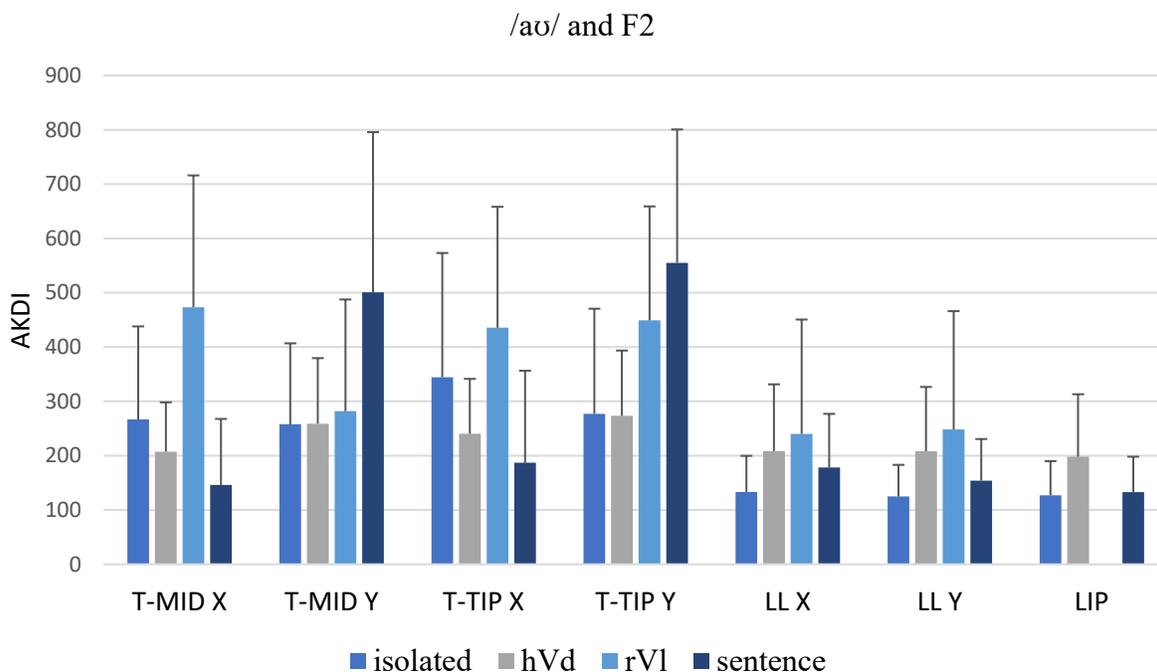| Sensor	Isolated		hVd		rVl		Sentence	
	M	SD	M	SD	M	SD	M	SD
T-MID X	266.6	171.5	207.3	90.9	473.2	242.9	146.0	121.6
T-MID Y	257.8	149.1	259.0	120.6	281.8	205.8	500.7	295.1
T-TIP X	344.2	229.0	240.4	101.1	435.6	222.6	186.9	169.4
T-TIP Y	277.0	193.6	273.4	120.1	449.2	209.8	554.9	245.8
LL X	133.1	66.6	208.4	123.0	240.2	210.5	178.2	98.8
LL Y	125.0	58.1	208.2	118.4	248.1	218.0	154.1	76.6
LIP	127.1	62.8	198.1	115.0	238.2	220.3	133.0	65.1

Figure 4. Mean AKDI and standard deviation of each /aʊ/ and F2 relationship.

Discussion

The primary purpose of this study was to investigate the effect of context on the relationship between the acoustics and kinematics of the diphthongs /aʊ/ and /aɪ/. Another purpose was to determine whether the results would differ by gender. There were no main or interaction effects of gender. The results of the analysis combining all participants show that context affects the acoustic and kinematic relationships in many instances. The single word rVl context and sentence context most often differed significantly from the baseline (isolated) condition.

The results of this study expand upon the work of McKell (2016), who reported on the baseline (isolated condition) relations between the acoustics and kinematics of the diphthongs /aʊ/ and /aɪ/. While it has long been known that coarticulation affects the production of individual speech sounds (Magen, 1997), it was not known how coarticulation might influence

the acoustic-kinematic relations. The findings of the current study suggest that coarticulation does have a significant effect on acoustic-kinematic relations. Magen (1997) suggested that it is not clear exactly how far coarticulatory effects can extend. The results of this study suggest that the more complex the phonetic context, the more the acoustic-kinematic relationship is affected. This is based on the observation that the rVl and sentence contexts yielded significantly different AKDI's compared to the baseline AKDI more often than was the case for the simple phrase hVd context.

Because of the nature of the task, McKell's (2016) findings did not take into account the effect of suprasegmental factors of speech. As explained by Nittrouer (1991), suprasegmental factors include variables such as rate and stress. The results of the present study support the hypothesis that suprasegmental factors may impact the acoustic and kinematic relations. The sentence context, which of all the conditions likely involved the most suprasegmental factors, had consistent significantly different AKDI values than the isolated diphthong context, which has no suprasegmental features.

/aɪ/

It is widely accepted that the frequency of F1 is inversely related to tongue height (Ferrand, 2015). The diphthong /aɪ/ moves from a central low position to a high front position. This tongue movement pattern should result in a strong inverse relation, and thus small AKDI, between F1 and the sensors tracking tongue height (T-MID Y and T-TIP Y). This also means there could potentially be a strong inverse relation between F1 and the sensors tracking lip aperture and lower lip vertical displacement (LIP and LL Y). Because F1 changes are less clearly linked to tongue advancement, it is less likely that there would be a small AKDI for F1 and the sensors tracking tongue advancement (T-MID X and T-TIP X). Forward movement of the sensor

tracking lower lip advancement (LL X) would be predicted to slightly lower the frequency of all formants as the vocal tract length increases. In the data, this expected pattern of smaller Y related AKDI's compared to X related AKDI's was observed only in the phrase hVd context and sentence context.

It has previously been demonstrated that F2 is directly related to tongue advancement (Ferrand, 2015). As the diphthong /aɪ/ moves from a central low position to a high front position, there is clear posterior to anterior movement. Therefore, a small AKDI between F2 and the sensors tracking tongue advancement (T-MID X and T-TIP X) would be predicted. Because F2 is not generally associated tongue height, it is less likely for there to be a small AKDI for the relations between F2 and sensors tracking tongue height (T-MID Y and T-TIP Y) or sensors tracking lip height and lip aperture (LL Y and LIP). This theoretical pattern was not consistently found in the data. For tongue movement, the X related AKDI values were smaller than the Y related AKDI values only in the isolated condition and single word rVl context.

For the diphthong /aɪ/ in the hVd context, only 3 of 14 acoustic-kinematic relationships had a significantly different AKDI compared to the isolated context. The reason for there being so few differences between the AKDI of the phrase hVd context and the isolated context is likely due to the minimally invasive coarticulatory effects of the sounds /h/ and /d/. Because /h/ is a voiceless glottal fricative that requires no specific articulatory shaping, it should have little effect on the following diphthong's acoustics or kinematics. Likewise, since /d/ is an alveolar stop that requires only the tongue tip to move, it also should have only a modest effect on neighboring sounds. Although the hVd word was produced in the phrase "I say hVd again," the surrounding /h/ and /d/ allowed the diphthong to be produced in a similar manner to the isolated production, and this only minimally influenced the AKDI.

It is noteworthy that in the phrase hVd context, only the tongue front (T-TIP X) but not the tongue mid (T-MID X) and F1 AKDI was significantly different from the isolated condition. This was unexpected because both sensors were tracking tongue movement. It could be speculated that the tongue front (T-TIP X) movement was exhibiting an anticipatory coarticulatory effect by beginning to move towards the final /d/ consonant during the diphthong. In the phrase hVd context, the F1 and tongue front (T-TIP X) relation was the only F1 and /aɪ/ relationship with a significantly different AKDI compared to the isolated context.

In the phrase hVd context the only /aɪ/ and F2 relations with significantly different AKDI values from the isolated condition were the F2 and mid tongue advancement (T-MID X) and F2 and tongue front advancement (T-TIP X) associations. Because it has previously been shown that F2 increases with tongue advancement, the production of two neighboring consonants appears to have influenced the kinematic-acoustic linkage by way of horizontal movement differences. The other articulatory and acoustic relations may not have led to significantly different AKDI values because of the minimally invasive nature of /h/ and /d/ on those aspects of the diphthong movement.

Because /r/ and /l/ are both liquid consonants, they have the potential to significantly influence neighboring sounds. The lip rounding associated with the articulation of "rile" and "raul" would slightly increase the length of the vocal tract, potentially decreasing the frequency of all formants. The strong coarticulatory influence of these sounds may explain the significant changes to the AKDI values for several articulators in comparison to the isolated diphthong condition.

In the single word rVl context for the /aɪ/ and F1 relationships, all tracked articulatory points besides lower lip advancement (LL X) presented a significantly different AKDI values

than in the isolated context. Lower lip advancement (LL X) may not have shown a significantly different AKDI in the phrase rVl context compared to the isolated context because lip advancement and F1 have not been reported previously as being closely linked.

In the single word rVl context for the /aɪ/ and F2 relationships, all tracked articulatory points besides tongue advancement (T-MID X) and lower lip advancement (LL X) presented a significantly different AKDI than in the isolated context. Many articulatory points and F2 relationships were significantly different than the AKDI of the diphthong produced in an isolated context. It appears likely that the strong influence of /r/ and /l/ on neighboring sounds led to a less straightforward connection between the kinematics and acoustics. It is not clear why mid tongue advancement (T-MID X) and F2 or lower lip advancement (LL X) did not have significantly different AKDI's than in the isolated condition when so many other acoustic/kinematic linkages were affected.

Across the /aɪ/ F1 and /aɪ/ F2 relationships for the sentence context, 12 of 14 relations had significantly different AKDI values than in the isolated context. This suggests that there are strong contextual effects on acoustic and kinematic associations. As Magen (1997) explained, coarticulatory effects can extend beyond neighboring sounds. It could be speculated that the sentence context AKDI was consistently significantly different from the isolated context AKDI because there were many surrounding sounds that had the potential to cause coarticulatory effects. The infiltration of neighboring sounds was apparent on the acoustic waveform as speakers would blend together the words "sight of the." For example, there was little evidence of a voiceless plosive /t/ but rather a continuation of voicing through the brief plosive. As all of these potential coarticulatory effects surrounded the word "sight," speakers' AKDI values understandably differed from the isolated productions of the diphthong /aɪ/.

In the sentence context for the /aɪ/ and F1 relationships, all tracked articulatory points except for mid tongue height (T-MID Y) showed a significantly different AKDI than the isolated context AKDI. Mid tongue height (T-MID Y) may not have varied significantly from the isolated context because, as Farnetani and Faber (1992) explained, the jaw contributes the gross movements but the tongue (likely the tongue tip in this scenario) provides the fine shaping. This would suggest that the movements of the mid tongue did not change significantly during the diphthong portion of the word "sight" compared to when simply producing /aɪ/ in isolation. However, the movement of the tongue tip, and potentially its relation to acoustics, may have been affected by the increased articulatory demands of sentence production.

In the sentence context for /aɪ/ and F2, similar to the /aɪ/ and F1 results, all tracked articulatory point except for mid tongue height (T-MID Y) had significantly different AKDIs than in the isolated condition. As noted above, the tongue tip likely played a stronger role in shaping the segments in the sentence than the tongue midpoint.

/aʊ/

During the diphthong /aʊ/ the tongue moves from a central low position to a high back position. As the tongue rises from a low to high position F1 should theoretically decrease. There should be a strong inverse relation and thus a small AKDI for F1 and the sensors tracking tongue height (T-MID Y and T-TIP Y). As the tongue retracts from a central position to a back position F2 should slightly decrease. A strong direct relationship and consequently a small AKDI should be found for F2 and the sensors tracking tongue advancement (T-MID X and T-TIP X).

In the phrase hVd context of the /aʊ/ and F1 connection, mid tongue height (T-MID Y) and tongue front height (T-TIP Y) were the only tracked articulatory points with significantly different ADKI values compared to the isolated condition. Because there should be a strong

inverse relationship between tongue height and F1, leading to a small AKDI, the changes in the hVd phrase context appear to be related to vertical movement differences and their impact on the connection between movement and the formant trajectories. For the other tracked articulatory points and their F1 relationships as well as the /aʊ/ and F2 relationships, none were significantly different in their AKDI values compared to the isolated context. This is likely because of the minimally invasive nature of the surrounding sounds /h/ and /d/ on these movement parameters.

In the single word rVl context of the /aʊ/ and F1 comparisons, mid tongue height (T-MID Y) and tongue front advancement (T-TIP X) were the only sensors that did not show a significantly different AKDI than in the isolated condition. One potential reason why tongue front advancement (T-TIP X) was not significantly different may be because there is not usually a strong linkage between F1 and tongue advancement. In the /aʊ/ and F1 associations across all contexts, the T-TIP X AKDI was never significantly different than in the isolated context.

In the single word rVl context of the /aʊ/ and F2 association, only the tracked articulatory points of mid tongue advancement (T-MID X) and tongue front height (T-TIP Y) showed significantly different AKDI's compared to the isolated condition. Because we would expect a small AKDI for F2 and sensors measuring tongue advancement, it would be anticipated that tongue advancement (T-MID X) would be significantly different in this context where /r/ and /l/ have such a strong coarticulatory influence. In isolation, there should be a small AKDI. The sound /r/ involves movement of the mid tongue and therefore coarticulation from this could cause a change in the diphthong movements and potentially their relation to the acoustics. The lip rounding could lead to the lowering of all formant frequencies as the vocal tract lengthens while producing "raul." The tongue tip height (T-TIP Y) may have been significantly different than when the diphthong was produced in isolation because when /aʊ/ is produced in isolation

there is little vertical movement of the tongue tip. But when adding the alveolar /l/ to the end of the diphthong, and presuming anticipatory coarticulation, the tongue tip becomes active when saying "raul." This action combined with the lowering of formants associated with increasing vocal tract length may have caused a significant change in the AKDI.

In the sentence context, the /aʊ/ and F1 comparisons all showed significantly different AKDI values except for mid tongue advancement (T-MID X) and tongue front advancement (T-TIP X). The fact that 5 of 7 of these tracked articulatory points in their relation to F1 showed significantly different AKDI values compared to the isolated condition is evidence of the strong effect of sentence context on the acoustic and kinematic relationship.

The /aʊ/ and F2 association in the sentence context resulted in 4 of 7 tracked articulator points showing significantly different AKDI values compared to the isolated condition. The sensors that were not significantly different were lower lip advancement (LL X), lower lip height (LL Y), and lip aperture (LIP). Lower lip height (LL Y) and lip aperture (LIP) would not be anticipated to differ significantly as there is no predicted connection between F2 and lip height.

Limitations of the Present Study and Directions for Future Research

One limitation of the current study is the simplicity of the method used to compare acoustics and kinematics. While the approach certainly reveals a general relationship between the measured kinematic movements and the diphthong formant trajectories, it does not account for any nonlinear relationships between movement and sound. Although the movement patterns of individual sensors on the articulators were being compared to formant trajectories to learn how the kinematics may have shaped the acoustics, there are other factors that could affect the formants, such as vocal tract length and individual differences in vocal tract shape.

Another limitation is some uncertainty whether the effect of context on AKDI values is due to the length of the phrase or the nature of the word in which the diphthong is produced. The phrase level diphthong was in an hVd context, while the single word diphthong was produced in an rVl context. It would be valuable to know whether the phrase length had an effect on the acoustic and kinematic relationship. This would require production of the diphthong in the same word context (rVl or hVd for example) and then producing that word in different phrase lengths.

This study analyzed only five samples of each diphthong in each context. As individuals' productions can vary with each sample, it would be beneficial to have a larger amount of productions from each speaker. This would ensure that the averages and data are as representative of the speakers' typical speech patterns as possible. A greater number of participants would allow stronger inferences to speakers in general.

Conclusion

The results of this study demonstrate that context has a significant impact on acoustic and kinematic relationships. As the sentence context resulted in the highest amount of significantly different AKDIs when compared to the isolated condition, the single word rVl context resulted in the second highest amount, and the phrase level hVd context resulted in the lowest amount, it can be presumed that the more complex the context the greater the effect on the acoustic and kinematic relationship. These data have significant implications for researchers who study acoustics to make inferences about speech movements. This impact of context shown in the current study suggests that caution is warranted in relying on acoustics to draw inferences about articulator movements in complex phonetic contexts. These results also suggest that a study of sounds produced in one context does not allow straightforward generalization to other contexts.

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

Farnetani, E., Faber, A. (1992). Tongue-jaw coordination in vowel production: Isolated words versus connected speech. *Speech Communication*, *11*, *401-410.*

Ferrand, C. T. (2015). *Speech science: An integrated approach to theory and clinical practice* (3rd ed.): Boston, MA: Pearson.

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.

Laaksonen, J.-P., Rieger, J., Happonen, R.-P., Harris, J., & Seikaly, H. (2010). Speech after radial forearm free flap reconstruction of the tongue: A longitudinal acoustic study of vowel and diphthong sounds. *Clinical Linguistics & Phonetics, 24*, 41-54.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of Acoustical Society of America, 78*, 1949-1957.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics, 25*, 187-205. doi:https://doi.org/10.1006/jpho.1996.0041

McKell, K. M. (2016). *The association between articulator movement and formant trajectories in diphthongs* (Unpublished master's thesis) Brigham Young University, Provo, Utah.

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)

Nittrouer, S. (1991). Phase relations of jaw and tongue tip movements in the production of VCV

    utterances. *Journal of Acoustical Society of America, 90*, 1806-1815.

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.

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

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

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 examined the relation of the first and second formants to tongue movement during diphthong production. The authors' purpose was to determine the accuracy with which acoustic measures represent lingual movement.

Method: Twenty native speakers of American English with normal speech, as determined by the experimenters, participated in this research. The speakers read aloud the sentence *the boy gave a shout at the sight of the cake* at normal loudness and rate, five times. While the participants were speaking, the acoustic signal was recorded. Kinematic measures were computed from a recording of the movements of a single magnetic sensor placed at the midline, 1 cm posterior of the tip of the tongue.

Results: As anticipated, /aʊ/, /aɪ/, and /eɪ/ all resulted in a negative correlation between the acoustic F1 measure and the kinematic measure of tongue elevation. Unexpectedly, only one of the twenty participants demonstrated strong negative correlation while producing the /ɔɪ/ diphthong. /aɪ/ involved the strongest correlation, while /ɔɪ/ involved the weakest correlation. /ɔɪ/ had the highest mean vertical displacement while /eɪ/ had the lowest mean vertical displacement. A strong positive correlation between F2 and anteroposterior tongue movement was anticipated. Each diphthong exhibited this strong positive correlation. /ɔɪ/ had the greatest anteroposterior displacement while /eɪ/ had the least anteroposterior displacement.
Conclusion: Many of the speakers' diphthongs revealed the expected associations between acoustic and kinematic measures, but there were also several unexpected patterns. Relations for F1, especially, exhibited several unexpected patterns that warrant further study. Individual speakers and individual diphthongs vary in the strength of their correlations between lingual movement and the first two formants. The widespread assumption of straightforward links between tongue movements and formants may not reflect the complexity of diphthong acoustics and kinematics, especially when analyzing diphthongs in the context of connected speech.

Relevance to the current work: The authors suggest that diphthongs' acoustic and kinematic properties are affected by coarticulation.

Farnetani, E., Faber, A. (1992). Tongue-jaw coordination in vowel production: Isolated words versus connected speech. *Speech Communication*, *11*, 401-410.

Objective: The authors investigated the positions of the tongue and jaw in the production of Italian vowels /i/ and /ɑ/ in various contexts. The aim of the study was to assess the two articulators' coordination, coarticulation, reduction, and compensation.

Method: The study analyzed data from one speaker of Standard Northern Italian. Data were collected with the Movetrack magnetometer system and electropalatography. A receiver coil was attached midline on the tongue dorsum to measure tongue movement and a receiver coil was attached midline on the lower incisors to measure jaw movement. The study considered the data of the tongue dorsum and the jaw's vertical and horizontal displacement. The participant read vowels /ɑ/ and /i/ in three utterance types; VCV pseudowords and words in isolation, CVCV

words in isolation, and CVCV words in pre-final phrase position in complex and natural sentences.

Results: The researchers determined that the tongue and jaw can function largely independent of each other. The articulators can work synergistically or in an opposing manner. While the jaw performs gross movements, the tongue provides more finite shaping of the vocal tract. Reduction caused by stress and utterance type is noted to affect both jaw and tongue movements.

Conclusion: The researchers concluded that compensation and change in the coordination of articulators occur in connected speech. They also concluded that a speaker attempts to conserve time and energy by partially centralizing articulator movements.

Relevance to the current work: Farnetani and Faber assessed vowels properties across speech contexts. The researchers found that the roles of the articulators adjust across speech contexts. As these roles adjust, the relation between the kinematic movements and acoustics may adjust across speech contexts as well.

Ferrand, C. T. (2015). *Speech science: An integrated approach to theory and clinical practice* (3rd ed.): Boston, MA: Pearson.

Relevance to the current work: Author addresses generally accepted rule of an association between the first two formants and tongue movement.

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: The paper summarized major approaches for acoustic analysis of dysarthric speech. The authors also identify equipment needed for modern speech-analysis and suggests various types of measurements for speech and voice analysis.

Relevance to the current work: The writers explain metrics reflective of speech subsystems. The authors also state that, as a general rule, F1 holds an inverse relationship to tongue height and F2 holds a direct relationship with tongue advancement. The current study will assess how accurate this general rule is across speech contexts.

Laaksonen, J.-P., Rieger, J., Happonen, R.-P., Harris, J., & Seikaly, H. (2010). Speech after radial forearm free flap reconstruction of the tongue: a longitudinal acoustic study of vowel and diphthong sounds. *Clinical Linguistics & Phonetics, 24*, 41-54.

Objective: The aim of this study was to describe the effect radial forearm free flap reconstruction of the tongue has on speech one year post surgery via acoustic analysis.

Method: Speech samples from 18 patients who had reconstructive surgery due to oral cancer were analyzed. One recording was made before the operation and the remaining recordings were made one, six, and 12 months following the operation.

Results: For males, significant formant changes were observed within participants one year post surgery compared to pre-surgery. However, the vowel space remained unchanged after the

surgery. The researchers propose the unchanged vowel space indicated no improvement in ability to produce intelligible vowel sounds. For females, no significant long-term changes were noted in the acoustics or size of vowel space.

Conclusion: While the males' vowel space remained unchanged the formants were observed to change. Unchanged vowel space indicated that the male speakers had no improvement in ability to produce vowel sounds, but rather, that male participants likely learned to compensate by using articulators other than the tongue.

Relevance to the current work: The researchers used acoustics to measure and compare speech conditions. The current study will analyze the relation of acoustics and kinematics across speech conditions to determine the reliability of acoustics as an independent form of measurement.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America. 78*, 1949-1957.

Objective: The purpose of the study was to assess phonetic and sentential effects on vowel duration, closure duration, and the C/V ratio as correlates to determine if word-final stops are voiced.

Method: The speech of three male and three female volunteers was assessed. Nine sets of minimal pair consonant-vowel-consonant words were analyzed. Each test word was embedded in one of four sentence frames and repeated twice, resulting in a total of 144 test sentences. Acoustic measurements were taken and vowel duration, postvocalic closure duration, and C/V ratios were extracted and statistically analyzed.

Results: Researchers' findings supported the existing evidence that vowels lengthen in phrase final positions. Vowel duration was found to be longer for words ending in voiced stops compared to words ending with voiceless stops. Vowel duration was a reliable indicator of voiced vs voiceless categories in all scenarios. Closure duration was a less consistent marker of voicing categories across contexts. C/V ratio did not reliably identify the voicing when the local phonetic environment changed or when it was not in a phrase-final position.

Conclusion: Of the three temporal attributes studied, vowel duration was the most reliable indicator of voicing across the various contexts of sentence position, local phonetic environment, place of articulation, and inherent vowel duration.

Relevance to the current work: The researchers used acoustics to assess the impact of various phonetic and sentence conditions on speech properties. The current study also addresses the effect of phonetic context on the association between acoustic and kinematic measures of speech.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics, 25*, 187-205. doi:https://doi.org/10.1006/jpho.1996.0041

Objective: The intent was to address the range and impact of coarticulation. The authors aimed to address the effect of linguistic units on coarticulation. The study specifically focused on: the distance coarticulatory effects extend, the potential role of stress facilitating these effects, and the strength of and nature of carry-over and anticipatory effects.

Method: The speech of four male speakers of American English, aged 27-37 years, was studied. Speakers were recorded producing 72 set phrases. The vowels /i/ and /ɑ/ were targeted.
Results: Coarticulatory effects were found to extend further than previously thought. The coarticulatory effects occasionally extended from the first vowel to the second vowel of a consonant-vowel-consonant-vowel word and vice versa, indicating that the coarticulatory effect can begin sooner, and last longer, than initially believed. This pattern however, was not noted for all speakers in all circumstances. The researchers attempted to discern which was greater, carry over or anticipatory effects, to determine a predictable pattern. It was found that each speaker differed in such patterns. A clear pattern of the effect of stress mediating coarticulatory effects was noted with only one speaker.

Conclusion: Coarticulatory effects can extend further than previous research had assumed. Whether the effects are typically anticipatory or a result of carry-over varies according to the speaker.

Relevance to the current work: The study speaks to the impact of coarticulation. The current study will also take into consideration the effects of coarticulation and how it will impact the relationship between acoustics and kinematics across range of contexts as coarticulation does not influence diphthongs in isolation but will impact diphthongs produced in other contexts.

McKell, K. M. (2016). *The association between articulator movement and formant trajectories in diphthongs* (Unpublished master's thesis). Brigham Young University, Provo, Utah.

Objective: The study examined the association between formant trajectories and tongue and lip movements in the diphthongs /aɪ/, /aʊ/, and /ɔɪ/.

Method: The speech of 17 native speakers of American English was examined. Electromagnetic sensors on their tongue and lips collected kinematic data which were time aligned with the acoustic data. F1 and F2 trajectories of the middle 50% of the diphthongs were compared to the kinematic data of the tongue and lips via absolute difference of z-scores along each track.

Results: Tongue movement may be the best predictor of F1 and F2 changes for the diphthong /aɪ/, likely due to the fact that the diphthong lacks lip rounding. Lip movement may be the best predictor of F1 and F2 changes in the diphthong /aʊ/, likely due to the substantial lip rounding of the offglide vowel. The diphthong /ɔɪ/ presented challenges in determining the relationship between the articulators and formant changes because the actions of the lips and tongue during this diphthong acted in opposite directions on F1.

Conclusion: The articulators that most contribute to F1 and F2 differ based on diphthong.

Relevance to the current work: Both studies assess the relation between acoustics and kinematics. The current study will adopt the novel method of data analysis used by McKell, and apply it to a wider range of speech contexts, as opposed to diphthongs in isolation.

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 study assessed the association between speech kinematics and acoustics in typical speakers. Speaking rate and loudness were varied to evaluate changes in both acoustic and kinematic measures. The authors anticipated that information from this study may contribute to the formulation of new therapy approaches that target articulatory performance patterns and intelligibility.

Method: 10 healthy, typical, native English speakers participated in the study. The speakers produced the sentence "Tomorrow Mia may buy you toys again" at a normal rate and loudness, at a fast rate, at a slow rate, and at a normal rate but loud volume. Kinematic measures were tracked using three-dimensional electromagnetic articulography. Data from the sensor placed on the midsagittal line, 6 cm posterior to the tip of the tongue, were analyzed due to its contribution to the movement of the diphthong /aɪ/. Acoustic data were gathered via a microphone placed 15 cm from the speaker's mouth.

Results: The study examined phonetic specification and phonetic variability. Kinematic data examination showed that slow speech resulted in larger lingual displacements than did typical, loud, and fast speech. Loud speech led to larger overall displacement than did typical and fast speech. Fast speech resulted in smaller lingual displacement than typical or loud speech. Acoustic data examination showed that during slow speech, vowel distances were larger than during typical, loud, and fast speech. Loud speech also caused vowel distances to be larger than during fast speech.

Conclusion: Lingual displacement was associated with changes in acoustic vowel distance. Conversely, kinematic and acoustic variability showed no positive linear correlation.

Relevance to the Current Work: Changes in the speaking condition affect the kinematic and acoustic relations, which is the focus of the present study.

Nittrouer, S. (1991). Phase relations of jaw and tongue tip movements in the production of VCV utterances. *Journal of Acoustical Society of America, 90*, 1806-1815.

Objective: Nittrouer analyzed the effect of various factors, such as rate and stress, on the relative timing of articulator movements.

Method: The speech of two adult males and two adult females was studied. The four utterances /tɑdɑt/, /tɑtɑt/, /tɛdɑt/, and /tɛtɑt/ were used in the carrier phrase "it's a ___ again". Four rate/stress patterns were introduced: slow/first syllable stressed, slow/first syllable unstressed, fast/first syllable stressed, and fast/first syllable unstressed. Each pattern was repeated 20 times by each participant. Acoustic and kinematic data were collected. Kinematic measures were collected via x-ray microbeam.

Results: A greater portion of the time was taken by the motion of the tongue tip raising during the short vowel cycles than during the long vowel cycles. Relationships such as this, between the duration of the vowel-consonant and the onset time of the tongue tip raising, were most determined by the effects of stress and rate. The researcher stated that while the impact of rate was evident in the data, it was not as strong as the impact of stress. Nittrouer reported that variability within conditions was low compared to the variability across conditions.

Conclusion: Suprasegmental factors influence spatiotemporal relationships among articulators. An array of factors contribute to the organization of speech.

Relevance to the current work: The study addressed the impact of suprasegmental factors on articulatory relations. This supports the current study's aim to assess acoustics and kinematics in an array of contexts, as sounds produced in isolation do not exhibit suprasegmental factors.

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.

Objective: The researchers analyzed vowel articulation in speakers with muscle tension dysphonia before and after participating in manual circumlaryngeal treatment.

Method: Recordings of 111 women with muscle tension dysphonia who showed improvement after receiving manual laryngeal reposturing and/or circumlaryngeal massage were analyzed. Recordings were of the second and third sentences from The Rainbow Passage. Quadrilateral vowel space area and vowel articulation index for each speaker were calculated for the vowels: /i, æ, ɑ, u/.

Results: Quadrilateral vowel space area and vowel articulation index were found to notably increase in the post-treatment samples as compared to the pre-treatment samples.

Conclusion: Speakers have improved acoustic vowel space, and ideally improved vowel articulation, after participating in manual circumlaryngeal therapy.

Relevance to the current work: Acoustic measures via formants were used to make inferences about articulatory movements.

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

Relevance to the current work: Stevens explains the quantal nature of speech. He explains that acoustics are not equally sensitive to all articulatory movements. The author suggests this nonlinear phenomenon is a factor in phonology in language. Stevens explains that phoneme boundaries may reflect boundaries between areas with articulatory-acoustic sensitivity. The author's attention to the nonlinear relationship between articulatory movements and acoustic measures supports the current study's purpose of further analyzing articulatory movement and acoustic relationships.

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: The purpose of this study was to examine the changes in kinematic and acoustic measures in a group of individuals with Amyotrophic Lateral Sclerosis (ALS). The researchers focused on the slope of the second formant transition and on tongue movement speed as its kinematic correlate.

Method: Thirty-one individuals diagnosed with ALS, and presenting with at least one symptom of bulbar involvement, participated in the study. They presented with a range of intelligibility and speaking rates. The ALS group was distributed into two-sub groups, based on speaking rate:

ALS with a normal rate (AN), and ALS with a slow speaking rate (AS). The control group consisted of 12 neurologically healthy speakers. All participants repeated the sentences 'Say doily again' and 'I love Seattle in the spring', five times. The tongue movements and formant frequencies during the opening of /doI/ and /jæ/ and the closing movement in /oI/ were evaluated. Participants with ALS repeated these sentences at their habitual speaking rate while participants without ALS repeated these sentences at their habitual speaking rate, as well as at half of their typical rate.

Results: Speakers without ALS, when speaking with a reduced speaking rate, decreased the speed of their articulatory movements yet increased the range of their articulatory movement. Thus, their F2 range increased when speaking rate decreased. The participants with ALS in the AS subgroup had a shallower F2 slope as well as slower kinematic movement than the control group. These speakers also had a weaker association between F2 range and tongue displacement. The characteristic most strongly correlated with intelligibility was the F2 slope, rather than movement measures.

Conclusion: Healthy speakers experienced an increased F2 range when speaking slowly, whereas individuals with ALS experience a reduced F2 range the slower they spoke. From this information, we can infer that there is a difference between a healthy, intentionally slowed speech system compared to a speech system with a slow rate caused by disease. The F2 slope is a suitable indicator of disease progression in speakers with ALS. However, in the early disease stages, kinematic measures may be more useful to track changes because there may be no compensatory strategies in place which might affect those measures, and because it directly measures only one organ system.

Relevance to the Current Work: Although in some instances acoustic and kinematic measures were related to each other as expected, in other instances the measures varied in relation from what was expected.