2023-08-11

The Effect of Speaking Context on Articulatory Kinematics in Habitual and Clear Speech

Lauren Elizabeth Clarke
Brigham Young University

The Effect of Speaking Context on Articulatory Kinematics in Habitual and Clear Speech

Lauren Elizabeth Clarke

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
Garrett Cardon
Dallin Bailey

Department of Communication Disorders

Brigham Young University

ABSTRACT

The Effect of Speaking Context on Articulatory Kinematics
in Habitual and Clear Speech

Lauren Elizabeth Clarke
Department of Communication Disorders, BYU
Master of Science

This study examined the effect of speaking contexts on articulatory kinematics in habitual and clear speech conditions. Ten male and 10 female participants (ages 18–29) completed speaking tasks in three contexts and two conditions. The contexts were word, phrase, and passage, with both mid-sentence and end of-sentence stimuli in the phrase and passage contexts. The two conditions were habitual and clear speech. Participants had sensors attached to the mid-tongue, jaw, lower lip, and upper lip, and an electromagnetic articulograph tracked their movements. Three tokens for each stimulus were analyzed for duration, displacement, and velocity. Articulatory coordination was measured through absolute and percent jaw contribution, and displacement correlations. Statistical analysis revealed significant changes across both conditions and contexts. Generally, the articulator movements were larger for clear versus habitual speech and decreased progressively in size from word to phrase to passage. Duration significantly increased in the clear speech condition and decreased from word to phrase to passage, which likely underlies the changes seen in the other measures. Percent jaw contribution to lower lip movement was significantly higher in the clear speech condition, percent jaw contribution to tongue movement was significantly higher for the passage compared to the other contexts, and jaw and lower lip correlations with the tongue were higher in the clear condition and lower in the passage context. Incidental rate variation and motor equivalence across speakers limit the degree to which we can interpret these results in terms of articulatory coordination. Overall, this study demonstrates significant changes in speech kinematics across contexts in both clear and habitual conditions, indicating that researchers should exercise caution when generalizing findings from studies using short, contrived stimuli.

Keywords: speech improvement, articulation (speech), motion

ACKNOWLEDGMENTS

I am grateful for the mentorship of Dr. Christopher Dromey throughout the research process and his encouragement in pursuing my goals and interests. I am also grateful for Dr. Garrett Cardon and Dr. Dallin Bailey and their willingness to provide insights and feedback as part of my thesis committee. I would also like to express appreciation for the graduate students that worked with me and assisted me through the data collection process and provided encouragement during setbacks and complications. Finally, I want to thank my friends and family who supported me throughout the completion of my thesis and showed enthusiasm for my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF THESIS STRUCTURE AND CONTENT

This thesis, *The Effect of Speaking Context on Articulatory Kinematics in Habitual and Clear Speech*, is written in a hybrid format that aligns with length and style requirements for submission to research journals in communication disorders. The initial pages comply with university submission requirements. The annotated bibliography is included in Appendix A. The Institutional Review Board Approval Letter is in Appendix B. Appendix C contains the reading passage stimulus used in this study.

Introduction

Speech therapy for motor speech disorders aims to improve speech production by changing articulatory movements. Consequently, a large body of research has been completed to measure and describe motor behaviors of the speech mechanism. Understanding the characteristics of articulation in healthy speakers can pave the way for studies of the underlying difficulties contributing to disordered speech. Moreover, examining how different methods of modifying speech, such as reduced rate or increased loudness, change articulatory movements can guide speech pathologists in selecting appropriate treatment targets. In many kinematic studies, diphthongs have been a common target due to their contribution to speech intelligibility and the extent of movement involved in their production.

While acoustic data are informative and more accessible clinically, measuring speech movements provides insights not attainable from acoustics alone. Consider the principle of motor equivalence. The same perceptual and even acoustic output can result from different combinations of articulatory movements (Hughes & Abbs, 1976). For example, both within and across speakers' productions of the rounded, high-back vowel /u/, the tongue may reach a range of posterior positions and the lips may protrude to varying degrees, but all productions could still be perceived as /u/ (Stevens, 1989; Weismer et al., 2003). Due to the nonlinear relationship between kinematics and acoustics, it is important to measure movements directly to understand how targeted interventions may affect articulation.

In recent years, the construct of "clear speech" has been a popular area of kinematic and acoustic research. Clear speech is a common therapy target involving over-enunciated speech and has been found to improve speech intelligibility (Whitfield et al., 2021). Investigators have found that using clear speech compared to habitual speech results in increased velocity, displacement, and range of motion of the tongue, lips, and jaw, as well as increased syllable and utterance duration, vocal intensity, and acoustic vowel contrast (Dromey, 2000; Kuruvilla-Dugdale & Chuquilin-Arista, 2017; Mefferd, 2017; Whitfield et al., 2021).

While speech kinematic research can provide valuable information regarding the movements underlying disorders and treatment approaches, current kinematic research has some limitations. One experimental design problem is that this line of inquiry often relies on highly controlled contexts, with the participants repeating a word or phrase with carefully chosen phonetic components to elicit specific articulatory movements. For example, researchers have often had participants say "buy Bobby a puppy" when targeting lower lip and jaw movement due to this sentence having a high number of bilabials (Dromey, 2000; Maner et al., 2000; Whitfield et al., 2021; Wisler et al., 2022). Repeating speech in contrived contexts may not be reflective of conversational speech for two main reasons. First, predetermined speaking contexts may not involve the same cognitive load as everyday communication. While conversational speech may at times involve short, simple phrases, it often involves longer, more complex utterances. Maner et al. (2000) found that the spatiotemporal index (STI), a measure related to articulatory instability, increased in participants as they produced longer, more complex utterances. This could suggest that articulation is less stable when linguistic processing demands increase, which would mean that a chosen experimental speech context could have different kinematic results if processing demands differ.

Secondly, the repetition of a stimulus and potential habituation to the task will reduce the attentional demands. Producing spontaneous language rather than repeatedly saying the same utterance would place greater cognitive demands on the speaker. Research is needed to understand how repetitive, contrived tasks compare with more naturalistic speaking contexts.

Furthermore, speaking under novel conditions, as frequently targeted in therapy (e.g., loud speech, clear speech, etc.), also requires increased attention due to unfamiliarity with the speaking condition. A number of studies have shown that increased attentional demands affect speech kinematics (Bailey & Dromey, 2015; Whitfield et al., 2021). Whitfield et al. (2021) found that accuracy on a visuomotor task decreased when using clear compared to habitual speech, suggesting that clear speech requires increased attentional demands due to being an unfamiliar speaking condition. For the purposes of this paper, speech modification techniques, such as clear speech or loud speech, will be referred to as speaking conditions while target productions of various phonetic complexity level, such as words versus sentences, will be referred to as speaking contexts.

While processing and attentional demands have been suggested to influence kinematic measures across contexts of increasing phonetic complexity and during unfamiliar conditions, Tasko and McClean (2004) examined movement strokes across speech contexts of increasing length and found no differences when participants were using typical speech. They did, however, find context effects on speed, distance, and duration when participants spoke under conditions of altered rate and loudness. Asking participants to consciously vary their rate and loudness when speaking is a relatively unfamiliar speaking condition and likely demands increased attention. This suggests that even if context effects are not perceptible in habitual speech, unfamiliar conditions such as clear speech may be influenced by context effects. However, current research regarding context effects is limited and does not include clear speech.

Overall, there is reason to believe a particular speech context and a clear speech condition could influence the kinematic results of a study. For research regarding motor speech assessments and interventions to be most valid, it is vital to understand how well the stimuli relate to everyday communication. For example, if a study investigates a clear speech condition to determine ease and effectiveness in intervention for dysarthric patients using multiple repetitions of a single word, do the findings reflect the efficacy of clear speech in naturalistic speech when cognitive and attentional loads are increased? Understanding what is similar and different across speech contexts and conditions will help research more directly apply to clinical practice. There is a need for more research comparing speech modification conditions (such as clear speech) across contexts of increasing complexity while accounting for repetitiveness and unfamiliarity of speaking conditions.

One difficulty in comparing speech across contexts is the challenge of using sufficiently sensitive measures to detect subtle kinematic changes. Tasko and McClean (2004) used movement strokes in their study to compare speaking contexts without needing a specific phonetic target. Strokes are defined as “the period between two successive local minima in the speed history of an articulator point” (Tasko & Westbury, 2002, p. 127). The defined stroke periods are then analyzed using measures such as the stroke distance (the sum of the segment lengths along the trajectory), stroke duration, and peak or boundary stroke speeds. Stroke metrics provide valuable information about movement patterns of the articulators, but they are limited to distance, time, and speed units. Strokes are generalizable to any length of speech task due to their lack of specificity. They do not capture direction of movement, relative timing and contribution of articulatory gestures, or the link between kinematics and acoustics. As a result, movement strokes provide insight on a global kinematic level but gloss over segmental differences in speech production such as the coordination of multiple speech gestures. There is a need for research comparing speech contexts using more sensitive measures that could detect small changes in articulation.

Examining a small unit of speech, such as a diphthong, requires analysis at a fine level of detail to consider small changes in kinematic features such as velocity, displacement, and timing. When analyzing speech on a segmental level, measures should be specific to the expected gestures used to produce the target sound. For example, when analyzing the diphthong /aɪ/, measures may include duration and magnitude of tongue and jaw movements and the relationship between vertical tongue position and change in the first formant (Dromey et al., 2013). Adams et al. (1993) computed a velocity profile for each set of kinematic sensor data that considered magnitude of change, peak velocity, duration, symmetry of movement, and number of velocity peaks for each gesture. Using these measures, the authors detected subtle changes in articulation under conditions of altered speaking rate.

Speech is a complex motor behavior that often requires the coordination of multiple speech gestures, and it is possible that altering the speech condition or context may alter the coordination of these gestures. Measures of coordination may include kinematic data from multiple articulators or examine coordination between kinematic and acoustic data. Coordination can be considered on multiple dimensions. Temporal coordination may be measured by calculating the correlation between the direction of movement for two articulators or by considering timing of maximum or minimum formant value, movement velocity, or sensor displacement (Bailey & Dromey, 2015; Weismer et al., 2003). Spatial coordination may include comparison of the variability of displacement magnitudes to determine relative contributions of specific gestures to an overall articulatory target (Hughes & Abbs, 1976; Westbury et al., 2002). Highly sensitive measures are more likely to detect subtle motor coordination changes across conditions and contexts that would not be observed in a broad measure such as movement strokes.

To compare detailed measures across contexts of increasing phonetic complexity, the target production must have a consistent phonetic context to control for the influence of coarticulation on connected speech. The speech contexts should have the same phonemes both before and after the target to ensure any kinematic changes observed are due to context effects rather than influenced by coarticulation.

Statement of the Problem

There is insufficient research to support the generalization of findings using short, contrived phrases to more naturalistic everyday speech. A large body of research in speech kinematics relies on the use of brief, highly controlled stimuli and often draws clinical inferences from the findings. There is a need for research into kinematic changes across speaking tasks in order to increase our confidence in the application of the findings to conversational speech, which is the ultimate goal of clinical intervention.

Statement of the Purpose

The purpose of this study is two-fold: (a) to determine if there are context effects in typical speech when using fine-grained kinematic measures to infer the generalizability of word-level kinematic stimuli to connected speech; and (b) to examine whether the expected kinematic differences between habitual and clear speech are present to the same degree across a variety of speech contexts.

Research Hypotheses

This study will address the following research hypotheses:

1. We predict that coordination will decrease in a clear speech condition and vary across utterance contexts.

2. We predict decreased segment duration and magnitude of movements in longer speech contexts, such as phrases or extended passages.

Method

A description of the participants, instrumentation, stimuli, procedure, kinematic analysis, and statistical analysis are provided below. Prior to participating in the study, participants were provided with an overview of the study and signed a consent form approved by the Brigham Young University Institutional Review Board.

Participants

Participants in this study were 20 young adults who were native speakers of Standard American English with no history of speech, language, or hearing disorders, as determined by an initial interview. There were 10 men (ages 18–29) and 10 women (ages 18–25). Individuals were recruited from the university community.

Instrumentation

Multiple sets of data were collected in each recording session, only a portion of which was used in this study. All recordings were completed in a single-walled sound booth with the participant positioned 30 cm from a condenser microphone (AKG C2000Ba). To allow for software measurement of speech intensity in dB sound pressure level (SPL) from the microphone signal, a calibration vowel was recorded at 50 cm and measured with a sound level meter. Articulatory kinematic data were collected with an NDI Wave electromagnetic articulograph (Northern Digital Inc. Waterloo, Ontario, Canada), which recorded data from the following 3 mm sensor coils: two reference sensors on eyeglass frames which served as the coordinate system origin and allowed for head movement correction; two coils at midline on the vermillion borders of the upper (UL) and lower lips (LL); two coils at midline at the tongue front (not used in this study) and mid-tongue (TM); and one coil attached to the mandibular central incisors to measure jaw movement (J). The sensor coils were attached using cyanoacrylate adhesive and small squares of Stomahesive (ConvaTec, Inc.) for the jaw sensor to protect tooth enamel. A computer outside the sound booth tracked and recorded the x, y, and z positions of each coil using the Wavefront software (Northern Digital Inc., 2017). The movement data were sampled at a rate of 100 Hz while the audio signal sampling rate was 22,050 Hz.

Speech Stimuli

The target word "type" was used to elicit the diphthong /aɪ/, which involves substantial tongue movement in an easy to segment context and is a frequent target in kinematic studies. Each stimulus was defined by condition and context. Conditions were habitual (HABIT) or clear (CLEAR), and contexts included a target word in isolation (WORD), the target word embedded in the middle of a phrase (PHR M) or at the end of a phrase (PHR E), and a reading passage with the target word both in the middle and at the end of sentences (PAS M or PAS E). In all M and E conditions, the target word was kept in the same phonetic context: "she's a type A person," or "the right type." This resulted in a total of 10 stimuli. The target word was put into a phrase to determine if the chosen surrounding phonetic context in each level of stimuli resulted in coarticulatory influences when compared to the word in isolation. Two phrases were used: "she's a Type A person," and "she insisted on having the right type."

The reading passage was used to examine potential kinematic changes as the target occurred in productions of increasing length and phonetic complexity. Each stimulus was presented on a screen for the participants to read after instruction.

For HABIT stimuli, the participants were asked to read the stimuli in a comfortable speaking voice, and for CLEAR stimuli, they were asked to read the stimuli with exaggerated speech clarity. Before presenting the CLEAR stimuli, the researcher told the participant, "Now I am going to have you read again, but this time I want you to speak as clearly as possible as you read. Imagine we are in a noisy room and enunciate every word so that I can understand what you are saying." In the contexts WORD, PHR M, and PHR E, the participants produced the stimulus five times for each condition, with pauses included between productions. In the contexts PAS M and PAS E, the target word was included three times each, so participants were only asked to read the passage once for each condition.

Procedure

After coils were attached to the articulators, the experimenter engaged the participant in conversation for five minutes to help them habituate to the sensation of speaking with sensors attached to minimize learning effects during data collection. After this period, all participants were presented with all HABIT stimuli first, then all CLEAR stimuli due to concern that the CLEAR condition could potentially influence how participants spoke in the HABIT condition. The order of contexts (WORD, PHR, and PAS) within each condition and the order of stimuli in the WORD and PHR contexts were randomized.

Kinematic Analysis

Segmentation

The kinematic recording of each participant was first roughly segmented into separate tokens for each stimulus using a custom Matlab (Mathworks, 2023) application. Each token was then finely segmented using a second custom Matlab application. Kinematic markers for fine segmentation were selected based on expected articulatory patterns at the beginning and end of the diphthong and the observed consistency of those landmarks across tokens. The first cursor was placed on the TM acceleration peak nearest to the corresponding TM velocity zero-crossing (see Figure 1). In cases in which one of the acceleration peaks was not clearly closer to the zero-crossing, the highest acceleration peak within the acoustic boundaries of the target word was used. For the second cursor, the highest LL velocity peak within the acoustic boundaries of the target word was used (see Figure 1). In cases where the signal was unusual and could not be segmented according to this procedure, the token was considered corrupted and replaced with another or omitted from analysis. Typically, three tokens from each participant were used for each of the five stimulus types (WORD, PHR M, PHR E, PAS M, and PAS E).

Kinematic Measures

Kinematic measures were selected to reflect expected low-to-high articulatory movement for the diphthong /aɪ/ and the expected larger magnitude of movements typically seen in a clear speech condition. These measures were the segment duration and the maximum vertical displacement and velocity for each marker (TM, J, LL, and UL). Coordination was measured through articulator displacement correlations and the contribution of J to TM and LL displacement. The selected measures of coordination included TM correlation with LL, TM correlation with J, LL correlation with J, and absolute and percent J contributions to TM and LL vertical movements.

Statistical Analysis

A linear mixed model (LMM) analysis in SPSS 28 was conducted to test for differences in the dependent variables across the conditions and contexts. Condition (HABIT and CLEAR) and context (WORD, PHR, and PAS) were fixed factors and speaker was the random factor. Speaker sex was included as a covariate. Bonferroni corrected multiple comparisons that were part of the LMM computation examined differences between the contexts.

Results

This study examined the effect of speaking contexts on kinematic measures in both habitual and clear speech conditions. Articulators reported in this study include the mid-tongue (TM), jaw (J), lower lip (LL), and upper lip (UL). Intra-rater reliability of segmentation is reported below. Effects of the HABIT and CLEAR speaking conditions are reported in Tables 1 and 2. Effects of the WORD, PHR, and PAS contexts are reported in Tables 1 and 3. Figures 2 and 3 illustrate examples of the changes seen across conditions and contexts.

Reliability Testing

Data from two randomly selected participants were segmented a second time and compared with the original results to measure intra-rater reliability. The mean correlation between the original and re-segmented data was .996, indicating excellent reliability.

Effects of Conditions on Kinematic Measures

Descriptive statistics for the two speaking conditions, HABIT and CLEAR, are found in Table 1. The results of LMM testing of the conditions are found in Table 2. There was a significant main effect for all kinematic measures except UL velocity and LL-J correlation. Figure 2 provides an example of a change across conditions. The segment duration increased in the CLEAR condition, as did the displacement and velocity for TM, J, and LL, as well as displacement for UL. TM correlation with LL and with J increased significantly in the CLEAR condition. The J contribution to TM and to LL displacement both significantly increased in the CLEAR condition, but only the J percent contribution to LL significantly increased in the CLEAR condition.

Effects of Contexts on Kinematic Measures

Table 1 displays the descriptive statistics for the three contexts—WORD, PHR, and PAS—and Table 3 displays the results of LMM testing and Bonferroni corrected multiple comparisons. There was a significant main effect for segment duration. Comparison analysis revealed that duration significantly decreased across tasks, with WORD having the longest duration, followed by PHR, and PAS having the shortest duration.

There were significant main effects for TM, J, and LL displacement which are visualized in Figure 3. Comparison analysis revealed that TM and J displacement significantly decreased from WORD to PHR and then to PAS stimuli. The LL displacement decreased significantly between WORD and PAS and between PHR and PAS. A significant main effect was also found for TM velocity, with comparison analysis revealing velocity decreasing significantly from WORD to PHR, and then to PAS.

Additional significant main effects were found for J contribution to TM and LL displacement and for J percent contribution to TM. For both J contribution to TM and LL, comparison analysis revealed significantly decreasing J contribution across tasks. The WORD tasks had the highest J contribution to TM and LL, followed by PHR, and PAS had the lowest J contribution. The percent contribution of J to TM increased significantly from WORD to PAS and from PHR to PAS.

Finally, there were significant main effects for TM correlation with both LL and with J. Concurrent contrasts revealed that for both variables, correlation decreased significantly from PHR to PAS and from WORD to PAS.

Discussion

The purpose of this study was to infer the generalizability of speech in shorter contexts to more naturalistic contexts in both habitual and clear speech conditions. Significant changes were found between conditions (HABIT and CLEAR) and across contexts (WORD, PHR, and PAS). Segment duration increased from HABIT to CLEAR and decreased from WORD, to PHR, and then to PAS. The articulator movements were generally larger for CLEAR than HABIT and decreased progressively in size from WORD to PHR to PAS.

Effects of Conditions on Kinematic Measures

Displacement increased in the CLEAR condition for all articulators, and velocity increased for all except for the UL. These findings are consistent with those of other studies (Dromey, 2000; Kuruvilla-Dugdale & Chuquilin-Arista, 2017; Mefferd, 2017; Whitfield et al., 2021). For example, Kuruvilla-Dugdale and Chuquilin-Arista (2017) found that speed and range of movement increased, and Whitfield et al. (2021) found that duration, velocity, and movement range increased during a clear speech condition.

Clear speech is considered to be exaggerated speech. It generally involves larger articulatory movements than habitual speech and results in increased intelligibility (Whitfield et al., 2021). Clear speech is often used as a therapy target when treating individuals with dysarthria since it helps patients compensate for the reduced movements frequently associated with dysarthria. Our study supports the utility of targeting clear speech, since speakers were able to produce significantly larger movements with minimal instruction.

Effects of Contexts on Kinematic Measures

Tasko and McClean (2004) examined task effects across speech contexts (a nonsense phrase, a sentence, a reading script, and a monologue). The authors found the nonsense phrase to have higher duration, displacement, and velocity of movement strokes than the other tasks. However, the stroke metrics for the other tasks were not significantly different from each other. This is different from the findings in the current study, which revealed significant differences across WORD, PHR, and PAS. TM velocity and TM and J displacement decreased from WORD to PHR and then to PAS. The PAS context also involved decreased LL displacement compared to the other two contexts. It is possible that differences between speaking contexts were not found by Tasko and McClean (2004) but were found in the current study because of methodological differences. We chose detailed segmental kinematic measures at the syllable level, while the stroke metrics used in Tasko and McClean (2004) were more general and did not offer the same degree of detail in analysis.

Measures of Coordination

We hypothesized that coordination would vary across speaking contexts and decrease in the CLEAR condition. Analysis at the syllable level limited the type of measures that could be used to examine coordination, but we attempted to measure this construct through displacement correlations of articulators along with the absolute and relative J displacement contributions to the movements of LL and TM. Several variables seemed to underly the significant changes in these measures across conditions and context.

Incidental Duration Changes

While speech rate was not deliberately targeted in this study, the increased duration in the CLEAR condition and decreasing duration from WORD to PHR to PAS indicate incidental rate variation (see Figure 4). This pattern of duration changes necessarily leads to corresponding changes in displacement and velocity. It is expected that both displacement and velocity would decrease with shorter segment durations.

Changes in the displacement across conditions and contexts would influence the displacement correlation measures. Greater range in the data increases the likelihood of finding a potential correlation. We found significant increases in the correlation between TM and J and between TM and LL from the HABIT to the CLEAR conditions, and decreased correlations between TM and J and between TM and LL in the PAS context compared with the other two contexts. The CLEAR condition had longer segment durations and larger displacements compared to HABIT, and the PAS context had shorter durations and smaller displacements than the other two contexts. It then follows that stronger correlations could be found for CLEAR compared to HABIT and weaker correlations would be found for PAS compared to the other contexts as a consequence of increased range as opposed to stronger coordination. It is possible that other factors may have influenced these correlations, but the influence of duration and displacement on the correlations makes it difficult to unambiguously interpret these results with respect to coordination.

Bailey and Dromey (2015) used UL-LL correlation to compare inter-articulator coordination during speech in isolation compared to speech during other tasks. They found an increase in lip coordination for the speech during concurrent tasks compared to the speech in isolation. This suggests that there may have been a change in correlation across speaking contexts similar to the change seen Bailey and Dromey (2015). However, determining a potential relationship between articulator correlations and speech contexts and conditions would require further analysis, such as controlling for token duration and displacement.

Similar to the correlations, absolute J contribution to LL and TM displacements is highly influenced by duration and displacement. For example, in the diphthong /aɪ/, the jaw helps create the low to high vowel movement by carrying the tongue upward, but the tongue also moves upward independently of the jaw. J contribution to TM considers how much of the tongue's upward movement is due to the upward J movement and how much the tongue moved independently of it. Absolute J contribution is simply the amount of displacement provided by the jaw. As overall articulator movements increased in the CLEAR condition and decreased from WORD to PHR to PAS, absolute J contribution increased in the CLEAR condition and decreased across WORD, PHR, and PAS. This is likely due to the corresponding changes in displacement since smaller overall displacement would result in smaller contribution to displacement. To control for changes in overall displacement, percent J contribution was calculated, and is discussed in the following section.

Motor Equivalence

Percent contribution accounts for overall displacement changes across conditions and contexts by considering the ratio of each articulator's displacement to determine relative contribution to the overall upward movement for the diphthong /aɪ/. We found that percent J contribution to LL increased from HABIT to CLEAR, and percent J contribution to TM was higher in the PAS context compared to WORD and PHR. Mefferd (2017) also examined relative contribution of the jaw to tongue movements and reported lower jaw contribution to decoupled tongue movements in a clear speech condition, which was not reflected in our results. The reason for the discrepant findings is unclear, but could be related to differences in the speech stimuli.

Hughes and Abbs (1976) also studied relative contributions of articulators and described changes in relative contributions as a reflection of motor equivalence both within and across speakers. Motor equivalence means that there are multiple combinations of articulatory movements that can result in the same acoustic and perceptual output. For example, in our study, it is generally expected that both the jaw and tongue move upward as the vowel transitions from low to high, but some speakers may rely on mostly on jaw movement to raise the tongue while other speakers may raise their tongue in addition to the jaw. Percent J contribution indirectly references motor equivalence because it will reflect this change. A possible explanation for why the change in relative jaw contribution found in Mefferd (2017) was not reflected in our results is motor equivalence across speakers. For example, it is possible that for some speakers J was rising as expected during /aɪ/, but TM was simultaneously falling relative to J, resulting in an overall upward TM movement due to the larger J contribution even if TM was moving in the opposite direction to J. As seen in Table 1, many of the kinematic measures used in this study had large standard deviations, which indicates substantial variation between speakers, indicating the influence of motor equivalence. As a result, interpretation of J contribution is limited without a detailed analysis of motor patterns in each token.

Limitations and Implications for Future Research

In this study, the brief diphthong segment limited which measures could be applied. While significant changes were found across both conditions and contexts, future research could use longer, more phonetically complex stimuli, such as multisyllable words, to allow for measures such as relative timing of the articulators. In the current study, a sophisticated analysis of coordination was not possible.

As seen in Table 1, large standard deviations were present for many of the measures, likely due to the influence of motor equivalence. A full token-by-token analysis of motor equivalence was beyond the scope of this study, but future research may include this level of analysis to examine motor equivalence effects across conditions and contexts. For example, it is possible that individuals may display consistent motor patterns in the WORD and PHR contexts but demonstrate a different articulation pattern in the PAS context. Further research is needed to fully understand the effects of speaking tasks on articulatory kinematics.

Conclusion

Significant changes in speech kinematics were found between habitual and clear speech conditions as well as across word, phrase, and passage speaking contexts. Changes in speaking rate likely contributed to many of these changes. Nevertheless, it is important for researchers to bear in mind these changes when selecting stimuli. Highly contrived, short phrases often used in research studies may not fully generalize to everyday connected speech. Future studies should continue to explore the extent to which lab speech can be generalized to more naturalistic contexts.

References

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research, 36*(1), 41–54. https://doi.org/10.1044/jshr.3601.41

Bailey, D. J., & Dromey, C. (2015). Bidirectional interference between speech and nonspeech tasks in younger, middle-aged, and older adults. *Journal of Speech, Language, and Hearing Research, 58*(6), 1637–1653. https://doi.org/10.1044/2015_JSLHR-S-14-0083

Dromey, C. (2000). Articulatory kinematics in patients with Parkinson disease using different speech treatment approaches. *Journal of Medical Speech-Language Pathology, 8*(3), 155–161.

Dromey, C., Jang, G.-O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication, 55*(2), 315–328. https://doi.org/10.1016/j.specom.2012.09.001

Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica, 33*(3), 199–221. https://doi.org/10.1159/000259722

Kuruvilla-Dugdale, M., & Chuquilin-Arista, M. (2017). An investigation of clear speech effects on articulatory kinematics in talkers with ALS. *Clinical Linguistics & Phonetics, 31*(10), 725–742. https://doi.org/10.1080/02699206.2017.1318173

Maner, K. J., Smith, A., & Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research, 43*(2), 560–573. https://doi.org/10.1044/jslhr.4302.560

Mathworks. (2023). *Matlab* (Version 2023a) [Computer software]. https://www.mathworks.com/

Mefferd, A. S. (2017). Tongue- and jaw-specific contributions to acoustic vowel contrast changes in the diphthong /ai/ in response to slow, loud, and clear speech. *Journal of Speech, Language, and Hearing Research, 60*(11), 3144–3158. https://doi.org/10.1044/2017_jslhr-s-17-0114

Northern Digital Inc. (2017). *Wavefront* (Version 2.2.1) [Computer software]. https://www.ndigital.com/

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*(1–2), 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7

Tasko, S. M., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research, 47*(1), 85–100. https://doi.org/10.1044/1092-4388(2004/008)

Tasko, S. M., & Westbury, J. R. (2002). Defining and measuring speech movement events. *Journal of Speech, Language, and Hearing Research, 45*(1), 127–142. https://doi.org/10.1044/1092-4388(2002/010)

Weismer, G., Yanusova, Y., & Westbury, J. R. (2003). Interarticulator coordination in dysarthria: An x-ray microbeam study. *Journal of Speech, Language, and Hearing Research, 46*(5), 1247–1261. https://doi.org/10.1044/1092-4388(2003/097)

Westbury, J. R., Lindstrom, M. J., & McClean, M. D. (2002). Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research*, *45*(4), 651–662. https://doi.org/10.1044/1092-4388(2002/052)

Whitfield, J. A., Holdosh, S. R., Kriegel, Z., Sullivan, L. E., & Fullenkamp, A. M. (2021). Tracking the costs of clear and loud speech: Interactions between speech motor control and concurrent visuomotor tracking. *Journal of Speech, Language, and Hearing Research, 64*(6S), 2182–2195. https://doi.org/10.1044/2020_JSLHR-20-00264

Wisler, A., Goffman, L., Zhang, L., & Wang, J. (2022). Influences of methodological decisions on assessing the spatiotemporal stability of speech movement sequences. *Journal of Speech, Language, and Hearing Research, 65*(2), 538–554. https://doi.org/10.1044/2021_JSLHR-21-00298

Tables

Table 1

Descriptive Statistics for Conditions and Contexts

Variable	HABIT						CLEAR					
	WORD		PHR		PAS		WORD		PHR		PAS	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Seg duration	128	26	117	27	107	31	152	55	137	41	125	33
TM disp	6.96	2.47	6.02	2.40	4.18	2.06	9.83	3.09	8.77	3.02	6.51	3.15
J disp	2.76	1.40	2.72	1.77	2.45	1.87	5.04	2.68	4.02	2.55	3.46	2.03
LL disp	4.75	1.46	4.59	1.88	4.08	1.98	6.98	3.52	6.25	2.96	5.30	2.33
UL disp	3.79	2.99	3.64	1.81	4.32	3.68	6.04	4.70	5.07	3.90	4.61	2.61
TM vel	93.1	29.0	88.5	32.1	68.7	32.8	127.6	32.2	116.2	37.2	94.3	37.4
J vel	43.2	17.8	45.8	27.3	41.9	25.8	57.6	30.0	56.6	31.3	55.2	28.9
LL vel	106.5	22.6	125.5	43.4	120.3	44.7	136.5	90.3	139.8	73.3	126.7	43.7
UL vel	5.3	29.3	–2.9	20.7	2.1	51.8	9.5	51.0	4.1	48.9	–0.3	39.0
J cont TM	2.14	1.08	2.10	1.37	1.89	1.44	3.89	2.07	3.11	1.97	2.68	1.57
J cont LL	3.01	1.53	2.97	1.94	2.67	2.04	5.49	2.93	4.39	2.78	3.78	2.22
J pct cont LL	35.4	26.0	41.7	40.0	64.8	82.8	40.8	20.2	37.4	25.5	50.5	48.9
J pct cont TM	63.1	25.8	63.3	27.1	64.9	27.8	80.2	28.9	70.3	28.6	70.7	25.0
TM LL corr	.69	.30	.58	.37	.42	.52	.71	.34	.68	.33	.63	.43
TM J corr	.44	.61	.35	.56	.17	.65	.61	.61	.53	.56	.43	.62
LL J corr	.80	.39	.86	.20	.82	.25	.80	.38	.84	.33	.87	.21

Note. PHR = phrase, PAS = passage, TM = mid-tongue, J = jaw, LL = lower lip, UL = upper lip, seg duration = segment duration, disp = displacement, vel = velocity, cont = contribution in mm, pct cont = percent contribution, corr = correlation. All measures are in the vertical direction.

Table 2

Linear Mixed Model Results for Habitual and Clear Conditions

Variable	Main Effects		
	df	*F*	*p*
Seg duration	1, 565.143	58.47	< .001
TM disp	1, 565.014	277.59	< .001
J disp	1, 565.072	105.30	< .001
LL disp	1, 565.052	99.57	< .001
UL disp	1, 565.120	27.46	< .001
TM vel	1, 565.042	203.73	< .001
J vel	1, 565.001	63.31	< .001
LL vel	1, 565.202	11.69	< .001
UL vel	1, 565.058	1.74	.188
J cont TM	1, 565.072	105.30	< .001
J cont LL	1, 565.072	105.30	< .001
J pct cont LL	1, 565.092	22.82	< .001
J pct cont TM	1, 565.012	3.69	.055
TM LL corr	1, 565.043	20.75	< .001
TM J corr	1, 565.023	30.86	< .001
LL J corr	1, 564.917	0.14	.711

Note. TM = mid-tongue, J = jaw, LL = lower lip, UL = upper lip, seg duration = segment duration, disp = displacement, vel = velocity, cont = contribution in mm, pct cont = percent contribution, corr = correlation. $p \leq .05$ is considered statistically significant.

Table 3

Linear Mixed Model Results for Word, Phrase, and Passage Contexts

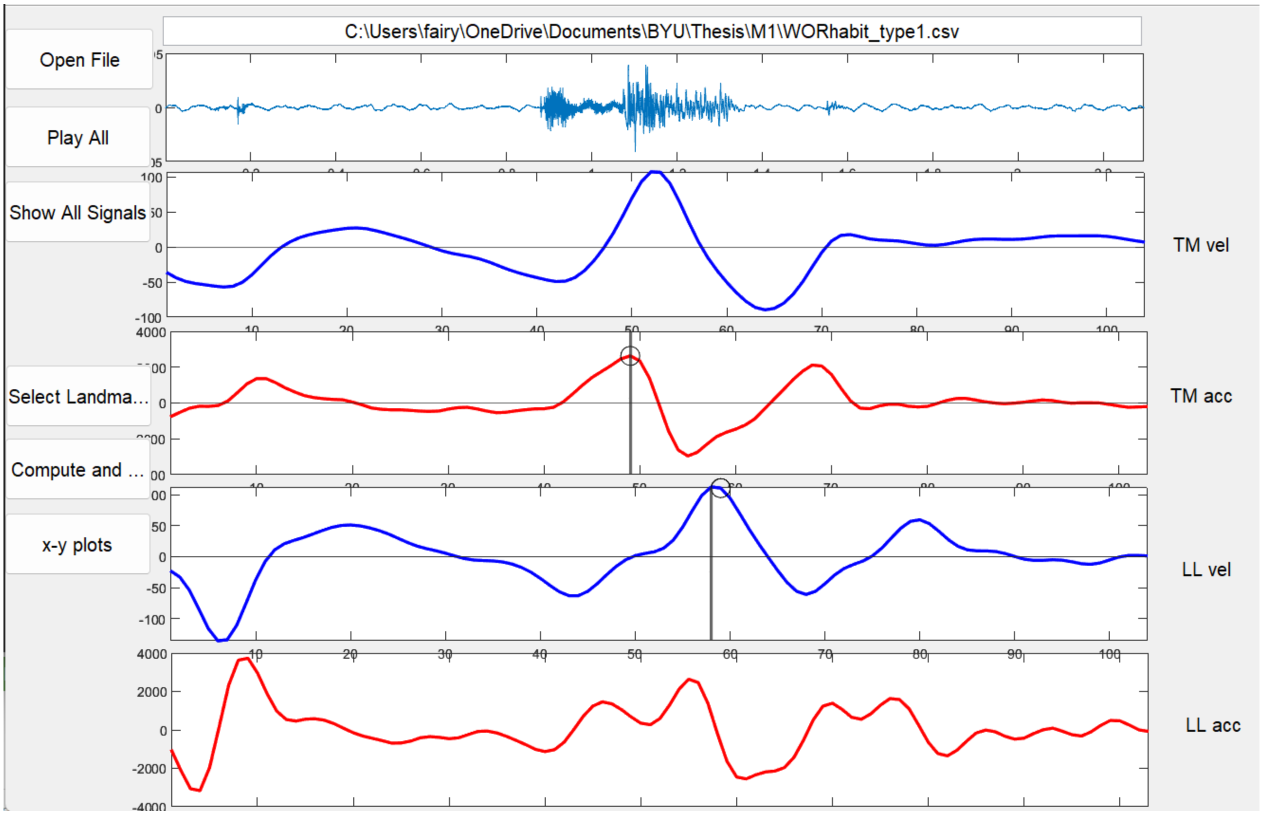
Variable	Main Effects			WORD PHR	PHR PAS	WORD PAS
	df	*F*	*p*	*p*	*p*	*p*
Seg duration	2, 565.403	23.95	< .001	< .001	< .001	< .001
TM disp	2, 565.076	130.50	< .001	< .001	< .001	< .001
J disp	2, 565.182	13.34	< .001	.013	.017	< .001
LL disp	2, 565.168	17.68	< .001	.128	< .001	< .001
UL disp	2, 565.246	1.84	.160	.169	1.000	.522
TM vel	2, 565.109	77.16	< .001	.010	< .001	< .001
J vel	2, 565.058	2.52	.081	1.000	.091	.457
LL vel	2, 565.484	2.90	.056	.162	.111	1.000
UL vel	2, 565.125	2.55	.079	.124	1.000	.118
J cont TM	2, 565.182	13.34	< .001	.013	.017	< .001
J cont LL	2, 565.182	13.34	< .001	.013	.017	< .001
J pct cont LL	2, 565.208	1.99	.138	.141	1.000	.511
J pct cont TM	2, 565.216	15.59	< .001	1.000	< .001	< .001
TM LL corr	2, 565.166	14.64	< .001	.146	< .001	< .001
TM J corr	2, 565.114	12.62	< .001	.346	< .001	< .001
LL J corr	2, 565.029	1.69	.186	.202	1.000	.806

Note. PHR = phrase, PAS = passage, TM = mid-tongue, J = jaw, LL = lower lip, UL = upper lip, seg duration = segment duration, disp = displacement, vel = velocity, cont = contribution in mm, pct cont = percent contribution, corr = correlation. $p \leq .05$ is considered statistically significant.

Figures

Figure 1

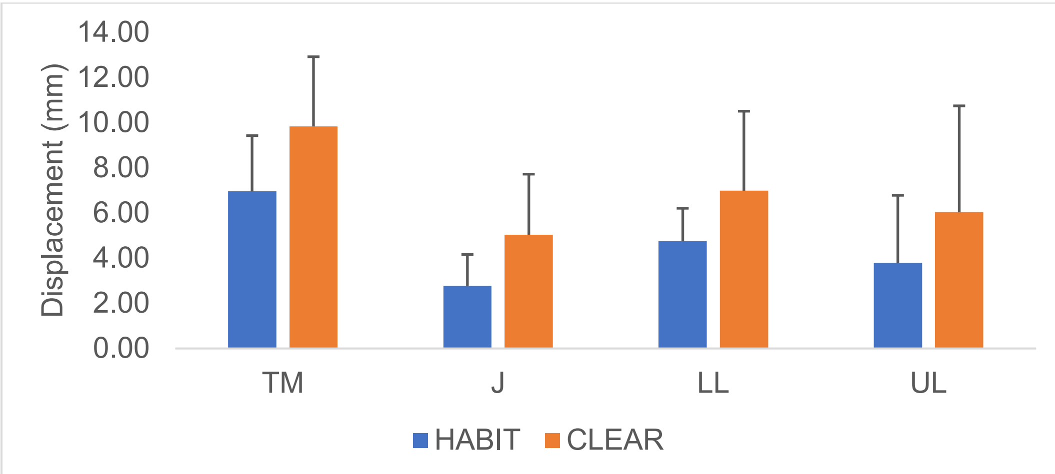
Example of Segmentation Points for the Target Diphthong



Note. TM vel = mid-tongue velocity, TM acc = mid-tongue acceleration, LL vel = lower lip velocity, and LL acc = lower lip acceleration.

Figure 2

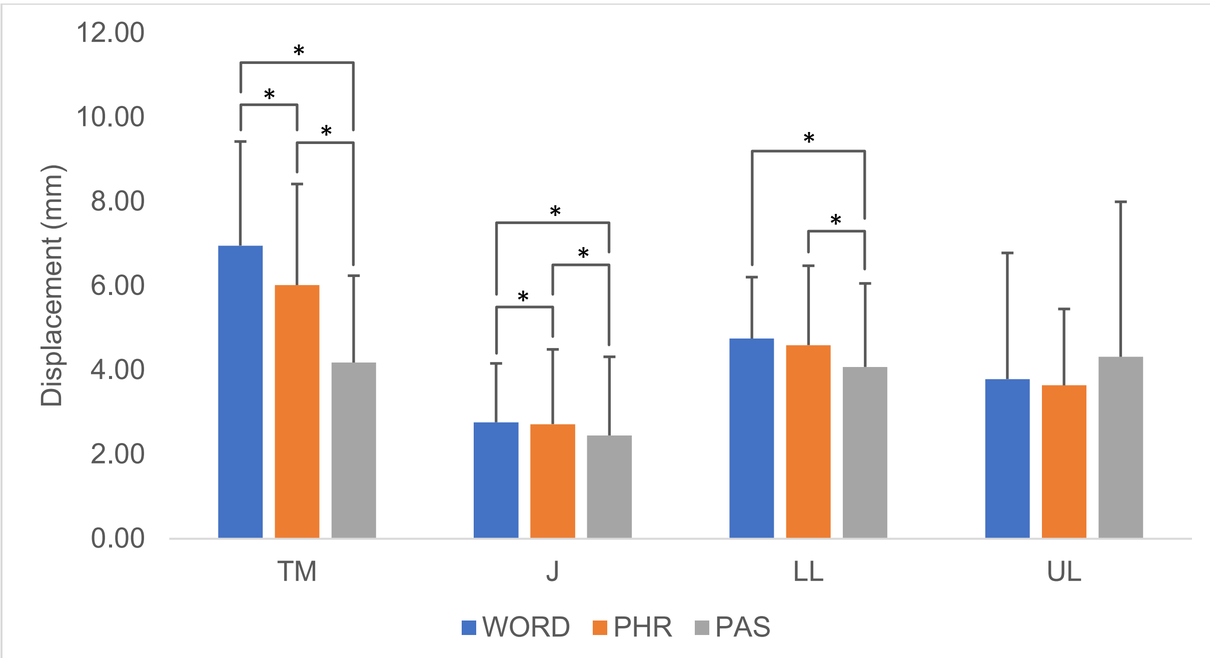
Changes in Mean Displacement Across Habitual and Clear Conditions in the Word Context



Note. TM = mid-tongue, J = jaw, LL = lower lip, UL = upper lip. All changes in displacement between the habitual and clear conditions were significant.

Figure 3

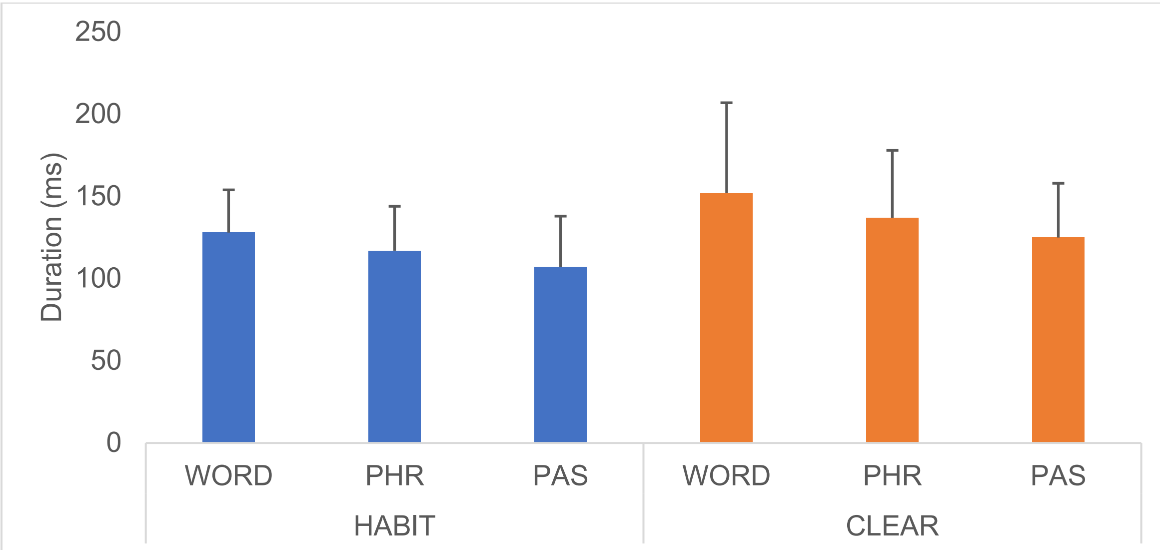
Changes in Mean Displacement Across Word, Phrase, and Passage Contexts in the Habitual Condition



Note. PHR = phrase, PAS = passage, TM = mid-tongue, J = jaw, LL = lower lip, UL = upper lip, * = significant difference between groups. Changes in TM and J displacements across contexts were significant. The change in LL displacement was significant only in the PAS context. Changes in UL displacement were not significant.

Figure 4

Changes in Segment Duration Across Conditions and Contexts



Note. PHR = phrase, PAS = passage. All changes in duration across conditions and contexts were significant.

APPENDIX A

Annotated Bibliography

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research*, *36*(1), 41–54. https://doi.org/10.1044/jshr.3601.41

Objective: This study investigated the effects of speaking rate on the velocity movement profiles of the lower lip and tongue. *Method:* One woman and four men between the ages of 19–35 years spoke the phrase "tap a tad above" 10 times each at five different speaking rates (habitual, four times as fast, two times as fast, half as fast, and a quarter as fast). An x-ray microbeam system was used to track two dimensional (inferior-superior and anterior-posterior) movements of pellets placed on the tongue tip and lower lip of each participant. The recordings were segmented into the closing and opening gestures in "tap a tad" and principal component analysis was used to determine a single dimension displaying the most data. It was determined that the opening gesture for the tongue tip resulted in the most consistent and extensive duration changes across subjects and rates, so all further analyses were done using this gesture. *Results:* As rate decreased, a variable calculated from max displacement, peak velocity, and total movement time increased. Additionally, two measures of asymmetry and the number of velocity peaks increased, and the time to peak velocity decreased. The measure of kurtosis was inconsistent, likely due to the asymmetry and multimodality of the data. *Conclusion:* The velocity profiles at fast rates of speech appear to be more consistent and more similar across speakers while the velocity profiles at slower rates appear to include more sub-movements and

variability. *Relevance to the current work:* Velocity profiles reflect kinematic changes across task conditions.

Bailey, D. J., & Dromey, C. (2015). Bidirectional interference between speech and nonspeech tasks in younger, middle-aged, and older adults. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1637–1653. https://doi.org/10.1044/2015_JSLHR-S-14-0083 *Objective:* This study examined how age and divided attention influence speech kinematics. *Method:* 60 adults from three age groups (younger, middle-aged, and older) participated, with 10 male and 10 female participants in each group. Participants said "I saw Patrick pull a wagon packed with apples" each time they heard a tone. This was done in isolation and while completing nonspeech tasks: a linguistic task (deciding if two words are semantically related), a cognitive task (deciding if two fractions are equal), and a manual motor task (placing pegs). The researchers used 10 perceptually correct tokens from each condition and measured displacement, lower lip spatiotemporal index (LL STI), duration, sound pressure level (SPL), and correlation between the upper and lower lips (UL-LL), as well as number of correct responses to each of the tasks. *Results:* Task type influenced speech kinematics. Duration, negative UL-LL correlation, and LL STI were greater during the linguistic and cognitive tasks compared to speech in isolation. LL displacement was lower, and SPL was higher in the manual motor task compared to the speech only condition. *Conclusion:* Task difficulty may have contributed to results, and it seems that similar tasks may interfere more than dissimilar tasks. Decreased stability during the linguistic and cognitive tasks may indicate reduced automaticity of speech. *Relevance to the current work:* Automatic or repeated speech may show kinematic differences compared to generative speech due to reduced attentional demands.

Dromey, C. (2000). Articulatory kinematics in patients with Parkinson disease using different speech treatment approaches. *Journal of Medical Speech-Language Pathology*, *8*(3), 155–161.

Objective: This study compared lip kinematics of participants with Parkinson disease (PD) in typical, loud, and hyperarticulate speech. *Method:* One female and nine male adults with PD (49–77 years old, mean age 67.5) wore a head-mounted strain-gauge system to record lip movements while they said "Buy Bobby a puppy" 20 times each in normal, loud, and hyperarticulate conditions. Measures included lower lip displacement, peak velocity, spatiotemporal index (STI), duration, and sound pressure level. *Results:* Displacement and velocity in the experimental conditions were larger than in normal speech but were not significantly different across the two conditions. Hyperarticulate speech resulted in higher STI values and an extra velocity peak in some speakers, indicating greater variability in this condition. *Conclusion:* Both loud and hyperarticulate speech seem to overcome the small movements and quietness of hypokinetic dysarthria. The greater variability seen in the hyperarticulate condition may be due to increased attentional demands and lack of familiarity with the task. *Relevance to the current work:* Kinematic differences between hyperarticulate and normal speech are seen at the sentence level.

Dromey, C., Jang, G.-O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication*, *55*(2), 315–328. https://doi.org/10.1016/j.specom.2012.09.001

Objective: This study compared changes in the first and second formants (F1 and F2) with predicted tongue movements. *Method:* 20 participants read the sentences "The boot

on top is packed to keep" and "The boy gave a shout at the sight of the cake" five times each in habitual speaking conditions. A jaw tracking instrument was used to track the movements of a small magnet attached near the tip of the tongue. Productions were segmented to isolate the diphthongs /ɔɪ/, /au/, /aɪ/, and /eɪ/ and analyzed to compare kinematic displacement data to formant frequency changes. *Results:* There was a negative correlation between F1 and vertical tongue movement for /aɪ/. There was a positive correlation between F2 and anteroposterior tongue movement, but the strength of this relationship varied. *Conclusion:* It is possible that the linearity of tongue movement during /aɪ/ contributed to the strong relationship between F1 and vertical tongue movement, or it is possible that lack of lip-rounding in this diphthong led to tongue position being the strongest contributor to F1. The reason for the F2 and vertical tongue movement correlation was less clear but may have been due to coarticulatory effects or speaking style. *Relevance to the current work:* Vertical tongue movement is a consistent and strong contributor to acoustic quality in the diphthong /aɪ/ and could likely be significant across conditions and contexts.

Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica*, *33*(3), 199–221. https://doi.org/10.1159/000259722

Objective: This study investigated the presence of motor equivalence in labial-mandibular coordination across two different speaking rates. *Method:* Six female participants repeated the syllables /hæbæb, hibib, hɛbɛb/ in a carrier phrase 10 times each at habitual and fast speaking rates. Kinematic data for the upper lip, lower lip, and jaw were recorded using a strain gauge system. *Results:* In productions where the vertical

distance between the upper and lower lip stayed relatively constant, the relative contributions of the lower lip and jaw in each production varied. There was not a significant change in absolute or relative magnitude of displacement for any articulators in the increased speaking rate condition. *Conclusion:* Labial-mandibular coordination is demonstrated by the adjustment of articulatory gestures to compensate for reduced movement of another articulator, which is referred to as motor equivalence. This coordination extends to altered speaking conditions such as maintaining absolute and relative displacement of articulators at a fast speaking rate. *Relevance to the current work:* The same speech production can be achieved through subtle kinematic changes.

Kuruvilla-Dugdale, M., & Chuquilin-Arista, M. (2017). An investigation of clear speech effects on articulatory kinematics in talkers with ALS. *Clinical Linguistics & Phonetics*, *31*(10), 725–742. https://doi.org/10.1080/02699206.2017.1318173

Objective: This study investigated how using clear speech affected measures of articulatory kinematics in speakers with amyotrophic lateral sclerosis (ALS). *Method:* The sample of speakers with ALS included five males and 2 females with an average age of 62.49 years old. This group was compared to a healthy control group of nine males and five females with an average age of 60.51 years old. Participants said "Say that I owe you a yoyo today" 10 times habitually and using clear speech. Data collected using electromagnetic articulography was used to calculate spatiotemporal index (STI), maximum speed, range of movement, and duration. *Results:* Significant differences in kinematic measures were observed between groups in both the habitual and clear conditions. Talkers with ALS had lower jaw movement variability habitually but greater variability in clear speech compared to controls. *Conclusion:* The kinematic changes between habitual and clear speech for speakers with ALS were limited compared to controls. Clear speech resulted in higher movement speed and greater range of movement compared to habitual, but this difference was observed primarily in the control group. In clear speech, speakers with ALS had lower max speed, reduced range of movement, and longer utterance durations compared to controls. *Relevance to the current work:* Clear speech is kinematically different from habitual speech but may require more sensitive measures when assessing individuals with communication disorders.

Maner, K. J., Smith, A., & Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research*, *43*(2), 560–573. https://doi.org/10.1044/jslhr.4302.560

Objective: This study examined how increased processing demands due to increased utterance length and complexity affected articulatory movement stability in speech motor performance. *Method:* Eight young adults and eight five-year-old children said the phrase "Buy Bobby a puppy" and four other sentences of varying syntactic complexity with this same phrase embedded in them. An Optotrak system was used to track lower lip movement, and these signals were used to calculate the spatiotemporal index (STI), which measures stability across repetitions. *Results:* The children had higher STIs (less stability) than the adults, and the children's STIs increased with longer, more complex utterances. The adult STIs either did not change or increased in the longer, more complex utterances. *Conclusion:* Articulatory motor stability seems to be affected by increased processing demands. *Relevance to current work:* If processing demands affect motor performance, it is possible different speech tasks will result in different motor behaviors.

Mefferd, A. S. (2017). Tongue- and jaw-specific contributions to acoustic vowel contrast changes in the diphthong /ai/ in response to slow, loud, and clear speech. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3144–3158. https://doi.org/10.1044/2017_jslhr-s-17-0114

Objective: This study investigated the relationship between slow, loud, and clear speech and decoupled movements of the tongue and jaw. *Method:* Kinematic data from 20 typical young adult speakers were collected using an electromagnetic articulograph. The speakers said “see a kite again” five times each in habitual, fast, slow, loud, and clear conditions. Measures included tongue and jaw displacement, and acoustic vowel contrast. *Results:* Jaw displacements were significantly larger during clear, loud, and slow speech than the habitual condition, and they were largest during clear speech. Decoupled tongue displacements were also significantly larger during the test conditions compared to the habitual condition. Slow speech had larger tongue displacements than clear speech which were in turn larger than loud speech. Composite posterior tongue movement was larger in test conditions compared to typical, and slow and clear speech resulted in larger composite movements than loud speech. Speech modification also resulted in larger acoustic vowel contrast, with slow speech resulting in the largest contrast, then clear speech, then loud speech. *Conclusion:* Speakers had significantly increased decoupled tongue displacement in slow speech which accounted for a change in acoustic vowel contrast. Increased decoupled tongue and jaw displacements during clear speech result in increased vowel distinctiveness. *Relevance to the current work:* We expect to see kinematic differences between habitual and clear speech at the syllable level.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1–2), 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7

Relevance to the current work: The author provides examples to support the notion that there is a nonlinear relationship between articulator positioning and acoustic output. While some movements drastically change formants and other acoustic features of speech, other articulatory changes will have little effect on acoustics. For example, the backed position of the tongue and rounding of the lips both contribute to a low second formant (F2), but since both gestures contribute to a low F2, the magnitude of each gesture may vary across productions while still resulting in the same acoustic output. This principle of motor equivalence indicates that speakers may alter their speech movements to achieve the same acoustic output under varying conditions.

Tasko, S. M., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research*, *47*(1), 85–100. https://doi.org/10.1044/1092-4388(2004/008)

Objective: This study looked at typical kinematic variation across speaking tasks and conditions. *Method:* Fifteen young adult males spoke a nonsense phrase, sentence, a short version of the Hunter script, and a spontaneous monologue about their jobs or hobbies. The nonsense phrase and sentence were spoken in habitual, soft, loud, fast, and slow conditions. An articulograph tracked sensors on the upper and lower lips (UL and LL), mandibular incisors, and tongue tip. These kinematic signals were divided into movement strokes. Measures of these strokes included stroke duration, peak speed, stroke distance. Spatial variation was determined using the standard distance (SDIS), which looks at deviation from the mean without regard to shape, and variance (VAR), which considers the spatial distribution of the points. *Results*: At habitual rate and loudness, the nonsense phrase was different across all measures while the other tasks were similar to each other. Tongue tip and LL exhibited more unidimensional spatial variation for the nonsense task compared to the other tasks. Tongue tip SDIS was larger for the sentence than for other tasks. Changes in loudness resulted in changes in the distance and speed of strokes and in the size of the spatial variation. Duration changes due to changing rate were more pronounced for the nonsense phrase than for the sentence. *Conclusion:* Linguistic tasks of different types seem to be similar kinematically, and shorter utterances may be used to represent connected speech. *Relevance to current work:* Movement strokes may not be sensitive to task differences.

Tasko, S. M., & Westbury, J. R. (2002). Defining and measuring speech movement events. *Journal of Speech, Language, and Hearing Research*, *45*(1), 127–142. https://doi.org/10.1044/1092-4388(2002/010)

Objective: This study described a method of segmenting speech movement signals into distinct units referred to as “strokes.” The authors define this measure as the period between two minima in the speed waveform for an articulatory sensor. *Method:* Eighteen healthy young adults read a slightly expanded version of the Hunter script, which was chosen to simulate the phonetic distribution of conversational speech. Each reading was recorded in four separate recordings and used to calculate each speaker’s oral reading rate. An x-ray microbeam system was used to track pellets placed on the tongue blade and dorsum (T1 and T4), mandibular incisors (MI), molars, and lower lip (LL). Kinematic data included stroke distance, duration, peak speed, and minimum speed. *Results:* Movement strokes do not map directly onto acoustic speech units, and generally there were different numbers of strokes for different articulators within an utterance. There was a significant number of small strokes that appeared in a variety of contexts. MI had the longest stroke durations, T1 and T4 had notably higher stroke duration, T1 had the highest peak speeds, and T1 and T4 had the highest minimum speeds. *Conclusion:* Strokes are a valuable measurement unit for kinematic analysis and can be used to measure movement regardless of speech units. *Relevance to the current work:* General speech kinematic features can be compared across speakers and conditions.

Weismer, G., Yanusova, Y., & Westbury, J. R. (2003). Interarticulator coordination in dysarthria: An x-ray microbeam study. *Journal of Speech, Language, and Hearing Research*, *46*(5), 1247–1261. https://doi.org/10.1044/1092-4388(2003/097)

Objective: This study compared the coordination of labial and lingual movements when speakers with Parkinson's (PD), speakers with amyotrophic lateral sclerosis (ALS), and control speakers produced the vowel /u/. *Method:* Controls included 14 men and 7 women, speakers with PD included 13 men and 5 women, and speakers with ALS included 5 men and 4 women. Speakers read the sentence "She had your dark suit in greasy wash water all year" twice at a habitual rate while kinematic data were collected using an x-ray microbeam with markers on the tongue dorsum and upper and lower lips. Additionally, acoustic recordings were collected through a microphone. Measures included movement of the second formant (F2), vowel duration, and positional maximum or minimum from kinematic time histories. *Results:* Global movement patterns for the movement from /s/ to /u/ were similar across groups, but detailed analysis revealed three things: (a) in PD samples, the timing of lip protrusion relative to the vowel nucleus was different compared to control or ALS samples; (b) in ALS samples, there was a time delay between the kinematic lip and tongue extrema not seen to the same extent in other groups; (c) the pattern of kinematic and acoustic extrema occurring approximately two thirds into the vowel was not consistent in ALS samples. *Conclusion:* These results indicate there may be articulatory discoordination for speakers with ALS. This analysis only examined timing measures rather than magnitudes, so it is possible further changes may be noted by considering spatial coordination. *Relevance to the current work:* Coordination measures show subtle changes across groups, but timing or spatial coordination alone may miss some changes.

Westbury, J. R., Lindstrom, M. J., & McClean, M. D. (2002). Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research*, *45*(4), 651–662. https://doi.org/10.1044/1092-4388(2002/052) *Objective:* This study investigated the effectiveness of a method of measuring tongue and lip movements decoupled from the jaw. *Method:* 44 young adults read the sentence “She had your dark suit in greasy wash water all year.” An x-ray Microbeam recorded movements of the central incisors, molars, lower lip, tongue blade, and tongue dorsum. The researchers used three different methods to calculate tongue and lip motion relative to the jaw: only-translation, only-rotation, and translation-rotation with estimated jaw rotation. These were compared to calculations using the translation-rotation model with measured jaw rotation. *Results:* Positional and speed errors were smallest when using the estimated rotation method. *Conclusion:* Calculations that do not account for rotation lead to greater errors than those that do not account for translation. Estimating rotation improves accuracy compared to methods that ignore rotation. *Relevance to the current* *work:* It is important to account for the rotational movement of the jaw when decoupling kinematic data from the lips, tongue, and jaw.

Whitfield, J. A., Holdosh, S. R., Kriegel, Z., Sullivan, L. E., & Fullenkamp, A. M. (2021). Tracking the costs of clear and loud speech: Interactions between speech motor control and concurrent visuomotor tracking. *Journal of Speech, Language, and Hearing Research*, *64*(6S), 2182–2195. https://doi.org/10.1044/2020_JSLHR-20-00264

Objective: This study identified kinematic and acoustic changes in clear and loud speech conditions in a dual-task context. *Method:* A sample of healthy young adult speakers—two males and 23 females—said the sentence “Buy Bobby a puppy,” first as a speech-only task then while doing a visuomotor task (tracking a target). In each context, participants repeated the sentence multiple times habitually, using loud speech, and using clear speech. A 3D passive-optical motion analysis system was used to track markers on the upper and lower lips (UL and LL) and mandible. Measures included lip aperture range of motion (LA range), speech rate, interval length, and speech intensity. *Results:* Compared to the habitual speech condition, speech intensity, LA range, peak LL velocities, and syllable duration increased in the loud condition. All these changes also occurred in the clear speech condition, along with increased intersyllable durations and reduced speaking rate. Compared to the speech-only condition, in habitual speech the visuomotor task reduced intensity and LA range, but accuracy stayed the same, while in the loud and clear styles the visuomotor task did not affect intensity or LA, but tracking accuracy decreased. *Conclusion:* Phonatory adjustments were larger for loud speech, and articulatory adjustments were greater for clear speech compared to habitual speech. It is likely that changes were not seen between tasks in the loud and clear conditions due to participants prioritizing a speaking style, which seems to require more attentional resources. *Relevance to the current work:* The higher attentional resources used for clear speech may reveal kinematic changes not seen in a habitual speaking style.

Wisler, A., Goffman, L., Zhang, L., & Wang, J. (2022). Influences of methodological decisions on assessing the spatiotemporal stability of speech movement sequences. *Journal of Speech, Language, and Hearing Research*, *65*(2), 538–554. https://doi.org/10.1044/2021_JSLHR-21-00298

Objective: This study examined the effects of the number of repetitions, stimulus length, and parsing errors on spatiotemporal index (STI) calculation. *Method:* Synthetic speech signals were created for each experiment. For comparing the number of repetitions, “true” STI was calculated using 10,000 repetitions (an arbitrarily high number to approximate actual STI). Calculations using two to 20 repetitions were compared to the true STI. Stimuli were generated from 0.2 seconds (“Buy”) to 2.0 seconds (“Buy Bobby a Puppy Buy Bobby a Puppy”) with three different levels of temporal instability. The timespans of the signals were artificially varied. The authors considered three different levels of stability, and they incrementally varied the onset and offset placements. *Results:* More repetitions led to better estimates and reduced variance, but the benefit becomes marginal as the number of repetitions increases. The conventional 10–20 repetitions seem acceptable for estimating STI. With no temporal instability, utterance length had minimal effect on STI, but if there is temporal instability, STI is more sensitive to temporal instability in longer versus shorter utterances. Larger parsing errors led to larger STI, with the strongest effect being seen in the more stable conditions. *Conclusion:* For best results when calculating STI, use the same number of repetitions across analyses,

maintain consistent length of stimuli, and choose an accurate and consistent method of parsing the signal. *Relevance to the current work:* This study design uses repetitive, contrived speech stimuli frequently found in research.

APPENDIX B

Institutional Review Board Approved Consent

Consent to be a Research Subject

Title of the Research Study: Connecting lab speech with everyday communication
Principal Investigator: Christopher Dromey, PhD
IRB ID#:

Introduction

This research study is being conducted by Professor Christopher Dromey, assisted by Lauren Clarke and Jessica Martin, all from the Department of Communication Disorders at Brigham Young University, to determine how speaking shorter or longer words and phrases affects how the tongue, lips, and jaw move. You were invited to participate because you are a native speaker of American 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 sit in a sound-treated booth in room 106 of the John Taylor Building where your speech will be recorded
- new, disposable small sensors (3 x 3 mm) will be attached with dental glue to your tongue, lips, and lower front teeth
- for 10 minutes you will either read aloud or chat with the researcher as you get used to the feeling of the sensors in your mouth
- you will read aloud words, phrases, and sentences that will be presented on a computer screen in front of you
- you will be asked to speak as you typically would and also with exaggerated clarity
- total time commitment will be 60 minutes in one recording session

Risks/Discomforts

You might feel uncomfortable having electromagnetic sensors attached to your tongue, lips, and lower teeth. It is possible that some of the dental glue will remain on the tongue surface for a few minutes after the experiment is over. This may feel odd, but it will feel normal again within a few minutes.

The researcher will view the surface of your tongue after removing the sensors to make sure that any traces of the glue are minimal. The single-use sensors will be thrown away after removal.

You might feel some fatigue; if so, you may take a break at any time during the study.

Benefits

There will be no direct benefits to you as a participant. However, we anticipate that the findings from this study will benefit the field of speech pathology by helping us design better treatments for people with speech problems.

Confidentiality

The research data will be kept on password protected computer and only the researchers will have access to the data. Before we analyze the recordings, all identifying information will be removed so that your name will not be linked to the recordings. Only summary data from groups of participants will be reported in publications and presentations. After the study, the de-identified data will be kept on a password-protected computer in the researcher's office for possible future analysis with new techniques.

Compensation

You will receive $15 for your participation, whether you finish the recording or not; compensation will not be prorated.

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 133 TLRB, 801-422-6461, 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 Human Research Protections Program by phone at (801) 422-1461; or by email: BYU.HRPP@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:________

APPENDIX C

Reading Passage Stimulus

You know my friend Kathy—she's a Type A person who's very decisive and confident. When a shoe store messed up her order, she drove two hours because she insisted on having the right type. It was clear from the start she's a Type A person. She never seemed to waver on even one choice she made. Well, when her husband proposed to her, she struggled to make a choice. In college, she was very picky when dating and always said she was waiting for the right type. Her mom said she had unrealistic standards, but Kathy insisted it was the one choice she needed to make carefully. One time a guy asked her out, and when Kathy asked where they were going for dinner, he couldn't make a choice. She was very annoyed and ended the date then and there. I was shocked, but I figured if she's a Type A person then maybe she needed a guy who was a little more assertive. Kathy met Liam and knew immediately he was the right type. They dated for five months, but Kathy said she wasn't ready to get married and they broke up. It was the one choice she regretted in the whole time I've known her. Not even a week later, Kathy called Liam. He said that he didn't want to get back together unless they got engaged, so she had to make a choice. Kathy thought about it for two weeks and agreed to marry him. She always says it was the best choice she's ever made.