

## Research Article

# Tongue- and Jaw-Specific Contributions to Acoustic Vowel Contrast Changes in the Diphthong /ai/ in Response to Slow, Loud, and Clear Speech

Antje S. Mefferd<sup>a</sup>

**Purpose:** This study sought to determine decoupled tongue and jaw displacement changes and their specific contributions to acoustic vowel contrast changes during slow, loud, and clear speech.

**Method:** Twenty typical talkers repeated “see a kite again” 5 times in 4 speech conditions (typical, slow, loud, clear). Speech kinematics were recorded using 3-dimensional electromagnetic articulography. Tongue composite displacement, decoupled tongue displacement, and jaw displacement during /ai/, as well as the distance between /a/ and /i/ in the F1–F2 vowel space, were examined during the diphthong /ai/ in “kite.”

**Results:** Displacements significantly increased during all 3 speech modifications. However, jaw displacements increased significantly more during clear speech than during loud and slow speech, whereas decoupled tongue

displacements increased significantly more during slow speech than during clear and loud speech. In addition, decoupled tongue displacements increased significantly more during clear speech than during loud speech. Increases in acoustic vowel contrast tended to be larger during slow speech than during clear speech and were predominantly tongue-driven, whereas those during clear speech were fairly equally accounted for by changes in decoupled tongue and jaw displacements. Increases in acoustic vowel contrast during loud speech were smallest and were predominantly tongue-driven, particularly in men. **Conclusions:** Findings suggest that task-specific patterns of decoupled tongue and jaw displacement change and task-specific patterns of decoupled tongue and jaw contributions to vowel acoustic change across these speech modifications. Clinical implications are discussed.

According to the hyper- and hypo-articulation theory (Lindblom, 1990), typical talkers produce speech on a continuum from hypoarticulation to hyperarticulation, depending on the listener’s needs. That is, when speech clarity demands are low, talkers reduce their articulatory efforts and produce speech with relatively small movement amplitudes. Such hypoarticulated speech yields low phonetic contrast as a consequence of poor kinematic and acoustic specification (Gay, 1978; Kuehn & Moll, 1976; Lee, Shaiman, & Weismer, 2016; Lindblom, 1963; Mefferd, 2015; Mefferd & Green, 2010). However, when speech clarity demands are high, talkers naturally increase their articulatory effort and produce speech that is characterized by large articulatory movements. Such hyperarticulated

speech inherits high phonetic contrast as the result of increased articulatory and acoustic specification (e.g., Leung, Jongman, Wang, & Sereno, 2016; Moon & Lindblom, 1994; Tasko & Greilick, 2010; Tjaden, Lam, & Wilding, 2013). Studies have shown that the quality of vowel production is particularly important and can account for changes in speech clarity and speech intelligibility in typical talkers (Ferguson & Kewley-Port, 2007; Ferguson & Quené, 2014; Lam & Tjaden, 2013).

Hyperarticulated speech is a common goal in speech behavioral interventions that target reduced phonetic contrast in speakers with low speech intelligibility. Studies on typical speakers have shown that hyperarticulated speech can be elicited in various ways; for example, by instructing talkers to speak as clearly as possible (e.g., Ferguson & Quené, 2014; Lam, Tjaden, & Wilding, 2012; Tasko & Greilick, 2010), by asking talkers to reduce their speaking rate (e.g., Edwards, Beckman, & Fletcher, 1991; Gay, 1978; Mefferd, 2015; Mefferd & Green, 2010; Turner, Tjaden, & Weismer, 1995), or more indirectly by asking talkers to

<sup>a</sup>Vanderbilt University Medical Center, Nashville, TN

Correspondence to Antje S. Mefferd: antje.mefferd@vanderbilt.edu

Editor-in-Chief: Julie Liss

Received March 31, 2017

Revision received June 12, 2017

Accepted June 15, 2017

[https://doi.org/10.1044/2017\\_JSLHR-S-17-0114](https://doi.org/10.1044/2017_JSLHR-S-17-0114)

**Disclosure:** The author has declared that no competing interests existed at the time of publication.

increase their vocal intensity (e.g., Darling & Huber, 2011; Mefferd & Green, 2010; Schulman, 1989; Tasko & McClean, 2004; Tjaden & Wilding, 2004). However, side-by-side comparisons suggest that these cues to hyperarticulate do not elicit comparable increases in phonetic contrast. For example, clear speech has been shown to elicit a significantly greater peripheral vowel space area than slow and loud speech in typical talkers (Tjaden et al., 2013). Furthermore, a more expanded vowel space area and greater acoustic vowel contrast was found during slow speech compared to loud speech in typical talkers (Mefferd & Green, 2010; Tjaden & Wilding, 2004).

The difference in magnitude of vowel acoustic change across slow, loud, and clear speech suggests a difference in magnitude of tongue displacement change in response to these speech modifications. Indeed, direct investigations on the strength of associations between articulatory displacement changes in response to these speech modifications and their acoustic consequences supported this notion for most talkers. In a recent study on typical talkers, for example, changes in acoustic vowel specification were moderately or even strongly associated with changes in kinematic vowel specification indexed by the specific tongue position in the vocal tract (Lee et al., 2016). Similarly, the degree of acoustic contrast change in response to rate and loudness manipulation was strongly associated with the extent of change in tongue displacement during these speech tasks (Mefferd, 2015; Mefferd & Green, 2010).

However, the tongue is anatomically coupled with the jaw. Tongue displacements, therefore, can contain contributions of the tongue as well as contributions of the jaw passively moving the tongue. It is important to consider these two components that make up the tongue composite movement, because they may achieve differential vowel acoustic consequences; that is, jaw movements can achieve gross adjustments of the degree of vocal tract constriction, whereas tongue movements independent of the jaw (i.e., decoupled tongue movements) can achieve more refined vocal tract configurations and manipulate the specific constriction location along the palate.

Only a few studies are currently available that have investigated how decoupled tongue and jaw displacement change in response to speaking rate, loudness, or speech clarity modulation. Outcomes of these studies suggest that increases in tongue composite displacement in response to slow, loud, and clear speech may have different underlying patterns of decoupled tongue and jaw displacements. For example, it has been shown that loud speech elicits predominantly displacement changes of the jaw than of the decoupled tongue (Tasko & McClean, 2004), whereas slow speech elicits predominantly displacement changes of the decoupled tongue than of the jaw (Hertrich & Ackerman, 2000; Perkell & Zandipour, 2002; Westbury & Dembowski, 1993). Finally, clear speech has been shown to elicit considerable increases in both tongue and jaw displacements (Tasko & Greilick, 2010).

Direct side-by-side comparisons of tongue- and jaw-specific changes across these three speech modifications

are, however, currently lacking, and it remains unknown if loud, clear, and slow speech elicit task-specific patterns of tongue and jaw displacement changes. It is also currently unknown to what extent each articulator contributes to the speech acoustic changes in response to these speech modifications. Such information is critical to better understand the articulator-specific demands placed on a talker when cued to implement slow, loud, or clear speech, particularly in the light of these speech modifications being commonly used as a therapeutic treatment approach to improve speech intelligibility in talkers with dysarthria (Yorkston, Hakel, Beukelman, & Fager, 2007). Currently, clear guidance on the selection of a specific speech modification as a treatment approach is lacking for many talkers with dysarthria. Thus, an improved understanding about articulator-specific changes in response to these speech modifications may be the first step to help better predict how talkers with articulator-specific impairments (i.e., tongue-dominant impairment vs. jaw-dominant impairment) respond to these three speech modifications. These insights may also eventually help predict which approach will maximize vowel acoustic contrast in these talkers. Even from a theoretical perspective, findings yield important insights about task-specific interarticulatory movement patterns and their impact on speech acoustics.

Therefore, the current study aimed to address two research questions: (a) What are the effects of slow, loud, and clear speech on decoupled tongue and jaw displacements? (b) How do decoupled tongue- and jaw-specific changes contribute to acoustic vowel contrast changes in response to slow, loud, and clear speech?

To formulate hypotheses about the effects of speech modification on decoupled tongue and jaw displacement, it is helpful to first discuss the relations between decoupled tongue, jaw, and tongue composite movement in a three-dimensional (3D) model (see Figure 1a). In this model, the amount of jaw displacement is indicated along the  $x$ -axis, and the amount of decoupled tongue displacement is indicated along the  $y$ -axis. The overall size of the tongue composite displacement is indicated by the distance from the origin to the circular isoline. As can be seen, if the tongue composite movement consists of only jaw movement, then the size of the tongue composite displacement is equal to the amount of jaw displacement along the  $x$ -axis. Similarly, if the same tongue composite displacement consists only of decoupled tongue movement, then the size of the tongue composite displacement is equal to the amount of decoupled tongue displacement along the  $y$ -axis. If decoupled tongue and jaw both contribute to the tongue composite movement, the extent of relative contribution of each articulator can be expressed by an angular displacement from the  $x$ -axis along the isoline of tongue composite movement size (or, in short, by an angle  $\theta$ ). The larger the angular displacement from the  $x$ -axis along the isoline, the more the decoupled tongue contributes to the tongue composite movement (see Chung, Kong, Edwards, Weismer, & Fourakis, 2012, for similar use of angular displacement to describe positional changes in acoustic vowel space).

**Figure 1.** (a) A model of the relations between decoupled tongue and jaw displacements and the resulting tongue composite displacement (circular isoline). (b) Hypothetical framework for predicted changes in decoupled tongue and jaw displacement, as well as the changes in tongue composite displacements for productions of the diphthong /ai/ in response to slow, loud, and clear speech. Changes in the angular displacement from the  $x$ -axis  $\theta$  indicate the relative contribution of the decoupled tongue and jaw to the tongue composite displacement along the isoline. The angle also indicates articulator-specific contributions to acoustic contrast change with an angle close to  $0^\circ$  indicating predominantly jaw-driven vowel acoustic change, an angle close to  $90^\circ$  indicating predominantly tongue-driven vowel acoustic change, and an angle around  $45^\circ$  indicating tongue- and jaw-driven vowel acoustic changes.

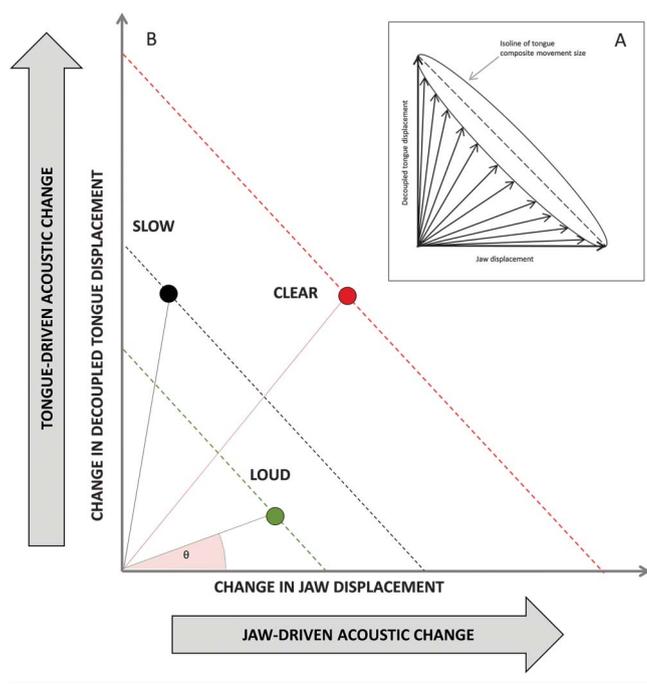


Figure 1b shows the hypothetical framework from a perspective in which the circular isoline of the tongue composite movement is projected as a straight line. This was done to reduce the visual complexity of the hypothetical framework. The  $x$ -axis and the  $y$ -axis in the hypothetical framework indicate jaw and decoupled tongue displacement change relative to typical speech, respectively. Predictions for the location of clear, loud, and slow speech within the described framework were made based on previous kinematic and acoustic findings. Specifically, because the magnitude of change in vowel acoustic contrast is associated with the size of the tongue composite movements (Mefferd & Green, 2010), the predicted intercepts for each speech modification with the  $x$ - and  $y$ -axes were based on previous studies reporting differences in the magnitude of change in acoustic vowel contrast or acoustic vowel space in response to slow, loud, and clear speech (Lee et al., 2016; Mefferd, 2015; Mefferd & Green, 2010; Tjaden et al., 2013; Tjaden & Wilding, 2004). Predictions for the angular displacement from the  $x$ -axis along the task-specific tongue composite movement isolines were made based on previous kinematic

findings of relative changes in tongue and jaw displacements in response to speaking rate, loudness, and speech clarity modifications (Hertrich & Ackermann, 2000; Tasko & Greilick, 2010; Tasko & McClean, 2004).

As can be seen in Figure 1b, talkers were expected to achieve the greatest increase in acoustic vowel contrast during clear speech (Tjaden et al., 2013); hence, the intercept of the tongue composite movement isoline with the  $x$ - and  $y$ -axes was set as the largest of all three speech modifications. The angle  $\theta$  was predicted to be approximately  $45^\circ$ , indicating equal increases in tongue and jaw displacements (Tasko & Greilick, 2010). Furthermore, slow speech was expected to elicit the second largest increase in acoustic vowel contrast (Tjaden et al., 2013), which is why the intercept of the tongue composite isoline with the  $x$ - and  $y$ -axes was set as second largest of all three speech modifications. The angle  $\theta$  was expected to be much larger than the angle of clear speech and closest to  $90^\circ$  of all three speech modifications. That is because speaking rate change has been shown to elicit more change in decoupled tongue displacement than in jaw displacement (e.g., Edwards et al., 1991; Hertrich & Ackermann, 2000). Furthermore, the decoupled tongue displacements were expected to increase to similar extents for slow and clear speech, presuming that both speech modifications would maximize decoupled tongue movements. Finally, loud speech was expected to produce the smallest increases in vowel acoustic contrast relative to slow and clear speech (Mefferd & Green, 2010; Tjaden et al., 2013). Therefore, the intercept of the tongue composite isoline with the  $x$ - and  $y$ -axes for loud speech was set as the smallest of all three speech modifications. Furthermore, the angular displacement was hypothesized to be the closest to the  $x$ -axis based on kinematic findings that loud speech effects more consistently showed increases in jaw displacements than in decoupled tongue displacements (Tasko & McClean, 2004).

With regard to tongue- and jaw-specific contributions to acoustic vowel contrast change, no previous studies provided a basis for formulating a hypothesis. However, we thought it was reasonable to hypothesize that the articulator that changes the most in response to a speech modification would account predominantly for the vowel acoustic changes. Thus, changes in acoustic vowel contrast in response to slow speech were expected to be predominantly tongue-driven, whereas those for loud speech were expected to be predominantly jaw-driven. Finally, decoupled tongue and jaw displacement changes were expected to contribute equally to acoustic vowel contrast change during clear speech.

## Method

### Participants

Twenty-four typical talkers participated in this study; however, only data of 20 talkers (nine men, 11 women) are presented here. The data of four participants were excluded because these participants did not perform the speech tasks correctly (see section of speech task verification). Talkers

ranged in age from 18 to 28. All participants passed a standard hearing screening (500, 1000, 2000, and 4000 Hz at 25 dB HL in both ears) and denied a history of neurological conditions or a previous diagnosis of a speech, language, or hearing impairment. Furthermore, all participants spoke with a Standard American English dialect.

### **Experimental Tasks**

All participants were asked to repeat “see a kite again” five times as they normally speak. Then they were asked to repeat the utterance five times in the following ways: as fast as possible, at half their typical speaking rate, twice as loud, and finally as clearly as possible using effortful articulation to overenunciate each word in the sentence. Furthermore, the instructions mentioned that, by overenunciating each word, their speech could become slower and/or louder than normal (Tjaden et al., 2013). Speech conditions were not counterbalanced or randomized. All participants completed the speech conditions in the same order.

Although fast speech was included in the protocol, the data were not further analyzed for this study because fast speech is not used as a speech behavioral modification approach in therapeutic interventions. The diphthong /ai/ embedded in “kite” was of interest in this study. The word “kite” was used because it provides clearly defined boundaries in the tongue kinematic and speech acoustic signal. The diphthong /ai/ was also chosen because it is mainly produced by tongue and jaw movements.

### **Data Collection and Processing**

Articulatory movements of the tongue and jaw were captured with a sampling rate of 250 Hz using a 3D electromagnetic articulograph (AG501, Carstens Medizintechnik GmbH). To acquire speech kinematic data, two sensors were attached with dental adhesive (Periacyrl 90, GluStitch, Inc.) to the sagittal midline of the tongue. One tongue sensor was placed approximately 4 cm posterior to the tongue tip; the other sensor was placed approximately 1.5 cm posterior to the tongue tip. Only the movement of the posterior tongue sensor was of interest to this study because the tongue segment to which this sensor was attached is associated with forming tongue palate constriction during vowels and diphthong productions (Wang, Samal, Rong, & Green, 2016). Furthermore, three sensors were attached with a small amount of putty (Stomahesive, ConvaTec) on the gumline of the lower teeth to track jaw movements. Specific landmarks were between the right and left canines and premolars, as well as between the lower central incisors. The center jaw sensor (at sagittal midline) was selected for this study because it was best suited to implement the tongue–jaw decoupling algorithm based on Westbury and colleagues (2002; see next section for details). An additional sensor was attached with dental adhesive to the sagittal midline of the upper lip as well as to the lower lip; however, their movements were not analyzed in this study. Finally, all participants wore

plastic goggles that had three sensors evenly spaced, with the head center sensor approximately aligning with the top nose bridge and the other two sensors placed to the right and left. These sensors were used as reference sensors to track head movements during speech production.

A short REST recording was completed in which a still shot of the sensor setup was taken. For 13 of the 24 participants, this was done by asking the talker to bring their teeth together, close their lips, and rest the tongue inside the mouth. For all other participants, three additional reference sensors were taped to the plastic bite plate provided by Carstens for a REST recording in which each participant was asked to hold the bite plate between their teeth. This bite plate became commercially available while the project was underway.

Raw data were converted into positional data using the CalcPos software provided by Carstens. Next, the NormPos software (Carstens) was used for head movement correction and transposing the kinematic data into a head-based coordinate system with the origin defined by either the center jaw sensor or the center bite plate sensors (located approximately 20 mm anterior to the jaw center sensor). For those participants who complete a REST recording without a bite plate, the right, center, and left jaw sensors were used as reference sensors to transpose all data into the local coordinate system. Otherwise the right, center, and left bite plate sensors were used to transpose the data. As a result, the origin was located 20 mm more posteriorly in the anterior–posterior dimension for those participants who did not have a bite plate during the REST recording compared to those whose REST file was recorded with the bite plate. Because only relative positional changes were used for this study, differences in the origin did not have an impact on any measures. All kinematic data were then smoothed with a 15-Hz LP filter in SMASH, a MATLAB-based software program (Green, Wang, & Wilson, 2013).

The acoustic signal was recorded at a sampling rate of 48 kHz and 16-bit resolution in a synchronized fashion with the kinematic data. The cable of a lavalier microphone (Audiotechnica AT899) was taped to the hood of the articulograph, with the microphone hanging in front of the participant’s head, creating a microphone-to-mouth distance of approximately 20 cm. A calibration tone was recorded for each participant and used as a reference to calculate the vocal intensity for each participant across all speech conditions.

### **Data Analysis**

Kinematic data were analyzed in SMASH. Onsets and offsets of the target /ai/ were determined based on the posterior tongue’s positional extrema (trough for /a/, peak for /i/) in the ventral-dorsal dimension. 3D posterior tongue sensor positions at the onset and offset of the diphthong /ai/ were used to calculate tongue composite displacement based on the Euclidean distance formula (Mefferd, 2015). Identical to the procedure for the posterior tongue displacement calculation, the center jaw sensor positions associated

with the onset and offset of the diphthong /ai/ were used to calculate jaw displacements at the lower central incisors. However, due to the anatomical coupling of the tongue and jaw, the tongue composite displacement includes the contributions of the jaw. Therefore, the extracted posterior tongue displacement and jaw displacement measures were submitted to a decoupling algorithm to determine relative contributions of jaw and tongue to the tongue composite displacement. Such a decoupling is, however, not simple. Jaw displacements recorded at the lower central incisors are known to be greater than those at a more posterior location (Westbury, Lindstrom, & McClean, 2002), that is, because of the jaw's rotational movements during speech production. For that reason, decoupling of the tongue and jaw could not be achieved by subtraction of anterior jaw displacement from posterior tongue displacement. Previous research has shown that a linear subtraction approach can introduce large errors (Westbury et al., 2002). Thus, decoupling has to take into account the jaw's pitch rotation.

Currently, there is only one empirically tested decoupling approach available for 3D kinematic data (see Henriques & van Lieshout, 2013). This approach, however, requires a special recording setup where three jaw sensors are embedded in thermoplastic molds that position the jaw sensors perpendicular to the labial surface of the frontal incisors so that the orientation of the sensor can be used to determine the pitch rotation. Because the kinematic data of this study were not recorded with such a special sensor setup, this approach could not be used in this study. Instead, our approach to account for jaw pitch rotation in estimating posterior jaw displacement was based on the decoupling algorithm developed by Westbury and colleagues (2002) for two-dimensional (2D) speech kinematic data.

On the basis of the notion that jaw movements during speech resemble an arc, the jaw's pitch rotation angle  $\alpha$  was estimated based on the extracted jaw displacement (jaw  $\text{displ}_{\text{measured}} = 3\text{D Euclidean distance between onset and offset}$ ). Per empirical evaluation by Westbury and colleagues, a jaw pitch rotation of  $0.52^\circ/\text{mm}$  jaw displacement can be assumed (see Westbury et al., 2002):

$$\alpha = \text{jaw displ}_{\text{measured}} * 0.52. \quad (1)$$

Next, the length of the radius from the center of rotation to the jaw sensor ( $r_{\text{total}}$ ) was calculated based on the circular function (arc length = radius \* central angle [in radians]):

$$r_{\text{total}} = \frac{\text{jaw displ}_{\text{measured}}}{\alpha \text{ (in radians)}}. \quad (2)$$

When applying Equation 1 to Equation 2, it can be seen that  $r_{\text{total}}$  was a constant radius of 110.18 mm used for all participants (Westbury et al., 2002). Next, the distance from the center of rotation to the posterior tongue location was estimated by subtracting the distance from

jaw sensor to posterior tongue sensor from  $r_{\text{total}}$ . This distance was calculated from posterior tongue and jaw center sensor locations during the REST recording. Once the length of the radius  $r_{\text{posterior jaw}}$  was determined, the jaw displacement at the posterior location ( $\text{displ}_{\text{jaw estimated}}$ ) was calculated:

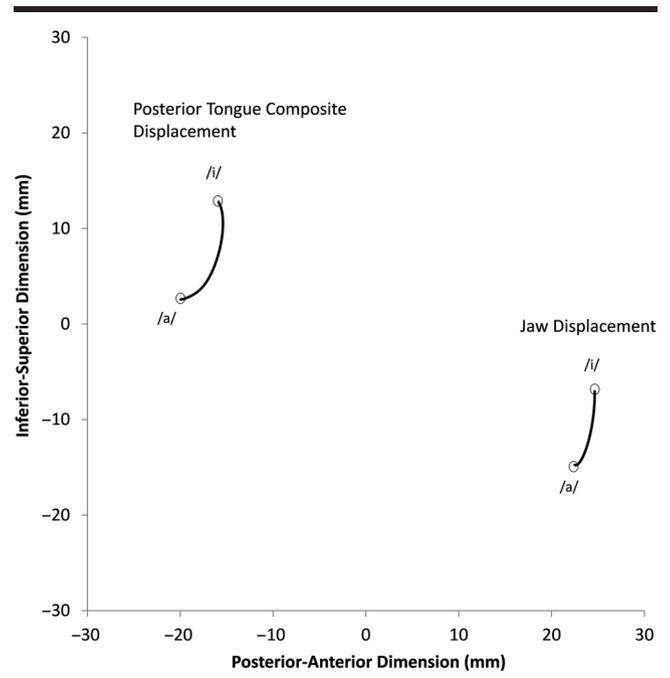
$$\text{displ}_{\text{jaw estimated}} = r_{\text{posterior jaw}} * \alpha \text{ (in radians)}. \quad (3)$$

The jaw displacement  $\text{displ}_{\text{jaw estimated}}$  was then subtracted from the extracted posterior tongue displacement to estimate the displacement of the tongue independent of the jaw ( $\text{displ}_{\text{iTongue estimated}}$ ). Throughout the rest of the manuscript, we will refer to  $\text{displ}_{\text{jaw estimated}}$  as jaw displacement.

$$\text{displ}_{\text{iTongue estimated}} = \text{displ}_{\text{totalPT}} - \text{displ}_{\text{jaw estimated}} \quad (4)$$

Figure 2 provides an example of this approach. Tongue composite movement and jaw movement of the diphthong production during one repetition of clear speech is shown in the sagittal plane. In this example, the jaw displacement was 10.02 mm, and the posterior tongue composite displacement was 11.33 mm. The distance between the central jaw sensor and the posterior tongue sensor was 41.94 mm. By using Equation 1,  $\alpha = 5.21$  ( $10.02 * 0.52$ ).

**Figure 2.** An example of jaw and posterior tongue composite movement during the diphthong /ai/ produced during clear speech by a female talker. Note that only movement in the sagittal plane was used in this figure although 3D kinematic data were recorded and 3D Euclidean distances between onset and offset (indicated by unfilled circles) were calculated for both articulators.



Then in Equation 2,  $r_{\text{total}}$  is determined to be 110.18 mm ( $r_{\text{total}} = 10.02/0.52 * 10.02 * \pi/180$ —degrees had to be converted into radians). Then  $r_{\text{posterior jaw}}$  was calculated as described using the REST recording. The  $r_{\text{posterior jaw}}$  was 68.24 mm (110.18 – 41.94 mm) and was then used in Equation 3 to determine  $\text{displ}_{\text{jaw estimated}}$ . Based on Equation 3, the displacement at the posterior jaw  $\text{displ}_{\text{jaw estimated}}$  was then determined to be 6.21 mm ( $68.24 * 5.21 * \pi/180$ —again, degrees had to be converted into radians). Finally, Equation 4 was used to subtract estimated posterior jaw displacement from the observed posterior tongue composite displacement (11.34 mm – 6.21 mm); hence, independent tongue displacement was estimated to be 5.13 mm for this diphthong production.

The described approach to estimate jaw pitch rotation and posterior jaw displacement during the diphthong production differs from Westbury et al. (2002) in one major way. Instead of reducing the recorded 2D jaw movements into one dimension by calculation of a principal component signal, in the current study, three dimensions were reduced into one dimension by calculation of the 3D Euclidean distance change of the jaw. The principal component signal of the jaw was not used, because it was deemed more parsimonious for this study to work with the already extracted 3D distance measures of the jaw rather than adding the calculation for the principal component signal to determine the relative change of jaw displacement from onset to offset based on the principal component signal. A comparison of jaw displacement based on the principal component signal (calculated from  $x$ ,  $y$ , and  $z$  data points) and 3D Euclidean distance measures in a pilot data set containing 60 data points (3 talkers  $\times$  4 speech conditions  $\times$  5 repetitions) showed an absolute mean difference of 0.22 mm (range: 0.00–1.78 mm). Furthermore, the possible impact of using 3D data rather than 2D data was also evaluated based on this pilot data set. The absolute mean difference between the 3D Euclidean measure and the 2D Euclidean measure was 0.10 mm (range: 0.00–0.49 mm). These small differences concur with previous reports that jaw movements during speech occur predominantly in the sagittal plane (e.g., Ostry, Vatskiotis-Bateson, & Gribble, 1997).

Finally, it is important to point out that this approach is not suited to generate a continuous trajectory of independent tongue movement during speech production. This approach was specifically developed for analyzing the articulatory movements during the diphthong /ai/, and in its current form this approach is likely limited to “decompose” the segment-length tongue composite displacements into its two components (independent tongue, jaw) by extracting observed anterior jaw and posterior tongue displacements, adjusting the jaw displacement based on the shorter radius and estimated pitch rotation of the jaw before subtracting jaw displacement from the posterior tongue composite displacement.

Because this algorithm has not been used before for 3D kinematic data, we evaluated its performance by comparing the estimated posterior jaw displacement to actual posterior jaw displacement in one typical talker. For this

talker, an additional sensor was attached to the lateral surface of the second right molar. The posterior tongue sensor was placed parallel to the posterior jaw sensor on the sagittal midline of the tongue. Kinematic data were collected during the same speech tasks that all participants completed for the study. Data were analyzed using the decoupling approach described above, as well as the direct approach (measuring the 3D Euclidean distance for the posterior jaw sensor based on the 3D positions at onset and offset of the target /ai/). For each measuring approach, 20 data points were obtained. The median absolute error was 0.19 mm, with a standard deviation of 0.23 mm and a range from 0.01 to 0.93 mm. In comparison, median errors reported by Westbury and colleagues for their decoupling algorithm were 0.51 mm, with a maximum error of 2.6 mm (Westbury et al., 2002).

### *Acoustic Data Analysis*

Acoustic vowel contrast calculation followed procedures described in Mefferd (2015). Briefly, using the spectrographic view in TF32 (Milenkovic, 2005), the diphthong onsets and offsets were identified using the acoustic characteristics of surrounding consonants as landmarks. The linear predictive coding algorithm of TF32 was used to identify the measurement points for the F1 and F2 values of /a/ and /i/. Specifically, for /a/ the F2 minimum and corresponding F1 values were selected, and for /i/ the F2 maximum and corresponding F1 values were selected. These formant values were used to calculate the 2D Euclidean distance between /a/ and /i/ in the F1–F2 vowel space.

### *Verification of Task Performance*

To verify that participants indeed increased their vocal intensity or reduced their speaking rate, vocal intensity and movement duration were measured. Vocal intensity was measured in TF32. Using the spectrographic view and waveform, the onset and offset of the diphthong /ai/ were identified (i.e., first and last glottal impulse, respectively). Mean dB values were extracted and reexpressed relative to the intensity of the calibration tone recorded for each participant. Diphthong durations were determined in the kinematic domain based on the previously defined kinematic onsets and offsets of the diphthong (posterior tongue sensor positional trough and peak in the vertical dimension). Finally, to verify that participants indeed increased speech clarity, acoustic vowel contrast was examined.

### *Statistical Analysis*

Task performance was evaluated by submitting each participant’s raw values of vocal intensity, diphthong durations, and vowel contrast to mixed linear model analyses (one for each dependent variable). Task performance was further examined on an individual basis to ensure that all participants indeed modified their speech according to instructions provided. Individual performance verification was important for the validity of the regression analysis

where raw data of typical performance and one speech modification were submitted. The within-group analyses were necessary to determine the effect sizes for rate, loudness, and clarity changes so that findings of this study could be interpreted within the context of previous work. Furthermore, within-group analyses provided insights on how intensity, duration, and vowel acoustic contrast changed across all speech tasks, not just relative to typical speech.

Speech modification effects on decoupled tongue and jaw displacement, as well as the composite tongue movement and acoustic vowel contrast, were determined by submitting each participant's raw values to mixed linear model analyses (one for each variable). The critical  $p$  value was set to .05, and in pairwise comparisons, the  $p$  value was adjusted for multiple comparisons using Bonferroni corrections. To determine the jaw-specific contributions to changes in acoustic vowel contrast, the estimated jaw displacements were regressed against acoustic vowel contrast measures associated with typical speech and an experimental condition (typical-clear, typical-slow, typical-loud). To determine tongue-specific contributions to changes in acoustic vowel contrast, the decoupled tongue displacements were regressed against acoustic vowel contrast measures in the same fashion. Finally, to evaluate the hypothetical framework, the mean decoupled tongue and jaw displacement changes were plotted for each speech modification in an  $x$ - $y$  plot. Using the mean change in tongue composite displacement as the intercept with the  $x$ - and  $y$ -axes, the isoline tongue composite displacement changes were plotted. The angular displacement from the  $x$ -axis was calculated for each speech modification. The isoline of tongue composite displacement and the angle  $\theta$  were used to evaluate the hypothetical framework descriptively.

### Reliability Measures

To determine intrameasurer reliability, 22% of the acoustic vowel contrast calculations were remeasured and recalculated by the same analyst after approximately 2 months of the initial completion of measurements. The mean absolute differences between initial and remeasured data were 22.9 Hz ( $SD = 35.1$  Hz) and 33.2 Hz ( $SD = 45.3$  Hz) for F1 and F2, respectively. Similarly, intermeasurer reliability was determined by remeasuring and recalculating 24% of the acoustic vowel contrast calculations. The mean absolute differences between the first analyst and the second analyst were 20.2 Hz ( $SD = 29.6$  Hz) and 18.9 Hz ( $SD = 30.6$  Hz) for F1 and F2, respectively. These results are comparable to those of previous acoustic studies (Lam, Tjaden, & Wilding, 2012; Lee et al., 2016; Tjaden et al., 2013).

## Results

### Prehypothesis Testing: Performance of Speech Modification

Inspection of individual performance patterns revealed that all participants increased their vocal intensity from typical speech to loud speech and increased their diphthong

durations from typical to slow speech. Four participants were excluded from the data set because their acoustic vowel contrast either decreased or remained similar in response to clear speech (CYF10 decreased by 77 Hz, CYF21 decreased by 56 Hz, CYM13 decreased by 69 Hz, CYM10 increased by 16 Hz). The acoustic data of these participants suggested that they already used clear speech during the typical speech condition because, relative to the group mean, their acoustic vowel contrasts were extremely large and comparable to those other talkers demonstrated during clear speech.

Table 1 displays the means and standard deviations of diphthong durations, vocal intensity, and acoustic vowel contrast measures for all four speech conditions based on the 20 remaining talkers. A significant main effect for speech task on diphthong durations was found,  $F(3, 142) = 56.493$ ,  $p < .001$ . Slow speech was associated with significantly longer diphthong durations compared to typical, loud, and clear speech ( $p < .001$ ). Furthermore, relative to typical speech, diphthong durations were significantly longer during clear and loud speech ( $p \leq .001$ ).

Speech task also had a significant effect on vocal intensity,  $F(3, 187) = 69.357$ ,  $p < .001$ . Pairwise comparisons showed that vocal intensity was significantly greater during loud speech compared to typical, clear, and slow speech ( $p < .001$ ). Furthermore, vocal intensity was significantly greater during clear speech than during typical speech ( $p = .004$ ) and slow speech ( $p < .001$ ).

Speech tasks had a significant effect on acoustic vowel contrast,  $F(3, 189) = 153.155$ ,  $p < .001$ . Specifically, loud, clear, and slow speech yielded significantly larger acoustic vowel contrast relative to typical speech ( $p < .001$ ). Furthermore, acoustic vowel contrast was significantly larger during slow and clear speech than during loud speech ( $p < .001$ ). Finally, acoustic vowel contrast also tended to be larger during slow speech compared to clear speech ( $p = .055$ ).

### Speech Modification Effects

Figure 3a displays the mean jaw displacements for each speech task. As can be seen, speech modifications had a significant effect on jaw movements,  $F(3, 205) = 63.345$ ,  $p < .000$ . That is, jaw displacements were significantly larger during clear, loud, and slow speech than during typical speech ( $p < .001$ ). Furthermore, jaw displacements were significantly larger during clear speech than during loud and slow speech ( $p < .001$ ).

As can be seen in Figure 3b, speech modifications had a significant effect on the decoupled tongue movements,  $F(3, 204) = 66.76$ ,  $p < .001$ . Relative to typical speech, decoupled tongue displacements were significantly larger during clear, loud, and slow speech ( $p \leq .001$ ). In addition, slow speech was associated with significantly larger tongue displacements than clear speech ( $p = .006$ ) and loud speech ( $p < .001$ ). Finally, tongue displacements were significantly larger during clear speech than during loud speech ( $p = .004$ ).

Figure 3c shows the mean composite movement of the posterior tongue for each speech condition. A significant

**Table 1.** Mean durations ( $\pm$  SDs), mean vowel intensities ( $\pm$  SDs), and vowel acoustic contrasts ( $\pm$  SDs) across speech conditions.

Measure	Typical		Loud		Clear		Slow	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Duration (s)	0.110	0.018	0.127	0.028	0.136	0.031	0.235	0.072
Vocal intensity (dB)	74.644	7.414	87.992	8.075	78.140	7.275	73.758	7.452
Contrast (Hz)	554.317	142.037	698.201	183.262	935.317	192.250	1015.267	147.449

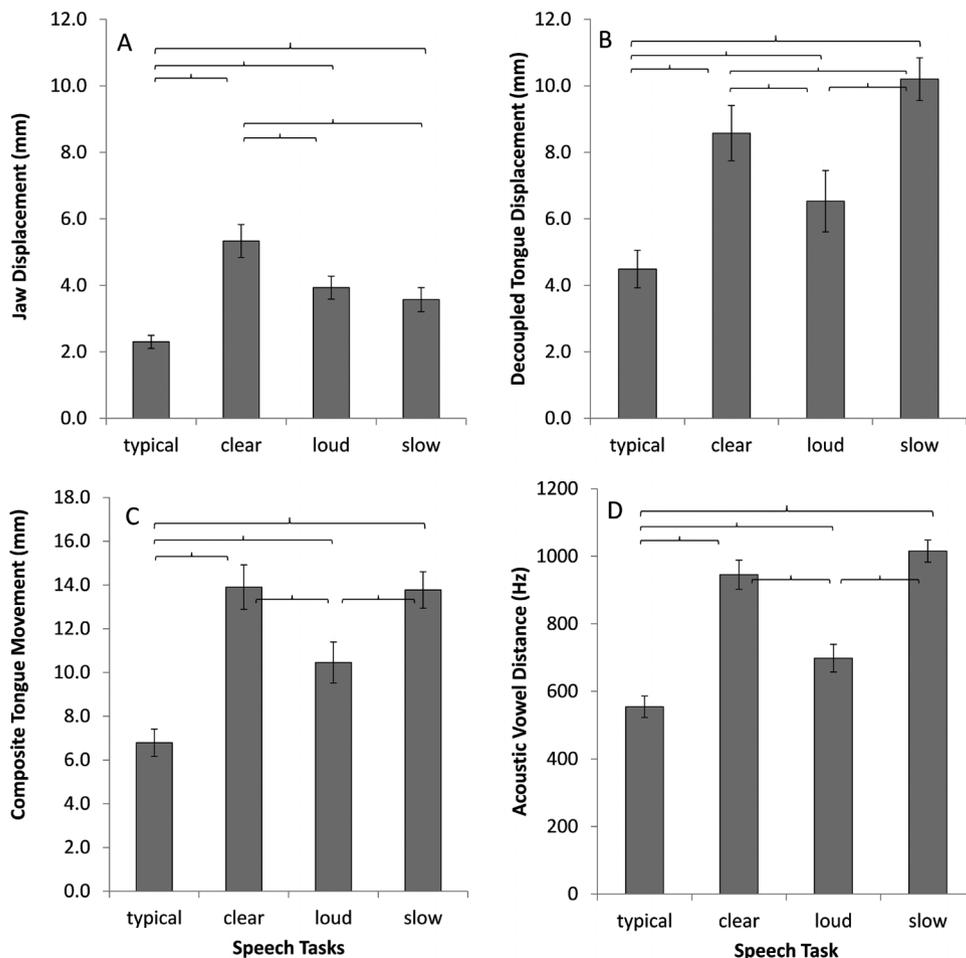
effect of speech modification on composite tongue movement was found,  $F(3, 192) = 90.86, p < .001$ , with slow, loud, and clear speech eliciting significantly larger composite movements than typical speech and slow and clear speech also eliciting significantly larger tongue composite movements than loud speech ( $p < .001$ ).

Figure 3d displays the mean acoustic vowel contrast for each speech task. As mentioned above, speech modifications had a significant effect on acoustic vowel contrast,  $F(3, 189) = 153.155, p < .001$ . Specifically, loud, clear, and slow speech yielded significantly larger acoustic vowel

contrast relative to typical speech ( $p < .001$ ). Furthermore, acoustic vowel contrast was significantly larger during slow and clear speech than during loud speech ( $p < .001$ ). Finally, acoustic vowel contrast also tended to be larger during slow speech compared to clear speech ( $p = .055$ ).

It should be noted that gender effects were investigated and were found to be significant. For all variables, men had, in general, larger values than women. These differences are to be expected due to known anatomical differences in vowel tract size, which are reflected in the vowel acoustics. Gender  $\times$  Task interactions, however, were not significant.

**Figure 3.** Group means ( $\pm$  SE) for decoupled tongue displacements (a), jaw displacements (b), composite movements of the tongue (c), and acoustic vowel contrast (d) across all speech conditions. Brackets indicate significant pairwise comparisons.



Therefore, only the findings for the across all talkers were presented for speech modifications effects.

### Articulator-Specific Contributions to Vowel Acoustic Change

The six panels of Figure 4 show the jaw- and tongue-specific contributions to change in acoustic vowel contrast in response to the three speech modifications (typical–clear, typical–loud, typical–slow). Separate analyses were completed for female and male talkers due to the gender-specific articulatory-to-acoustic mappings (e.g., Mefferd & Green, 2010). Changes in jaw displacement from typical to clear speech accounted for 33.1% of the variance in acoustic vowel contrast in women and 33.4% in men ( $p < .001$ ), whereas changes in decoupled tongue displacement accounted for 25.9% of change in acoustic vowel contrast in women and 33.2% in men ( $p < .001$ ). Furthermore, changes in jaw displacement from typical speech to loud speech accounted for 4.2% of the variance in acoustic vowel contrast in women ( $p = .033$ ) and 9.4% in men ( $p = .003$ ). However, changes in tongue displacement accounted for 19.8% of change in acoustic vowel contrast in female talkers and 49.0% in male talkers ( $p < .001$ ). Finally, changes in jaw displacement from typical to slow speech accounted for 20.0% of the variance in acoustic vowel contrast in women and 29.5% in men ( $p < .001$ ). Finally, changes in tongue displacement

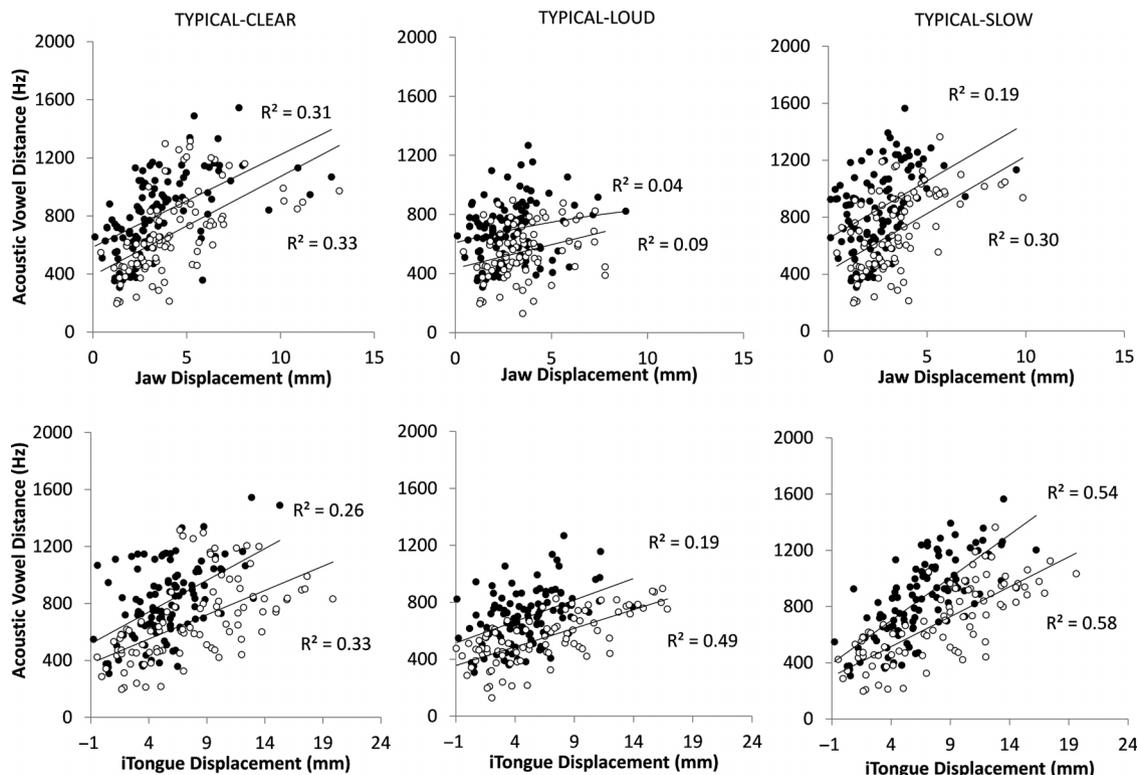
accounted for 53.8% of change in acoustic vowel contrast from typical to slow speech in women and 58.0% in men ( $p < .001$ ).

### Testing the Hypothetical Framework

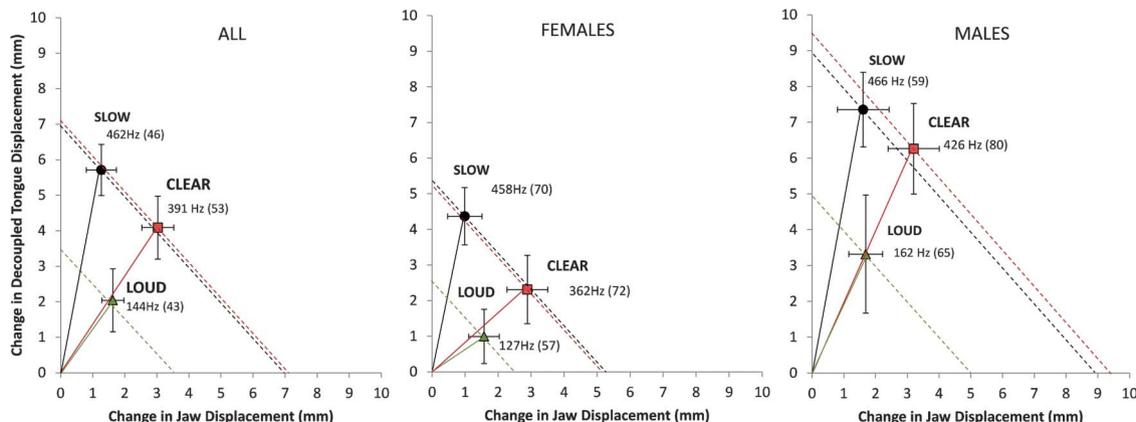
The three panels of Figure 5 show the findings for the hypothetical framework with the changes in decoupled tongue and jaw displacements relative to typical speech in an  $x$ - $y$  graph. Furthermore, the mean change of the tongue composite displacement was drawn (colored dashed lines) for each speech modification. The corresponding mean change in acoustic vowel contrast was provided for each speech modification in a text box next to the kinematic data. Figure 5a shows the overall group means, whereas Figures 5b and 5c show means for women and men, respectively.

As can be seen in Figure 5a, slow speech was associated with the largest increase in decoupled tongue displacement in combination with the smallest increase in jaw displacements; however, jaw displacements were not statistically different from loud speech. These relative changes in decoupled tongue and jaw displacement resulted in slow speech being positioned at an angle of  $78.46^\circ$  from the  $x$ -axis on an isoline almost identical to that of clear speech. However, the increase in acoustic vowel contrast in response to slow speech tended to be larger than the increase in acoustic vowel contrast during clear speech ( $p = .055$ ). Finally,

**Figure 4.** Acoustic vowel contrast changes as a function of jaw displacement changes (top) and decoupled tongue displacement changes (bottom). Filled circles are female talkers; unfilled circles are male talkers.



**Figure 5.** Testing the hypothetical framework across all participants (a), in female talkers (b), and in male talkers (c). Mean decoupled tongue and jaw displacements ( $\pm SE$ ) for each speech modification and the tongue composite displacement (dashed colored lines).



increases in decoupled tongue displacement contributed to a greater extent to changes in acoustic vowel contrast than did jaw displacement increases during slow speech.

Clear speech elicited the second largest increase in decoupled tongue displacement in combination with the largest increase in jaw displacements across all three speech modifications. Although clear speech was positioned on an isoline of tongue composite displacement change that was comparable to that of slow speech, its angle from the  $x$ -axis was smaller ( $53.47^\circ$ ) and tended to elicit a smaller increase in acoustic vowel contrast. Finally, decoupled tongue and jaw displacement increases were found to contribute relatively equally to the increases in acoustic vowel contrast.

Finally, loud speech was associated with the smallest increase in decoupled tongue displacement of all three speech modifications in combination with the smallest increase in jaw displacements. Yet, the difference in jaw displacement increase relative to slow speech was nonsignificant. The pattern of decoupled tongue and jaw displacement increase observed for loud speech resulted in an angular displacement from the  $x$ -axis that was smaller than that of slow speech; however, it was similar to that of clear speech ( $51.37^\circ$ ). However, articulatory performance of loud speech was associated with the smallest tongue composite displacement change and yielded the smallest change in acoustic vowel contrast across all three speech modifications. The increases in decoupled tongue displacement contributed to a greater extent to changes in acoustic vowel contrast than did the increases in jaw displacement during loud speech.

Some interesting sex-specific findings should be pointed out as well, although these findings are preliminary due to the small number of talkers within each group (11 female talkers, nine male talkers). In general, speech modifications had similar effects on tongue and jaw displacement and vowel acoustic contrast in female and male talkers; however, displacement changes were overall larger in male talkers than in female talkers, which resulted in larger tongue composite displacements in male talkers compared to female talkers. Furthermore, angular

displacements from the  $x$ -axis for loud and clear speech were larger in male talkers than in female talkers (see Table 2). These gender differences were also evident in articulator-specific contributions to acoustic vowel changes. Increases in decoupled tongue displacements accounted for smaller portions of the acoustic vowel contrast change in clear and loud speech in female talkers compared to those in male talkers. Furthermore, increases in jaw displacements accounted for a fewer portion of the acoustic vowel contrast change in slow speech in female talkers than in male talkers. Table 2 provides a summary of the kinematic and acoustic measures of interest for female and male talkers, as well as across all participants.

## Discussion

This study aimed to determine (a) how decoupled tongue and jaw displacements change in response to slow, loud, and clear speech and (b) how displacement changes of the decoupled tongue and jaw contribute to vowel acoustic changes in response to these speech modifications. Previously observed changes in vowel acoustics and tongue composite movements, as well as relative changes in decoupled tongue and jaw displacements, during slow, loud, and clear speech (e.g., Hertrich & Ackermann, 2000; Mefferd & Green, 2010; Tasko & Greilick, 2010; Tasko & McClean, 2004; Tjaden et al., 2013) provided the basis for a testable hypothetical framework. The discussion will first focus on decoupled tongue and jaw displacement changes in response to the three speech modifications and will then address outcomes for tongue- and jaw-specific contributions to acoustic vowel contrast changes within the context of the proposed framework. Finally, clinical implications, study limitations, and future directions will be discussed.

### Speech Modification Effects

For slow speech, decoupled tongue- and jaw-specific displacement changed as hypothesized. Findings of this study

**Table 2.** Mean increases in jaw displacement (estimated at the posterior location), independent tongue displacement (iTongue), and tongue composite displacement (all in mm), as well as the mean increase in acoustic vowel contrast (in Hz), relative to typical speech for the group, female talkers, and male talkers.

Group	Speech modification	Increase in jaw displacement	Increase in iTongue displacement	Increase in tongue composite displacement	Increase in acoustic vowel contrast	Angle $\theta$	Jaw contribution to acoustic change	Tongue contribution to acoustic change
All	Loud	1.63	2.04	3.67	144	51.37	—	—
	Clear	3.03	4.09	7.12	391	53.47	—	—
	Slow	1.27	5.71	6.98	462	77.46	—	—
Female	Loud	1.58	0.99	2.57	127	37.95	4	19
	Clear	2.89	2.31	5.20	362	33.08	31	26
	Slow	0.99	4.37	5.36	458	85.51	19	54
Male	Loud	1.69	3.32	5.01	162	63.02	9	49
	Clear	3.20	6.26	9.46	426	62.92	33	33
	Slow	1.61	7.35	8.97	466	77.64	30	58

Note. Furthermore, the angle  $\theta$  (in degrees) and the relative contributions of the jaw and tongue to acoustic vowel contrast change (in %) for female and male talkers are presented.

concur with those of previous studies that have examined rate effects on tongue and jaw movements during sentence productions (e.g., Edwards et al., 1991) and in diadochokinetic tasks (e.g., Hertrich & Ackermann, 2000). These studies found that the displacement of the primary articulator (tongue, lower lip) increased in response to slow speech, whereas jaw displacements remained relatively stable (Hertrich & Ackermann, 2000) or decreased (Edwards et al., 1991).

Although a relatively large number of kinematic studies have examined articulatory displacement changes in response to loud speech (e.g., Darling & Huber, 2011; Dromey & Ramig, 1998; Huber & Chandrasekaran, 2006; Schulman, 1989), relative changes of the tongue and jaw have rarely been investigated during loudness manipulations. Tasko and McClean (2004), for example, reported that jaw displacements increased more consistently than decoupled tongue displacements across a variety of speech materials. We hypothesized that jaw displacements would increase to a greater extent than decoupled tongue displacements in our target utterance; however, this hypothesis could not be confirmed. Instead, increases in decoupled tongue displacement were comparable to those of the jaw. Schulman (1989) suggested that loud speech could be considered a natural perturbation, and the increases in decoupled tongue displacements may be adaptive responses similar to those previously observed during bite block perturbations. Because the notion of tongue displacement changes as adaptive responses can be best evaluated in the context of the vowel acoustic changes, we will come back to this notion in the next section. Finally, findings of fairly equal increases in decoupled tongue and jaw displacements in response to clear speech were predicted and are in congruence with previous findings by Tasko and Greilick (2010).

### Evaluating the Hypothetical Framework

The hypothetical framework was developed based on previous studies showing different magnitudes of change in vowel acoustic contrast in response to slow, loud, and

clear speech. Specifically, clear speech was expected to elicit greater increases in acoustic vowel contrast than slow and loud speech based on previous findings by Tjaden and colleagues (2013). Furthermore, based on previously observed strong linear associations between changes in acoustic vowel contrast and changes in tongue composite displacements (e.g., Mefferd & Green, 2010), clear speech was also expected to elicit greater increases in tongue composite movements than slow and loud speech. However, the tongue composite movements were almost identical for slow and clear speech, and acoustic vowel contrast tended to be larger during slow speech compared to clear speech ( $p = .055$ ). Methodological differences between previous acoustic studies and the current one may explain the discrepant acoustic findings. Specifically, the current study examined acoustic vowel contrast during diphthong productions, whereas Tjaden and colleagues (2013) examined acoustic vowel space changes based on monophthongs.

As hypothesized, distinctly different patterns of tongue- and jaw-specific changes were observed for slow speech and clear speech. That is, jaw displacements increased significantly more during clear speech, whereas decoupled tongue displacements increased significantly more during slow speech. Furthermore, as predicted, changes in acoustic vowel contrast were accounted for in task-specific ways for slow and clear speech. That is, acoustic changes during slow speech were predominantly tongue-driven, whereas those during clear speech were driven by both articulators. The difference between slow and clear speech was more pronounced in female talkers than in male talkers (angular displacement differed by  $52.43^\circ$  for slow and clear speech in women, whereas in men, the angular displacement between these two speech conditions only differed by  $14.72^\circ$ ). This suggests potential sex-specific interarticulatory performance patterns. However, because of the number of female and male talkers, these observations can only be considered preliminary. It is currently difficult to explain these sex-related differences in performance patterns. Anatomical differences in

the vocal tract size between female and male talkers may play a role.

Loud speech was expected to elicit predominantly increases in jaw displacement resulting in composite tongue movements that contain proportionally more jaw movements than decoupled tongue movements. Therefore, the tongue- and jaw-specific performance pattern for loud speech was expected to be distinct from those of clear and slow speech. To our surprise, this hypothesis was not confirmed. Loud speech and clear speech were, in fact, almost identical with regard to their angular displacement from the  $x$ -axis. However, loud and clear speech significantly differed in the magnitude of change in the displacement each task elicited.

Despite similar patterns of change in decoupled tongue and jaw displacements, acoustic findings suggest task-specific articulatory adjustments for loud and clear speech. As can be seen in Figure 6, F1 values of /i/ increased during loud speech but decreased during clear speech. Furthermore, F2 values of /i/ only increased considerably during clear speech. The lowered F1 as well as raised F2 values of /i/ during clear speech suggest “goal-directed” tongue elevation and advancement to enhance the vowel specificity of /i/ as demanded by the task (increased speech clarity). The upward shift of F1 in both vowels during loud speech, however, concurs with the findings of a lowered jaw position for both vowels in response to the increased vocal intensity demand (Schulman, 1989). It has been speculated that the jaw lowering across all vowels is implemented by the talker to reduce friction noise that may result from the increased subglottal and supraglottal pressure and airflow during loud speech (Schulman, 1989).

Interestingly, the increase in F1 values for both vowels was more common in women than in men, suggesting that male talkers compensated for their lowered jaw during /i/ by increasing decoupled tongue displacement, whereas female talkers did not. The larger angular displacement from the  $x$ -axis for loud speech in male talkers as well as the greater

amount of acoustic vowel contrast change accounted for by the decoupled tongue in male talkers support this notion. If a higher jaw position was responsible for the lower F1 value of /i/ in male talkers, then the increase in jaw displacement would have likely been greater in male talkers during loud speech.

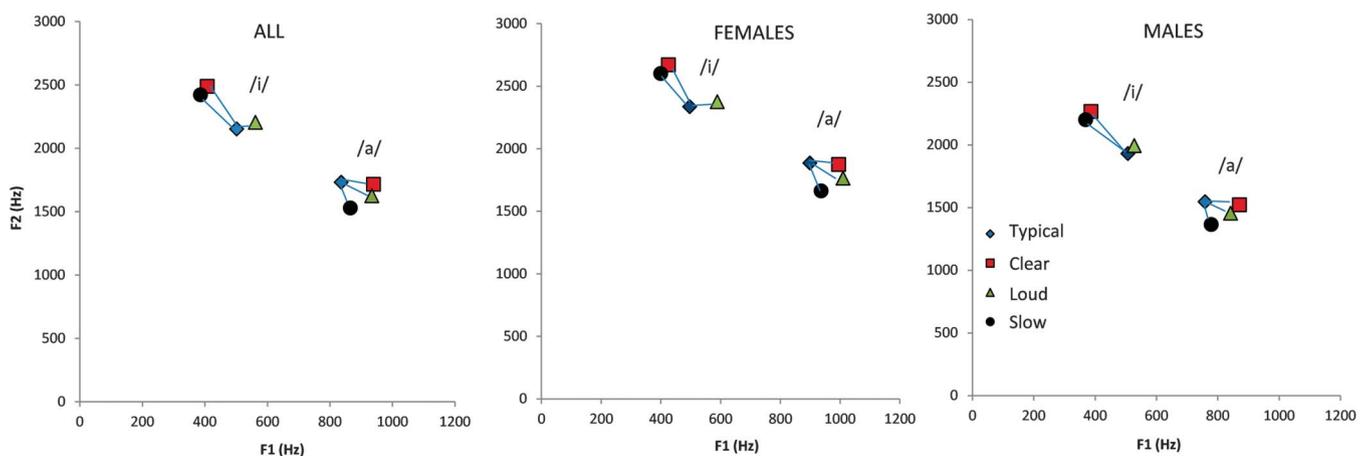
The observed differences in the F1–F2 pattern for /i/ across clear and loud speech are discrepant with findings by Tjaden and colleagues (2013). In their study, potential task-specific changes in vowel positions within the F1–F2 vowel space were examined by absolute angle measures as well as Euclidean distance measures. However, no significant differences were found in their study for the absolute angle measure across loud and clear speech. One reason for this discrepancy might be that acoustic measures were based on monophthongs in the study by Tjaden and colleagues, whereas diphthong productions were used in the current study.

### Task Performance

Although clear speech is commonly produced with increased loudness, as well as a slower speaking rate (e.g., Ferguson & Quené, 2014; Lam et al., 2012; Picheny, Durlach, & Braidá, 1986; Tjaden et al., 2013), the magnitude of change is much smaller in clear speech. In the current study, the relative change in vocal intensity in response to clear speech was similar to previously observed changes in vocal intensity during this speech modification (Tjaden et al., 2013). However, durational changes in response to clear speech were smaller compared to those observed in other studies. For example, Tasko and Greilick (2010) reported that formant transition duration increased by approximately 50 ms during clear speech. In the current study, however, movement duration from /a/ to /i/ only increased on average by 26 ms.

Increases in vocal intensity during loud speech were comparable to those previously reported by Darling and Huber (2011) and Tjaden et al. (2013). Finally, durational

**Figure 6.** Task-dependent changes of /a/ and /i/ positions in the F1–F2 vowel space. Mean F1 and F2 values for all talkers (a), female talkers (b), and male talkers (c) are provided. Error bars are omitted to improve readability of the figure. Blue lines indicate change from typical speech to the specific speech modification.



changes in response to slow speech were difficult to compare to previous studies, because in this study we used duration of tongue displacement during the production of /ai/, whereas other studies have used other approaches to quantify the rate changes (e.g., syllables per second, sentence durations). However, the durational changes of the movement transitions in the current study indicate that talkers followed the task instructions and approximately doubled diphthong durations during slow speech.

### **Clinical Implications**

Although currently it is unknown how tongue and jaw displacements change in response to slow, loud, and clear speech in talkers with dysarthria, findings from the current study suggest that speech modifications that maximize decoupled tongue displacement in impaired talkers may increase acoustic vowel contrast, at least for /a/ and /i/, more than those that elicit predominantly jaw displacement changes. Increased acoustic vowel contrast is associated with improved speech intelligibility (e.g., Connaghan & Patel, 2017; Kim, Hasegawa-Johnson, & Perlman, 2011; Turner et al., 1995); hence, increased acoustic vowel contrast is highly desirable for these talkers.

It is also clinically relevant that changes in tongue composite movements did not parallel changes in acoustic vowel contrast in response to slow and clear speech. Specifically, acoustic vowel contrast tended to be smaller during clear speech than during slow speech, whereas tongue composite movements were rather similar for the two speech conditions. Such discrepancies between task effects on tongue kinematics and vowel acoustics were particularly evident in female talkers and may be due to the difference in the task-specific tongue–jaw performance pattern for slow and clear speech. This suggests that the strength of association between changes in tongue composite displacement and changes in acoustic vowel contrast may become weaker as the relative contribution of the jaw to the tongue composite movement increases. One explanation for such differences in articulatory-to-acoustic mappings may be that the decoupled tongue can achieve more refined vocal tract configurations for high vowels such as /i/ compared to the jaw. Therefore, predominantly jaw-driven changes in tongue composite movements may result in less distinct vocal tract configurations than predominantly tongue-driven changes in tongue composite movements. This observation is particularly relevant for talkers with dysarthria who deal predominantly with tongue-related motor impairments and may move their jaw to a greater extent than typical talkers during speech production (e.g., talkers with amyotrophic lateral sclerosis).

### **Limitations**

In this study, the focus was exclusively on the relative displacement change during one diphthong. Other speech materials should be evaluated in the future. However, the diphthong /ai/ was well suited as a first step because it involved minimal lip rounding and potential task-specific

trade-offs between lip rounding and tongue back-raising in addition to jaw contributions. Furthermore, tongue- and jaw-specific contributions to changes in acoustic vowel contrast of /ai/ should be studied in other phonetic contexts in future studies to determine how generalizable the findings of the current study are for the diphthong /ai/ across a variety of phonetic contexts.

Despite all efforts to produce most accurate estimates of decoupled tongue and jaw displacements, the execution of a decoupling algorithm will always introduce a certain degree of error. Decoupled tongue movements were negative in six out of 400 raw values with the maximum negative value being  $-0.961$  mm (see Figure 4). The decoupled movement of the tongue should not, however, yield a negative value. The comparison of the measured and estimated jaw movement at the location of the posterior tongue sensor in one typical talker suggested that the decoupling algorithm used in this study performed similarly to those previously used for 2D kinematic data as well as those more recently developed for 3D kinematic data (Henriques & van Lieshout, 2013). A close inspection of the cases that yielded negative decoupled tongue values revealed no indication of any aberrant movement patterns that could explain the results. In three instances, negative independent tongue movements were noted during typical speech; in one instance, it was observed during clear speech, and in two instances, this occurred during loud speech. There is no reason to believe that these estimates of decoupled tongue displacements are any less accurate than those that were positive. Furthermore, statistical findings do not change whether or not these data points are included or removed. Therefore, negative decoupled tongue displacement values were not removed from the data set.

Finally, speech conditions were not counterbalanced or randomized across participants to control for a potential order effect. However, durational and intensity changes observed in response to speech modifications in the current study were congruent with those reported in previous studies where speech conditions were controlled for potential order effects (Darling & Huber, 2011; Tjaden et al., 2013). Nevertheless, the order of the speech modifications should be counterbalanced or randomized in future studies to improve scientific rigor.

### **Summary and Future Directions**

Outcomes of this study indicate that talkers increase decoupled tongue displacement significantly more during slow speech than during clear and loud speech, whereas increases in jaw displacement are comparable to those of loud speech and smaller than those of clear speech. Furthermore, decoupled tongue displacements predominantly account for changes in acoustic vowel contrast during slow speech. In addition, findings of this study suggest that both decoupled tongue and jaw displacements increase significantly more during clear speech than during loud speech; however, both speech modifications result in similar tongue–jaw performance patterns underlying tongue composite

movement. However, vowel positions in the F1–F2 vowel space indicate that the increase in decoupled tongue displacement and jaw displacement during clear speech results in increased vowel distinctiveness, particularly during /i/, whereas those during loud speech predominantly increase F1 in both vowels and result in relatively small increases in acoustic vowel contrast. Future studies are warranted to investigate articulator-specific changes that occur in response to these three speech behavioral modifications in talkers with dysarthria. Such insights will improve our knowledge about the articulator-specific mechanisms that underlie vowel acoustic changes in these talkers. Because these changes in acoustic vowel contrast are associated with changes in speech intelligibility, such insights will help elucidate articulator-specific contributions to intelligibility loss and recovery and will even inform clinical decisions on treatment selection.

## Acknowledgments

This research was supported by start-up funds from Vanderbilt University Medical Center and Grant R03DC015075 from the National Institute on Deafness and Other Communication Disorders, awarded to the author. I would like to thank Brett Myers, Ellen Hart, Sophie Mouros, Jaclyn Fitzsimmons, Mary Jo Bissmeyer, and Randy Hiroshige for their assistance with data collection and analysis. Special thanks also to my colleague Daniel Ashmead for his input on the decoupling algorithm and Kris Tjaden for inspiring conversations and suggestions throughout this project. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

## References

- Carstens Medizinelektronik GmbH.** (2014). CalcPos. [Computer Software]. Bovenden, Germany: AG501.
- Carstens Medizinelektronik GmbH.** (2014). NormPos. [Computer Software]. Bovenden, Germany: AG501.
- Chung, H., Kong, E. J., Edwards, J., Weismer, G., & Fourakis, M.** (2012). Cross-linguistic studies of children's and adult's vowel spaces. *The Journal of the Acoustical Society of America*, *131*(1), 442–454. <https://doi.org/10.1121/1.3651823>
- Connaghan, K., & Patel, R.** (2017). The impact of contrastive stress on vowel acoustics and intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research*, *60*(1), 38–50. [https://doi.org/10.1044/2016\\_JSLHR-S-15-0291](https://doi.org/10.1044/2016_JSLHR-S-15-0291)
- Darling, M., & Huber, J. E.** (2011). Changes to articulatory kinematics in response to loudness cues in individuals with Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, *54*, 1247–1259. [https://doi.org/10.1044/1092-4388\(2011\)10-0024](https://doi.org/10.1044/1092-4388(2011)10-0024)
- Dromey, C., & Ramig, L. O.** (1998). Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. *Journal of Speech, Language, and Hearing Research*, *41*, 1003–1018. <https://doi.org/10.1044/jslhr.4105.1003>
- Edwards, J., Beckman, M. E., & Fletcher, J.** (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America*, *89*, 369–382. <https://dx.doi.org/10.1121/1.400674>
- Ferguson, S. H., & Kewley-Port, D.** (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*, 1241–1255. [https://doi.org/10.1044/1092-4388\(2007\)087](https://doi.org/10.1044/1092-4388(2007)087)
- Ferguson, S. H., & Quené, H.** (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *135*(6), 3570–3584. <https://doi.org/10.1121/1.4874596>
- Gay, T.** (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America*, *63*(1), 223–230. Retrieved from <http://asa.scitation.org/journal/jas>
- Green, J. R., Wang, J., & Wilson, D. L.** (2013). SMASH: A tool for articulatory data processing and analysis. *Interspeech*, 1331–1335.
- Henriques, R., & van Lieshout, P.** (2013). A comparison of methods for decoupling tongue and lower lip from jaw movements in 3D articulography. *Journal of Speech, Language, and Hearing Research*, *56*, 1503–1516. [https://doi.org/10.1044/1092-4388\(2013\)12-0016](https://doi.org/10.1044/1092-4388(2013)12-0016)
- Hertrich, I., & Ackermann, H.** (2000). Lip–jaw and tongue–jaw coordination during rate-controlled syllable repetitions. *The Journal of the Acoustical Society of America*, *107*(4), 2236–2247. <https://dx.doi.org/10.1121/1.428504>
- Huber, J. E., & Chadrasekanan, B.** (2006). Effects of increased sound pressure level on lower lip and jaw movements. *Journal of Speech, Language, and Hearing Research*, *21*(2), 173–187. [https://doi.org/10.1044/1092-4388\(2006\)098](https://doi.org/10.1044/1092-4388(2006)098)
- Kim, H., Hasegawa-Johnson, M., & Perlman, A.** (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, *63*(4), 187–194. <https://doi.org/10.1159/000318881>
- Kuehn, D. P., & Moll, K. L.** (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, *4*, 303–320. Retrieved from <https://www.journals.elsevier.com/journal-of-phonetics/>
- Lam, J., & Tjaden, K.** (2013). Acoustic-perceptual relationships in variants of clear speech. *Folia Phoniatrica et Logopaedica*, *65*(3), 148–153. <https://doi.org/10.1159/000355560>
- Lam, J., Tjaden, K., & Wilding, G.** (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, *55*, 1807–1821. [https://doi.org/10.1044/1092-4388\(2012\)11-0154](https://doi.org/10.1044/1092-4388(2012)11-0154)
- Lee, J., Shaiman, S., & Weismer, G.** (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America*, *139*(1), 426–440. <https://doi.org/10.1121/1.4939894>
- Leung, K., Jongman, A., Wang, J., & Sereno, Z.** (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America*, *140*(1), 45–58. <https://doi.org/10.1121/1.4954737>
- Lindblom, B.** (1963). A spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, *35*, 1773–1781. <https://doi.org/10.1121/1.2142410>
- Lindblom, B.** (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht, the Netherlands: Kluwer.
- Mefferd, A. S.** (2015). Articulatory-to-acoustic relations in talkers with dysarthria: A first analysis. *Journal of Speech, Language, and Hearing Research*, *58*, 576–589. [https://doi.org/10.1044/2015\\_JSLHR-S-14-0188](https://doi.org/10.1044/2015_JSLHR-S-14-0188)
- Mefferd, A. S., & Green, J. R.** (2010). The articulatory-to-acoustic relationship in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1206–1219. [https://doi.org/10.1044/1092-4388\(2010\)09-0083](https://doi.org/10.1044/1092-4388(2010)09-0083)

- Milenkovic, P.** (2005). Time–frequency analysis software program for 32-bit Windows [Computer software]. Retrieved from <http://userpages.chorus.net/cspeech/>
- Moon, S.-J., & Lindblom, B.** (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, *96*, 40–55.
- Ostry, D., Vatiskiotis-Bateson, E., & Gribble, P.** (1997). An examination of the degrees of freedom of human jaw motion in speech and mastication. *Journal of Speech, Language, and Hearing Research*, *40*, 1341–1351. <https://doi.org/10.1044/jslhr.4006.1341>
- Perkell, J., & Zandipour, M.** (2002). Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements. *The Journal of the Acoustical Society of America*, *112*(4), 1642–1651. <https://dx.doi.org/10.1121/1.1506368>
- Picheny, M., Durlach, N., & Braida, L. D.** (1986). Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434–445. <https://doi.org/10.1044/jshr.2904.434>
- Schulman, R.** (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America*, *85*(1), 295–312. Retrieved from <http://asa.scitation.org/journal/jas>
- Tasko, S., & Greilick, K.** (2010). Acoustic and articulatory features of diphthong productions: A speech clarity study. *Journal of Speech, Language, and Hearing Research*, *53*, 84–99. [https://doi.org/10.1044/1092-4388\(2009/08-0124\)](https://doi.org/10.1044/1092-4388(2009/08-0124))
- Tasko, S. T., & McClean, M. D.** (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research*, *47*, 85–100. [https://doi.org/10.1044/1092-4388\(2004/008\)](https://doi.org/10.1044/1092-4388(2004/008))
- Tjaden, K., Lam, J., & Wilding, G.** (2013). Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of Speech, Language, and Hearing Research*, *56*, 1485–1502. [https://doi.org/10.1044/1092-4388\(2013/12-0259\)](https://doi.org/10.1044/1092-4388(2013/12-0259))
- Tjaden, K., & Wilding, G.** (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research*, *47*, 766–783. [https://doi.org/10.1044/1092-4388\(2004/058\)](https://doi.org/10.1044/1092-4388(2004/058))
- Turner, G. S., Tjaden, K., & Weismer, G.** (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, *38*(5), 1001–1013. <https://doi.org/10.1044/jshr.3805.1001>
- Wang, J., Samal, A., Rong, P., & Green, J. R.** (2016). An optimal set of flesh points on tongue and lips for speech-movement classification. *Journal of Speech, Language, and Hearing Research*, *59*, 15–26. [https://doi.org/10.1044/2015\\_JSLHR-S-14-0112](https://doi.org/10.1044/2015_JSLHR-S-14-0112)
- Westbury, J. R., & Dembowski, J.** (1993). Articulatory kinematics of normal diadochokinetic performance. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, *27*, 13–36. Retrieved from <http://www.umin.ac.jp/memorial/rilp-tokyo/>
- Westbury, J. R., Lindstrom, M., & McClean, M.** (2002). Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research*, *45*, 651–662. [https://doi.org/10.1044/1092-4388\(2002/052\)](https://doi.org/10.1044/1092-4388(2002/052))
- Yorkston, K., Hakel, M., Beukelman, D., & Fager, S.** (2007). Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review. *Journal of Medical Speech-Language Pathology*, *15*(2), 11–36. Retrieved from [https://www.pluralpublishing.com/journals\\_JMSLP.htm](https://www.pluralpublishing.com/journals_JMSLP.htm)