
Acoustic and Articulatory Features of Diphthong Production: A Speech Clarity Study

Stephen M. Tasko

Western Michigan University

Kristin Greilick

Sydney, Australia

Purpose: The purpose of this study was to evaluate how speaking clearly influences selected acoustic and orofacial kinematic measures associated with diphthong production.

Method: Forty-nine speakers, drawn from the University of Wisconsin X-Ray Microbeam Speech Production Database (J. R. Westbury, 1994), served as participants. Samples of clear and conversational productions of the word *combine* were extracted for analysis. Analyses included listener ratings of speech clarity and a number of acoustic and articulatory kinematic measures associated with production of the diphthong /aɪ/.

Results: Key results indicate that speaking clearly is associated with (a) increased duration of diphthong-related acoustic and kinematic events, (b) larger F1 and F2 excursions and associated tongue and mandible movements, and (c) minimal evidence of change in formant transition rate.

Conclusions: Overall, the results suggest that clarity-related changes in diphthong production are accomplished through larger, longer, but not necessarily faster diphthong-related transitions. The clarity-related adjustments in diphthong production observed in this study conform to a simple model that assumes speech clarity arises out of reduced overlap of articulatory gestures.

KEY WORDS: clear speech, diphthong, acoustics, speech movement, gesture theory

A broad goal of speech research has been to provide a principled account of the variation observed in the physical processes associated with sound production. An important source of variation is the adaptation that speakers make to maintain speech clarity when conversing in noisy environments, over weak mobile phone connections, or with persons who suffer from hearing loss. Beyond the practical motivations of making everyday communication easier, studying clarity-related variation can also be used to address a range of basic questions about the underlying organization of the speech motor control system. Those acoustic and physiologic features of a specific sound class that systematically vary with speech clarity are more likely to play a primary role in the planning, execution, and perception of that sound (e.g., Ferguson & Kewley-Port, 2002). In addition, speech clarity variation might also serve to reveal more global organizational features of the speech motor system (Bradlow, 2002; Matthies, Perrier, Perkell, & Zandipour, 2001; Perkell, Zandipour, Matthies, & Lane, 2002).

These practical and theoretical issues have led to a growing corpus of speech clarity studies. As a whole, the studies suggest that speaking

clearly involves adjustment at the segmental and suprasegmental levels. A ubiquitous finding is that clear speech is slower than conversational speech, which is due to a combination of increased articulation and pause time (Bond & Moore, 1994; Cutler & Butterfield, 1990, 1991; Ferguson & Kewley-Port, 2002, 2007; Liu & Zeng, 2006; Moon & Lindblom, 1994; Picheny, Durlach, & Braidá, 1986). Clear speech has also been reported to be produced with greater overall intensity levels than conversational speech, although intensity adjustments may be greater for sounds such as obstruents (Picheny et al., 1986).

Studies examining general features of clear speech have noted that vowels tend to have longer durations and an expanded formant space (Bond & Moore, 1994; Picheny et al., 1986). More recent studies of clarity-related changes in vowels have confirmed and expanded on these basic findings. Bradlow (2002) studied 12 monolingual English and nine bilingual Spanish–English talkers (speaking English and Spanish) as they produced various consonant–vowel (CV) combinations embedded in words during conversational and clear speaking conditions. Both English and Spanish speakers showed similar consonant context effects on vowel formants for both clear and conversational speech even though the clearly produced vowels were much longer in duration. Bradlow suggested the maintenance of coarticulation in the clear condition was intentional and served listener-oriented objectives. Second, there was a clarity-related increase in the size of the vowel formant space regardless of language. Vowel space expansion in both the crowded English and the sparser Spanish vowel space was interpreted as support for the notion that speakers apply a global clear speech strategy that also includes segments unlikely to be confused.

In a series of studies, Ferguson (Ferguson, 2004; Ferguson & Kewley-Port, 2002, 2007) evaluated acoustic and intelligibility features of vowels produced during clear and conversational conditions. Using data from a single talker, Ferguson and Kewley-Port (2002) found that during clear speech, back and front vowels were produced with lower and higher second formant (F2) values respectively, whereas all vowels showed an elevated first formant (F1), presumably due to a lower jaw position. The authors noted that only those vowels in more crowded regions of the vowel space showed more dynamic vowel formants in the clear condition, suggesting that such transformations are not global, but vowel specific. Ferguson (2004) examined the degree of talker variability for clear and conversational speech by developing a data set of 41 male and female speakers producing English vowels in a /bVd/ context. Vowel intelligibility varied widely across the talker group for both clear and conversational productions. An interesting result was

that not all speakers showed an obvious clear speech improvement in vowel intelligibility. Intelligibility improvement ranged from -12% to $+33\%$, indicating that an instruction to speak clearly does not assure a talker will effect an intelligibility improvement, at least for vowels. Ferguson and Kewley-Port (2007) followed up this work with a study aimed at evaluating the acoustic factors that differentiate speakers with a large, clear speech benefit from those with no clear speech benefit. Using the Ferguson (2004) database, the authors compared six talkers who showed no clear speech benefit with six talkers who showed a large, clear speech benefit on a number of acoustic measures. Results indicated that the large benefit group showed larger increases in vowel duration, a greater expansion of steady-state vowel space, a greater overall increase in F1, and higher F2 values for front vowels. Two measures of vowel formant dynamics did not differ for the large-benefit and no-benefit groups.

Comparatively little is known about the effect of speech clarity on diphthong production. Given that diphthongs involve a marked articulatory and acoustic transition (Gay, 1970), it is unclear how findings obtained for monophthongs would generalize to diphthongs. Wouters and Macon (2002) is one of the only studies we are aware of that specifically evaluated clarity-related adjustments to diphthong production. The authors used a spectral rate of change measure based on the first three formants to compare liquid-vowel, vowel-liquid, and diphthong transitions produced by a single talker across range of speech conditions, including clear speech. The authors found that for liquid transitions, spectral rate of change was greater for clear speech. This result was similar to Moon and Lindblom's (1994) finding for clearly produced glides, although a different approach for measuring transition rate was used. In contrast to the liquid transition, Wouters and Macon found that clearly produced diphthongs showed a reduced spectral rate of change. These findings suggest that clarity-related adjustments in acoustic transitions may be phoneme specific.

Given that one of the most common findings in clear speech studies is a reduced speaking rate, it may be helpful to review how speaking rate influences diphthong production. Gay (1968) examined the formant characteristics of a range of diphthongs produced by five adult men in a variety of CVC contexts at fast, conversational, and slow speaking rates. He found that for a given diphthong, the F1 and F2 values at transition onset and the F2 transition rate were minimally affected by speaking rate. However, the formant offset values were clearly rate dependent. Slow speech rate was associated with more extreme F1 and F2 offset values, whereas for the fast rate, there appeared to be a truncation of the formant transitions.

Weismer (1991) also examined rate-induced variations in diphthong production and provided clear evidence that the duration and extent of diphthong-related formant transitions were not independent. On the basis of the analysis of three healthy adult talkers speaking at slow, habitual, and fast rates, Weismer found a strong positive relationship between F2 transition duration and extent for the diphthong /aI/. However, the best fit line for the duration-extent relationship was quadratic, suggesting that shifting from fast to slow rate may not be a strictly linear process.

Tjaden and Weismer (1998) tested the hypothesis that rate modification may involve a relatively simple process of varying the degree of overlap in the core articulatory gestures used to form the utterance. Fast-speaking rates would involve more gestural overlap, whereas slow-speaking rates would be achieved by reducing the amount of overlap. With such a strategy, scaling adjustments to individual gestures would be unnecessary because the overlapping of gestures would serve to truncate or expand the amount of gesture realized in production. Using this framework, Tjaden and Weismer argued that for CV contexts, formant values at vowel onset would be closer to theoretical locus values for the consonant at slow rates (due to less overlap) and farther from locus values for fast rates (due to greater overlap). Results from eight speakers across a range of CV conditions (including the diphthong /aI/) produced along a rate continuum provided only partial support for the hypothesis. A wide range of speaker variation was observed, making generalization difficult. In particular, the success of the model for predicting the formant features of diphthongs was highly variable.

Weismer and Berry (2003) evaluated second formant patterns for a number of vowels, including the diphthong /oI/, as six healthy talkers recited /bVC/ words in a carrier phrase across a continuum of speaking rates. Two results relevant to the present study were found. First, the F2 frequency of the schwa vowel preceding the target word had a significant negative relation to target vowel duration for monophthongs but not for the diphthong /oI/. This relation between preconsonantal vowel formant values and postconsonantal vowel duration was interpreted as evidence for rate-induced variations in anticipatory coarticulation. More specifically, slowed speech was associated with reduced anticipatory coarticulation. Second, attempts to use relatively simple temporal and spectral rescaling approaches failed to account for rate-induced variations in /oI/ formant patterns. In other words, there does not appear to be a simple rate transform to which all speakers adhere. For example, some speakers varied the duration of the initial formant steady-state, whereas others varied the maximum rate of change of the formant transition.

Overall, the results of speaking rate studies that have included diphthongs indicate that changes in durational aspects of the sound are typically accompanied by changes in the duration and extent of the acoustic transition, and, although the form of this transition may be systematic (Weismer, 1991), it does not necessarily conform to simple speaker-wide rules (Weismer & Berry, 2003).

The goals of this study were threefold. The first goal was to describe how speech clarity impacts the acoustic and articulatory kinematic features of diphthong production in a large group of healthy speakers. A second goal was to compare these clarity results with published speech rate data to determine whether the two speech modes operate in a similar way for diphthong production. A third goal was to attempt to account for clarity changes for diphthongs using a relatively simple model of gestural overlap.

Method Speakers

Speakers were selected from the University of Wisconsin X-ray Microbeam Speech Production Database (XRMB-SPD; Westbury, 1994). This publicly available database includes the acoustic signal and synchronous midsagittal-plane motions of eight fleshpoints recorded from 57 neurologically and communicatively healthy, native English-speaking young adults performing a range of speech and nonspeech oral activities. Not all speakers could be used in this study. For technical reasons, most speakers in the data set have some missing data. In some cases, a particular record was not recorded. In other cases, a record exists, but a portion of the data was not usable. Each speaker's data set was inspected to establish that the records containing the relevant speech samples exist and to ensure that the records of interest were relatively complete. Only those speakers whose records of interest were either absent or contained large amounts of missing data were excluded from the study. This process yielded 49 speakers. The participant pool was weighted toward female speakers (29 women and 20 men). The median speaker age was 20.9 years (range = 18.3–37.0 years). The vast majority of speakers spent their early years in the midwestern states, so it might be reasonably assumed that most speak an upper midwest dialect of American English. However, dialect was not formally assessed. Additional details of chosen speakers' demographic and physical characteristics can be found in the XRMB-SPD Handbook (Westbury, 1994).

Data acquisition and processing. Briefly, the x-ray microbeam system records articulator motion by directing

a narrow high-energy x-ray beam to track the midsagittal position of a number of 2- to 3-mm diameter gold pellets glued to various structures within and around the oral cavity. A total of eight pellets were recorded. Four pellets were positioned on the midline surface of the tongue. Two pellets were positioned on the mandible at the gum line between the central incisors and between the first and second molar on the left side. The final two pellets were positioned midline at the vermilion border of the upper and lower lips. Pellets were sampled at rates ranging from 40 to 160 Hz, and the simultaneous sound pressure level signal was recorded at 22 kHz. Following acquisition, the position histories for each pellet underwent a series of processing steps. One of the processing steps involved reexpressing the data in an anatomically based Cartesian coordinate system in which the horizontal (or x) axis is located along the maxillary occlusal plane, and the vertical (or y) axis is orthogonal to the x -axis where the central maxillary incisor meets the maxillary occlusal plane. The position histories of the pellets were low-pass filtered at 10 Hz and resampled at 145 Hz. For a complete description of the data acquisition and processing steps, see Westbury (1994). Pellet motion reflects movement of discrete flesh-points within and around the vocal tract. For the present study, analysis was limited to motion of two fleshpoints on the tongue (hereafter T2 and T3) and a single fleshpoint on the mandibular incisor (hereafter MI).

Speech tasks. Speakers performed an oral reading of the sentence “combine all the ingredients in a large bowl” under clear and conversational conditions. To elicit the clear condition, speakers were directed to “repeat this sentence five times, very distinctly and clearly, as if you are trying to make someone understand you in a noisy environment. Do not pause between words.” Unless technically unfeasible, the first complete production of the test sentence was selected for analysis. Otherwise, the second complete production was used. The conversational production of the test sentence was not collected in isolation. Rather, the sentence was part of a long list of sentence-level stimuli that compose the task set. Therefore, for the conversational condition, the speaker was prompted to refer to the general instruction “... to recite the speech stimuli at a comfortable rate and loudness ...”

The word *combine*, which contains the diphthong /aI/, was extracted from the sentence for analysis. Because it was the initial word in the sentence, it was relatively easy to extract and retain acceptable auditory quality for use in a perceptual study. For each speaker, a single token of the clear and conversational productions of the word was selected for inclusion in the analysis. Using only a single replicate for analysis raises the possibility that it is not representative of the subject’s typical performance. However, this choice was made in order to maximize the number of participants and still

limit the duration of the listening study (described below) to a reasonable length.

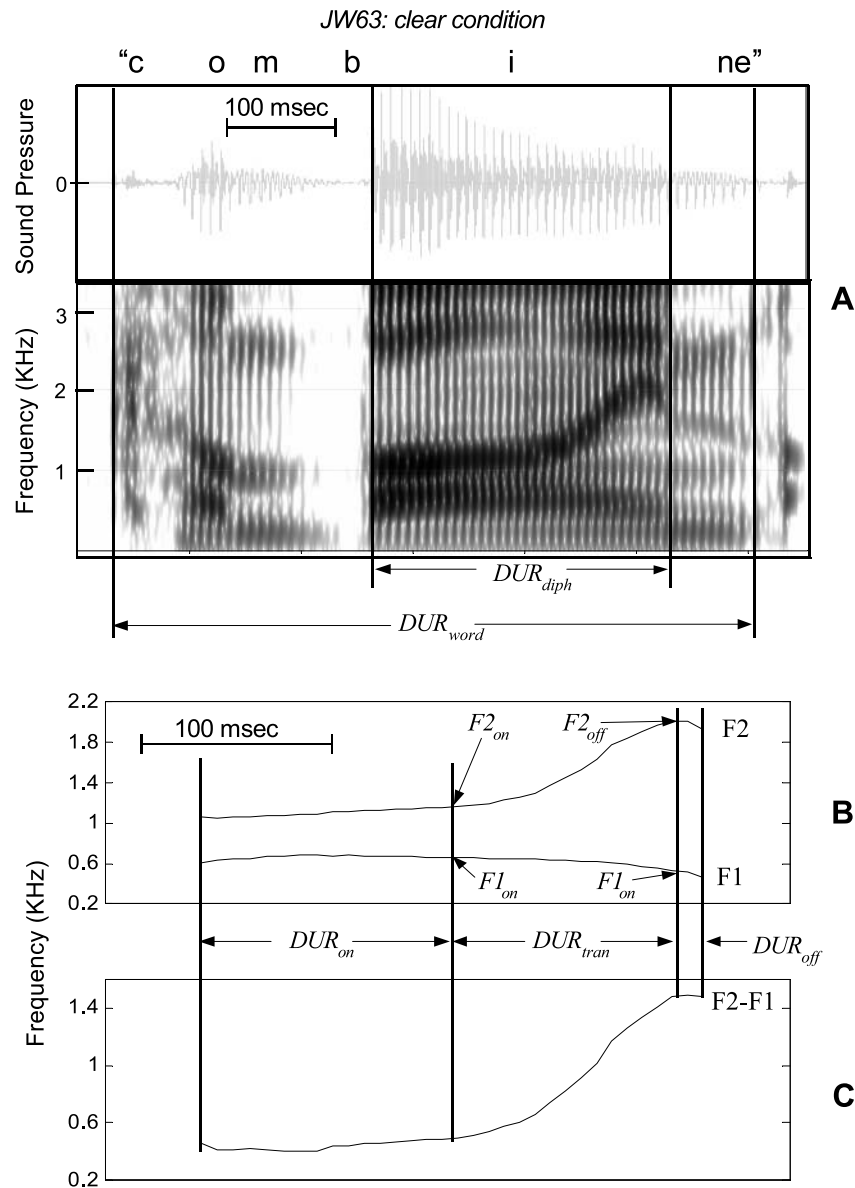
Acoustic, Articulatory Kinematic, and Auditory Perceptual Analysis

Acoustic analysis focused primarily on the diphthong /aI/ contained in the test word *combine*. First, the acoustic durations of the entire word (DUR_{word}) and the diphthong (DUR_{diph}) were measured using a combined waveform-spectrogram display of TF32, an acoustic software analysis package (Milenkovic, 2000). An example is shown in Figure 1A. *Word onset* and *word offset* were defined by the /k/-related release burst and the final glottal pulse of /n/, respectively. *Diphthong onset* was defined as the time of the first glottal pulse following plosive release. *Diphthong offset* was defined as the time of a sudden drop in signal amplitude associated with nasal production. Duration measures were made on two separate occasions by two different graduate students trained in acoustic phonetic segmentation. Interrater reliability was assessed by calculating a Pearson correlation between the two judges’ measures. For both duration measures, the correlation coefficient was greater than .99, suggesting a good correspondence between the judges’ measures.

Formant Estimation and Measurement

Once the temporal extent of the diphthong was identified, the first (F1) and second (F2) formant histories were estimated using a multistep process. *SpeechTool*, a locally developed software package (Gayvert & Hillenbrand, 2000), was used to generate a linear predictive coding (LPC)-based spectrum at 10-ms intervals for the entire duration of the diphthong. At each interval, all spectral peaks were automatically identified by the software program. At this point, no attempt was made to identify the peaks as corresponding to a particular formant. Following the peak-picking algorithm, the time histories of the spectral peaks were overlaid on an LPC-based spectrogram, and each token was examined visually and, if necessary, edited by Kristin Greilick. Those peaks that corresponded to the F1 and F2 regions in the spectrogram were extracted for further analysis. For example, Figure 1B shows the F1 and F2 values extracted from the signal in Figure 1A. These F1-F2 histories were used to determine the onset and offset frequencies, duration, extent, and rate of F1 and F2 transitions within the diphthong. There are a variety of ways to measure formant transitions. For example, Weismer and colleagues (e.g., Weismer, Kent, Hodge, & Martin, 1988) have popularized a method in which transition onset or offset are operationally defined as the start or end of a 20-Hz formant change over a 20-ms window. F1 and F2 are measured separately and, therefore, may have different onset and

Figure 1. Panel A shows the sound pressure waveform and wide-band spectrogram of the test word *combine*. The duration of the test word (DUR_{word}) and diphthong (DUR_{diph}) are marked. Panels B and C plot the F1-F2 history and the F1-F2 separation history for the diphthong. F1 and F2 transition onset was determined by at least a 20-Hz change in the formant separation history over 20 ms. F1 and F2 transition offset was identified when the transition in the separation history dropped below the 20-Hz change over 20-ms threshold. These time values were used to segment the formant history into a diphthong onset (DUR_{on}), diphthong transition (DUR_{tran}), and diphthong offset (DUR_{off}). F1 and F2 frequencies at transition onset and offset were also extracted. These values were used to determine the F1 and F2 transition extents ($F1_{extent}$, $F2_{extent}$) and rates ($F1_{rate}$, $F2_{rate}$).



offset times. This approach has been used with such regularity that it could be considered a measurement convention. We were interested in deriving single onset and offset time for the diphthong transition, while at the same time recognizing that both F1 and F2 frequencies exhibit substantial change during /aI/ production. To accomplish this, instead of measuring F1 and F2 change separately,

the difference between F1 and F2 was plotted, generating a formant separation history (see Figure 1C). Because the transition in /aI/ is characterized by an increase in F1-F2 spectral separation, this provides a single time history that takes into account both F1 and F2 frequency change. A custom written MATLAB routine was used to generate the formant separation history and its first-order time

derivative. *Diphthong transition onset* was defined as the onset of a sustained (threshold: 20 Hz over 20 ms) increase in F2-F1 velocity, whereas the transition offset was defined as the point when the velocity drops below this threshold.¹ Temporal measures included the duration of the diphthong-related formant transition (DUR_{tran}) and the duration of the diphthong preceding (DUR_{on}) and following (DUR_{off}) the primary formant transition. F1 and F2 values were extracted at transition onset ($F1_{on}$ and $F2_{on}$) and offset ($F1_{off}$ and $F2_{off}$). These variables were then used to derive the transition extent of F1 ($|F1_{off} - F1_{on}|$) and F2 ($|F2_{off} - F2_{on}|$) and rate of transition of F1 ($F1_{extent}/DUR_{tran}$) and F2 ($F2_{extent}/DUR_{tran}$).

Articulatory Kinematic Measures

Articulatory kinematic analysis focused on diphthong-related motion of the T2, T3, and MI fleshpoints. Figure 2A and 2B, respectively, plot the spectrogram and fleshpoint speeds associated with target word production. The speed history for each fleshpoint is the magnitude of the “x” and “y” velocity vectors ($[(dx/dt)^2 + (dy/dt)^2]^{1/2}$), which were determined using a 3-point central difference method of differentiation. Figure 2C plots the mid-sagittal, two-dimensional motion paths for each of the fleshpoints for the duration of the diphthong transition. Note that during the diphthong transition, the speed for each fleshpoint is generally characterized by an acceleration-deceleration sequence that is unimodal and bell-shaped. These speed histories were used to define the onset and offset of movement. Movements were identified as the period bounded by two successive minima in the speed history (Tasko & Westbury, 2002). Minima correspond to the moment-of-sign change in the first-order time derivative of the speed history (not shown). The filled and unfilled symbols in each speed history mark the onset and offset of the transition-related movement, respectively. These same symbols are used to mark the onset and offset in the fleshpoint motion paths in Figure 2C. Note that during the diphthong transition, the fleshpoints move in a superiorly oriented direction toward the speaker’s palate. Once the movement boundaries were determined, a number of kinematic measures were selected for analysis. The duration (DUR_{kin}), peak speed (SPD_{peak}), and path distance (DIS) were determined for each movement. In addition, the “x” and “y” position and fleshpoint speed were measured at movement onset and offset. The path distance covered by each movement was extracted from the time integral of the speed history.

¹Analysis was also performed using the more conventional method of identifying F1 and F2 formant transitions separately, popularized by Weismer et al. (1988). The results using this method were nearly identical to the data reported here except that when using the approach of Weismer and colleagues, the F2 transition rate was significantly reduced for the clear speech condition.

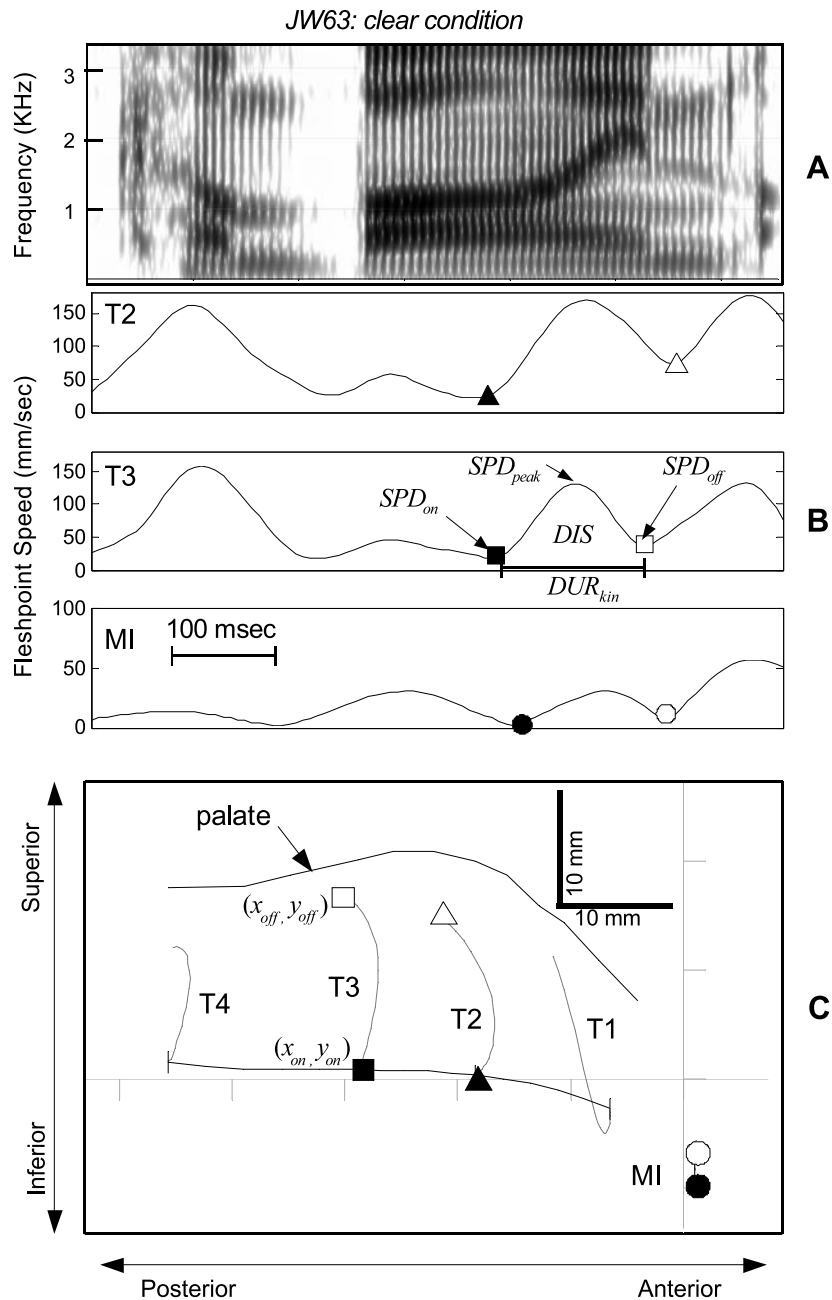
Auditory Perceptual Scaling of Speech Clarity

Ferguson (2004) provided evidence that individual talkers may be quite variable in the degree to which they attempt to produce clear speech. Therefore, an auditory-perceptual evaluation was conducted to attain a listener panel rating of the clarity difference between each speaker’s clear and conversational production of the test word. The listener panel consisted of 30 undergraduate and graduate students from the Western Michigan University Department of Speech Pathology and Audiology. All were native English speakers with normal hearing. Each listener performed the evaluation alone while seated in a quiet room in front of a PC computer, which randomly presented the speech stimuli through an amplifier (Realistic MPA-30) and loudspeaker (Paradigm Titan v.3). The PC computer used a 16-bit sound card to convert the digitized stimuli to analog waveforms. The loudspeaker was located about 100 cm from the listener, and signals were presented at a sound pressure level of 70–75 dBA. Stimulus presentation and response recording were controlled by Alvin (Hillenbrand & Gayvert, 2005), an open source, windows-based program for controlling listening experiments. Alvin allows the listener to control the rate of stimulus presentation and response. Each listener was presented a series of trials. Each trial contained a pair of clear and conversational productions of the test word produced by one of the speakers. The test words were separated by a 1,000-ms pause. While the stimulus pair was presented, the computer screen displayed a horizontally oriented “slider scale,” with the slider button positioned at the scale’s midpoint. Each end of the slider scale was labeled *Sample 1* (left side) and *Sample 2* (right side). The following instructions were read to the listeners prior to commencing the experiment.

You will be presented with a series of paired words read by a number of different speakers. For each presentation, the words will have differing degrees of clarity. Your task is (1) to indicate whether sample one or sample two is clearer, and (2) by how much. You will use the mouse to move the marker from the middle of the scale towards the presentation that is clearest. You will indicate the degree of clarity difference by how close you move the marker to the speech sample, the closer to either end, the larger the clarity difference. Try to make each choice match the clarity, as you perceive it.

Listeners were allowed to play each stimulus pair as many times as they wished. No practice trials were offered. A total of 196 signal pairs were presented (49 speakers presented four times). Both the trial order and the word order within each trial were randomized for each judge. The entire listening experiment lasted approximately 20–30 min.

Figure 2. Panel A shows the wide-band spectrogram of the test word. Panel B shows the T2, T3, and MI fleshpoint speed histories. Panel C shows the midsagittal two-dimensional motion paths of the tongue and MI fleshpoints. The horizontal axis represents the maxillary occlusal plane, and the vertical axis represents a line that is orthogonal to the horizontal axis at the central maxillary incisor. The solid and open symbols in the speed histories respectively mark the speed minima that were used to define movement onset and offset. The same symbols are used in panel C to mark the spatial location of the fleshpoints at movement onset and offset. Articulatory kinematic variables selected for this study are shown for the T3 fleshpoint.



The Alvin program recorded the listener response to each stimulus pair as an integer ranging from -500 to $+500$. For example, a score of -500 would indicate that the listener perceived the first sample to be much clearer than the second sample, in which a rating of $+500$ would

indicate that the listener perceived the second sample to be much clearer than the first sample. A rating of zero would indicate that the listener could not distinguish the clarity of the two samples. Given the random stimulus order, responses were rescaled so that ratings with

positive integers would always be associated with the stimulus in the clear conditions. For each stimulus pair (i.e., speaker), a listener rating was derived on the basis of the mean of that listener's four ratings. A panel rating was then derived on the basis of the mean rating across the 30 listeners. Interrater reliability was assessed using the intraclass correlation coefficient or ICC (2, k) (Shrout & Fleiss, 1979). The ICC (2, k) for the listener group was 0.97.

Statistical Analysis

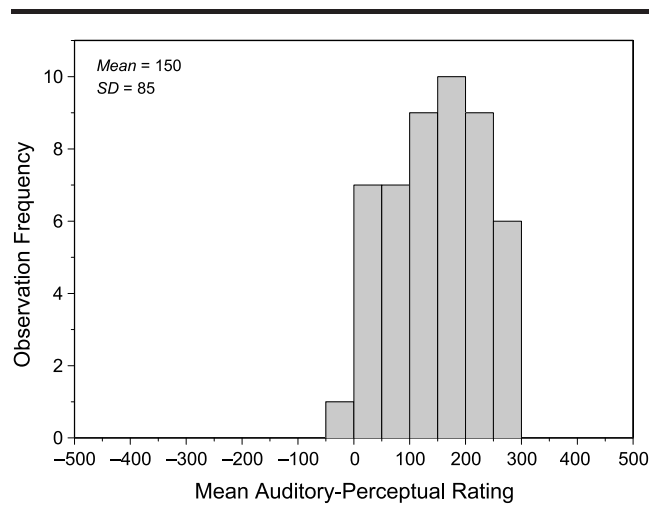
Comparisons between clear and conversational conditions were assessed using a series of individual repeated measures analyses of variance (ANOVAs). For acoustic measures, speech clarity was considered a within-subjects factor, and gender was treated as a between-subjects factor. For the articulatory kinematic measures, a more complex ANOVA model was used in which clarity and articulator fleshpoint identity served as within-subjects factors, and gender was treated as a between-subjects factor. Because the articulator factor contains more than two levels, a Huynh-Feldt correction was used to adjust the degrees of freedom. Gender was considered a factor in the statistical model because it is well known that males and females differ with regard to absolute formant values. In addition, Simpson (2001) has reported that males and females exhibit differences in articulatory kinematic measures during diphthong production. When all measures were tallied, a total of 22 separate ANOVAs were performed. In order to hold the experimentwide Type I error rate to less than 5%, the p values were adjusted for multiple comparisons ($.05/22 = p < .002$).

Results

Auditory Perceptual Ratings of Speech Clarity

Recall that the listener's task in the auditory perceptual experiment was to rate the difference in clarity between the conversationally and clearly produced test words for each of the speakers. Therefore, the listener may judge the clear test word to have greater clarity, equivalent clarity, or less clarity than the conversationally produced test word. Figure 3 plots a frequency histogram of the mean panel ratings for the speaker pool. It may be seen in the figure that the listener panel showed a distribution of mean ratings that are typically greater than zero, indicating that the clear productions were judged to have greater clarity than the conversational productions. The mean panel rating for the speaker group is 150 with a standard deviation of 85. We performed a one-sample t test to evaluate the hypothesis that the panel ratings of the speaker pool were statistically

Figure 3. Frequency histogram of mean panel rating for speaker pool. The ratings reflect the perceived difference in clarity between the conversational and clear test words. Positive mean ratings indicate the listener panel considered the clear token to sound clearer than the conversational token. Negative mean ratings indicate that the listener panel considered the conversational token to sound clearer than the clear token. Values near zero suggest no difference between the two tokens.



different from zero (i.e., no judged clarity difference). Results were significant, $t(48) = 12.29, p < .001$. However, it must be noted that the panel ratings for individual speakers varied widely, spanning from -10 to $+299$. Panel ratings for a number of speakers were near zero, suggesting minimal differences in perceived clarity of the clear and conversational productions.

Acoustic Results

The top panel of Figure 4 is a bar plot that compares the duration of test word, the diphthong within the test word, as well as the diphthong components for the clear and conversational speech conditions. The error bars reflect standard error. It can be observed that, with the exception of the diphthong offset duration, all segments are longer in the clear speech condition. Table 1 summarizes the results of a series of repeated measures ANOVAs performed on these data. The results support the observations in the top panel of Figure 4. There was neither a significant gender effect nor a significant Clarity \times Gender interaction. Both the word duration and the diphthong duration increased during clear speech. Did they increase to a similar degree? We completed a further analysis in which we compared clear and casual diphthong duration as a proportion of the word duration. During conversational speech, the diphthong occupied 45% of the total word duration, whereas during clear speech, it occupied 46% of the word. This similarity

Figure 4. The top panel is a bar plot comparing mean acoustic segment durations for clear and conversational conditions. The bottom panel is a bar plot comparing F1 and F2 onset and offset frequencies for clear and conversational conditions. Error bars reflect the standard error about the mean. * $p < .002$.

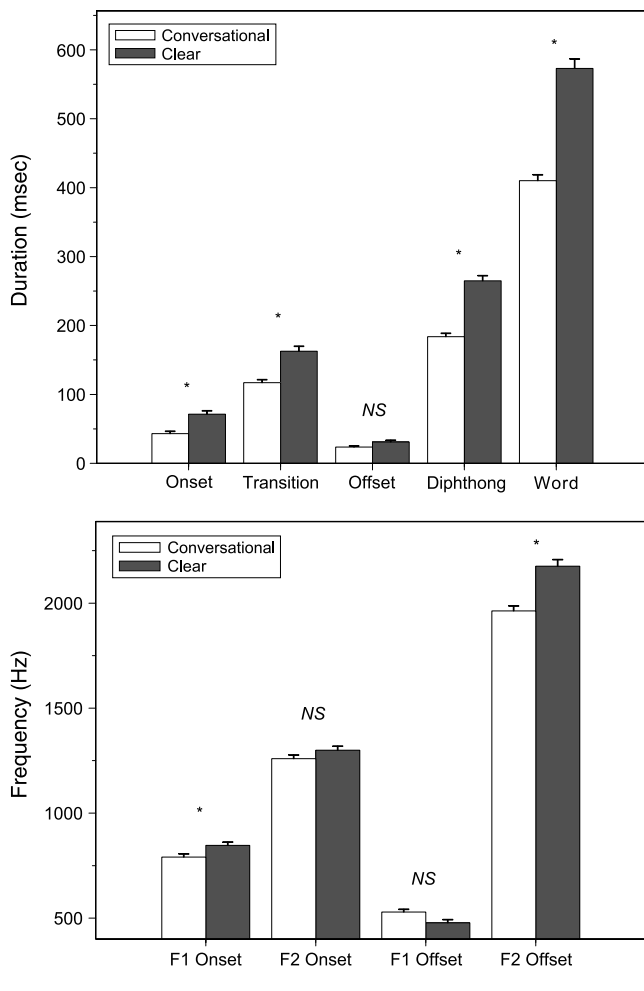


Table 1. Summary of repeated measures analyses of variance performed on temporal measures.

Factor	df	F values				
		DUR_{on}	DUR_{tran}	DUR_{off}	DUR_{diph}	DUR_{word}
Clarity	1	33.2*	70.2*	5.5	118.9*	152.1*
Gender	1	1.7	2.9	0.0	0.6	1.5
Clarity × Gender	1	0.0	3.2	2.1	0.8	3.9
error	47					

Note. DUR_{on} = duration of diphthong onset; DUR_{tran} = duration of diphthong transition; DUR_{off} = duration of diphthong offset; DUR_{diph} = duration of diphthong; DUR_{word} = duration of test word.

* $p < .002$.

suggests that the speaker group rescaled the word and diphthong duration at a similar amount.

Table 2 summarizes the ANOVA results for the transition-related measures of F1 and F2. First, the gender main effect was significant for many of the analyses, which is not surprising given the gender-related differences in the vocal tract size. More important for the goals of this study, there were no significant Clarity × Gender interactions. Therefore, the results reported below are for the speaker group as a whole. The left side of the table summarizes the F1 and F2 frequencies at transition onset and offset. Significant clarity main effects were only observed for F1 onset and F2 offset. These results are displayed graphically in the bottom panel of Figure 4. Here, it can be observed that F1 is significantly higher in the clear condition at transition onset, but there is no clarity-related difference at offset. F2, however, shows no clarity-related difference at transition onset but is significantly higher in the clear condition at transition offset. The right side of Table 2 highlights the analyses of the features of the formant transition. There was a significant clarity main effect for F1 and F2 extent. Both F1 and F2 showed significantly larger transition extents in the clear versus conversational speech condition. Finally, there were no clarity-related differences in the rate of F1 and F2 transitions, suggesting that the extent and duration of the transition scaled to roughly similar degrees.

Articulatory Kinematic Results

Table 3 summarizes the ANOVA results for the range of articulatory measures. The effects of most interest to the study questions are the speech clarity main effects and interactions between speech clarity and articulator type and gender. Although all of the articulator main effects were significant, this is a wholly expected finding given that each articulator fleshpoint has a unique position within the vocal tract and a different operating range (Tasko & Westbury, 2002) and therefore is not discussed further. There were no significant gender main effects or Clarity × Gender interactions. Therefore we collapsed men and women into one group for graphical display. Figure 5 summarizes the mean duration, distance, and peak speed of the transition-related movement for the conversational and clear speech conditions. The speaker group produced larger and longer articulatory movements in the clear speech condition. This is true for both tongue and mandible fleshpoints, although a significant Clarity × Articulator interaction for movement distance suggests that the clarity-related increase in movement distance was not equivalent for all fleshpoints. For peak fleshpoint speed, the clarity main effect was not significant, but there was a significant Clarity × Articulator interaction. Only the T3 fleshpoint showed a significant

Table 2. Summary of repeated measures analyses of variance performed on the formant transition measures.

Factor	df	F values							
		<i>F1_{on}</i>	<i>F2_{on}</i>	<i>F1_{off}</i>	<i>F2_{off}</i>	<i>F1_{extent}</i>	<i>F2_{extent}</i>	<i>F1_{rate}</i>	<i>F2_{rate}</i>
Clarity	1	11.6*	5.5	7.0	72.2*	13.9*	39.3*	0.5	8.0
Gender	1	76.2*	103.3*	23.6*	66.7*	15.1*	1.5	17.1*	5.7
Clarity × Gender	1	0.3	1.6	0.2	0.0	0.4	0.6	0.3	4.2
error	47								

Note. *F1_{on}* and *F2_{on}* = F1 and F2 onset; *F1_{off}* and *F2_{off}* = F1 and F2 offset; *F1_{extent}* and *F2_{extent}* = F1 and F2 transition extents; *F1_{rate}* and *F2_{rate}* = F1 and F2 transition rates.

**p* < .002.

increase in peak articulatory speed. There was not a significant clarity main effect for the speed at movement onset, but a significant clarity main effect and Clarity × Articulator interaction for speed at movement offset. The lower right panel of Figure 4 shows these results. Pairwise comparisons revealed that the speed of T2 at movement offset was significantly slower for clear speech.

Figure 6 plots the mean midsagittal positions of the three fleshpoints at the onset and offset of movement for the conversational and clear speech conditions. A line connects the onset and offset. This is not the actual trajectory path, but represents the Euclidean distance between the two points. The broken line marks the conversational speech condition, and the solid line marks

the clear speech condition. The ellipses that surround the onset and offset points reflect the standard error in each dimension. This plot combined with the results of Table 3 reveal some interesting patterns in the data. First, at movement onset, there was not a significant clarity main effect for horizontal fleshpoint position. For vertical fleshpoint position, the clarity main effect was significant. In Figure 6, it can be seen that all three fleshpoints begin the articulatory transition at a lower position in the clear speech condition. At transition offset, the clarity main effect and the Clarity × Articulator interaction was significant for both the horizontal and vertical fleshpoint positions. Figure 6 reveals that both tongue fleshpoints were higher and more anterior in the clear speech condition. MI did not show this pattern.

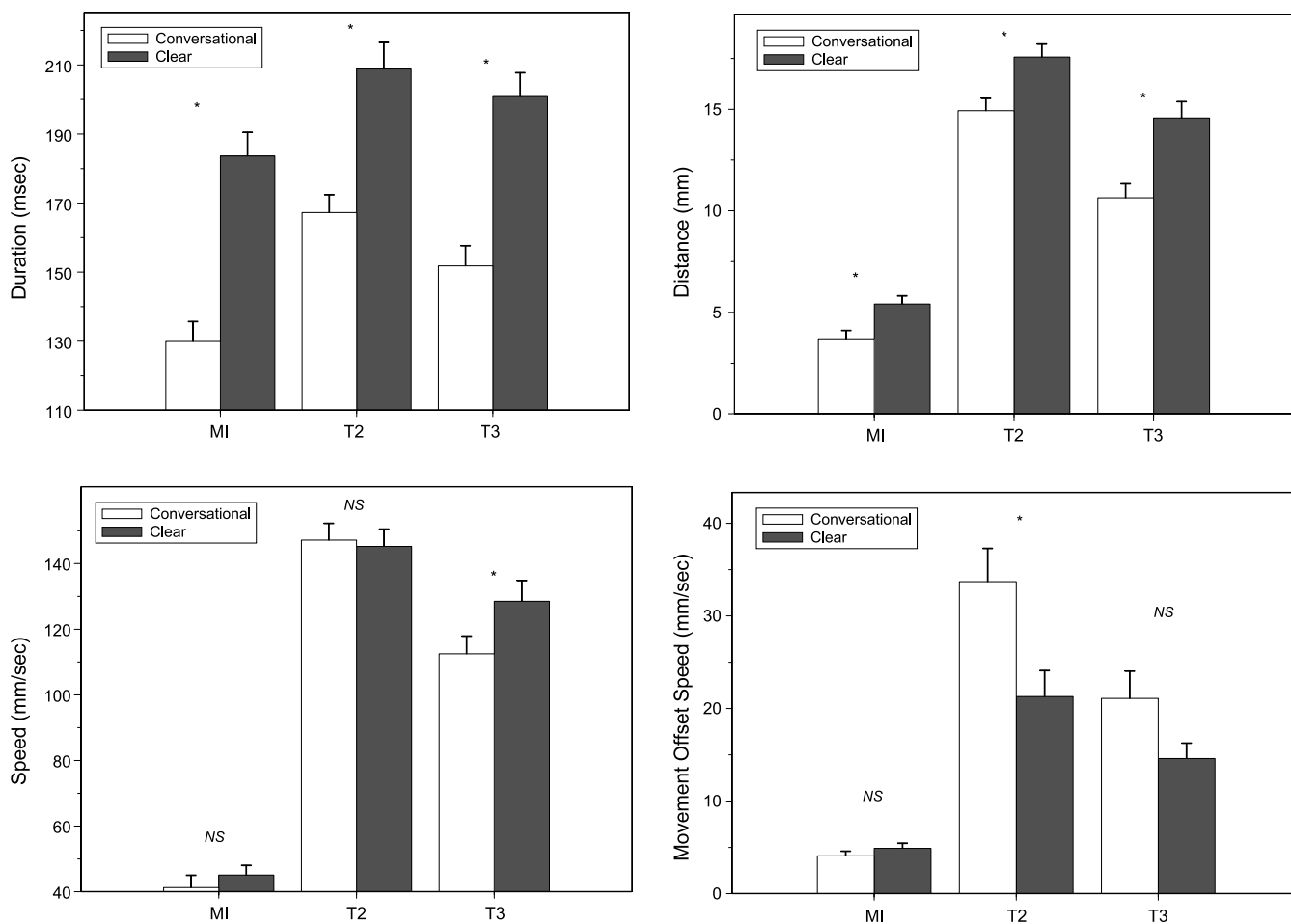
Table 3. Summary of repeated measures analyses of variance performed on the articulatory kinematic measures.

Factor	df	F values								
		Position at movement onset		Position at movement offset		Movement characteristics				
		<i>x_{on}</i>	<i>y_{on}</i>	<i>x_{off}</i>	<i>y_{off}</i>	<i>DUR_{kin}</i>	<i>DIS</i>	<i>SPD_{peak}</i>	<i>SPD_{on}</i>	<i>SPD_{off}</i>
Clarity (C)	1	1.7	13.7*	19.3*	26.3*	62.7*	42.1*	2.5	0.51	16.2*
C × G	1	0.51	0.01	3.5	0.04	0.03	0.66	1.0	0.15	5.5
error (C)	37									
Articulator (A)	2	2387.7*	334.8*	2589.2*	1223.4*	29.6*	230.9*	243.4*	16.8*	57.8*
A × G	2	5.0	2.2	2.6	2.1	0.86	8.4*	7.5*	3.5	3.2
error (A)	74									
C × A	2	1.35	0.52	15.3*	19.9*	1.4	11.3*	10.1*	1.3	13.2*
C × A × G	2	0.22	0.97	6.3	1.1	0.12	2.7	4.8	0.2	2.4
error (C × A)	74									
Gender (G)	1	0.06	1.7	0.37	0.28	1.6	3.1	1.3	3.9	2.8
error (G)	37									

Note. *x*: position with respect to the central maxillary incisor; *y*: position with respect to maxillary occlusal plane. *DUR_{kin}* = duration of kinematic measure; *DIS* = distance; *SPD* = speed.

**p* < .002.

Figure 5. These four bar plots compare clear and conversational speech conditions for movement duration, movement distance, peak movement speed, and speed at movement offset. * $p < .002$.

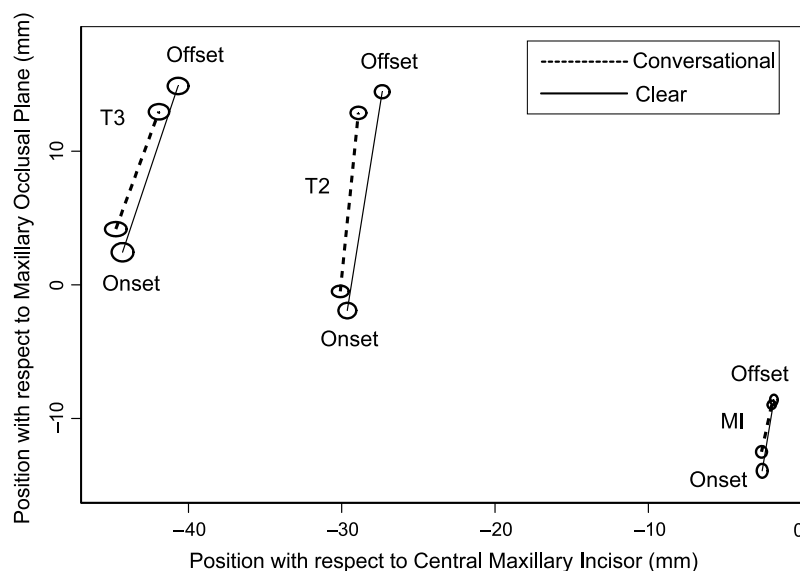


Subgroup analysis based on clarity judgments. One concern is that some speakers in the group who were not perceived by the listener panel to have made a substantive speech clarity adjustment may serve to obscure clarity-related differences in the data. To address this issue, a subset of 34 speakers was drawn from the larger pool. The inclusion criterion for the subset was a minimum of a 100-point conversational clear difference in the mean listener rating. We submitted this subset to a statistical analysis that was identical to that used on the full group. The results were identical to the full group with the exception of two comparisons. In the subgroup, there was a significant clarity main effect for the F1 offset frequency, $F(1, 32) = 16.70, p < .002$, and the main effect for speed at transition offset was near but did not reach significance, $F(1, 25) = 10.5, p = .003$. However, the Clarity \times Articulator interaction remained significant, $F(2, 50) = 9.9, p < .002$. In the next section, we examine a smaller subgroup of speakers.

F2 and T3 "Patterns" During Diphthong Production

In summary, the main results suggest that clarity-related adjustments in diphthong production are generally characterized by a "scaling up" of a number of discrete temporal and spatial measures. We also wanted to evaluate for clarity-related differences in the continuous time-varying patterns of the diphthong production. Because this would involve examining individual speaker data, instead of using the full data set, we focused attention on the 10 speakers the listener panel judged to exhibit the largest difference in clear and conversational productions. Furthermore, we limited analysis to the F2 history and the T3 fleshpoint motion and speed over the entire duration of the diphthong. Because the speakers are quite variable in the absolute timing of the diphthong gestures, we performed a simple linear time normalization so all speaker formant/movement histories had a

Figure 6. This x-y plot shows the spatial positions of T2, T3, and MI at onset and offset of diphthong-related articulatory transitions for conversational and clear speech conditions. Lines connecting the mean onset and offset positions that mark the conversational (broken) and clear (solid) speech conditions are based on Euclidean distance and not on the actual path of the fleshpoints. The size of the ellipses around the mean positions reflects the standard error in each spatial dimension.



common overall duration. Similarly, individual speakers showed substantial differences in overall magnitude of speech events. Therefore, formant/movement histories for a given speaker underwent a z transformation so that all traces had a common mean (zero) and standard deviation (one). The goal was to minimize variation captured in the previous analyses and highlight clarity-related variations in formant/movement history patterns. Figure 7 shows the results of this analysis. Solid fine lines represent traces for individual speakers. The heavy gray line marks the average trace for the group. The top two panels shows the F2 histories, the middle two panels show the vertical position histories for T3, and the bottom two panels plot the T3 speed histories. The left column plots the conversational productions, and the right column plots the clear productions. There are a number of observations worth noting. First, the individual speakers showed remarkable similarity in the F2 and T3 vertical position traces for both conditions (although there is one speaker who showed a unique F2 pattern for the conversational condition). Second, the pattern of the F2 history is very similar to the pattern of the T3 vertical position. The conversational condition is characterized by lack of a clear initial steady-state and a more constant slope over the course of the diphthong. In contrast, the clear condition shows a relatively “flat” region for the first third of the diphthong followed by a large increase in slope for the remaining two thirds of the diphthong’s duration. Finally, the T3 speed history (bottom panel) also shows some noteworthy clarity-related differences. First,

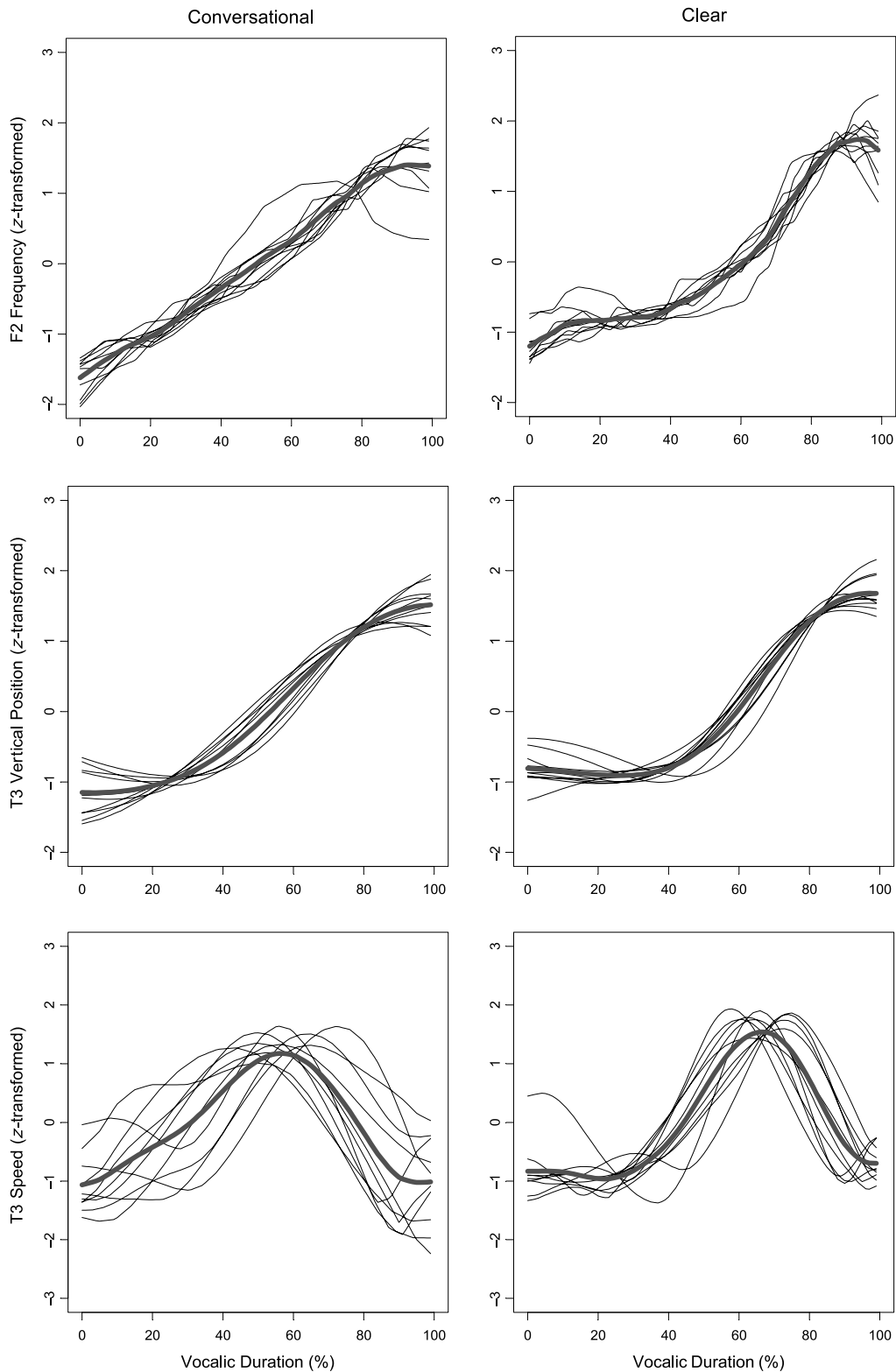
the conversational condition shows more cross-speaker variation as compared with the clear condition. Second, for many speakers, a steady acceleration of the fleshpoint has already begun at the onset of the diphthong in the conversational condition. In the clear condition, the prominent acceleration associated with the transition occurs much later. Individual speakers typically show minimal speed change, indicating relatively little movement during this period. This finding supports the observations in the F2 and T3 vertical position plots. For this speaker group, the clearly produced diphthong is characterized by a prolonged initial steady-state in which there is either little movement or movement unrelated to the diphthong transition.

Discussion

Comparison of Present Results to Other Clear Speech Studies

Consistent with previous studies, we also found that clear speech was characterized by an increased duration of acoustic and articulatory kinematic events. Specifically, all measured intervals, with the exception of the diphthong offset, were longer in the clear speech condition. However, it is important to point out that the offset is the most variable part of the diphthong and is frequently absent in natural speech (Lehiste & Peterson, 1961). It should be noted that although the diphthong is absolutely longer when spoken clearly, speakers spend an

Figure 7. This figure plots selected diphthong-related acoustic and kinematic traces for the 10 speakers who demonstrated the largest clarity rating differences. Each solid line represents a single token produced by a single speaker. The heavy gray traces represent the average trace for the group. The traces were submitted to a linear time normalization procedure to eliminate individual speaker differences in overall duration. Also, each trace underwent a z transformation to reduce individual speaker differences in overall scaling. The top row plots F2 histories across the two speaking conditions. The plots in the middle and bottom rows compare the vertical (“y”) position and speed of the T3 fleshpoint, respectively.



equivalent proportion of the total word duration in the clear (46%) and conversational conditions (45%). In other words, clarity-related adjustments in diphthong duration could simply be arising out of word-level timing changes. In addition to the temporal adjustments, speakers made a number of clarity-related changes to the spectral and kinematic characteristics of the diphthong transition. Kinematic analysis revealed that during clear speech, the tongue (T2 and T3)-mandible (MI) complex was more inferiorly oriented at transition onset and moved farther during the transition. T2 and T3 fleshpoint movements ended at a more superior and anterior position, whereas the MI ended movement at a position similar to conversational speech. Results from the formant analysis were similar to the kinematic results. During clear speech, F1 was higher at transition onset, both F1 and F2 had larger excursions during the transition, and F2 was higher at transition offset. The lower spatial position/higher F1 position at transition onset is consistent with clarity-related adjustments observed in vowels (Ferguson & Kewley-Port, 2002, 2007). We did not find a significant clarity-related change in formant transition rate. This result is at odds with Wouters and Macon (2002), who found that, compared with conversational mode, clearly produced diphthongs exhibited a lower spectral rate of change. However, the large methodological differences between the two studies make such a direct comparison difficult. Peak movement speed did not systematically vary with speech clarity. Only T3 showed greater peak movement speed in the clear speech condition. These equivocal findings are consistent with previous studies in which the articulatory kinematic features of clear speech have been evaluated, albeit in a different phonetic context (Matthies et al., 2001; Perkell et al., 2002). Finally, we failed to find any gender-related interactions in the acoustic and kinematic measures. This was a little surprising since Simpson (2001) reported marked gender differences in diphthong-related tongue movements using the same x-ray microbeam data set (but different speech samples) used in our study. Although methodological details are different, it is difficult to reconcile our results with those of Simpson.

Comparing Present Results With Speaking Rate Studies of Diphthongs

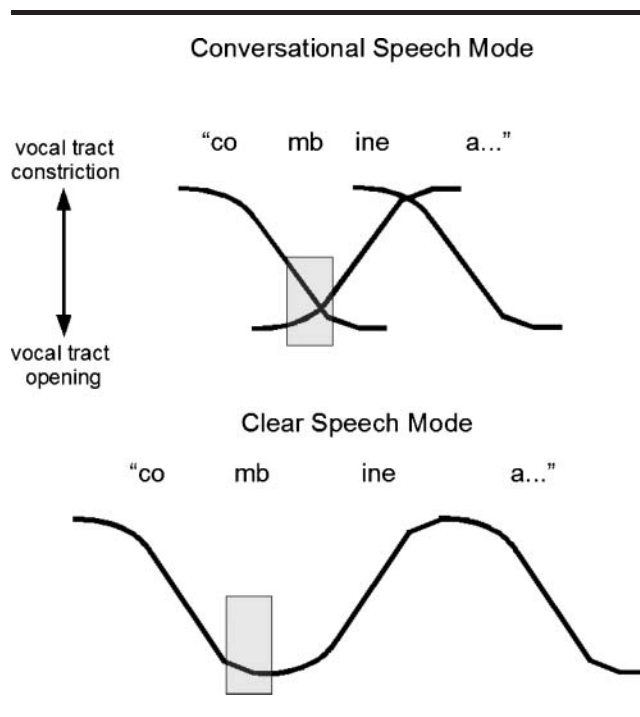
One of the strongest findings in the present study was that durational measures of diphthong production were increased for the clear speech condition. In other words, clear speech is slower than conversational speech. This observation begs the question: Are clearly produced diphthongs the same as slowly produced diphthongs? In some key respects, the answer seems to be yes. The clarity-related increases in formant extent and duration with minimal change in formant transition rate are

similar to the results of Gay (1968) and Weismer (1991). However, we did observe reliable clarity-based differences in F1-onset frequency, which is inconsistent with Gay's findings. It should be noted that envisioning clear diphthongs as slow diphthongs does not necessarily bring us closer to an explanation for how the clarity transformation is achieved. Rate transformation is not a simple matter, and there appears to be a variety of strategies that speakers may use (Tjaden & Weismer, 1998; Weismer & Berry, 2003).

A Proposed Mechanism for Diphthong Clarity Change

It may be possible to interpret the present results using the gestural overlap model outlined in the introduction (Tjaden & Weismer, 1998; Weismer & Berry, 2003). In this model, shorter overall durations can be achieved by increasing the degree of gestural overlap, whereas increasing duration would be achieved by reducing the degree of gestural overlap. In other words, durational changes may be accomplished by varying the phasing of gestures without any rescaling of the gestures themselves. Recall that the diphthong /aI/ was drawn from the word *combine* that was contained in the phrase "combine all the ingredients in a large bowl." Figure 8

Figure 8. This figure outlines how a gestural sliding model might help explain the clarity-related variations in acoustic and articulatory kinematic measures. The thick sigmoid-shaped traces represent hypothetical vocal tract constriction and opening gestures associated with utterance. The gray box represents the time interval during which bilabial closing and opening gestures occur.



provides a simple illustration of how a gesture overlap model might help explain the present data. The sigmoid-shaped traces in the figure represent hypothetical tongue blade gestures. Downward moving traces represent vocal tract opening gestures, and upward moving traces represent vocal tract constriction gestures. The top plot represents a conversational mode of speech. In this plot, the first trace in the sequence is a vocal tract-opening gesture associated with the following series of events: release from the velar plosive, production of the mid-central schwa vowel, and continued motion toward the open vocal tract needed for the onset of the diphthong. Following this opening gesture is a vocal tract constriction gesture associated with the diphthong transition and ending with a constricted position associated with the diphthong offglide (and tongue-tip elevation for the alveolar nasal). Following this constriction gesture is another vocal tract-opening gesture that is associated with transition toward the low, back vowel in the word *all*. The gray box in each plot marks the hypothetical time interval associated with lip closing-opening gesture sequence for the /m-/b/ combination. In this example of conversational speech, the tongue gestures overlap one another. The result is that each gesture is truncated in time and distance by the onset of the next gesture. This results in a shorter overall duration of the sequence and reduced articulatory excursions. The bottom plot illustrates how the model operates in a clear speech mode. Note that the size, duration, and sequence of the tongue gestures are identical to the conversational speaking mode. The only difference in the two plots is the amount of gestural overlap. For the clear speech mode, there is very little gestural overlap. This change in gestural phasing predicts some differences in diphthong features for clear and conversational productions. First, the reduced gestural overlap will result in an increase in the size and extent of the diphthong transition without any change in the rate of the transition. These predictions match both the majority of acoustic and articulatory kinematic results (see Figures 4 and 5), although the model does not predict the clarity-related increase in the peak speed of the T3 fleshpoint (see Figure 5). Second, because the reduced overlap reveals more of the onset and offset of the constriction gesture, the tongue will start the transition in a lower position and end it in a higher position. Furthermore, movement speed at transition onset and offset should be lower because as the overlap is minimized, more of the low-speed portion of the gesture is revealed. Our data provide partial support for these predictions. There is acoustic and kinematic evidence that the transition begins and ends in more extreme positions (see Figures 4 and 6). However, movement speed was only smaller in magnitude for movement offset in the clear speech condition (see Figure 5). Finally, the reduction in gestural overlap also allows the bilabial constriction-release gesture sequence to occur

earlier in the tongue constriction gesture. As a result, a greater proportion of the early component of the constriction gesture follows the bilabial release gesture, which would predict a longer formant steady-state. This was observed in the plots of Figure 7.

This interpretation does come with a number of caveats. First, the gestural overlap model predicts group data. It is probable that there are individual speakers who show patterns that are distinct from the group. However, Figure 7 suggests that those speakers showing the greatest perceptually based clarity effect exhibit a relatively consistent pattern of change in diphthong production. Second, this model requires that clear speech should also be slow speech. There is evidence that speakers can speak quickly and clearly (Krause & Braida, 1995). Although this does not invalidate our interpretation, it does imply that there may be multiple strategies to achieve speech clarity. Third, the gestural overlap account predicts that clear speech should be associated with reduced coarticulation. This view is at odds with some previous clear speech studies of vowels (Bradlow, 2002; Matthies et al., 2001) as well as with rate studies of diphthongs (Tjaden & Weismer, 1998; Weismer & Berry, 2003). Finally, our results are based on a single phonetic context. It is not immediately apparent how our results generalize to other diphthongs and other contexts, which is not to suggest that findings should be identical to the present results. Predictions based on the gestural overlap model would differ in their details due to a variety of contextual factors. The key point is that within such a model, such predictions can be made and tested.

Acknowledgments

Portions of this work were part of a master's thesis completed by the second author at Western Michigan University. Some of these results were presented at the 2004 American Speech-Language-Hearing Association Convention in Philadelphia, PA.

References

- Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication, 14*, 325–337.
- Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 241–273). Berlin, Germany and New York: Mouton de Gruyter.
- Cutler, A., & Butterfield, S. (1990). Durational cues to word boundaries in clear speech. *Speech Communication, 9*, 485–495.
- Cutler, A., & Butterfield, S. (1991). Word boundary cues in clear speech: A supplementary report. *Speech Communication, 10*, 335–353.

- Ferguson, S. H.** (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *116*, 2365–2373.
- Ferguson, S. H., & Kewley-Port, D.** (2002). Vowel intelligibility in clear and conversational speech for normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*, 259–271.
- 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.
- Gay, T.** (1968). Effect of speaking rate on diphthong formant movements. *The Journal of the Acoustical Society of America*, *44*, 1570–1573.
- Gay, T.** (1970). A perceptual study of American diphthongs. *Language and Speech*, *13*, 65–88.
- Gayvert, R. T., & Hillenbrand, J. M.** (2000). SpeechTool (Version 1.65). [Computer software]. Kalamazoo, MI: Western Michigan University.
- Hillenbrand, J. M., & Gayvert, R. T.** (2005). Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research*, *48*, 45–60.
- Krause, J. C., & Braida, L. D.** (1995). The effects of speaking rate on the intelligibility of speech for various speaking modes. *The Journal of the Acoustical Society of America*, *98*, 2982.
- Lehiste, I., & Peterson, G. E.** (1961). Transitions, glides and diphthongs. *The Journal of the Acoustical Society of America*, *38*, 268–277.
- Liu, S., & Zeng, F.-G.** (2006). Temporal properties of clear speech perception. *The Journal of the Acoustical Society of America*, *120*, 424–432.
- Matthies, M., Perrier, P., Perkell, J. S., & Zandipour, M.** (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research*, *44*, 340–353.
- Milenkovic, P.** (2000). Time-frequency analysis for 32-bit Windows. [Computer software]. Madison: University of Wisconsin.
- Moon, S.-J., & Lindblom, B.** (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, *96*, 40–55.
- Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H.** (2002). Economy of effort in different speaking conditions. I: A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America*, *112*, 1627–1641.
- Picheny, M. A., Durlach, N. I., & 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*, 434–446.
- Shrout, P. E., & Fleiss, J. L.** (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.
- Simpson, A. P.** (2001). Dynamic consequences of differences in male and female vocal tract dimensions. *The Journal of the Acoustical Society of America*, *109*, 2153–2164.
- Tasko, S. M., & Westbury, J. R.** (2002). Defining and measuring speech movement events. *Journal of Speech, Language, and Hearing Research*, *45*, 127–142.
- Tjaden, K., & Weismer, G.** (1998). Speaking-rate-induced variability in F2 trajectories. *Journal of Speech, Language, and Hearing Research*, *41*, 976–989.
- Weismer, G.** (1991). Assessment of articulatory timing. In J. Cooper (Ed.), *Assessment of speech and voice production: Research and clinical applications. NIDCD Monograph Volume 1* (pp. 84–95). Bethesda, MD: National Institute on Deafness and Other Communication Disorders.
- Weismer, G., & Berry, J. J.** (2003). Effects of speaking rate on second formant trajectories of selected vocalic nuclei. *The Journal of the Acoustical Society of America*, *113*, 3362–3378.
- Weismer, G., Kent, R. D., Hodge, M., & Martin, R.** (1988). The acoustic signature for intelligibility test words. *The Journal of the Acoustical Society of America*, *84*, 1281–1291.
- Westbury, J. R.** (1994). *X-ray microbeam speech production database user's handbook* [Software manual]. Madison: University of Wisconsin, Waisman Center.
- Wouters, J., & Macon, M. W.** (2002). Effects of prosodic factors on spectral dynamics. I: Analysis. *The Journal of the Acoustical Society of America*, *111*, 417–427.

Received June 15, 2008

Accepted July 1, 2009

DOI: 10.1044/1092-4388(2009/08-0124)

Contact author: Stephen M. Tasko, Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, MI 49008. E-mail: Stephen.tasko@wmich.edu.