Acoustic characteristics of clearly spoken English tense and lax vowels[a)]

Keith K. W. Leung[b)]
Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

Allard Jongman
Department of Linguistics, The University of Kansas, Lawrence, Kansas 66044, USA

Yue Wang
Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

Joan A. Sereno
Department of Linguistics, The University of Kansas, Lawrence, Kansas 66044, USA

(Received 9 October 2015; revised 7 June 2016; accepted 8 June 2016; published online 5 July 2016)

Clearly produced vowels exhibit longer duration and more extreme spectral properties than plain, conversational vowels. These features also characterize tense relative to lax vowels. This study explored the interaction of clear-speech and tensity effects by comparing clear and plain productions of three English tense–lax vowel pairs (/i-ɪ/, /ɑ-ʌ/, /u-ʊ/ in /kVd/ words). Both temporal and spectral acoustic features were examined, including vowel duration, vowel-to-word duration ratio, formant frequency, and dynamic spectral characteristics. Results revealed that the tense−lax vowel difference was generally enhanced in clear relative to plain speech, but clear-speech modifications for tense and lax vowels showed a trade-off in the use of temporal and spectral cues. While plain-to-clear vowel lengthening was greater for tense than lax vowels, clear-speech modifications in spectral change were larger for lax than tense vowels. Moreover, peripheral tense vowels showed more consistent clear-speech modifications in the temporal than spectral domain. Presumably, articulatory constraints limit the spectral variation of these extreme vowels, so clear-speech modifications resort to temporal features and reserve the primary spectral features for tensity contrasts. These findings suggest that clear-speech and tensity interactions involve compensatory modifications in different acoustic domains. © *2016 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4954737]

[ZZ] Pages: 45–58

I. INTRODUCTION

It is well established that speakers are capable of modifying their speech styles depending on context. The hyper- and hypo-articulation (H & H) theory indicates that speech is produced along a hypo- or hyper-articulation continuum in response to listeners' need and output constraints (Lindblom, 1990). In a noisy environment or when listeners have hearing impairments, a clear, hyper-articulated speaking style is adopted in order to enhance intelligibility (Summers *et al.*, 1988). To achieve this goal, clear (relative to plain conversational) speech involves reorganization of articulatory gestures, characterized by a more enunciated speaking manner involving a greater degree of speech articulator movement which leads to corresponding changes in acoustic features (Moon and Lindblom, 1994). In previous studies, the acoustic characteristics of hyper-articulated clear speech have been shown to differ from plain conversational speech in many respects. For vowels specifically, plain-to-clear speech modifications primarily involve increased duration, fundamental frequency (f_0) and intensity, and more peripheral formant frequencies (associated with an expanded vowel space), as well as relative, dynamic temporal and spectral changes (Cooke and Lu, 2010; Ferguson and Kewley-Port, 2002, 2007; Hazan and Baker, 2011; Kim and Davis, 2014; Krause and Braida, 2004; Lu and Cooke, 2008; Smiljanić and Bradlow, 2005). These acoustic features are also what characterize the tense–lax vowel distinction in English. For instance, tense vowels also involve longer durations and more peripheral formant frequencies than lax vowels (Clopper *et al.*, 2005; Hillenbrand *et al.*, 1995; Picheny *et al.*, 1986). One subsequent question that has not been thoroughly explored is how clear-speech vowel modifications may interact with tensity effects. This study thus compares the acoustic characteristics of plain-to-clear vowel modifications for tense and lax vowels in English.

A. English vowels produced in clear and plain speech

Vowels produced in clear and plain speech differ in the temporal domain. Previous studies examining English

---

[a)]A portion of this work was presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 2015.
[b)]Electronic mail: kwl23@sfu.ca

vowels produced in controlled segmental contexts (Ferguson and Kewley-Port, 2002, 2007) or excised from natural sentential contexts (Hazan and Baker, 2011; Kim and Davis, 2014; Lam *et al.*, 2012; Lu and Cooke, 2008; Picheny *et al.*, 1986; Smiljanić and Bradlow, 2008) consistently reveal that vowel duration increases in clear speech relative to plain conversational speech. Plain-to-clear vowel modification has also been explored in terms of relative duration to take into account contextual and speaking rate variations. For example, Tasko and Greilick (2010) found that although the duration of the diphthong /aɪ/ and the word in which it was embedded both increased in clear speech, the word-to-vowel ratio remained constant across speaking styles, indicating that speakers rescale word and vowel duration to a similar extent in clear and plain speech. Similarly, Smiljanić and Bradlow (2008) showed that the percentage vowel lengthening in clear speech relative to conversational speech (clear-conversational/conversational) also remained stable for tense and lax vowels. These findings suggest relational invariance for vowel duration across speaking styles to maintain the length contrast.

Clear and plain vowels also differ in the spectral domain as revealed by changes in vowel space. Clearly produced vowels are characterized by vowel space expansion as compared to plain vowels, involving increases in (1) vowel space perimeter calculated by the sum of the Euclidean distances between adjacent point vowels (Ferguson and Kewley-Port, 2007), (2) area defined as the Euclidian area covered by the polygon formed by the mean of vowel categories (Smiljanić and Bradlow, 2005), and (3) dispersion obtained from the mean of each vowel token's distance from the central point of a speaker's vowel space (Cooke and Lu, 2010; Smiljanić and Bradlow, 2005).

In line with the expansion of the vowel space, measurements of individual formant frequencies (F1, F2, F3) also reveal clear-speech effects. Previous studies generally suggest that, across vowels, F1 increases in clear speech relative to conversational speech (Ferguson and Kewley-Port, 2002; Huber *et al.*, 1999; Lu and Cooke, 2008), consistent with the articulatory findings that clear vowels are produced with a larger jaw opening (Kim and Davis, 2014). The overall increase in F1 may be the byproduct of increased intensity (resulting from increased vocal effort) found for clear speech, which also requires a larger jaw opening (Huber *et al.*, 1999; Krause and Braida, 2004). However, recent evidence has shown that the F1 of some high vowels (e.g., /i/, /ɪ/, /ʊ/) does not change across speaking styles (Ferguson and Quené, 2014), presumably due to the small degree of vertical jaw displacement for high vowels. For F2, while Kim and Davis (2014) found an overall increase in F2 for vowels produced in noise (clear speech) compared to those produced in quiet (conversational speech), examination of individual vowels demonstrated vowel-dependent changes (Ferguson and Kewley-Port, 2002; Lu and Cooke, 2008). For example, Ferguson and Kewley-Port (2002) showed that clearly produced front vowels (/i, ɪ, e, ɛ, æ/) had a higher F2 compared to conversationally produced vowels, but the back vowels (/ɑ, ʌ, o, ʊ, u/) had a lower F2 in clear speech than in conversational speech. These patterns indicate that front vowels became more fronted in clear speech while clear back vowels were produced with more tongue retraction, which corresponds to the vowel space expansion directions. However, an exception to the general pattern of F2 stretching was reported in Lu and Cooke (2008), in that F2 for /i/ and /ɪ/ tended to be lower (less fronted) in clear relative to conversational speech. This may be due to the constraints on variability in speech style for the most peripheral vowels, such as the high-front /i/ (Granlund *et al.*, 2012). Taken together, the results from vowel space and individual formant patterns show that the change in F2 is possibly the main contributing factor to the expansion of the vowel space in clear speech. In terms of F3, previous studies do not provide a conclusive picture about how it changes in clear speech. It has been shown that F3 either remains unchanged (Krause and Braida, 2004), increases (Kim and Davis, 2014), or decreases (Lu and Cooke, 2008) in clear speech. The large variance in the number (three to 10) and type (American versus Australian and British English, rounded versus unrounded) of vowels used in these studies has made the comparison less straightforward.

In addition to static spectral changes at a fixed temporal vowel point, the dynamicity of formant frequencies also contributes to clear-speech features. Previous studies revealed that speakers generally made their clear vowels more dynamic than plain vowels (Ferguson and Kewley-Port, 2002, 2007; Ferguson and Quené, 2014; Moon and Lindblom, 1994; Wouters and Macon, 2002). However, the degree of vowel dynamicity varies among individual vowels, with the intrinsically more dynamic vowels showing greater spectral change in clear speech (Assmann and Katz, 2005; Ferguson and Kewley-Port, 2007; Hillenbrand and Nearey, 1999). For example, Ferguson and Kewley-Port (2007) examined dynamic formant movements of clear and plain vowels in terms of spectral change and spectral angle at the vowel offset portion relative to onset. The spectral change value was obtained from the sum of the F1 and F2 formant frequency changes between the 20% and 80% time points of a vowel (Hillenbrand *et al.*, 1995). Spectral angle was computed as the sum of the angles of the F1 and F2 formant contours at the 80% point with reference to the frequency value at the 20% point. The results indicated that spectral change was greater in clear than in plain speech for the five most inherently dynamic vowels included in their study. However, spectral angle was reduced in clear speech compared to plain speech, arguably because the increase in spectral change values in clear speech was offset by the increase in vowel duration, leading to the reduction of spectral angle (Ferguson and Kewley-Port, 2007).

Plain-to-clear vowel modifications also involve differences in f_0 and intensity. Previous studies examining both individual vowels and vowels in sentential contexts unanimously found an increase in mean, median, maximum and/or minimum f_0 and intensity in clear speech compared to conversational speech (Cooke and Lu, 2010; Hazan and Baker, 2011; Kim and Davis, 2014; Krause and Braida, 2004; Lu and Cooke, 2008).

In summary, previous work on clear speech indicated that, in the temporal domain, vowels are longer in clear speech than in plain speech, with relational vowel duration being stable across speaking styles. In the spectral domain, plain-to-clear modifications in formant frequencies appear to be vowel-dependent. While F1 is generally higher (associated with greater vertical mouth opening) in clear speech relative to plain speech, this difference is more prominent for low vowels than for high vowels. The F2 in clear (relative to plain) speech tends to be higher for front vowels but lower for back vowels. These patterns are consistent with the vowel space analysis findings that clearly produced vowels yield an expanded vowel space compared to plainly produced vowels. In addition to the static spectral changes, clear-speech vowels also involve greater dynamic formant movements than plain vowels. Moreover, plain-to-clear vowel modifications are associated with an increase in f_0 and intensity.

B. English tense and lax vowels

For the English tense and lax vowel contrast, the primary acoustic cue is spectral difference (Hillenbrand *et al.*, 1995; Reetz and Jongman, 2009). Tense vowels (e.g., /i/, /ɑ/, /u/) are produced with more extreme articulatory movements than lax vowels (e.g., /ɪ/, /ʌ/, /ʊ/), thus involving more peripheral formant frequencies and as a result the vowel space covered by tense vowels is typically larger than that formed by their lax counterparts (Hillenbrand *et al.*, 1995). For dynamic formant movements, lax vowels have a shorter target position (steady-state formants) as well as longer and slower articulatory movements into and especially away from the target position (onglide/offglide) than tense vowels, indicating lax vowels' more dynamic formant trajectories (Lehiste and Peterson, 1961; Watson and Harrington, 1999).

Duration is only a secondary cue to the English tense and lax vowel distinction as vowel length contrast is not phonemic in English (Hillenbrand *et al.*, 2000; Reetz and Jongman, 2009; Smiljanić and Bradlow, 2008). However, it still represents an important tensity difference since tense vowels are typically longer than their lax vowel counterparts, presumably resulting from the longer excursions for the articulators to reach the more extreme tense vowel target positions (Hillenbrand *et al.*, 1995; Watson and Harrington, 1999). In terms of relative temporal characteristics, previous research has shown that the magnitude of durational change for tense and lax vowels may differ under the influence of contextual factors such as consonantal context and speaking rate (Port, 1981). For instance, compared to tense vowels, lax vowels (being intrinsically short) may be more resistant to further compression when speaking rate increases (Gopal, 1990).

C. Interactions of clear-speech and tensity effects

The review above shows that clear-speech and tensity effects share similar acoustic properties. Clear vowels are typically longer than plain vowels and so are tense relative to lax vowels, presumably resulting from the longer excursions for clear or tense vowels to reach their more extreme articulatory targets. Spectrally, these more extreme articulatory movements for clear and tense vowels lead to more peripheral formant patterns and subsequently expanded vowel space, as compared to their plain and lax counterparts. In addition, there is evidence indicating an increased dynamic formant movement for clear vowels relative to plain vowels. Formant dynamicity also contributes to tensity effects, although lax vowels appear to involve more dynamic formant trajectories than tense vowels. Given these similarities, it is possible that the vowel tensity contrast interacts with plain-to-clear speech modification in production. However, this interaction was rarely explored in previous research.

The few studies examining the clear-speech effects for tense and lax vowels are inconclusive. In the temporal domain, while some studies found a similar degree of plain-to-clear duration increase for tense and lax vowels and thus a similar degree of tense–lax contrast across speaking styles (Lam *et al.*, 2012; Roesler, 2013; Smiljanić and Bradlow, 2008), there was also evidence that tense vowels extended more than lax vowels in clear speech, resulting in greater tense–lax contrasts in clear relative to plain speech (Picheny *et al.*, 1986). Results from spectral measures also revealed inconsistent patterns. In terms of vowel space, some studies found similar plain-to-clear speech vowel space expansion for tense and lax vowels (Lam *et al.*, 2012; Picheny *et al.*, 1986; Roesler, 2013), yet Krause and Braida (2004) observed that (at a normal speaking rate) the expanded vowel space in clear speech was more evident for tense than lax vowels. However, analyses of individual formant patterns (Picheny *et al.*, 1986) and dynamic formant change (Lam *et al.*, 2012) also indicated greater changes for lax than tense vowels in clear speech.

So far no research has accounted for these discrepancies by directly addressing tensity effects in clear speech. It should be noted that although the above-reviewed studies all involved tense and lax vowel comparisons in clear speech, tensity was not the main focus of these studies: Picheny *et al.* (1986), as one of the pioneering studies, examined general clear-speech effects rather than focusing on tense–lax comparisons *per se*; Krause and Braida (2004) tackled clear-speech effects as a function speaking rate; Lam *et al.* (2012) aimed at comparing the characteristics of different types of clear speech; and Smiljanić and Bradlow (2008) focused on the temporal aspects of clear speech. Although Roesler (2013) focused on tensity effects in clear speech, the study was conducted with static measurements at the sentential level.

Thus, research is needed to systematically examine the acoustic correlates of tense and lax vowels in clear-speech modifications. Such research has significant theoretical implications for unraveling the underlying mechanisms governing clear-speech productions. As stated previously, clear speech is a hyper-articulated type of speech mode intended to improve intelligibility; it thus involves phoneme-extrinsic, quantitative modifications to enhance acoustic contrasts (Moon and Lindblom, 1994). This is different than vowel tensity contrasts that bear intrinsic acoustic features to mark phonemic distinctions. Such contrasts need to be stable across speaking styles in order to maintain phonemic "norms" (Ohala, 1995). Thus, the interaction of clear-speech and tensity effects may involve a trade-off between the intention of making phoneme-extrinsic modifications to enhance overall acoustic salience of vowels and the need for preserving phoneme-intrinsic features to maintain phonological distinctions of vowel categories. Indeed, there have been predictions that clear-speech productions are guided under the parallel principles of "contrast enhancement" and "maintenance of phonemic norms," and that different acoustic dimensions may be independently controlled by these principles (Smiljanić and Bradlow, 2008). Moreover, it has been speculated that, when these two principles are in conflict (e.g., under articulatory constraints), extrinsic, deliberate modifications may succumb to the more essential need for category maintenance (Hazan and Markham, 2004). As clear-speech and vowel tensity effects bear similar acoustic characteristics in multiple acoustic dimensions, exploring their interactions provides a unique testing case to identify the extent to which the implementation of the above-mentioned principles may reflect modifications in different acoustic dimensions—independently and/or jointly.

D. The present study

The present study addresses the issue of tensity and clear-speech interaction through acoustic analyses of a representative set of clearly and plainly produced English tense and lax vowels. The general hypothesis is that the acoustic distance between tense and lax vowels is enlarged in clear speech relative to plain speech, since a clear speaking style aims to increase acoustic distinctions between segments and thus enhance speech intelligibility. Moreover, we predict that clear-speech modifications are constrained by intrinsic acoustic properties of tense and lax vowels, given the primary need to maintain a phonemic tensity contrast. These predictions point to an interaction between clear speech and tensity effects; however, questions remain open as to how tense and lax vowels differ in the amount and manner of modification to the enlarged tense–lax distance in clear speech.

Specifically, two alternative predictions can be made if speaking style interacts with vowel tensity. First, the plain-to-clear modification of tense vowels may be greater than that of lax vowels. In the temporal domain, duration modifications may be greater for tense vowels than for lax vowels, because tense vowels involve longer steady-state and more stable acoustic features relative to lax vowels (Lehiste and Peterson, 1961) and thus may be more easily extended. Spectrally, tense vowels as more peripheral vowels have more room to move in the vowel space than lax vowels, since expansion of lax vowels would shorten the distance from their tense counterparts and consequently undermine phonemic tense–lax distinctions. Thus, it is conceivable that the vowel space expansion in clear speech would be greater for tense vowels than for lax vowels. However, it is also possible that certain tense vowels would show less plain-to-clear speech modifications than lax vowels. Despite the fact that tense vowels are more "free" to expand in the vowel space than lax vowels, articulatory constraints may limit the extent of such modifications, particularly for the most peripheral tense vowels (e.g., /i/). If a tense vowel already involves an extreme articulatory configuration, there will be limited room for further displacement and therefore limited room for plain-to-clear speech modification. Additionally, as reviewed previously, tense vowels show less dynamic spectral changes than lax vowels (Lehiste and Peterson, 1961). This intrinsic constraint may limit the extent of dynamic spectral clear-speech modifications in tense vowels, as compared to the inherently more dynamic lax vowels. Thus, we predict that clear-speech modifications for tense and lax vowels may differ as a function of their intrinsic acoustic properties.

II. METHODS

A. Participants

Eighteen speakers (10 female) were recruited from the undergraduate and graduate population at Simon Fraser University. Participants were native speakers of Western Canadian English aged 17–30 (mean: 21.7). This English dialect exhibits /ɑ/ and /ɔ/ merger (Clopper *et al.*, 2005), and thus speakers should produce the vowel in "cod" as the target vowel /ɑ/ of this study, instead of /ɔ/ in other varieties of English. They reported normal hearing and no history of speech or language disorders.

B. Materials

Six English words "keyed," "kid," "cod," "cud," "cooed," and "could" were used in this study. They carry three pairs of American English tense and lax vowels (/i-ɪ/, /ɑ-ʌ/, and /u-ʊ/). In previous research, /ɑ-ʌ/ has been used as a tense-lax pair along with the canonical tensity pairs /i-ɪ/ and /u-ʊ/ (e.g., Gopal, 1990; Lam *et al.*, 2012). In Western American (Canadian) English, /ʌ/ has been shown to be /ɑ/'s closest lax counterpart both in terms of spectral (Clopper *et al.*, 2005) and temporal (Gopal, 1990) features, similar to how /i-ɪ/ and /u-ʊ/ differ as tense-lax pairs. In particular, the lax vowels /ɪ/, /ʌ/, and /ʊ/ have more centralized positions in a vowel space than their tense counterparts /i/, /ɑ/, and /u/. On the basis of their relative positions in the vowel space, these vowels pairs will henceforth be referred to as high front (/i-ɪ/), low (/ɑ-ʌ/), and high back (/u-ʊ/) vowels. The common /kVd/ consonantal context ensures that the coarticulation effect on vowel formant movement and durational contrast is constant (Moon and Lindblom, 1994). The production of each token was recorded in isolation in plain and clear speaking styles.

C. Procedures

The participants' speech was recorded digitally in a sound-attenuating booth in the Language and Brain Laboratory at Simon Fraser University, using Sonic Foundry Sound Forge 6.4 at a sampling rate of 48 kHz. A Shure KSM microphone was placed at a 45 degree angle, about 20 cm away from the speaker's mouth. Participants were seated at a comfortable distance from the computer screen where prompts, instructions, and feedback were displayed. Before recording began, participants were asked to read the six words aloud to become familiar with the stimuli.

The recording session began with a warm-up session. Speakers produced five repetitions of each word in each of two blocks, in response to prompts appearing on a monitor. Programs were designed using MATLAB to provide prompts and feedback. In the first block, speakers read the words in the plain speaking style. They were asked to speak naturally, as the way they speak in daily conversation. In the second block, speakers read the words in the clear speaking style. They were instructed to speak clearly, as if they were talking to a hearing-impaired or elderly person. The warm-up sessions served to familiarize speakers with the interface and materials and to allow them to rehearse the two styles. The productions from the warm-up sessions were not included in the current analyses.

For the elicitation sessions, the procedure followed the one developed by Maniwa *et al.* (2009). Participants were told that we were testing a speech recognition computer program which was actually a simulated interactive computer program that seemingly attempted to perceive and recognize the tokens produced by a speaker, developed using MATLAB. Participants were instructed to speak naturally first, as if in casual conversation, when a prompt showed up on the screen. Then, the program would "guess" and indicate on the screen what they produced. The participant would then indicate whether the guess was correct by clicking a box on the screen. If the guess was considered correct, the program would move on to the next stimulus. Otherwise, the program would instruct the participant to repeat the stimulus as clearly as possible. In the acoustic analyses of such "incorrect guess" trials, the productions in response to the initial prompts served as the "plain speech," whereas the repeated productions were the "clear speech." To ensure distinct productions of plain versus clear speech, participants were instructed not to adjust their natural production to avoid repetitions because it is important to speak as naturally as possible in order to test the computer program's ability to recognize natural as well as clear speech.

A total of 148 productions were obtained per speaker, 74 per speaking style in 74 elicitation trials described above {[15 repetitions $\times$ 2 words ("keyed" and "cod") + 11 repetitions $\times$ 4 words ("kid," "cud," "cooed," and "could")] $\times$ 2 styles}. There were equal numbers of plain and clear productions. The prompts were presented in three blocks (24 randomly selected trials in the first block and 25 each in the other two) and speakers took a 3-min break after each block. The order of prompts and responses was the same for each participant.

D. Acoustic analyses

The acoustic analyses involve both static and relative, dynamic measures of vowels in both temporal and spectral domains. Nine metrics previously shown to characterize clear-speech vowel modification patterns were used: (1) Vowel duration; (2) Vowel-to-word duration ratio; (3) Vowel formant frequencies at steady state; (4) Vowel space perimeter; (5) Dispersion; (6) Spectral change; (7) Spectral angle; (8) f_0 at midpoint; and (9) Intensity at midpoint. For each metric, a total of 2664 instances [74 productions $\times$ 2 styles $\times$ 18 speakers] were measured. All measurements were obtained using Praat (Boersma and Weenink, 2013).

1. Static and relative temporal measurements

Word and vowel duration were measured based on the following methods. Word duration measures were made from the waveform, from the beginning of the burst of /k/ to the final point of the release of /d/. Vowel duration measures were made from the spectrogram and waveform. The vowel onset was defined as the onset of voicing as shown by strong vertical striations in the spectrogram, and the onset of periodicity of the waveform. The vowel offset was taken at the closure of /d/, corresponding to a cessation of high-frequency energy (Watson and Harrington, 1999).

Based on these word and vowel duration measurements, static and relative temporal metrics were obtained. In addition to vowel duration, the static analyses also include (1) absolute durational difference between clear and plain vowels: $DUR_{clear} - DUR_{plain}$ (Ferguson and Kewley-Port, 2007; Krause and Braida, 2004), and (2) relative durational difference—the percentage change in clear vowel duration relative to plain vowel duration: $(DUR_{clear} - DUR_{plain})/DUR_{plain}$ (Smiljanić and Bradlow, 2008). The relative duration metric consisted of the vowel-to-word duration ratio (Tasko and Greilick, 2010).

2. Static spectral measurements

a. Formant frequency values at steady state. The frequency values of F1, F2, and F3 were taken from the vowel midpoint. As a normalization procedure, the values were converted to critical-band rate in Bark using the following formula:

$$Z = [26.81/(1 + 1960/f)] - 0.53,$$

where Z is the critical-band rate in Bark and f is the raw frequency value in Hertz (Traunmüller, 1990). Each talker's mean F1, F2, and F3 values in Bark were obtained for each vowel and speaking style.

b. Vowel space perimeter. Previous studies examined vowel space expansion in clear speech based on F1 and F2 values (e.g., Bradlow *et al.*, 1996; Ferguson and Kewley-Port, 2007; Picheny *et al.*, 1986). However, this does not reflect the change in F3 between speaking styles. In the present study, the two rounded vowels (/u/ and /ʊ/) produced in clear and plain speaking styles may involve a change in F3 values, which cannot be represented in an F1 $\times$ F2 vowel space diagram. To capture the values of F1, F2, and F3 on a two-dimensional vowel space chart, the Bark Difference Metric was used in this study (Syrdal and Gopal, 1986). Z_3-Z_2 was used to model vowel advancement (i.e., Bark-converted F3 minus Bark-converted F2). This metric was also dependent on F3 values. A smaller F3, indicating more lip rounding, would reduce the values of Z_3-Z_2. Z_1-Z_0 was used to model vowel height (i.e., Bark-converted F1 minus Bark-converted f_0).

To measure the size of a talker's vowel space, this study followed the perimeter measure used in Ferguson and Kewley-Port (2007). The perimeter of the triangular vowel space was measured as the sum of Euclidean distances between adjacent vowels, separately for tense and lax vowels. As mentioned above, the Bark Difference Metric was used in creating the talker's vowel space and the coordinates of each vowel used to calculate the perimeter were its Z_3-Z_2 and Z_1-Z_0 values (refer to Fig. 5 in Sec. III C 2). Each talker's vowel space perimeter was obtained for each speaking style. The perimeter of clearly spoken vowels was predicted to be greater than that of naturally produced vowels, indicating an expansion of the vowel space.

c. Dispersion. The mean of each vowel token's distance from the central point was measured following previous studies (Bradlow *et al.*, 1996; Cooke and Lu, 2010; Kim and Davis, 2014; Smiljanić and Bradlow, 2005). Increased dispersion is associated with vowel expansion in clear speech. Instead of using raw formant frequency values, the Bark Difference Metric was also used for dispersion measurement. For each speaker, each vowel token's Euclidean distance from the centroid of the speaker's triangular vowel space was first obtained. Vowel space dispersion was then calculated as the mean of these distances for each talker separately for each speaking style and for tense and lax vowels.

3. Relative spectral measurements

To explore whether vowel formants would be more dynamic in clear speech, metrics were employed to assess the formant information in the vowel offset portion relative to the onset portion for individual vowel tokens following Ferguson and Kewley-Port (2007). Since F3 was included in the vowel space analyses, the dynamic or relative metrics were modified to include F3 values as well. The frequency values of F1, F2, and F3 were first obtained at the 20% and 80% time points of the vowel portion and converted to Bark. Then, the values were submitted to two relative metric formulas.

a. Spectral change (λ). Spectral change represents the sum of the F1, F2, and F3 absolute frequency shift:

$$\lambda = |Z_{1(80)} - Z_{1(20)}| + |Z_{2(80)} - Z_{2(20)}| + |Z_{3(80)} - Z_{3(20)}|,$$

where $Z_{1(20)}$, $Z_{1(80)}$, $Z_{2(20)}$, $Z_{2(80)}$, $Z_{3(20)}$, and $Z_{3(80)}$ are the Bark-converted F1, F2, and F3 values at 20% and 80% of the vowel portion. Increased formant movement would be associated with larger λ values. Each speaker's mean spectral change value was obtained for each vowel and speaking style.

b. Spectral angle (Ω). Spectral angle represents the sum of the absolute values of the F1, F2, and F3 angles. A higher Ω value reflects increased formant movement. A formant showing no change from the 20% point to the 80% point would have a spectral angle of 0 radian. Following Ferguson and Kewley-Port (2007), the angle θ_{F1} (in radians) for F1, θ_{F2} for F2 and θ_{F3} for F3 at the 20% point was calculated as the arctangent of the difference between the formant frequencies at the 20% and 80% points divided by the duration between these two points scaled to deciseconds:

$$\theta_{Fn} = \arctan[(Zn_{(80)} - Zn_{(20)})/d],$$

where n represents the number of the formant and

$$d(\text{in deciseconds}) = (\text{time}_{80} - \text{time}_{20})/100.$$

The spectral angle was calculated as sum of the absolute values of the F1, F2, and F3 angles:

$$\Omega = |\theta_{F1}| + |\theta_{F2}| + |\theta_{F3}|.$$

It should be noted that this metric does not capture the direction of formant movement (Ferguson and Kewley-Port, 2007). Each speaker's mean spectral angle was obtained for each vowel and speaking style.

4. Fundamental frequency and intensity

The f_0 and intensity were measured at the vowel midpoint using Praat. f_0 was measured in Hz using the autocorrelation algorithm (Boersma, 1993), with the pitch range between 75 and 500 Hz. Intensity was obtained using the mean energy method in Praat (Boersma and Weenink, 2013). Mean f_0 and intensity were obtained for each vowel and speaking style.

III. RESULTS

For all metrics except vowel space perimeter and dispersion, the mean data for each metric were first separately submitted to mixed-design analyses of variance (ANOVAs) with Style (clear, plain), Tensity (tense, lax), and Vowel type [high front (/i/ and /ɪ/), high back (/u/ and /ʊ/), and low (/ɑ/ and /ʌ/)] as within-subject factors, and Gender (male, female) as a between-subject factor. None of the metrics showed any significant interaction of Style by Gender [$Fs(1,16) < 1.80$, $ps > 0.198$] or Style by Tensity by Gender was obtained [$Fs(1,16) < 2.75$, $ps > 0.116$].[1] These metrics were subsequently analyzed by repeated measures ANOVAs with Style, Tensity, and Vowel type as factors collapsed across Gender. A similar case applies to the vowel space perimeter and dispersion metrics. Since the mixed-design ANOVAs with Style and Tensity as within-subject factors, and Gender as a between-subject factor yielded no significant interaction of Style by Gender [$Fs(1,16) < 1.75$, $ps > 0.204$] or Style by Tensity by Gender [$Fs(1,16) < 3.93$, $ps > 0.065$], the data were submitted to repeated measures ANOVAs with Style and Tensity as factors, and collapsed across Gender. Greenhouse–Geisser correction was used when Mauchly's test of sphericity yielded a significant result. For brevity, only significant main effects and interactions involving Style are reported.

A. Static temporal results

The Style × Tensity × Vowel analysis of variance (ANOVA) results for vowel duration (as well as the relative measure of vowel-to-word ratio) are presented in Table I.

For vowel duration, there was a significant main effect of Style, with clear vowels (280 ms) being generally longer than plain vowels (219 ms). There were also significant interactions of Style × Tensity and Style × Vowel type. For the

TABLE I. Summary of the repeated measures ANOVA results for the static and relative temporal measurements. Abbreviations for factors: S = Style; T = Tensity; VT = Vowel Type. Degrees of freedom (df) are in parentheses. Significant results are in boldface.

Metric	S (1,17)	S X T (1,17)	S X VT (2,34)	S X T X VT (2,34)
Vowel duration	**$F = 12.0$, $p = 0.003$**	**$F = 18.9$, $p < 0.001$**	**$F = 11.7$, $p = 0.001$**[a]	$F = 0.486$, $p = 0.620$
Vowel-to-word ratio	$F = 3.35$, $p = 0.085$	**$F = 21.9$, $p < 0.001$**	**$F = 7.24$, $p = 0.002$**	**$F = 7.82$, $p = 0.002$**
Follow-up analyses[b]				
— high front	$F = 0.718$, $p = 0.409$	**$F = 27.0$, $p < 0.001$**	n/a	n/a
— low	**$F = 7.91$, $p = 0.012$**	**$F = 5.25$, $p = 0.035$**	n/a	n/a
— high back	$F = 0.393$, $p = 0.539$	**$F = 16.4$, $p = 0.001$**	n/a	n/a

[a]Greenhouse–Geisser corrected, df = (1.3,21.3).
[b]Vowel-to-word ratio: Style × Tensity ANOVAs for each vowel type.

Style × Tensity interaction, follow-up paired samples t tests with Style as the factor separately for tense and lax vowels showed that clear tense (355 ms) and lax vowels (205 ms) were both longer than plain tense (272 ms) and lax vowels (165 ms) (tense: $p = 0.002$; lax: $p = 0.009$). To further explore this interaction, the mean clear and plain vowel duration differences ($DUR_{clear} - DUR_{plain}$) for tense and lax vowels were compared through paired samples t tests. Tense vowels (82 ms) yielded a greater clear and plain speech difference than lax vowels (39 ms) [$t(17) = 4.35$, $p < 0.001$], indicating a greater plain-to-clear vowel lengthening for tense vowels than for lax vowels. Figure 1a displays plain-to-clear vowel modifications (in terms of vowel duration difference) for tense and lax vowels. For the Style × Vowel type interaction, further analyses yielded longer vowel durations in clear speech (high front: 277 ms; low: 261 ms; high back: 301 ms) than in plain speech (high front: 211 ms; low: 215 ms; high back: 231 ms) for all vowel types ($ps < 0.005$). To further explore this interaction, paired samples t tests were conducted to compare the mean clear and plain vowel duration differences ($DUR_{clear} - DUR_{plain}$) for the three vowel types (level of significance adjusted to $p = 0.017$). Results showed that high front (66 ms) and high back vowels (70 ms) had no significant difference in plain-to-clear modification of vowel duration [$t(17) = 1.44$, $p = 0.167$]. However, low vowels (47 ms) had a smaller vowel duration increase than high front [$t(17) = -3.43$, $p = 0.003$] and high back vowels [$t(17) = -3.64$, $p = 0.002$]. Figure 1(b) displays plain-to-clear vowel modifications (in terms of vowel duration difference) for high front, high back, and low vowels.

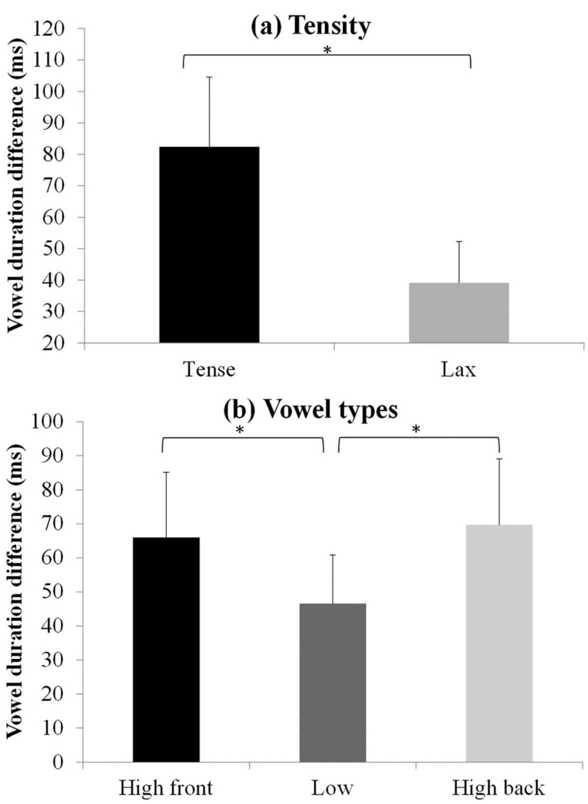

FIG. 1. Mean vowel duration difference between clear and plain speech ($DUR_{clear} - DUR_{plain}$) for (a) tense and lax vowels, and (b) three vowel types (high-front, low, and high-back). (*) indicates a statistically significant difference ($p < 0.05$). Error bars represent one standard error.

Additionally, the ratio of the percentage lengthening of clear vowel duration to their plain duration [$(DUR_{clear} - DUR_{plain})/DUR_{plain}$] was computed for tense and lax vowels in order to examine whether the extent of duration modification differs as a function of vowel tensity. A repeated measures ANOVA with Tensity and Vowel type as factors revealed a significant main effect of Tensity [$F(1,17) = 16.3$, $p = 0.001$] but a non-significant interaction of Tensity and Vowel type [$F(1.3,22.4) = 0.117$, $p = 0.804$]. Tense vowels (31%) had a greater percentage lengthening than lax vowels (23%) (Fig. 2).

Overall, the vowel duration results indicated that tense vowels exhibited greater degree of plain-to-clear lengthening than lax vowels, and high vowels showed greater degree of lengthening than low vowels.

B. Relative temporal results

For vowel-to-word duration ratio, as shown in Table I, the ANOVA revealed significant interactions of Style × Tensity, Style × Vowel type, and Style × Tensity × Vowel type. As further analyses, repeated measures ANOVAs with Style and Tensity as factors for each Vowel type were carried out. A significant interaction of Style × Tensity was obtained for all Vowel types. Subsequent analyses showed that for the tense vowels /i/ and /u/, clear productions yielded a higher vowel-to-word ratio (/i/: 0.566; /u/: 0.592) than plain productions (/i/: 0.547; /u/: 0.574) (/i/: $p = 0.003$; /u/: $p = 0.033$). In contrast,

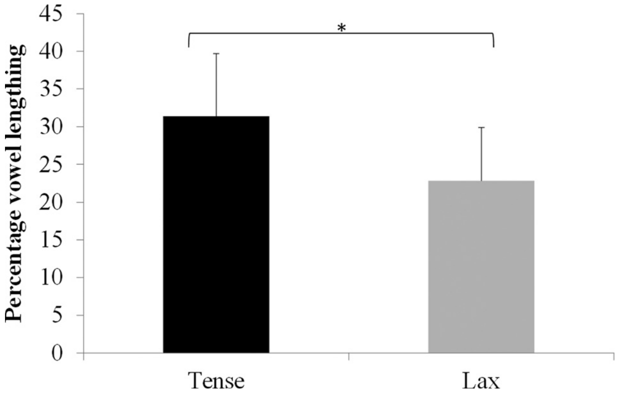

FIG. 2. Percentage vowel lengthening between clear and plain speech [$(DUR_{clear} - DUR_{plain})/DUR_{plain}$] for tense and lax vowels. (*) indicates a statistically significant difference ($p < 0.05$). Error bars represent one standard error.

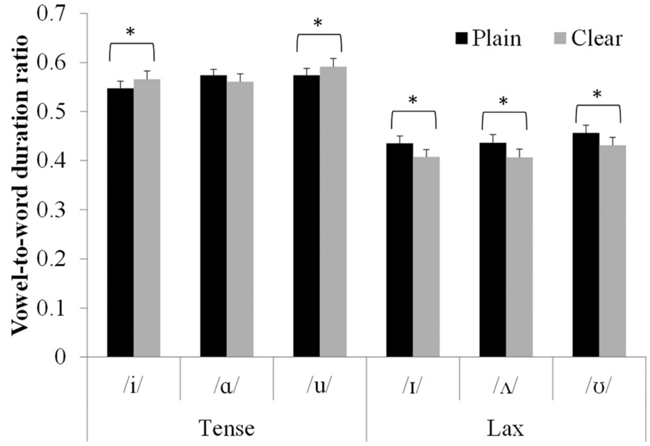

FIG. 3. Mean vowel-to-word duration ratio for six vowels in plain and clear speech, grouped by tensity. (*) indicates a statistically significant difference ($p < 0.05$). Error bars represent one standard error.

the lax vowels /ɪ/, /ʌ/, and /ʊ/ yielded a lower ratio in clear speech (/ɪ/: 0.407; /ʌ/: 0.406; /ʊ/: 0.431) than in plain speech (/ɪ/: 0.435; /ʌ/: 0.436; /ʊ/: 0.456) (/ɪ/: $p = 0.003$; /ʌ/: $p = 0.004$; /ʊ/: $p = 0.002$). In general, vowel-to-word ratio increased in clear relative to plain speech for tense vowels but decreased for lax vowels, as displayed in Fig. 3.

C. Static spectral results

The statistical results of static spectral measurements are summarized in Table II.

1. Steady-state formant frequencies

For F1, the interaction of Style × Vowel type was significant. Further analyses with Style as a factor for each Vowel type suggested that only the F1 of low vowels was significantly different. Clearly spoken low vowels yielded a significantly higher F1 (6.71 Bark), indicating a more lowered production, than plainly produced low vowels (6.62 Bark) ($p = 0.036$).

For F2, there was a significant main effect of Style. On average, clear vowels had a lower F2 (11.6 Bark) than plain vowels (11.7 Bark). There were significant interactions of Style × Tensity and Style × Vowel type. To follow up, repeated measures ANOVAs with Style and Tensity as factors for each Vowel type were conducted. A significant interaction of Style × Tensity was found for high front and low vowels. Further analyses with Style as a factor separately for tense and lax vowels in each vowel type showed that, for high front vowels, only /ɪ/ yielded a higher F2 (more fronted) in clear speech (13.3 Bark) than in plain speech (13.2 Bark) ($p = 0.017$). For low vowels, the clearly produced /ɑ/ (9.43 Bark) had a lower F2 (more retracted) than

TABLE II. Summary of the repeated measures ANOVA results for static spectral measurements. Abbreviations for factors are the same as in Table I. Degrees of freedom are in parentheses. Significant results are in boldface.

Metric	S (1,17)	S × T (1,17)	S × VT (2,34)	S × T × VT (2,34)
Static formant frequencies				
F1	$F = 0.272, p = 0.609$	$F = 1.64, p = 0.218$	**$F = 4.03, p = 0.027$**	$F = 0.833, p = 0.413$[a]
F2	**$F = 6.17, p = 0.024$**	**$F = 4.65, p = 0.046$**	**$F = 6.49, p = 0.010$**[b]	$F = 0.635, p = 0.536$
Follow-up analyses[c]				
— high front	**$F = 4.83, p = 0.042$**	**$F = 9.15, p = 0.008$**	n/a	n/a
— low	$F = 2.85, p = 0.110$	**$F = 5.35, p = 0.033$**	n/a	n/a
— high back	**$F = 8.03, p = 0.011$**	**$F = 8.03, p = 0.011$**	n/a	n/a
F3	**$F = 5.71, p = 0.029$**	$F = 0.694, p = 0.416$	$F = 1.21, p = 0.310$	**$F = 5.50, p = 0.009$**
Follow-up analyses[d]				
— high front	**$F = 8.18, p = 0.011$**	**$F = 13.7, p = 0.002$**	n/a	n/a
— low	$F = 4.26, p = 0.055$	$F = 2.47, p = 0.135$	n/a	n/a
— high back	$F = 1.66, p = 0.215$	$F = 1.83, p = 0.194$	n/a	n/a
Vowel space				
Perimeter	**$F = 7.81, p = 0.012$**	$F = 0.113, p = 0.741$	n/a	n/a
Dispersion	**$F = 5.07, p = 0.038$**	$F = 0.327, p = 0.575$	n/a	n/a

[a]Greenhouse–Geisser corrected, df = (1.5,24.8).
[b]Greenhouse–Geisser corrected, df = (1.5,24.8).
[c]F2: Style × Tensity ANOVAs for each vowel type.
[d]F3: Style × Tensity ANOVAs for each vowel type.

the plainly spoken /ɑ/ (9.58 Bark) ($p = 0.037$). For high back vowels, a significant main effect of Style was obtained, where the F2 was lower (more retracted) in clear speech (10.5 Bark) than in plain speech (10.7 Bark).

For F3, a significant main effect of Style was observed. On average, clear vowels had a higher F3 (15.13 Bark) than plain vowels (15.07 Bark). In addition, the interaction of Style × Tensity × Vowel type was significant. To further investigate the three-way interaction, the F3 values for each Vowel type were separately submitted to repeated measures ANOVAs with Style and Tensity as factors. High front vowels yielded a significant interaction of Style × Tensity. Further analyses showed that only for /ɪ/, F3 was higher (less rounded) in the clear (15.3 Bark) than plain (15.2 Bark) production ($p = 0.001$). For low vowels, the main effect of Style was marginally significant, with clear low vowels having a higher F3 (15.1 Bark) than plain low vowels (15.0 Bark). For high back vowels, no significant effect of Style or interaction of Style × Tensity was obtained.

Taken together, clear speech effects were observed in the static formant frequency results, with the direction of plain-to-clear modifications resulting in more peripheral formant patterns in clear speech. The tensity and style interactions were not consistent, with clear-speech modifications favoring the tense vowel for the /ɑ-ʌ/ pair, while favoring the lax vowel for /i-ɪ/. Figure 4 displays the mean Bark-converted formant frequencies for all vowels in plain and clear speech.

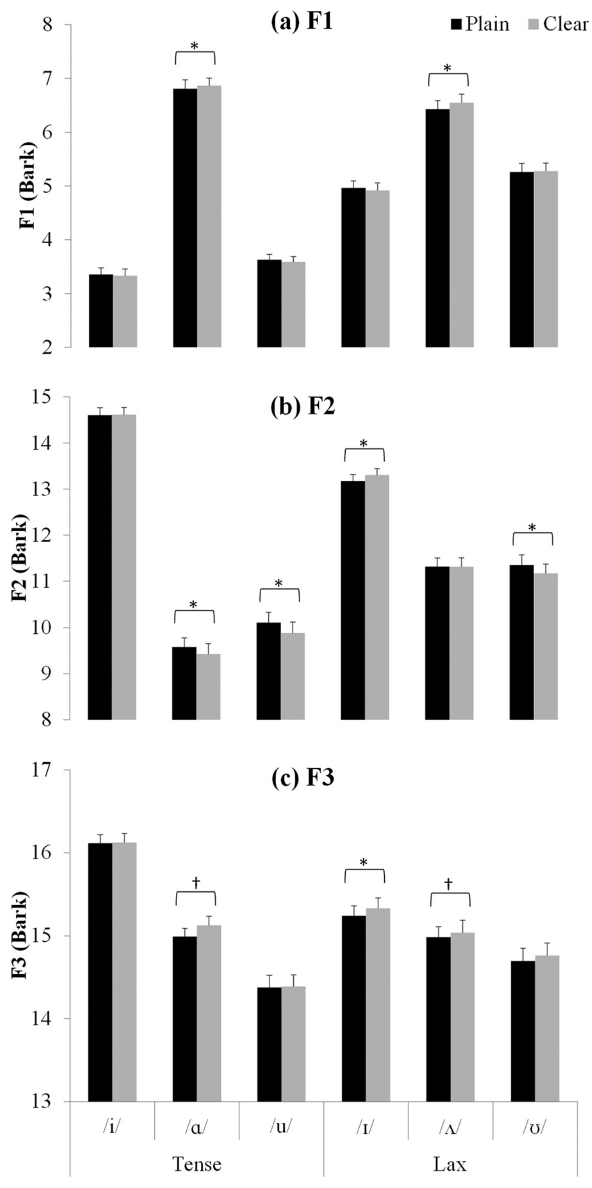

FIG. 4. Mean Bark-normalized formant frequency values (a: F1, b: F2, c: F3) for six vowels in plain and clear speech, grouped by tensity. (*) indicates a statistically significant difference ($p < 0.05$). A marginally significant difference is represented by a dagger. Error bars represent one standard error.

2. Vowel space measurements

The vowel space diagrams are shown in Fig. 5. The ANOVA for vowel space perimeter revealed a significant effect of Style. Clear vowels (8.96 Bark) had a longer perimeter than plain vowels (8.40 Bark). A significant effect of Style was also found for vowel space dispersion. Clear vowels (1.76 Bark) had a greater dispersion than plain vowels (1.68 Bark). No significant interaction of Style × Tensity was found for perimeter or dispersion. Therefore, vowel space expansion was observed in clear speech, but the degree of expansion was the same across vowel tensity.

D. Relative spectral results

The statistical results of the relative spectral measurements including spectral change and spectral angle are presented in Table III.

For spectral change, ANOVA yielded a significant main effect of Style (and a significant main effect of tensity [$F(1,17) = 12.7$, $p = 0.002$]). Moreover, there were significant interactions of Style × Tensity and Style × Vowel type. Overall, spectral change was greater in clear speech (1.94 Bark) than in plain speech (1.82 Bark). Lax vowels (2.11 Bark) had a greater spectral value than tense vowels (1.65 Bark). To follow up the interaction effects, repeated measures

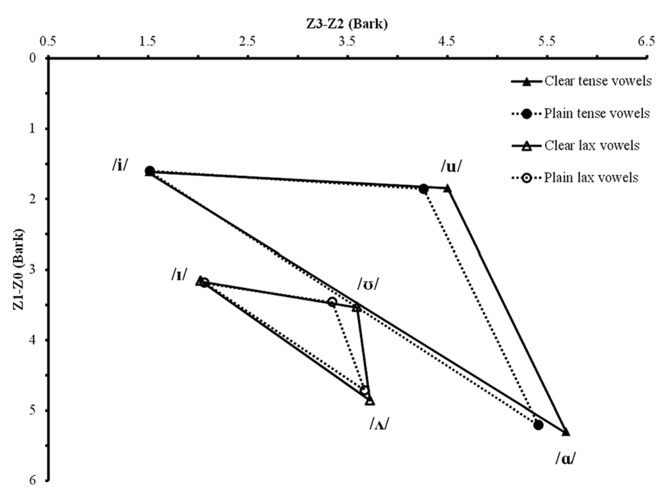

FIG. 5. Vowel space diagram for tense (filled symbols) and lax (open symbols) vowels in plain (dotted line, circle symbols) and clear (solid line, triangle symbols) speaking styles. The x axis represents vowel height and the y-axis represents vowel backness.

TABLE III. Summary of the repeated measures ANOVA results for relative spectral measurements. Abbreviations for factors are the same as in Table I. Degrees of freedom are in parentheses. Significant results are in boldface.

Metric	S (1,17)	S × T (1,17)	S × VT (2,34)	S × T × VT (2,34)
Spectral change	**$F = 4.95, p = 0.040$**	**$F = 15.7, p = 0.001$**	**$F = 4.10, p = 0.025$**	$F = 0.532, p = 0.539$[a]
Follow-up analyses[b]				
— high front	$F = 2.03, p = 0.173$	**$F = 6.06, p = 0.025$**	n/a	n/a
— low	$F = 3.67, p = 0.073$	**$F = 8.69, p = 0.009$**	n/a	n/a
— high back	**$F = 6.58, p = 0.020$**	$F = 4.29, p = 0.054$	n/a	n/a
Spectral angle	$F = 1.54, p = 0.231$	$F = 0.445, p = 0.514$	$F = 1.08, p = 0.327$[c]	$F = 0.292, p = 0.680$[d]

[a]Greenhouse–Geisser corrected, df = (1.5,24.9).
[b]Spectral change: Style × Tensity ANOVAs for each vowel type.
[c]Greenhouse–Geisser corrected, df = (1.3,21.7).
[d]Greenhouse–Geisser corrected, df = (1.5,24.8).

ANOVAs with Style and Tensity as factors were carried out for each Vowel type separately. Significant interactions of Style × Tensity were obtained for high front and low vowels. For high back vowels, the interaction of Style × Tensity was marginally significant. The spectral change for the tense vowel /i/ was smaller for clear speech (0.77 Bark) than for plain speech (0.90 Bark) ($p = 0.022$). In contrast, for the lax /ʌ/ and /ʊ/, clear tokens (/ʌ/: 2.95 Bark; /ʊ/: 2.55 Bark) yielded a greater spectral change than plain tokens (/ʌ/: 2.58 Bark; /ʊ/: 2.21 Bark) (/ʌ/: $p = 0.002$; /ʊ/: $p = 0.003$). Together, the spectral change measure revealed that tense vowels became less dynamic while lax vowels became more dynamic in clear relative to plain speech. The mean spectral change data of all vowels in both speaking styles are presented in Fig. 6.

For spectral angle, the main effect of Style was not significant, neither were the interaction of Style × Tensity nor Style × Tensity × Vowel type.

E. Correlation of relative temporal and spectral modifications

The above results reveal that tense vowels involve a plain-to-clear increase in vowel-to-word ratios (relative temporal metric), whereas lax vowels involve a plain-to-clear increase in spectral change values (relative spectral metric). There appears to be an inverse relationship between temporal and spectral modification as a function of tensity. To examine this potential trade-off effect, the plain-to-clear differences in spectral change values and those in vowel-to-word ratios for each speaker and vowel were compared in a correlation analysis. The results revealed a significant negative correlation [$r = -0.299, p = 0.002$], indicating that greater temporal modifications are accompanied by smaller spectral modifications (i.e., tense vowels), and vice versa (i.e., lax vowels). As the scatter plot (Fig. 7) shows, though with some overlapping, tense vowel tokens tend to cluster in the top left panel (with larger values in the temporal domain and smaller values in the spectral domain) while lax vowel tokens tend to cluster in the bottom right panel (with larger values in the spectral domain and smaller values in the temporal domain).

F. Fundamental frequency and intensity

For f_0, the ANOVAs revealed no main effect of Style [$F(1,17) = 0.580$, $p = 0.457$]. No significant interaction of Style × Tensity [$F(1,17) = 0.054$, $p = 0.820$] or Style × Tensity × Vowel type [$F(1.5,25.5) = 0.876$, $p = 0.400$] was obtained. For Intensity, there was a significant main effect of Style [$F(1,17) = 7.12$, $p = 0.016$]. Clear vowels (64.7 dB) were generally louder than plain vowels (63.6 dB).

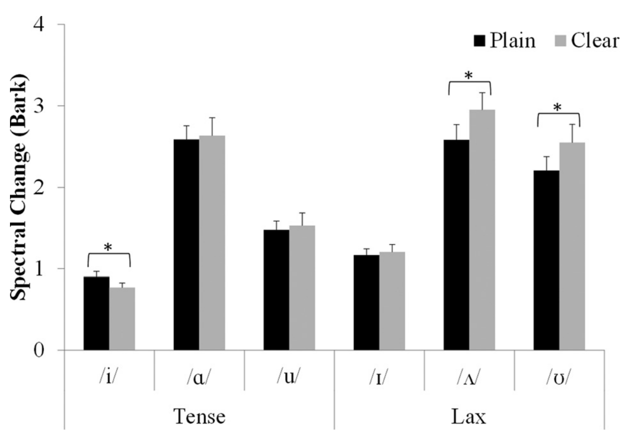

FIG. 6. Mean spectral change values for six vowels in plain and clear speech, grouped by tensity. (*) indicates a statistically significant difference ($p < 0.05$). Error bars represent one standard error.

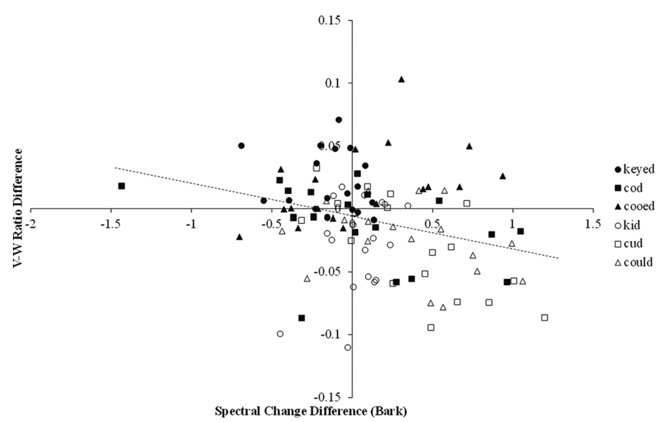

FIG. 7. Spectral change and vowel-to-word (V-W) ratio differences between clear and plain speech. Each point represents the mean values of each vowel produced by each speaker. Tense and lax vowels are represented by filled and open markers, respectively. The dashed line represents the best fit for all data points.

IV. DISCUSSION

This research examined the interaction of clear-speech and tensity effects on the acoustic characteristics of English vowels, investigating differences as a result of the enhancement of acoustic distance between the tense and lax vowel categories in clear relative to plain speech. Two alternative predictions were proposed in terms of how tense and lax vowels may change differently in clear speech. Clear speech could demonstrate either greater or smaller modifications for tense vowels than for lax vowels, presumably in keeping with the intrinsic acoustic differences between tense and lax vowels and their articulatory constraints.

A. Temporal modifications

The overall finding of longer durations in clear than plain speech across vowels agrees with previous research (Ferguson and Kewley-Port, 2002; Kim and Davis, 2014; Lu and Cooke, 2008). The findings on the effects of tensity support the general prediction that tense and lax vowel distinctions are enhanced in clear relative to plain speech. In terms of the extent of modification, the static and relative measurement results consistently support the overall hypothesis of greater changes for tense than lax vowels. First, the static temporal results in terms of both the absolute ($DUR_{clear} - DUR_{plain}$) and relative [$(DUR_{clear} - DUR_{plain})/DUR_{plain}$] durational change show greater plain-to-clear duration increases for tense vowels than for lax vowels. Since tense vowels are also intrinsically longer than lax vowels, this creates a greater tensity difference in clear speech than in plain speech. These patterns are consistent with the previous findings for clear tense and lax vowels (e.g., Ferguson and Kewley-Port, 2002, cf. tense and lax vowel durations reported in Table III, p. 263; Picheny *et al.*, 1986), supporting the prediction that clear-speech modifications are motivated by intrinsic vowel characteristics. First, since tense vowels are inherently long with stable spectral targets (Lehiste and Peterson, 1961), they can be maintained and extended without much deviation in quality. On the other hand, lax vowels are characterized as being short and spectrally dynamic; further lengthening would alter their spectral-temporal representations as well as shrinking the temporal distance between tense and lax vowels, thus reducing phonemic contrastivity.

Consistent with the static data, the relative temporal results show that the vowel-to-word duration ratio increased for tense vowels (/i/, /u/) but decreased for lax vowels (/ɪ/, /ʌ/, /ʊ/) in clear speech, resulting in a greater difference in vowel-to-word ratio between tense and lax vowels and thus greater tense–lax contrast in clear speech than in plain speech. These results also imply that a tense vowel lengthens relatively more than the word (/kVd/) that carries the tense vowel, whereas a lax vowel lengthens relatively less than the word itself. Consequently, a tense vowel generally becomes more prominent in a word in clear speech relative to plain speech. In contrast, a clear lax vowel occupies a relatively shorter portion of a word than a plain lax vowel. It is conceivable that an increase in ratio can be found for tense vowels only because the articulatory trajectory, and therefore vocal effort, involved in producing tense vowels is greater compared to lax vowels, leading to a relatively greater stretching of the vowel portion than the word itself (Nooteboom, 1997). On the other hand, as discussed earlier, since the dynamic nature of lax vowels limits their capacity to lengthen, they exhibited smaller clear-speech temporal modifications as compared to tense vowels. As for the lack of a clear-speech vowel-to-word ratio change for /ɑ/ (in contrast to the increased ratios for /i/ and /u/), it could be due to the constraints of lengthening this low vowel in the context of stop consonants (/k/, /d/) that require a high tongue position is involved. By contrast, high vowels (/i/, /u/) do not involve such contrasting transitions from and into the adjacent stops. As has been observed previously, carry-over and anticipatory effects of high-position contextual stop productions may undermine the intrinsic lengthening of a low vowel due to jaw lowering (Lindblom, 1990).

It should also be noted that the current temporal results are not in agreement with some of the previous studies in which no reliable tense–lax differences in plain-to-clear vowel lengthening were observed, either in terms of static absolute or relative durations (Roesler, 2013; Smiljanić and Bradlow, 2008) or relative durational ratios (Tasko and Greilick, 2010). These discrepancies may be due to the differences in the contexts in which target vowels were elicited in these studies. While the current research involved target vowels embedded in isolated monosyllabic words, vowels in these other studies occurred in sentential contexts (often in word-medial, non-prepausal positions). This may influence temporal measures because the average durations of vowels preceding non-prepausal word-final consonants were more constrained than those of vowels preceding prepausal word-final consonants (Crystal and House, 1988). The monosyllabic context (equivalent to prepausal word-final position) in this study may allow room for duration variations, leading to the observation of tense–lax differences in clear speech. Moreover, target vowels in the current study involved a single context (/kVd/), while vowels in the sentential-context studies were analyzed across a variety of consonantal contexts (Roesler, 2013; Smiljanić and Bradlow, 2008). Previous research has indicated that, in slow-rate speech, tense vowels may lengthen more than lax vowels in some consonantal contexts but not others (Gopal, 1990). In sum, it appears that the monosyllabic, single consonantal context in this study allowed more room for durational variation, resulting in effects that were not found in previous studies.

B. Spectral modifications

In addition to the temporal effects, the results also show clear-speech and tensity interactions in the spectral domain. The relative spectral results (spectral change measure) revealed a greater tensity difference in spectral change in clear speech than in plain speech. Specifically, the spectral change values for the lax vowels /ʌ/ and /ʊ/ were higher in clear than plain speech, whereas the value for the tense vowel /i/ was lower in clear than plain speech. As a result, lax vowels became relatively more dynamic compared to their tense vowel counterparts in clear relative to plain speech. The current lax vowels in general had higher spectral change values than tense vowels, as previously claimed (Lehiste and Peterson, 1961). Therefore, the difference between tense and lax vowels in terms of vowel dynamicity increased in clear speech relative to plain speech, supporting our prediction of a greater tensity distinction in clear speech. These results are consistent with previous findings that spectral change values significantly increase from plain to different types of clear speech for lax vowels but not for tense vowels (Lam *et al.*, 2012). The acoustic results also find support in intelligibility studies which show that the intelligibility of /ʌ/ and /ʊ/ was more affected by dynamic formant information compared to /i/ (Assmann and Katz, 2005; Hillenbrand and Nearey, 1999). Thus, findings from the current and previous studies consistently show that lax vowel productions involve greater plain-to-clear modifications in dynamic spectral changes than tense vowel productions. These patterns are presumably due to the intrinsic differences between tense and lax vowels, with tense vowels being more stable and less susceptible to variation and lax vowels being more dynamic and prone to variation. Unlike the spectral change results, the spectral angle measure did not yield any significant effect of speaking style or interaction of speaking style and vowel tensity. It should be noted that this measure includes duration in the calculation; thus, the effects of spectral angle could be neutralized given both the spectral change value and vowel duration increase in clear speech, as was also claimed by Ferguson and Kewley-Port (2007).

In terms of static spectral features, the clear-speech effects for both tense and lax vowels are characterized by vowel space expansion as shown in the vowel space perimeter and dispersion results, consistent with previous findings (Ferguson and Kewley-Port, 2007; Smiljanić and Bradlow, 2005). However, tense and lax vowels did not differ in the amount of vowel space expansion in clear speech (see also Lam *et al.*, 2012; Picheny *et al.*, 1986; Roesler, 2013). The vowel space results do not show a greater tense–lax distance in clear (than plain) speech. This indicates that the tense vowels are constrained from extending further in the vowel space (Granlund *et al.*, 2012). Individual formant changes may reflect such constraints.

Results from individual formant frequency changes also revealed that tense and lax vowels changed in similar manners, with the exception of /i/. Specifically, both tense and lax low vowels (/ɑ/ and /ʌ/) yielded an F1 increase in clear (relative to plain) speech, indicating a greater jaw opening. Both tense and lax clear high back vowels (/u/ and /ʊ/) showed an F2 decrease, suggesting a greater tongue retraction. The lax high front vowel /ɪ/ increased F2 and F3 in clear speech, suggesting a greater tongue advancement. F3 increase was found for the tense–lax vowel pair /ɑ/-/ʌ/. These results are compatible with the present vowel space expansion results as well as previous findings (Ferguson and Kewley-Port, 2002; Ferguson and Quené, 2014; Lu and Cooke, 2008). It should be noted that, across formants, the tense vowel /i/ did not exhibit any modification from plain to clear speech. While the resistance to F1 increase was expected for high vowels such as /i/ (Ferguson and Quené, 2014), no increase in F2 or F3 was found either, suggesting the absence of further tongue advancement or lip-spreading for the front vowel /i/ in clear-speech modifications. The absence of clear-speech effects for /i/ is presumably due to its extreme articulation. As the results showed, the F1 was already low (lowest among all the vowels), and the F2 and F3 values were already high (highest across vowels) in plain speech productions, suggesting the extreme high front position of its articulation in the vocal tract. It is thus conceivable that articulatory constraints may have limited the capacity for further articulatory excursions in clear speech. Indeed, previous research has shown similar constraints for extreme vowels such as /i/ (Granlund *et al.*, 2012). Taken together, given that the formant patterns for the majority of the vowels exhibited similar plain-to-clear modifications for tense and lax vowels, these formant results did not support the general prediction of a greater tense–lax distinction in clear relative to plain speech. However, the finding of different modification patterns for /i/ and /ɪ/ due to the absence of change for /i/ favors the hypothesis that articulatory constraints limit displacement.

The last point to note in terms of spectral modifications is related to f_0. In contrast to previous work, no plain-to-clear modification was found for f_0. This may be due to a difference in elicitation context. Previous studies which obtained an increase in f_0 mostly examined f_0 range and mean values in sentential context (Cooke and Lu, 2010; Kim and Davis, 2014; Lu and Cook, 2008). Since sentence-length materials have greater f_0 variability caused by contextual effects (Krause and Braida, 2004), it is possible that speakers changed their intonation pattern in clear speech and brought about the changes in overall f_0.

C. Compensatory clear-speech modifications

The above results in individual acoustic domains appear to show that our overall prediction was partially supported, since a tensity difference in clear-speech modifications was only evident in some acoustic attributes. However, further examination of the combined effects of different acoustic features revealed that these acoustic attributes collaborate in a complementary manner to serve different speech production functions, maintaining phoneme-intrinsic (tensity) categorical distinctions on the one hand and making phoneme-extrinsic (clear-speech) modifications on the other. Thus, our hypothesis was supported in that tense–lax distinctions were indeed enhanced in clear speech, with modifications utilizing different acoustic attributes for tense and lax vowels.

Specifically, the relative results revealed different plain-to-clear modifications for tense and lax vowels in the temporal and spectral domain. Tense vowels showed significant temporal modifications with a vowel-to-word duration ratio increase in clear speech, but their spectral change values either reduced or remained unchanged across speaking styles. In contrast, lax vowels demonstrated significant spectral modifications with an increase in spectral change values in clear speech, while their temporal change was not evident (as shown by a decreased vowel-to-word duration ratio). These results suggest a trade-off in the use of acoustic cues for modifications, as supported by the follow-up correlation analysis showing that spectral change and vowel-to-word duration ratio values were inversely correlated. The scatter plot (Fig. 7) illustrates that tense vowels involved greater (positive) vowel-to-word duration ratio modifications but smaller (negative) spectral change modifications, as compared to lax vowels which showed the reverse pattern. The static results revealed similar trade-off patterns. While tense vowels showed greater plain-to-clear lengthening than lax vowels, the spectral modifications for the most peripheral tense vowel /i/ were restricted.

These compensatory temporal and spectral clear-speech modifications can be attributed to the intrinsic acoustic differences between tense and lax vowel categories. As reviewed previously, tense vowels are intrinsically long with extended steady-state spectral patterns, while lax vowels are inherently short with dynamic spectral patterns due to their long offglides (Hillenbrand *et al.*, 1995; Lehiste and Peterson, 1961). It is thus conceivable that tense vowels rather than lax vowels are more prone to lengthening in clear-speech modifications, as stretching the spectrally stable tense vowels would largely involve quantity changes, whereas stretching short and dynamic lax vowels would result in vowel quality shifts. Likewise, the dynamic lax vowels rather than tense vowels have more capacity for spectral change in clear speech, since excessive spectral variations may alter the intrinsic steady-state nature of tense vowels.

The current patterns can be interpreted in the context of H & H-based accounts for phonetic variation, which claim that speech production is guided by the parallel principles of "maintenance of phonemic norms" and "contrast enhancement" (Lindblom, 1990; Ohala, 1995; Smiljanić and Bradlow, 2008). Notably, Smiljanić and Bradlow (2008) took these claims to explain their findings of similar plain-to-clear temporal modifications in tense and lax vowels, arguing that the stable durational modification is for the purpose of maintaining the phonemic norms of tense and lax vowels. As the present study involved both temporal and spectral analyses, our findings provide new evidence to extend the claims made by Smiljanić and Bradlow (2008) based on temporal results. Specifically, while the current results support the notion that clear-speech modifications are governed by the principles of "norm maintenance" and "contrast enhancement," our findings further demonstrate that the two principles are realized through compensatory modifications in the temporal and spectral acoustic dimensions and triggered and constrained by articulation. Specifically, "norm maintenance" involves preserving intrinsic phonemic quality for the purpose of keeping categorical (tensity) distinctions, and is thus more essential than "contrast enhancement," which involves phoneme-extrinsic, quantitative modifications (in clear speech) to enhance acoustic distinctiveness.

Indeed, the present results showed that primary acoustic features were preserved to maintain phonemic tensity contrasts, while clear-speech modifications tended to resort to those acoustic attributes that would result in quantitative changes. This is evident from the current static temporal and spectral results, showing that clear-speech modifications were realized by greater tense than lax vowel lengthening, whereas static formant changes for tense and lax vowels were similar. Since tense and lax vowels are primarily differentiated by spectral differences, preserving the spectral distance between these vowels is essential for tense and lax vowels to remain contrastive. Thus, when articulatory constraints prevent peripheral tense vowels from expanding further in the vowel space, clear-speech modifications (which are extrinsic to vowel categorization) resort to the temporal domain that is secondary to the tensity contrast. The current relative measurement results consistently indicate the relation between these two principles. Tense vowels mainly involved temporal modifications in clear speech to preserve the stability of spectral features that characterize tense vowels, while lax vowel modifications primarily lie in the spectral domain to maintain the "short" nature of lax vowels. Taken together, the current findings suggest that clear-speech modifications of tense and lax vowels involve interrelated coordination of acoustic attributes used to preserve phonological categorical contrasts and those used to make quantitative modifications to enhance acoustic salience of speech sounds.

V. CONCLUDING REMARKS

The current findings and theoretical implications have practical applications for informing how critical acoustic attributes can be selectively utilized in hyper-articulated speech in different communicative contexts and how they are weighted in perception to enhance intelligibility by different listener populations (e.g., between clinicians and hearing-impaired listeners, between language instructors and learners, or between care-givers and infants). It is argued that the acoustic cues available to different populations and the speech enhancement strategies they respond to can be quantitatively and qualitatively different (Lam *et al.*, 2012; Smiljanić and Bradlow, 2009). For example, while durational modifications resulting in enhanced acoustic saliency between tense and lax vowels facilitate intelligibility for native listeners in degraded listening conditions or for hearing-impaired audiences (Picheny *et al.*, 1986; Uchanski, 1988), such clear-speech strategies may be detrimental for non-native listeners who would first need to establish awareness of critical spectral cues that mark phonemic tensity differences (Smiljanić and Bradlow, 2009). Similarly, studies on infant-directed speech (IDS) showed that English-speaking mothers utilize spectral rather than durational qualities to signal the category membership for tense and lax vowels, which helps infants to establish these phonemic tensity categories (Werker *et al.*, 2007). On the other hand, although vowel space expansion has been similarly found in clear speech (Ferguson and Kewley-Port, 2002) and IDS (Kuhl, 1997), the IDS modifications which result in overlap of different vowel categories (e.g., /i, ɑ, u/) are found to inhibit intelligibility (Kirchhoff and Schimmel, 2005). These findings are in keeping with the theoretical undertaking of hyper-articulation, indicating that the development of speech enhancement algorithms needs to work under the parallel principles underlying cue-and category-level adaptation on the basis of communicative needs.

ACKNOWLEDGMENTS

This study was funded by a research grant from the Social Sciences and Humanities Research Council of Canada (SSHRC Insight Grant No. 435-2012-1641). We thank Sylvia Cho, Courtney Lawrence, and Xiaojie Zhao from the Language and Brain Lab at Simon Fraser University for their assistance. Special thanks to Dr. Ann Bradlow for her input on measurement methods, and to Dr. Zhaoyan Zhang and the two anonymous reviewers for their valuable suggestions.

[1]For brevity, presentation of the non-significant results involving Gender has been combined: F's and p's refer to the results from multiple sets of ANOVAs, one for each metric. Maximum F and minimum p values are reported.

Assmann, P. F., and Katz, W. F. (**2005**). "Synthesis fidelity and time-varying spectral change in vowels," J. Acoust. Soc. Am. **117**, 886–895.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," IFA Proc. **17**, 97–110.

Boersma, P., and Weenink, D. (**2013**). "Praat: Doing phonetics by computer," Version 5.3.60. www.praat.org (Last viewed 9/3/2015).

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Clopper, C. G., Pisoni, D. B., and de Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Cooke, M., and Lu, Y. (**2010**). "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," J. Acoust. Soc. Am. **128**, 2059–2069.

Crystal, T. H., and House, A. S. (**1988**). "Segmental durations in connected-speech signals: Current results," J. Acoust. Soc. Am. **83**, 1553–1573.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Ferguson, S. H., and Kewley-Port, D. (**2007**). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," J. Speech Lang. Hear. R. **50**, 1241–1255.

Ferguson, S. H., and Quené, H. (**2014**). "Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **135**, 3570–3584.

Gopal, H. S. (**1990**). "Effects of speaking rate on the behavior of tense and lax vowel durations," J. Phonetics **18**, 497–518.

Granlund, S., Hazan, V., and Baker, R. (**2012**). "An acoustic-phonetic comparison of the clear speaking styles of Finnish-English late bilinguals," J. Phonetics **40**, 509–520.

Hazan, V., and Baker, R. (**2011**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130** 2139–2152.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**, 3013–3022.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). "Formants of children, women, and men: The effects of vocal intensity variation," J. Acoust. Soc. Am. **106**, 1532–1542.

Kim, J., and Davis, C. (**2014**). "Comparing the consistency and distinctiveness of speech produced in quiet and in noise," Comput. Speech. Lang. **28**, 598–606.

Kirchhoff, K., and Schimmel, S. (**2005**). "Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition," J. Acoust. Soc. Am. **117**, 2238–2246.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E., Sundberg, U., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **277**, 684–686.

Lam, J., Tjaden, K., and Wilding, G. (**2012**). "Acoustics of clear speech: Effect of instruction," J. Speech, Lang., Hear. Res. **55**, 1807–1821.

Lehiste, I., and Peterson, G. E. (**1961**). "Transitions, glides, and diphthongs," J. Acoust. Soc. Am. **33**, 268–277.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht, the Netherlands), pp. 403–439.

Lu, Y., and Cooke, M. (**2008**). "Speech production modifications produced by competing talkers, babble, and stationary noise," J. Acoust. Soc. Am. **124**, 3261–3275.

Maniwa, K., Jongman, A., and Wade, T. (**2009**). "Acoustic characteristics of clearly spoken English fricatives," J. Acoust. Soc. Am. **125**, 3962–3973.

Moon, S., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.

Nooteboom, S. (**1997**). "The prosody of speech: Melody and rhythm," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Malden, MA), pp. 640–673.

Ohala, J. J. (**1995**). "Clear speech does not exaggerate phonemic contrast," in *Proceedings of the 4th European Conference on Speech Communication and Technology, Eurospeech'95*, pp. 1323–1325.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Port, R. F. (**1981**). "Linguistic timing factors in combination," J. Acoust. Soc. Am. **69**, 262–724.

Reetz, H., and Jongman, A. (**2009**). *Phonetics: Transcription, Production, Acoustics, and Perception* (Blackwell, Chichester, UK), pp. 182–207.

Roesler, L. (**2013**). "Acoustic characteristics of tense and lax vowels across sentence position in clear speech," unpublished Master's thesis, University of Wisconsin-Milwaukee, Milwaukee, WI.

Smiljanić, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**, 1677–1688.

Smiljanić, R., and Bradlow, A. R. (**2008**). "Stability of temporal contrasts across speaking styles in English and Croatian," J. Phonetics **36**, 91–113.

Smiljanić, R., and Bradlow, A. R. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Lang. Linguist. Compass **3**, 236–264.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.

Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.

Tasko, S. M., and Greilick, K. (**2010**). "Acoustic and articulatory features of diphthong production: A speech clarity study," J. Speech Lang. Hear. Res. **53**, 84–99.

Traumüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**, 97–100.

Uchanski, R. M. (**1988**). "Spectral and temporal contributions to speech clarity for hearing impaired listeners," unpublished Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (**2007**). "Infant-directed speech supports phonetic category learning in English and Japanese," Cognition **103**, 147–162.

Wouters, J., and Macon, M. W. (**2002**). "Effects of prosodic factors on spectral dynamics. I. Analysis," J. Acoust. Soc. Am. **111**, 417–427.