

ACOUSTIC CHARACTERISTICS OF CLEARLY SPOKEN ENGLISH TENSE AND LAX VOWELS

Keith King Wui Leung¹, Allard Jongman², Yue Wang¹, Joan A. Sereno²

1. Language and Brain Lab, Department of Linguistics, Simon Fraser University, Canada

2. KU Phonetics and Psycholinguistics Lab, Department of Linguistics, University of Kansas, USA

kwl23@sfu.ca, jongman@ku.edu, yuew@sfu.ca, sereno@ku.edu

ABSTRACT

The acoustic features of clearly produced vowels have been widely studied, but a less explored area concerns the difference in the adaptations of tense and lax clear vowels. This study explored the clear production of three pairs of English tense and lax vowels (/i-I/, /ɑ-A/, /u-U/) to determine whether tense vowels show a larger clear versus conversational speech difference than lax vowels. Vowel space, individual formant frequency values, dynamic formant information and vowel duration of tense and lax vowels were examined. Results suggest there was more conversational-to-clear vowel lengthening for tense vowels than for lax vowels. However, an opposite effect was found for spectral measures. Lax vowels yielded greater vowel space expansion, formant frequency change, and dynamic formant movement than tense vowels in clear speech.

Keywords: Clear speech, English tense and lax vowels, acoustic analysis

1. INTRODUCTION

Speakers naturally adopt a clear, hyperarticulated speaking style when speaking in a noisy environment or with hearing-impaired listeners, in order to produce speech that is more audible and intelligible. This involves modifications of acoustic characteristics of speech sounds. A number of studies have investigated the acoustic properties of clearly produced English vowels.

In terms of duration, clear vowels are generally longer than conversational vowels [1, 2, 3]. For spectral changes, vowel space was found to be expanded in clear speech compared to conversational speech in terms of perimeter [1] and dispersion [2] measurements. Furthermore, formant frequency changes for individual vowel tokens reveal which formants particularly contribute to vowel space expansion in clear speech. For the general conversational-to-clear modifications, F1 was found to increase in clear speech regardless of vowel category [3], indicating that speakers produced clear vowels with a larger mouth opening.

The F2 of clearly produced front vowels became higher, while that of clearly produced back vowels became lower, indicating more tongue fronting of front vowels and more tongue retraction of back vowels in clear speech [3]. The F3 of clearly produced vowels was lower than that of naturally produced vowels [4], suggesting that speakers' lips were more rounded in clear speech. In addition to the steady-state spectral features, clearly produced vowels exhibit greater dynamic formant movement (in terms of spectral change and spectral angle) than naturally produced vowels [1].

Together, these results indicate that the production of clear speech involves greater degrees of articulatory movement, resulting from greater mouth opening, more peripheral tongue positions, and greater degree of lip protrusion; and therefore also takes longer to produce. Interestingly, these more "extreme" articulatory features for clear speech are also what characterize tense vowels in English relative to lax vowels. Acoustically, tense vowels also involve more peripheral formant frequencies and longer durations than lax vowels [5]. One subsequent question is whether and how the clear speech effects observed previously may interact with tensivity effects. It is conceivable that tense vowels show more conversational-to-clear modifications than lax vowels since tense vowels involve longer articulatory excursions (i.e., more room for variability). However, little research has explored this issue. The only report addressing the tensivity factor involved a small sample size and lacked statistical analysis, but it did suggest a more expanded vowel space and increased duration for tense vowels than for lax vowels in clear speech [6].

Therefore, the current study focuses on the acoustic characteristics of clear speech in English tense and lax vowels. Based on the abovementioned studies, we hypothesized that tense vowels would demonstrate a greater clear versus conversational speech difference than lax vowels. More specifically, tense vowels would have greater vowel space expansion, formant frequency change, dynamic formant movement and vowel lengthening than lax vowels in clear speech.

2. METHOD

2.1. Participants

Ten native speakers (5 males) of Western Canadian English aged 17-30 (mean: 22.4) were recruited. This English dialect exhibits /a/ and /ɔ/ merger [7], and thus speakers should produce the vowel in “cod” as the target vowel /a/, instead of /ɔ/. They reported no history of speech or hearing impairments.

2.2. Materials

Six English words “keyed, kid, cod, cud, coed” and “could” carrying the target vowels /i/, /ɪ/, /a/, /ʌ/, /u/ and /ʊ/, respectively, in the context of /kVd/ were used. The production of each token was recorded in isolation in conversational and clear speaking styles.

2.3. Procedures

The stimuli were recorded in a sound-attenuating booth at a sampling rate of 48 kHz. A Shure KSM microphone was placed at a 45 degree angle, about 20cm away from the speaker’s mouth. MATLAB programs were designed to provide prompts, which were displayed on a computer screen.

For the elicitation sessions, 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. They were instructed to speak naturally first when a prompt showed up on the screen. Then, the program would “guess” and indicate on the screen what they produced. If a participant indicated that the guess was correct by clicking a box on the screen, the program would move on to the next stimulus. Otherwise, the program would instruct the participant to repeat the stimulus as clearly as possible (see Maniwa et al. [8] for details of this procedure). In the acoustic analyses, the productions in response to the initial prompts served as the “conversational speech”, whereas the repeated productions constituted the “clear speech”. A total 144 productions (12 repetitions X 6 words X 2 styles) were obtained for each speaker.

2.4. Acoustic Analyses

All measurements were obtained using Praat [9].

2.4.1. Formant frequencies at steady state

The frequency values of F1, F2 and F3 were taken from the vowel midpoint. The values were converted to critical-band rate in Bark:

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

where Z is the critical-band rate in Bark and f is the raw frequency value in Hertz [10].

2.4.2. Vowel space

The vowel space was represented using the Bark Difference Metric [11]. $Z_3 - Z_2$ was used to model vowel advancement (i.e., Bark-converted F3 minus Bark-converted F2). $Z_3 - Z_1$ was used to model vowel height (i.e. Bark-converted F3 minus Bark-converted F1) [12]. “Perimeter” of vowel spaces was measured as the sum of Euclidean distances between adjacent vowels [1]. “Dispersion” was calculated as the mean of each vowel token’s Euclidean distance from the centroid of speaker’s triangular vowel space [2].

2.4.3. Dynamic formant movement

Based on Ferguson & Kewley-Port [1], “spectral change (λ)” was obtained from the sum of F1, F2 and F3 absolute frequency shift:

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

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, Bark-converted F2 and Bark-converted F3 at 20% and 80% of the vowel portion.

“Spectral angle (Ω)” represented the sum of the absolute values of the F1, F2 and F3 angles.

$$(3) \quad \Omega = |\theta_{F1}| + |\theta_{F2}| + |\theta_{F3}|$$

The angle of each formant was obtained by

$$(4) \quad \theta_{Fn} = \arctan[(Z_{n(80)} - Z_{n(20)})/d],$$

where n represents the number of the formant and d (in deciseconds) = (time₈₀ - time₂₀)/100

2.4.4. Vowel Duration

Vowel duration measures were made from the spectrogram and waveform. Vowel onset was defined as the onset of voicing as shown by strong vertical striations in the spectrogram, and the onset of periodicity in the waveform. Vowel offset was taken at the closure of /d/, corresponding to a cessation of high-frequency energy.

3. RESULTS

Mean data for all metrics, except for vowel space measures, were separately submitted to mixed-design ANOVAs with Style (clear, conversational), Tensity (tense, lax), and Vowel [high front (/i/ and /ɪ/, high back (/u/ and /ʊ/), low (/a/ and /ʌ/)] as within-subject factors, and Gender (male, female) as a between-subject factor. For vowel space measures, data were submitted to mixed-design ANOVAs with Style (clear, conversational) and Tensity (tense, lax) as within-subject factors, and Gender (male, female) as a between-subject factor.

3.1. Formant frequencies at steady state

For F1 and F3, the ANOVAs revealed no significant effect of Style or any interaction related to Style. For F2, the interaction of Style by Tensity [F(1,8) = 5.39, $p = .049$] and Style by Vowel [F(2,16) = 6.90, $p = .007$] were significant. Repeated measures ANOVAs with Style and Tensity as factors were conducted for each vowel pair as follow-up analyses. A significant interaction of Style by Tensity was obtained only for high front vowels [F(1,9) = 11.5, $p = .008$]. Post-hoc pairwise comparisons revealed that only the F2 of clearly produced /i/ (14.7 Bark) was higher than that of naturally produced /i/ (13.3 Bark) ($p = .008$). For high back vowels, there was a significant main effect of Style [F(1,9) = 9.18, $p = .014$], but the interaction of Style by Tensity was not significant. Neither the main effect of Style nor the interaction of Style by Tensity was significant for low vowels.

3.2. Vowel space

For perimeter, the ANOVA revealed a significant main effect of Style [F(1,8) = 5.52, $p = .047$]. The overall perimeter of clear vowels (10.0 Bark) was longer than that of conversational vowels (9.49 Bark). The interaction of Style by Tensity was marginally significant [F(1,8) = 4.70, $p = .062$]. Further analyses with Style as a factor separately for tense and lax vowels revealed a significant difference for lax vowels only ($p = .035$), with lax vowels yielding a significantly shorter perimeter (5.63 Bark) in conversational speech than in clear speech (6.37 Bark) (Figure 1). For dispersion, the main effect of Style [F(1,8) = 4.39, $p = .069$] and the

interaction of Style by Tensity [F(1,8) = 4.50, $p = .067$] approached significance. The near-significant interactions were likely to be caused by the more noticeable increase in dispersion for lax vowels in clear than in conversational speech.

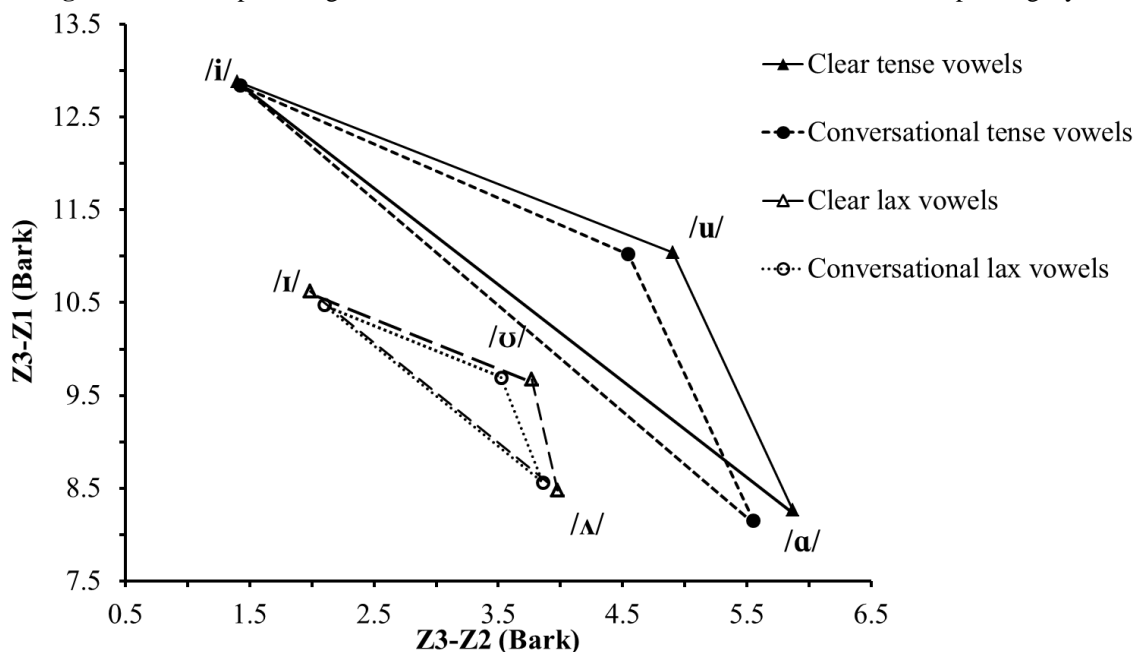
3.3. Dynamic metrics

For spectral change, the ANOVA yielded a significant interaction of Style by Tensity [F(1,8) = 8.69, $p = .019$]. Further analyses with Tensity as a factor separately for each speaking style yielded a significant difference only for clear speaking style ($p = .006$), suggesting that tense vowels had a smaller spectral change (1.58 Bark) than lax vowels (2.29 Bark) in clear speech. Moreover, further analyses with Style as a factor separately for tense and lax vowels revealed a significant difference for lax vowels only ($p = .003$). Clearly produced lax vowels had a greater spectral change (2.29 Bark) than conversationally produced ones (1.94 Bark). For spectral angle, the ANOVA revealed no significant effect of Style or any interaction related to Style.

3.4. Vowel duration

There was a significant interaction of Style by Vowel by Tensity [F(2,16) = 7.53, $p = .005$]. As a follow-up of the three-way interaction, repeated measures ANOVAs with Style and Tensity as within-subject factors were conducted separately for each vowel pair. The interaction of Style by Tensity was significant for all vowel pairs [high front: F(1,9) = 15.4, $p = .004$; low: F(1,9) = 5.50, $p = .044$; high back: F(1,9) = 20.1, $p = .002$]. The clear productions of all vowels (/i/: 334ms; /ɪ/: 177ms; /a/:

Figure 1: Vowel space diagram for tense and lax vowels in clear and conversational speaking styles



304ms; /ʌ/: 177ms; /u/: 360ms; /ʊ/: 193ms) were longer than their conversational productions (/i/: 263ms; /ɪ/: 151ms; /ɑ/: 262ms; /ʌ/: 157ms; /u/: 287ms; /ʊ/: 162ms) (/i/: $p = .001$; /ɪ/: $p = .001$; /ɑ/: $p = .003$; /ʌ/: $p = .001$; /u/: $p = .001$; /ʊ/: $p = .004$). The duration increases for each vowel pair were further compared in t tests. The results showed that the increase was greater in magnitude for tense vowels than for lax vowels [/i/ (71ms) vs. /ɪ/ (26ms): $t(9) = 3.92$, $p = .004$; /ɑ/ (42ms) vs. /ʌ/ (20ms): $t(9) = 2.35$, $p = .044$; /u/ (73ms) vs. /ʊ/ (31ms): $t(9) = 4.49$, $p = .002$].

4. DISCUSSION

The results of this study showed that F2 and vowel duration yielded greater conversational-to-clear modifications for tense vowels. For vowel duration results, not only were clearly produced vowels on average longer than conversationally produced vowels, consistent with previous findings [1, 2, 3], but also the lengthening of vowels from conversational to clear speaking style was greater for tense vowels than for lax vowels. This clearly supported the hypothesis of this study.

For F2 results, clearly spoken high front vowels had a *higher* F2 than conversationally produced ones, reflecting greater tongue advancement in clear speech. On the other hand, high back vowels had a *lower* F2 in clear speech than in conversational speech, indicating greater tongue retraction in clear speech. These patterns are in line with previous findings of a more peripheral F2 [3] and more stretched vowel space [1, 2] in clear speech. However, detailed analyses show that these general patterns may not apply to the tense and lax vowels to the same degree. For high front vowels, the non-significant effect of speaking style for /i/ suggested that the F2 values of the conversational and clear productions were not different. An increase in F2 was found only for the lax vowel /ɪ/. For high back vowels, the F2 lowering for both the tense /u/ and lax /ʊ/ contributed to the clear speech effects. Consequently, more lax vowels than tense vowels actually demonstrated conversational-to-clear F2 modifications (/ɪ/, /ʊ/ vs. /u/), suggesting that speaking style had a stronger effect on lax vowels than on tense vowels, which did not support the hypothesis of the present study.

The findings of vowel space measures were similar to the F2 results. Since there was no significant effect of speaking style for F1 and F3, any vowel space size change was mainly caused by the changes in F2 between conversational and clear speech. This became apparent by making reference to the vowel plots (Figure 1), where vowel space

expansion in clear speech occurred mainly along the horizontal axis which represents the difference between Bark-converted F3 and F2. As a result, vowel space measurement results should be closely linked to the F2 findings. In fact, vowel space perimeter results showed that both tense and lax vowel perimeter increased in clear speech, which can also be seen in Figure 1, but there was also a tendency for that increase to occur mainly in lax vowels. The same trend was also observed in dispersion results. Both vowel space measures, therefore, suggested that vowel space expansion in clear speech was greater for lax vowels than for tense vowels, similar to the F2 results.

To address the lack of clear speech effects for /i/, we may consider its position in the vowel space. As the highest and most front vowel in English, /i/ already occupies a peripheral position in the vowel space, thus there is less room for this vowel to move further away from the neutral position of the vowel space compared to its lax vowel counterpart /ɪ/. The articulatory excursion for /i/ may be longer, but it does not necessarily leave more room for variability due to its extreme target position. Therefore, /i/ tends to be more spectrally stable and less susceptible to acoustic variability, corroborating previous findings [13]. However, it then raises the issue of why the high back tense vowel /u/ appeared to have noticeably more conversational-to-clear speech formant modifications than /i/. One possibility is that the /u/ fronting typically exhibited in Canadian English and West Coast American English [7, 14] leaves more room for speakers to further retract their tongue when they enunciate the vowel. Consistent with the static measures, the spectral change metric also demonstrated a greater spectral change only for clearly produced lax vowels as compared to naturally produced lax vowels. These findings could possibly be explained by the relatively less extreme articulatory positions for lax vowels which may allow more dynamic spectral changes [15].

Taken together, these findings suggest clear speech effects for both tense and lax vowels, as demonstrated by modifications in the temporal domain for tense vowels and in the spectral domain for lax vowels, presumably due to articulatory constraints.

5. ACKNOWLEDGMENTS

We thank Sylvia Cho, Beverly Hannah, Lisa Tang, and Xiaojie Zhao for their assistance. We also thank Dr. Ann Bradlow for her input on measurement methods. This research has been funded by a SSHRC Insight Grant (435-2012-1641).

6. REFERENCES

- [1] Ferguson, S. H., Kewley-Port, D. 2007. Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J. Speech Lang. Hear. Res.* 50, 1241-1255.
- [2] Bradlow, A., Torretta, G., Pisoni, D. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20, 255-272.
- [3] Ferguson, S. H., 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.
- [4] Lu, Y., Cooke, M. 2008. Speech production modifications produced by competing talkers, babble, and stationary noise. *J. Acoust. Soc. Am.* 124, 3261-3275.
- [5] Kondaurova, M. V., Bergeson, T. R. 2012. Effects of deafness on acoustic characteristics of American English tense/lax vowels in maternal speech to infants. *J. Acoust. Soc. Am.* 132, 1039-1049.
- [6] 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. *J. Speech Lang. Hear. Res.* 29, 434-446.
- [7] Clopper, C. G., Pisoni, D. B., de Jong, K. 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *J. Acoust. Soc. Am.* 118, 1661-1676.
- [8] Maniwa, K., Jongman, A., Wade, T. 2009. Acoustic characteristics of clearly spoken English fricatives. *J. Acoust. Soc. Am.* 125, 3962-3973.
- [9] Boersma, P., & Weenink, D. 2013. *Praat: doing phonetics by computer [Computer program]. Version 5.3.60.* <http://www.praat.org/>
- [10] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.* 88, 97-100.
- [11] Syrdal, A. K., 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.
- [12] Thomas, E. R., Kendall, T. 2007. *NORM: The vowel normalization and plotting suite.* <http://ncslaap.lib.ncsu.edu/tools/norm/> [13]
Hillenbrand, J.M., Clark, M.J., & Nearey, T.M. 2001. Effects of consonant environment on vowel formant patterns. *J. Acoust. Soc. Am.* 109, 748-763.
- [14] Schneider, E.W. 2006. English in North America. In B.B. Kachru, Y. Kachru, & C.L. Nelson, *The Handbook of World Englishes*. Malden, MA; Oxford: Blackwell.
- [15] Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099-3111.