EFFECT OF SPEECH RATE ON A VOWEL CONTRAST AND IMPLICATIONS FOR SECOND LANGUAGE TRAINING

Miki Shrosbree[a], *Tetsu Narita*[b], *Izabelle Grenon*[b] *& Mikio Kubota*[b]

[a]Sophia University, Japan; [b]Seijo University, Japan
mikishros@gmail.com; naritets@hotmail.co.jp; igrenon@seijo.ac.jp; mkubota@seijo.ac.jp

ABSTRACT

The present study evaluates talkers' ability to consciously modify the duration and Euclidean distance of the English high front vowel contrast in a predictable manner by changing their speech rate. Our results concur with previous studies showing that vowel duration is highly sensitive to changes in speech rate, and predictably decreases as speech rate increases. Although some changes in the contrastive distance between the English high front tense and lax vowels were observed, and partly reached statistical significance, a closer look at individual data failed to reveal any reliable pattern, such as a consistent decrease in the vowel distance as speech rate increases. Implications for the design and improvement of second language training programs with difficult non-native contrasts are discussed.

Keywords: speaking rate, speech contrast, production, second language training, English

1. INTRODUCTION

The multi-talker high-variability paradigm is a training program used to improve second language (L2) learners' perception of a difficult L2 contrast. It consists of presenting isolated words uttered by various speakers and contrasting the target sounds in different linguistic contexts through a forced-choice identification task. Although this method significantly improves L2 learners' perception of L2 contrasts, the learners generally fail to achieve native-like accuracy [5, 6, 12]. Accordingly, attempts have been made to improve the paradigm through acoustic manipulation of the training stimuli, either by increasing the variability of unreliable acoustic cues [12] or increasing the contrast of the most reliable acoustic cue [5]. The perceptual fading technique, in particular, consists of presenting L2 learners with fully enhanced exemplars of the target speech sounds at training onset and gradually reducing the contrastiveness of the stimuli presented as training progresses [5]. If we were to use this technique with a vowel contrast, for instance, we would want to present vowels that are most contrastive (i.e. with the longest Euclidean distance) at the first training session, and vowels that are least contrastive (i.e. with the shortest Euclidean distance) at the last training session. Given that acoustic manipulation can be overly time-consuming, finding an easier way to prepare enhanced (most contrastive) and unenhanced (least contrastive) stimuli may be advantageous for the future design of such paradigms.

Vowel duration is highly sensitive to changes in speech rate and usually decreases as speech rate increases [2, 3]. While the vowel space of unstressed vowels is compressed at fast speech rate [9], some studies with stressed vowels reported no correlation between speech rate and the size of vowel space, whether speech rate was varied within [2, 3] or between talkers [11]. Nevertheless, spectral changes have been observed to occur with changes in vowel duration through the use of different speaking styles (clear versus citation speech) [7]. Crucially, the above-mentioned studies were designed to evaluate whether changes in speech rate *naturally* affect the production of vowels. It is yet unclear if we can purposefully *induce* predictable changes in vowel production by using a word-reading task and giving specific instructions to the talkers. In addition, since the perceptual distance between frequencies is distorted at low frequencies (around first formant of vowels), a measure of the spectral distance between vowels may be misleading about their actual perceptual distance, which may be better represented with the use of a sensory scale (bark).

Accordingly, the current study evaluates whether recording talkers at different speech rates using a word list format may suffice to induce an increase in the variability of misleading cues while enhancing the most robust ones. Specifically, we investigated whether a change in speech rate may induce predictable changes in the Euclidean distance (in bark) between the English vowels /i/ and /ɪ/, while increasing their duration variability.

Since spectral differences are the most robust cues used by native English speakers to contrast these vowels, while vowel duration is mostly ignored [4], the most desirable outcome for the design of a perceptual fading training paradigm would be to see the perceptual distance—the crucial acoustic cue—decrease as speech rate increases, and to have the duration of the vowels—the unreliable cue—vary from equally long to equally short.

2. METHODOLOGY

2.1. Participants

Participants were 4 native Canadian English speakers (2 females, 2 males), aged between 17 and 37.

2.2. Materials

The corpus consisted of 38 word pairs contrasting the English vowels /i/ and /ɪ/ in various linguistic contexts, as traditionally used in the design of a multi-talker high-variability paradigm [12]. Twenty were simple (CVC) minimal pairs (*heal-hill*) and 18 were complex monosyllabic (CVCC, CCVC) or disyllabic pairs (*peeking-picking*).

2.3. Recordings

The talkers were recorded directly to computer at a sampling rate of 22 KHz. The word pairs were read from a printed list. The researcher supervising the recordings was watchful of any list reading effect, and made the talkers repeat any word pairs, in reverse order, whenever necessary. The list of words was first recorded at a speech rate that felt most comfortable to each talker, which we refer to as the normal speech rate. The second time the list was read, talkers were instructed to read each word carefully, as if to pronounce it clearly for someone who did not clearly hear the difference in vowel quality, and to try to make the vowel duration equally long for each word of the same pair. Finally, the third time the list was read, talkers were instructed to read the list faster than their normal speech rate, as if speaking in a hurry or excitedly, and to try to make the vowel duration equally short for words of the same pair.

2.4. Acoustic measurements

Vowel duration was measured from the first pitch period to the last of each vowel, as determined through visual inspection of the wideband spectrogram and speech waveform in Praat [1], using standard criteria [8]. The first (F1) and second formants (F2) were measured by hand corrected LPC analyses in Praat.

2.5. Calculation of vowel distance

Since we were most interested in the perceptual distance between the vowels, not the distance between their acoustic realizations, all the formants were converted in bark, a tonotopic sensory scale [13]. The formant values in Hz were converted in bark using the formula proposed by Traunmüller [10] and reported in (1) below, where f is the formant converted (i.e. F1 or F2 in Hz).

(1) $$z(bark) = \left\lfloor \frac{26.81f}{1960+f} \right\rfloor - 0.53$$

The Euclidean distance d between vowels of the same word pair was calculated using the transformed Pythagorean theorem in (2):

(2) $$d = \sqrt{[(F1_1 - F1_2)^2 + (F2_1 - F2_2)^2]}$$

where $F1_1$ - $F1_2$ corresponds to the difference between the F1 value of the tense vowel and that of its lax counterpart, while $F2_1$ - $F2_2$ is the difference between the F2 values of the same vowel pair.

3. RESULTS

Table 1 provides a summary of the mean duration of the tense and lax vowels, and the average distance between the vowels (in bark and in Hz) at slow, normal and fast speech rate for each talker.

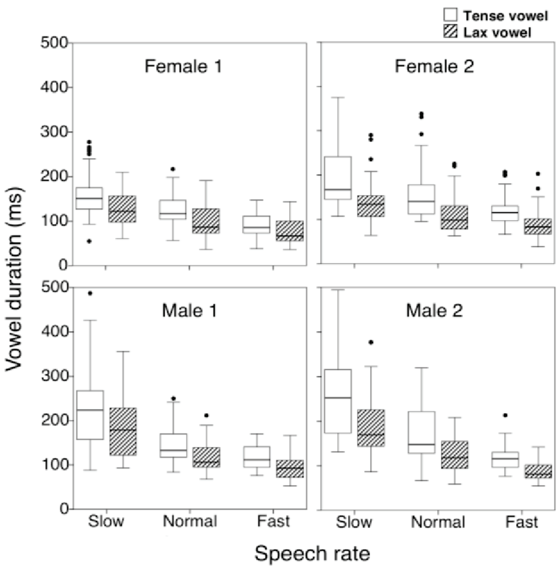
Table 1: Mean vowel distance (in bark and hertz) and vowel duration (tense and lax) at slow, normal and fast speech rate for each talker.

	Distance (bark)	Distance (Hz)	Duration /i/ (ms)	Duration /ɪ/ (ms)
Female 1				
Slow	2.26	589	162	127
Normal	2.11	542	129	100
Fast	1.91	502	90	76
Female 2				
Slow	2.52	722	200	139
Normal	2.44	669	161	112
Fast	2.48	706	123	90
Male 1				
Slow	1.42	284	228	188
Normal	1.52	336	147	119
Fast	1.68	389	116	94
Male 2				
Slow	2.21	409	265	190
Normal	2.84	507	171	121
Fast	1.95	445	117	87

A multivariate analysis of variance (MANOVA) with speech rate as the independent variable and the duration of the tense vowel, duration of the lax vowel and distance in bark as the dependent variables, reports a significant effect of speech rate on the production of the vowels (F[6, 904] = 29.786, p < .001). Separate analyses of variance (ANOVA) reveal a significant effect of speech rate on the duration of the tense vowel (F[2, 453] = 98.329, p < .001), and on the duration of the lax vowel (F[2, 453] = 90,613, p < .001), and a less significant effect of speech rate on the distance between the vowels (F[2, 453] = 3.920, p < .05). Post-hoc tests with Bonferroni correction confirmed a significant effect of change in speech rate (from slow to normal and from normal to fast) on the duration of both vowels (p < .001). However, the effect of speech rate on the vowel distance was significant only from normal to fast speech (p < .05).

We can see in Figure 1 that the duration of the tense and lax vowels decreases steadily as speech rate increases, and consistently for each of our talkers. Although the duration of the tense vowel is, on average, systematically longer than its lax counterpart, as reported in Table 1, the duration of the vowels highly overlap, even within the production of the same talker, as illustrated in Figure 1.

Figure 1: Boxplots of the tense and lax vowel duration as a function of speech rate for each talker.



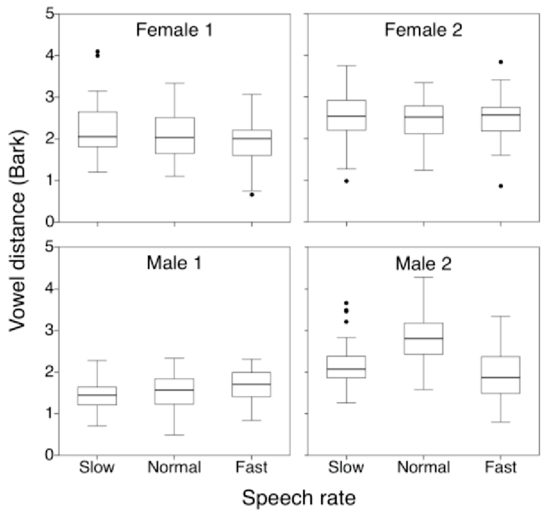
The fact that the tense and lax vowels are almost perfectly correlated ($r = .90$), as reported in Table 2 below, suggests that both vowels are equally affected by a change in speech rate. This is somewhat surprising in our study, since our talkers were specifically asked to try to equate the duration of the vowels at slow and fast speech rate. Thus, although each talker was able to alter the duration of both vowels by modifying their speech rate, they were unable to equate the duration of the vowels at slow or fast speech rate.

Table 2: Intercorrelations between dependent variables (all talkers).

	1	2	3	4
1. Distance (bark)	—	**.79**	-.01	-.08
2. Distance (Hz)		—	-.07	-.15
3. Duration tense vowel			—	**.90**
4. Duration lax vowel				—

Finally, even though the effect of speech rate on the vowel distance (in bark) was slightly significant from normal to fast speech rate, an analysis of individual data, as illustrated in Figure 2, suggests that the effect of speech rate on the contrastive distance between vowels was not always in the direction anticipated. For instance, Table 1 and Figure 2 show a slight decrease in the vowel distance as speech rate increases for Female 1, but the opposite pattern for Male 1, for whom the distance between the vowels increases as his speech rate increases. And for Male 2, the distance between the vowels is higher at normal speech rate than at either slow or fast speech rate.

Figure 2: Boxplots of vowel distance (in bark) as a function of speech rate for each talker.



To sum up, although speech rate somewhat affected the contrastive distance between the vowels, the impact of speech rate on vowel distance is mostly unpredictable: the distance between the tense and lax vowels does not reliably decrease as speech rate increases. On the other hand, a gradual increase in speech rate triggers a systematic decrease in the duration of both the tense and lax vowels, which may provide an additional confound to the already overlapping distribution of the tense and lax vowel duration, and may be useful in the design of an adaptive computer-based L2 training program, as discussed in more details in the next section.

4. DISCUSSION

The present study examined whether an increase in speech rate (from slow to normal to fast) could induce a gradual decrease in the perceptual distance of the (Canadian) English /i/ and /ɪ/ contrast, while considerably increasing vowel duration variability. Our results demonstrate that while the distance between the high front vowels—the most robust cue to distinguish these vowels—does not reliably decrease as speech rate increases, the duration of vowels—an unreliable cue—consistently decreases as speech rate increases.

Accordingly, recording talkers at different speech rates could potentially be an efficient way of creating stimuli with vowels varying in duration from short to long for the purpose of a multi-talker high-variability training paradigm such as the one designed by Wang and Munro [12]. It is relevant to note, though, that despite being instructed to do so, talkers were unable to equate the duration of the vowels at slow and fast speech rate. However, given that the duration of the tense and lax vowels highly overlap at each speech rate and for each talker, the additional variability induced by a change in speech rate may suffice to draw L2 learners' attention away from vowel duration and help them focus on the most robust cue, in this case, spectral differences.

Our results suggest that changing speech rate is not a reliable way of manipulating spectral differences between vowels. Although some changes were observed in the Euclidean distance of the vowels (in bark and in hertz), these changes are too unpredictable to provide a mean to create stimuli reliably varying from most contrastive to least contrastive by modifying speech rate from slow to fast. While the distance between vowels decreased as speech rate increased for at least one speaker (Female 1), the reverse was observed for Male 1, and other patterns occurred for Female 2 and Male 2. Although our results do not support the use of speech rate as a way to create fully enhanced and unenhanced stimuli, they highlight the fact that different talkers may produce more contrastive vowels than others. Hence, instead of manipulating vowel formants for the purpose of a perceptual fading training paradigm, it may be relevant to carefully plan the order of presentation of each talker. Since only one talker is typically presented per training session for this kind of training paradigm, capitalizing on the contrastiveness in vowel distance naturally occurring between talkers may provide a time-saving way of preparing an adaptive program that targets to present the most enhanced (contrastive) tokens at training onset.

5. ACKNOWLEDGMENTS

Research supported by a JSPS Postdoctoral Fellowship for Foreign Researchers (P10768) to Izabelle Grenon.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2007. *Praat: Doing phonetics by computer* [computer program].

[2] Fourakis, M. 1991. Tempo, stress, and vowel reduction in American English. *J. Acoust. Soc. Am. 90*, 1816-1827.

[3] Gay, T. 1978. Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Am. 63*, 223-230.

[4] Grenon, I. 2010. *The Bi-Level Input Processing Model of First and Second Language Perception*. Doctoral dissertation. Univ. of Victoria. Canada.

[5] Iverson, P., Hazan, V., Bannister, K. 2005. Phonetic training with acoustic manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *J. Acoust. Soc. Am. 118*, 3267-3278.

[6] Logan, J.S., Lively, S.E., Pisoni, D.B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *J. Acoust. Soc. Am. 89*, 874-886.

[7] Moon, S.J., Lindblom, B. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *J. Acoust. Soc. Am. 96*, 40-55.

[8] Peterson, G.E., Lehiste, I. 1960. Duration of syllable nuclei in English. *J. Acoust. Soc. Am. 32*, 693-703.

[9] Stetson, R.H. 1951. *Motor Phonetics: A Study of Speech Movements in Action.* Amsterdam: North-Holland.

[10] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am. 88*, 97-100.

[11] Tsao, Y.C., Weismer, G., Iqbal, K. 2006. The effect of intertalker speech rate variation on acoustic vowel space. *J. Acoust. Soc. Am. 119*, 1074-1082.

[12] Wang, X., Munro, M.J. 2004. Computer-based training for learning English vowel contrasts. *System 32*, 539-552.

[13] Zwicker, E. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *J. Acoust. Soc. Am. 33*, 248.