The Effects of Speaking Rates and Vowel Length on Formant Movements in Japanese

Yukari Hirata and Kimiko Tsukada
Colgate University and University of Alabama at Birmingham

1. Introduction

This study examined the extent to which variation in speaking rates and phonemic vowel length affect formant frequencies and movements in Japanese vowels. It is known that vowel targets are not always reached at fast rate, i.e., vowel targets might be undershot at fast rate. Vowel duration is one of the factors that determine the degree of formant undershoot (Moon & Lindblom, 1994; Lindblom, 1963). Effect of speaking rates on formant movements has been examined in English (Gay, 1978), Swedish (Engstrand, 1988), Dutch (Van Son & Pols, 1992), and Korean (Magen & Blumstein, 1993). However, very few studies have examined how formants of Japanese vowels vary as a function of speaking rates with the exception of Imaizumi, Kiritani, Hirose, Togami, & Shirai (1987).

Japanese has five short (/i e a o u/) and five long (/i: e: a: o: u:/) vowels, and vowel length is phonemic, e.g., /i/ ‘stomach’ vs. /i:/ ‘good.’ The duration ratio of short to long vowels is 1 : 2.4-3.2 (Han, 1962; Tsukada, 1999; Ueyama, 2000).

The first question in the present study was whether long vowels are more likely than short vowels to reach their targets or resist coarticulation with neighboring segments. If this were the case, formants of long vowels would occupy peripheral areas in the F1-F2 vowel space, compared to those of short vowels. The second question was whether formants of each vowel systematically vary as a function of speaking rates. If vowel duration is a factor that determines the degree of formant undershoot, as suggested by Moon & Lindblom (1994) and Lindblom (1963), formants of vowels spoken at slower speaking rates would occupy more peripheral areas in the acoustic vowel space than those spoken at faster rates.

2. Method

2.1 Participants

Two male native speakers of Japanese participated in the present experiment. The participants were 35 and 31 years old, and were students at the University of Alabama at Birmingham. They were from Ibaraki and Gifu prefectures, respectively, and reported speaking standard Japanese. The participants arrived in the United States at the age of 32 and 28, and had lived in the United States for three and two years, respectively. Their self-reported proficiency of spoken English was the level of being comfortable in daily conversation. Both reported that they had difficulty understanding English when it was spoken fast or with strong dialects. They reported using English and Japanese for roughly 60% and 40% of the time, respectively.

2.2 Stimuli

Stimuli were 10 nonwords including five short and five long Japanese vowels, i.e., /mimi/-/mi:mi/, /meme/-/me:me/, /mama/-/ma:ma/, /momo/-/mo:mo/, and /mumu/-/mu:mu/. The vowels in the first syllables of these nonwords had a pitch accent, and those vowels were measured for this study. The order of the nonwords was randomized. They were written in a handout with Japanese *katakana* orthography. Two filler tokens were added at the beginning and the end of each list at each speaking rate to minimize the effect of list intonation. A carrier sentence /sokowa ____ to kaite arimasu/ ‘___ is written there.’ was also written in each page of the handout, and participants were asked to read the nonwords with the carrier sentence.

© 2004 Yukari Hirata and Kimiko Tsukada. Proceedings of the 2003 Texas Linguistics Society Conference, ed. Augustine Agwuele, Willis Warren, and Sang-Hoon Park, 73-85. Somerville, MA: Cascadilla Proceedings Project.

2.3 Procedure

Each speaker produced the stimuli at three speaking rates. Instructions for producing the three rates were those used in Port (1977), and were written in a handout. The "normal" rate was defined as a tempo that is relaxed and comfortable for the speaker. The "slow" rate was defined as the slowest tempo possible that the speaker can produce while keeping a sentence flowing together, i.e., without obviously inserting breaks between words. The "fast" rate was defined as the fastest tempo possible without making an excessive number of errors. Beyond these instructions, the rates spoken by each speaker were self-determined. General instructions were given to the participants in Japanese, but the definition of the three rates was given both in English and in Japanese so that no misinterpretation should occur. Prior to recording, the speakers practiced reading sentences with a wide range of speaking rate.

The speakers read the lists of stimuli at the normal, then the slow, and finally the fast rate. At each rate, the speakers read the same stimuli three times in a differently randomized order. The materials were recorded using a digital audio tape recorder (Sony TCD-D10PROII) with a stereo microphone (Shure SM48) in a sound booth. The recorded utterances were digitized at a 22.05 kHz sampling rate with an automatic 11 kHz low pass filter setting and a 16-bit quantization using the Kay Elemetrics' Computer Speech Lab (CSL 4300B).

2.4 Analysis

Acoustic analysis was conducted on a total of 180 tokens (2 speakers x 10 vowels x 3 rates x 3 repetitions). The vowels of interest were phonetically labeled using EMU speech database system. The first two formant frequencies at the midpoint and the formant trajectory over the entire vowel duration were analyzed. The first two formant center frequencies were automatically tracked in ESPS/Waves (The settings were: 12th order linear predictive coding analysis, cosine window, 49-ms frame size, and 5-ms frame shift).

3. Results

3.1 F1 and F2 midpoint values

Figure 1 shows effect of vowel length on the F1 and F2 midpoint values of all five vowel types. The vertical axis represents F1, and the horizontal axis represents F2. The radius of the ellipse is 2.4 times the standard deviations of the mean, covering 95% of the data points. The vowel symbols indicate the centroids and they are the average values of those distributions. The circles with broken lines represent five short vowels averaged across three rates, and the circles with solid lines represent five long vowels averaged across three rates. Effect of vowel length was shown in the long vowels occupying a more peripheral portion of the F1-F2 vowel space than the short counterparts. For example, the long /i/ and /e/ were more fronted (peripheral) than the short counterparts in the dimension of F2. There was an overlap between the short and the long /a/, but the long /a/ also occupied the higher F2 and higher F1 region, indicating that the long /a/ was produced with a lower tongue position and/or wider jaw opening. There was also an overlap between the short and the long /o/, but the short /o/ also occupied the inner portion of the vowel space. The long /u/ occupied slightly higher F2 and lower F1 region than the short /u/.

Figure 2 shows effect of speaking rates for the five vowel types. The overall rate effect appeared to be less consistent and smaller than the vowel length effect shown in Figure 1. That is, we do not observe that the vowels spoken at slower rates occupy more peripheral portions in the vowel space. It is shown in the following section that this apparent lack of overall rate effect is, at least partially, due to the interaction of the effects of rate and vowel length.

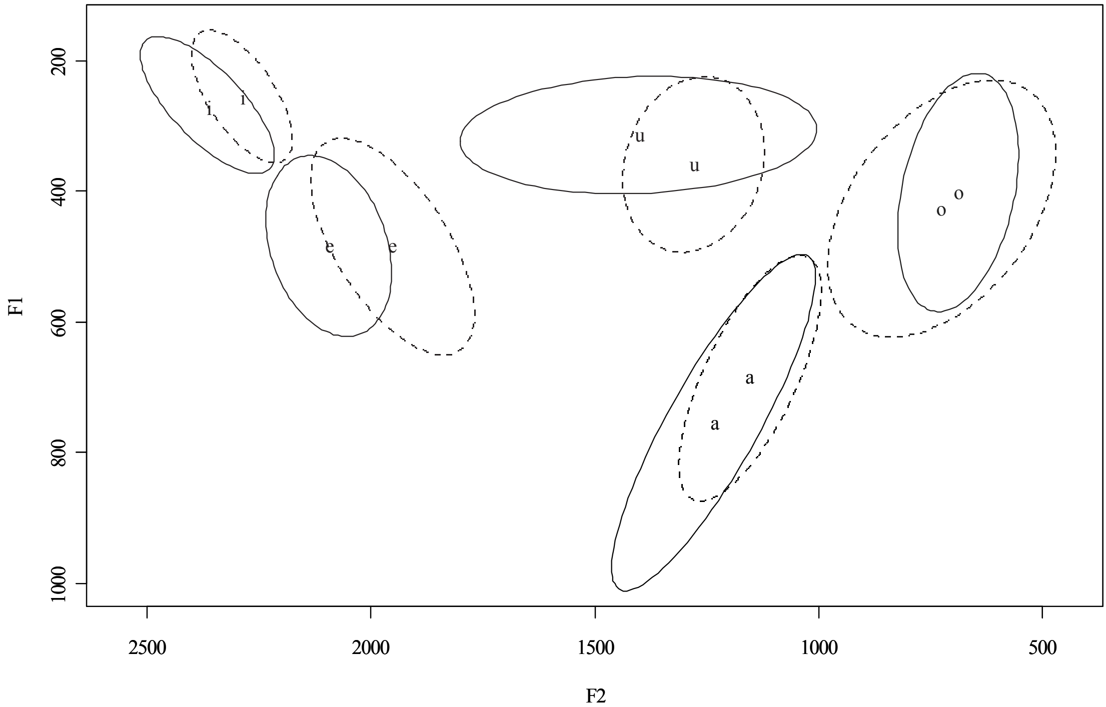

Figure 1. Effects of vowel length on the F1 and F2 midpoint values of all five vowel types. The vertical axis represents F1, and the horizontal axis represents F2 in Hz. The radius of the ellipse is 2.4 times the standard deviations of the mean, covering 95% of the data points. The vowel symbols indicate the centroids and they are the average values of those distributions (n=18 per vowel, per length). The circles with broken lines represent five short vowels averaged across three rates, and the circles with solid lines represent five long vowels averaged across three rates.

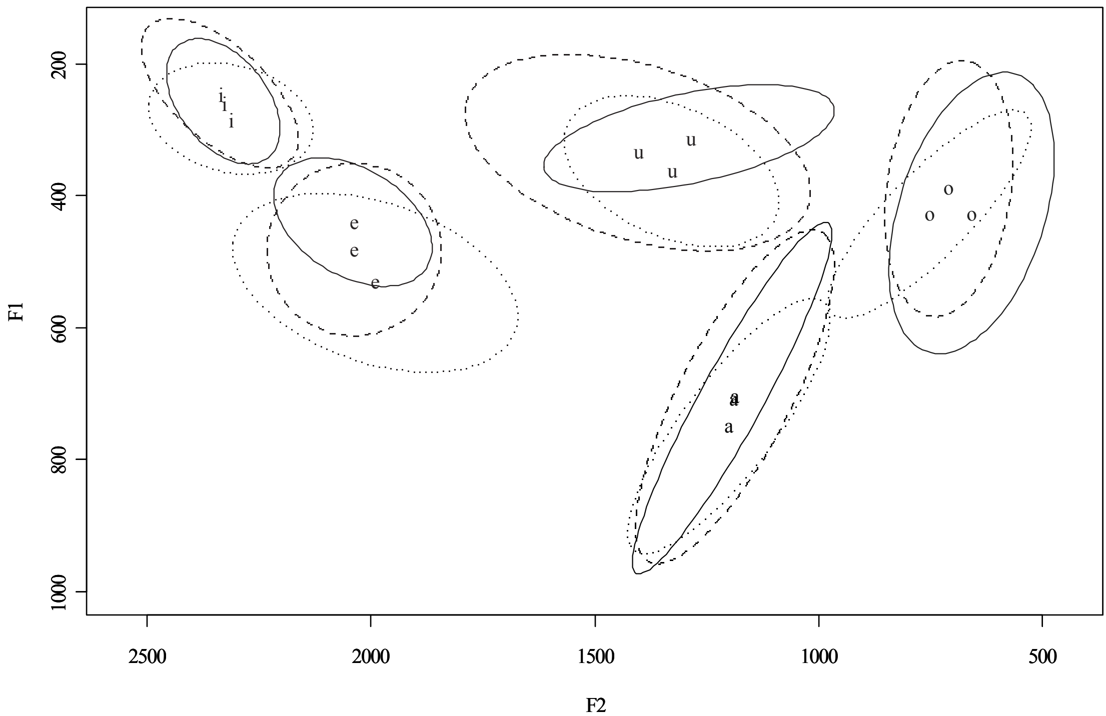

Figure 2. Effects of speaking rates on the F1 and F2 midpoint values of all five vowel types. The vertical axis represents F1, and the horizontal axis represents F2 in Hz. The radius of the ellipse is 2.4 times the standard deviations of the mean, covering 95% of the data points. The vowel symbols indicate the centroids and they are the average values of those distributions (n=12 per vowel, per rate). The circles with solid lines represent the vowels spoken at the slow rate, averaged across length. The circles with broken lines represent the vowels spoken at the normal rate, and the circles with dotted lines represent the vowels spoken at the fast rate, averaged across length.

3.2 Formant trajectories for individual vowels

In Figure 3, each of the formant trajectories was linearly interpolated, and time was normalized in order to linearly stretch or compress the trajectories such that they were of equal duration. The left panel of this figure shows effect of vowel length for the vowel pair /i/. The broken lines represent the trajectories of the short vowel averaged across rates/speakers/repetitions. The solid lines represent the trajectories of the long vowel. The important thing to look at in each panel is how apart the broken and the solid lines are. When the lines are clearly apart, it shows a clear effect of length. The high vowel /i/ showed an effect of vowel length on F2. The F2 was approximately 100 Hz higher for the long vowel than the short vowel throughout the entire duration.

The right panel shows effect of speaking rates. The solid lines represent the vowels spoken at the slow rate, averaged across length/speakers/repetitions, and the broken and dotted lines represent those of the normal and the fast rates, respectively. The three lines of F2 are hardly distinguishable because they are overlapping. There was little difference among the formant trajectories across three rates, showing limited effects of speaking rates. Thus, this result reinforces the notion of point vowels.

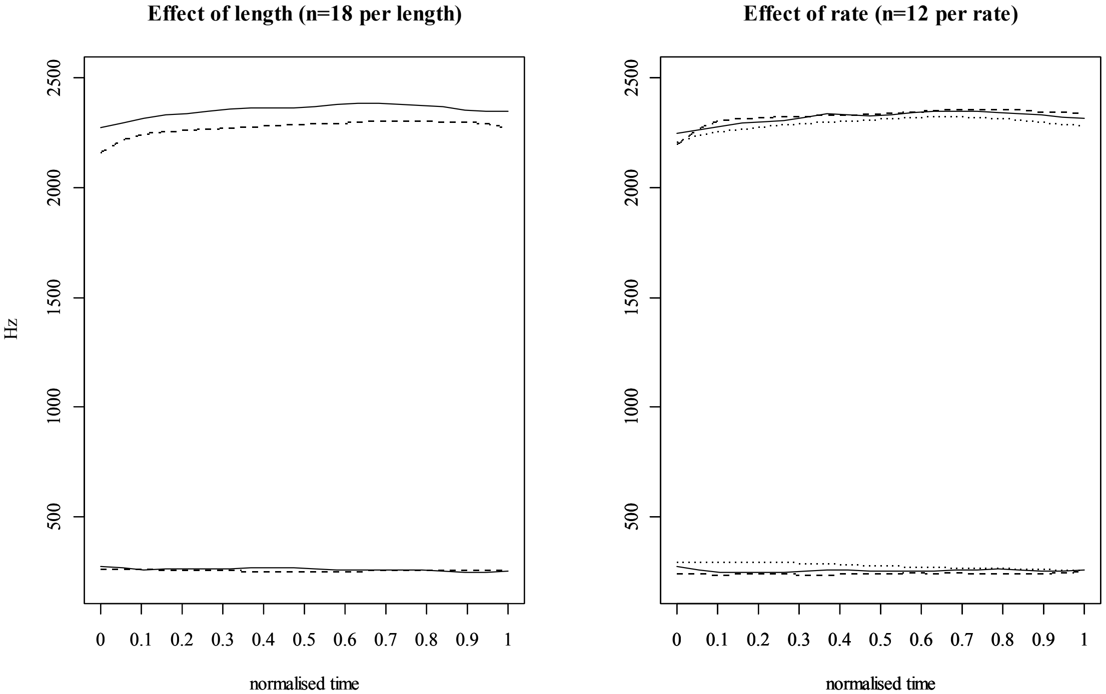

Figure 3. Effects of length and speaking rates on the F1 and F2 trajectories of vowel [i]. The horizontal axis represents normalized time from zero to one, and the vertical axis represents formant frequency in Hz. In the left panel, solid lines represent F1 and F2 trajectories of the long vowel, and broken lines represent F1 and F2 trajectories of the short vowel, averaged across rates/speakers/repetitions. In the right panel, solid, broken, and dotted lines represent trajectories of slow, normal, and fast [i], respectively, averaged across length/speakers/repetitions.

Effect of vowel length interacted with effect of speaking rates for /e/, /o/, and /a/. Figure 4 shows the F1 and F2 trajectories of /e/. As shown in the left panel, the effect of vowel length was that the F2 of the long vowel was approximately 200 Hz higher than that of the short vowel. Compared to the effect of vowel length, effect of speaking rates on F2 was smaller: the F2 of the slow rate, as shown in the solid line, was about 50 Hz higher than that of the fast rate, as shown in the dotted line. The effect of rate was found in F1 as well: the F1 for the slow rate was lower than the F1 for the fast rate. This panel shows the effect of rates on /e/ averaged across length, but in order to examine the interaction of the two effects, the short and the long vowels were plotted separately in Figure 5.

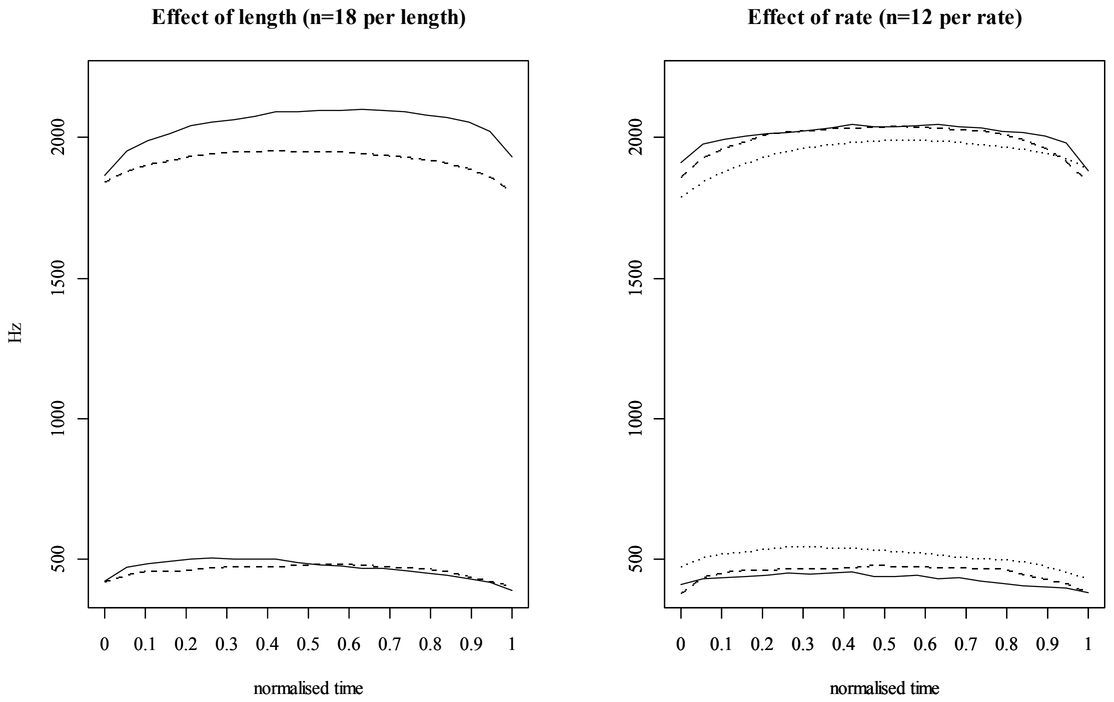

Figure 4. Effects of length and speaking rates on the F1 and F2 trajectories of vowel [e]. In the left panel, solid lines represent F1 and F2 trajectories of the long vowel, and broken lines represent F1 and F2 trajectories of the short vowel, averaged across rates/speakers/repetitions. In the right panel, solid, broken, and dotted lines represent trajectories of slow, normal, and fast [e], respectively, averaged across length/speakers/repetitions.

The left panel of Figure 5 shows the long vowel spoken at three rates, and the right panel shows the short vowel spoken at three rates. For F2, effect of speaking rates was found only for the short vowel: the F2 of the short vowel spoken at the slow rate was about 100 Hz higher than that spoken at the fast rate. However, rate change did not affect the F2 of the long vowel. Because of the interaction of the rate and the vowel length effects, the difference between the F2s of the long and the short vowels was smaller for slower rates. That is, the difference between the long and the short vowels spoken at the slow rate (as indicated by the solid lines in the two panels) was smaller than the difference between the long and the short vowels spoken at the fast rate (as indicated by the dotted lines). For F1, rate effect was shown in both the short and the long vowels: the F1 was lower for the slow rate than the fast rate.

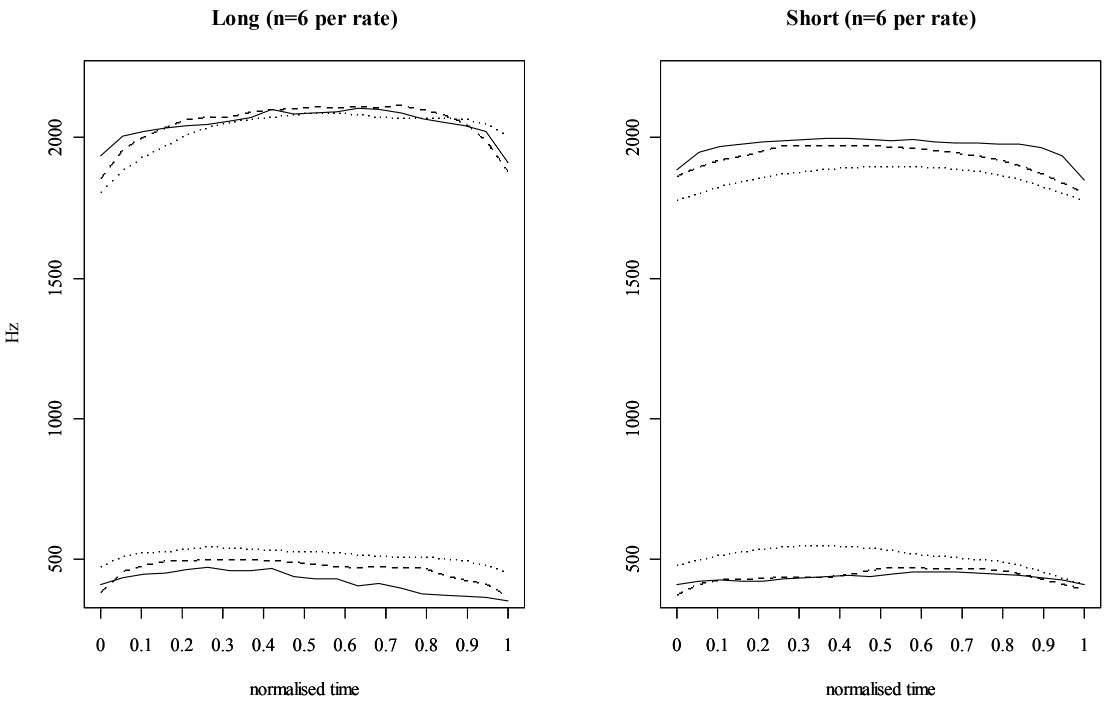

Figure 5. Interaction between length and rate effects for [e]. In the left panel, solid, broken, and dotted lines represent trajectories of the long [e] spoken at slow, normal, and fast rates, respectively. In the right panel, solid, broken, and dotted lines represent trajectories of the short [e] spoken at slow, normal, and fast rates, respectively.

An interaction of rate and vowel length effects was also found for /o/, but we first look at each effect separately. In Figure 6, effect of vowel length for this back vowel was that F2 was lower for the long vowel than the short vowel, which is the opposite of non-back vowels such as /e/ shown earlier. Effect of rate was shown in the lower F2 for slower rates. The F2 of the slow rate was approximately 100 Hz lower than the F2 of the fast rate.

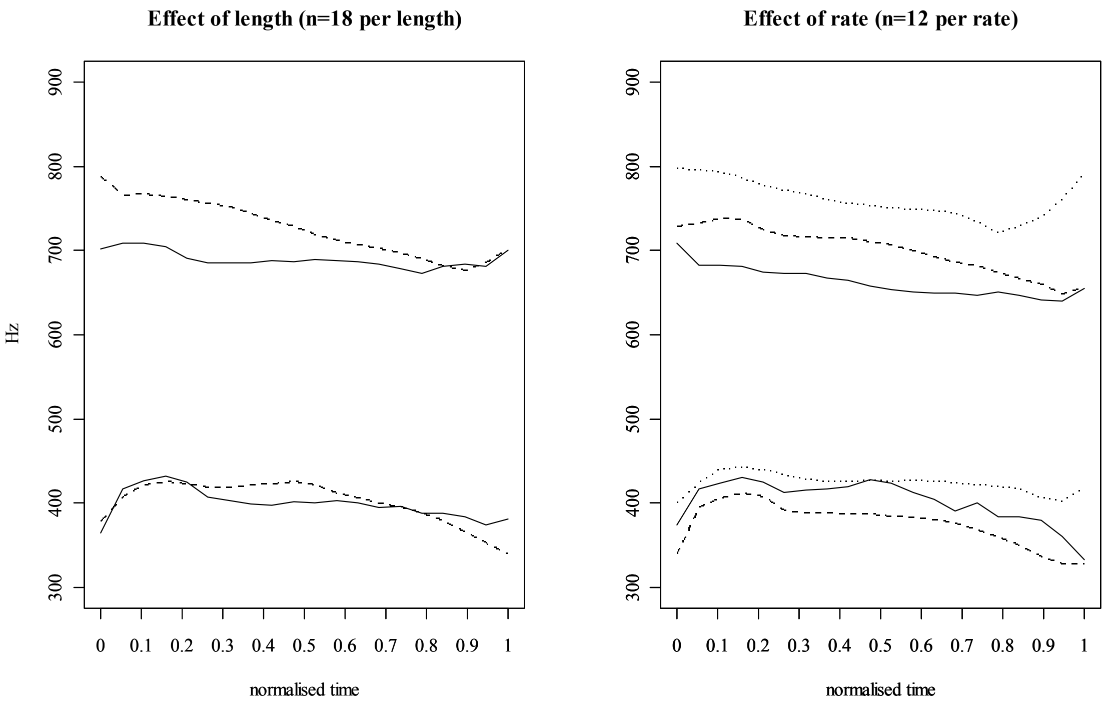

Figure 6. Effects of length and speaking rates on the F1 and F2 trajectories of vowel [o]. In the left panel, solid lines represent F1 and F2 trajectories of the long vowel, and broken lines represent F1 and F2 trajectories of the short vowel, averaged across rates/speakers/repetitions. In the right panel, solid, broken, and dotted lines represent trajectories of slow, normal, and fast [o], respectively, averaged across length/speakers/repetitions.

When we look at the short and the long vowels separately in Figure 7, we can see that the F2 of the long vowel did not change as rate changed, but that the F2 of the short vowel was approximately 200 Hz lower for the slow rate than the fast rate, indicating the effect of rate. A small effect of rate was found in the F1 of the short vowel as well: the F1 was lower for slower rates. The F1 of the fast rate, shown with a dotted line, was consistently higher and towards the center of the vowel space.

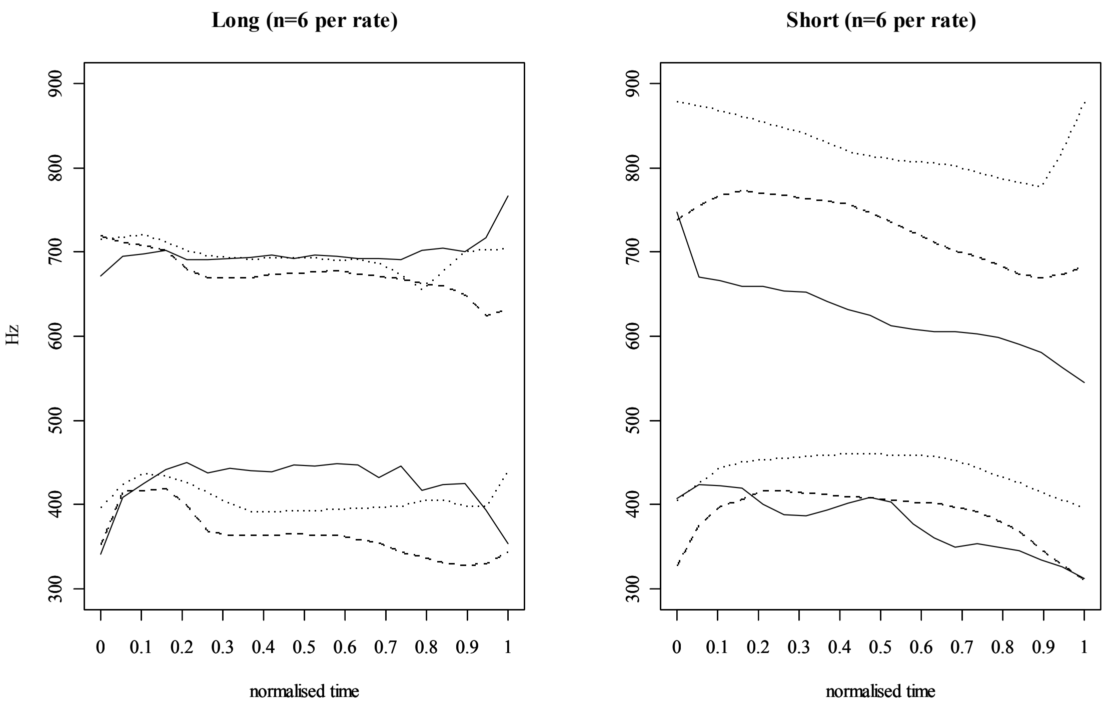

Figure 7. Interaction between length and rate effects for [o]. In the left panel, solid, broken, and dotted lines represent trajectories of the long [o] spoken at slow, normal, and fast rates, respectively. In the right panel, solid, broken, and dotted lines represent trajectories of the short [o] spoken at slow, normal, and fast rates, respectively.

An interaction of rate and vowel length effects was also found for the low vowel /a/, but we first look at each effect separately. In Figure 8, effect of vowel length was shown clearly in the higher F2 for the long vowel than the short vowel. Similarly, F1 was higher for the long vowel than the short vowel. In contrast, the overall effect of rate was smaller, as shown in the overlapping trajectories across three rates in the left panel.

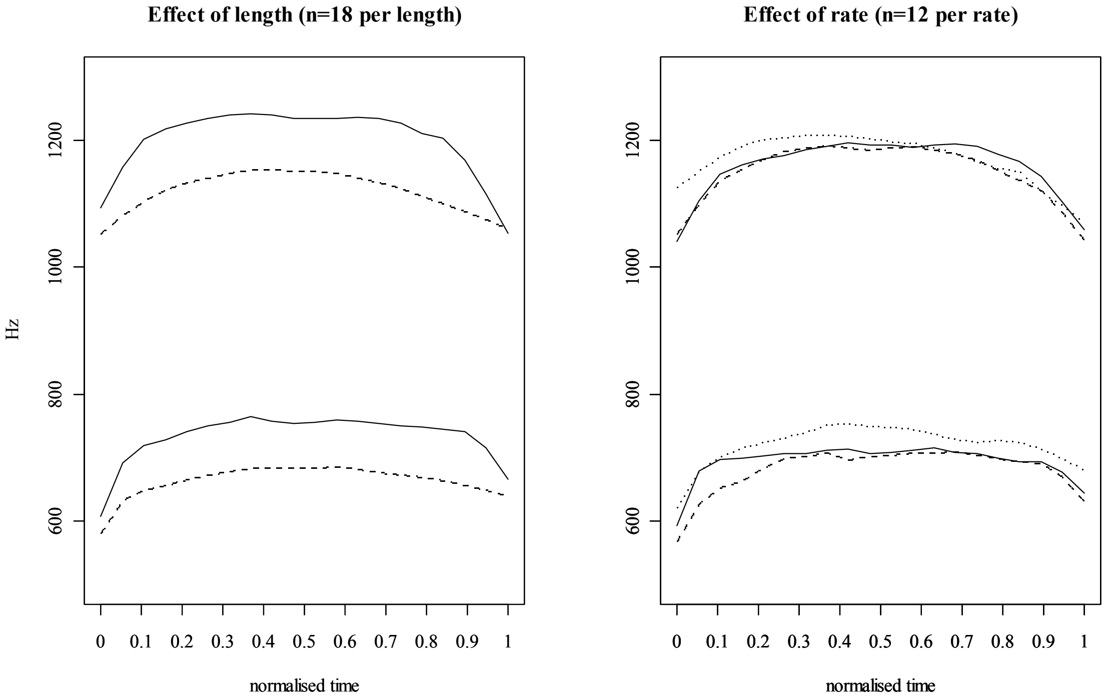

Figure 8. Effects of length and speaking rates on the F1 and F2 trajectories of vowel [a]. In the left panel, solid lines represent F1 and F2 trajectories of the long vowel, and broken lines represent F1 and F2 trajectories of the short vowel, averaged across rates/speakers/repetitions. In the right panel, solid, broken, and dotted lines represent trajectories of slow, normal, and fast [a], respectively, averaged across length/speakers/repetitions.

Figure 9 shows differential effects of rate on the short and the long /a/. For the short vowel, F2 was higher at the slow than the fast rate, but for the long vowel, F2 was lower at slower rates. As a result, the difference in F2 between the long and the short vowels was smaller as rate decreased. That is, the F2 difference between the long and the short vowels for the slow rate (as indicated by the solid lines) was smaller than the difference for the fast rate (as indicated by the dotted lines). A differential effect of rate was found for F1. For the long vowel, F1 was lower for the slower rates, which is opposite to what we would expect. However, for the short vowel, F1 of the slow rate was highest, that of the fast rate was the second highest, and that of the normal rate was the lowest.

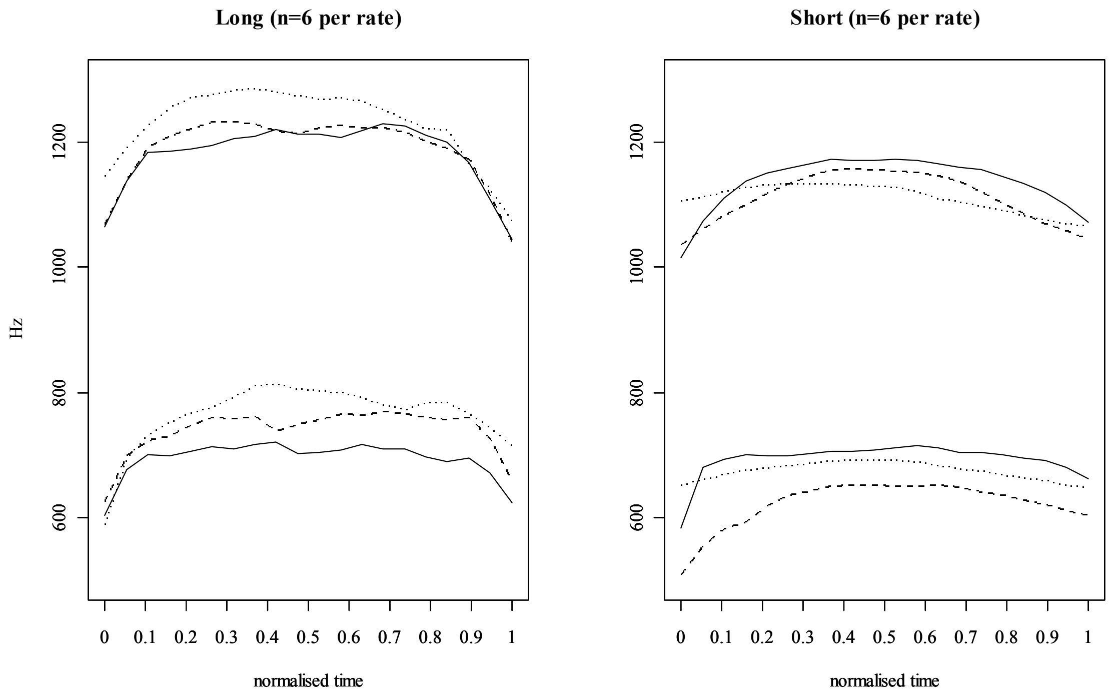

Figure 9. Interaction between length and rate effects for [a]. In the left panel, solid, broken, and dotted lines represent trajectories of the long [a] spoken at slow, normal, and fast rates, respectively. In the right panel, solid, broken, and dotted lines represent trajectories of the short [a] spoken at slow, normal, and fast rates, respectively.

For /u/, effect of vowel length is shown in the left panel of Figure 10. As in the pairs of /i/, /e/, and /a/, the F2 of the long vowel was higher than that of the short vowel. Effect of rate on their F2 was not clear, as it was highest for the normal rate, the second highest for the fast, and the lowest for the slow rates.

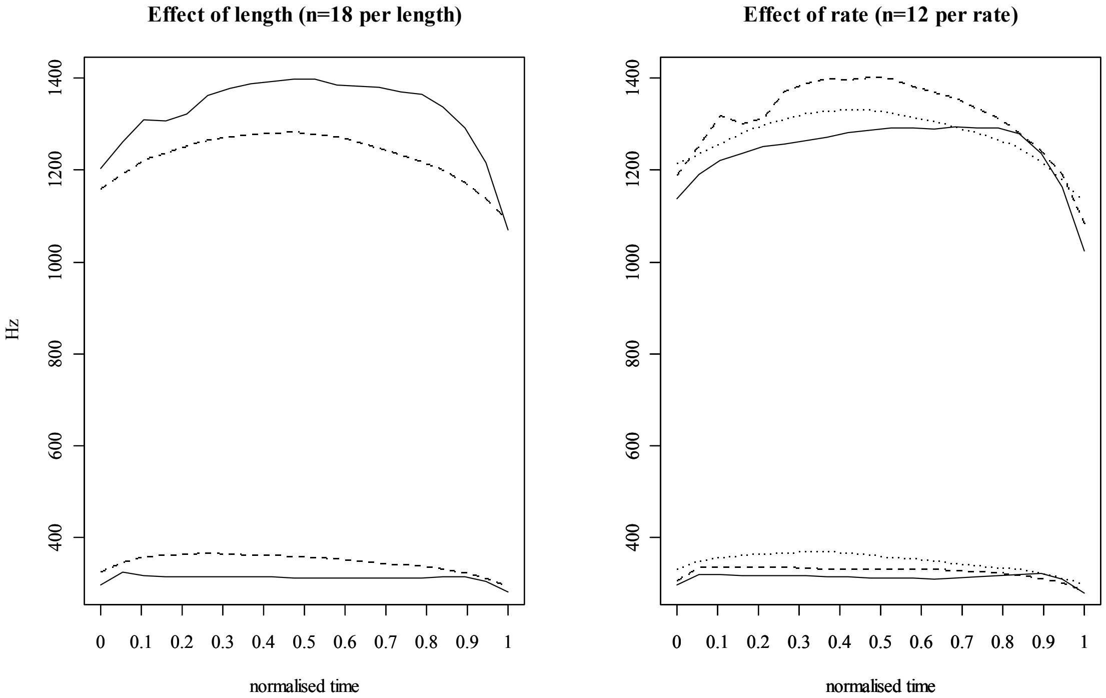

Figure 10. Effects of length and speaking rates on the F1 and F2 trajectories of vowel [u]. In the left panel, solid lines represent F1 and F2 trajectories of the long vowel, and broken lines represent F1 and F2 trajectories of the short vowel, averaged across rates/speakers/repetitions. In the right panel, solid, broken, and dotted lines represent trajectories of slow, normal, and fast [u], respectively, averaged across length/speakers/repetitions.

4. Summary

To summarize the effects of vowel length, the long vowels occupied a more peripheral portion of the F1-F2 vowel space than the short vowels did. This supports a suggestion that long vowels resist coarticulation to a greater extent than short vowels do.

To summarize the effects of speaking rates, the short mid vowels /e/ and /o/ were most susceptible to change in speaking rates, suggesting that the degree of coarticulation is greater for these vowels than the other vowels. For /e/, F2 was higher and F1 was lower for slower speech. For /o/, both F1 and F2 were lower for slower speech. The high vowel /i/ was least affected by rate changes, thus reinforcing the notion of point vowel.

Effect of speaking rates interacted with effect of vowel length in /e/, /o/ and /a/. For /e/ and /o/, only the short member showed rate effects. For /a/, as rate decreased, the difference in F2 between the short and the long vowels was smaller.

These results based on two speakers are tentative, and we are planning for further investigation. Our future work is to compare the triplet (e.g., /mama/ vs. /ma:ma/ vs. /mama:/), and also to include analysis of real words. We will also examine patterns of rate effects realized by different individuals, as they may realize a change in speaking rates in different manners.

References

EMU speech database system. Available: http://emu.sourceforge.net/ [Last consulted 7/28/2003].

Engstrand, O. (1988) Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *Journal of the Acoustical Society of America (JASA)*, 83(5), 1863-1875.

Gay, T. (1978) Effect of speaking rate on vowel formant movements. *JASA*, 63(1), 223-230.

Han, M. (1962) The features of duration in Japanese. *Study of Sounds,* 10, 65-81.

Imaizumi, S., Kiritani, S., Hirose, H., Togami, S., & Shirai, K. (1987) Preliminary report on the effects of speaking rate upon formant trajectories. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, 21, 147-151.

Lindblom, B. (1963) Spectrographic study of vowel reduction. *JASA*, 35(11), 1773-1781.

Magen, H. S. & Blumstein, S. E. (1993) Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics*, 21, 387-409.

Moon, S. J. & Lindblom, B. (1994) Interaction between duration, context, and speaking style in English stressed vowels. *JASA*, 96(1), 40-55.

Tsukada, K. (1999) *An Acoustic Phonetic Analysis of Japanese-accented English.* Unpublished doctoral dissertation, Macquarie University.

Ueyama, M. (2000) *Prosodic transfer: an acoustic study of L2 English vs. L2 Japanese.* Doctoral dissertation, University of California at Los Angeles.

Van Son, R. J. J. H. & Pols, L. C. W. (1992) Formant movements of Dutch vowels in a text, read at normal and fast rate. *JASA*, 92, 121-127.

Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception

edited by Augustine Agwuele, Willis Warren, and Sang-Hoon Park

Cascadilla Proceedings Project   Somerville, MA   2004

Copyright information

Proceedings of the 2003 Texas Linguistics Society Conference:
Coarticulation in Speech Production and Perception
© 2004 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 1-57473-402-4 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Hirata, Yukari and Kimiko Tsukada. 2004. The Effects of Speaking Rates and Vowel Length on Formant Movements in Japanese. In *Proceedings of the 2003 Texas Linguistics Society Conference*, ed. Augustine Agwuele et al., 73-85. Somerville, MA: Cascadilla Proceedings Project.

or:

Hirata, Yukari and Kimiko Tsukada. 2004. The Effects of Speaking Rates and Vowel Length on Formant Movements in Japanese. In *Proceedings of the 2003 Texas Linguistics Society Conference*, ed. Augustine Agwuele et al., 73-85. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1069.