

# Tongue movement kinematics in long and short Japanese consonants

Anders Löfqvist<sup>a)</sup>

Haskins Laboratories, New Haven, Connecticut 06511 and Department of Logopedics,  
Phoniatrics & Audiology, Clinical Sciences, Lund, Lund University, Lund, Sweden

(Received 22 November 2006; revised 30 March 2007; accepted 30 March 2007)

This paper examines tongue movements in stop and fricative consonants where the duration of the oral closure/constriction for the consonant is varied for linguistic purposes. Native speakers of Japanese served as subjects. The linguistic material consisted of Japanese word pairs that only differed in the duration of the lingual consonant, which was either long or short. Recordings were made of tongue movements using a magnetometer system. Results show a robust difference in closure duration between the long and short consonants. Overall, the path of the tongue movement during the consonant was longer for the long than for the short consonant. All speakers decreased the speed of the tongue movement during the long consonant. These adjustments in tongue movements were most likely made to maintain the contact between the tongue and the palate for the closure and constriction. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2735102]

PACS number(s): 43.70.Aj, 43.70.Bk [BHS]

Pages: 512–518

## I. INTRODUCTION

This paper examines tongue movements in stop and fricative productions where the duration of the oral closure/constriction for the consonant is varied for linguistic purposes, using speakers of Japanese. In Japanese, the ratio of closure duration for long and short consonants is about 2:1 (Beckman, 1982; Han, 1994; Hirata and Whiton, 2005). There is an extensive body of acoustic studies of the Japanese sound system with particular emphasis on the role of the mora for speech timing. A mora is traditionally regarded as a unit of timing in Japanese, but its existence and acoustic manifestations are debated, and the reader is referred to Warner and Arai (2001) for a review. The long consonants in Japanese are sometimes referred to as “geminate,” and such a consonant contributes one mora; it is also traditionally assumed that a mora boundary occurs in the long consonant (Vance, 1987). However, the primary focus of this paper is not on mora timing but on speech motor control, capitalizing on the length distinction in Japanese sound structure to study how tongue movements are controlled when the duration of a consonant is changed.

If the duration of the oral closure for the consonant is increased, a speaker is still constrained to maintain the contact between the tongue and the palate to make the closure or constriction for the consonant. To do this, the speaker can in principle use two strategies for controlling the tongue movement, both of which involve modulating the speed of the tongue movement. One strategy would be to momentarily stop the tongue from moving. Alternatively, the speaker could slow down the tongue movement for the long consonant. Of these two possibilities, the second one is the most likely, since a large body of research on tongue movements in speech suggests that the tongue hardly ever stops moving. For example, the material presented by Mooshammer *et al.*

(1995) showed that the tongue body is moving during the closure for a velar stop (see also Houde, 1968; Perkell, 1969). The amount of tongue movement during the closure varied with vowel context. Thus, the tongue body movement during the closure was about 1 mm when the preceding vowel was /i/, and between 4 and 10 mm when the preceding vowel was /u/ or /a/.

Studies of tongue movements during velar stop consonant production have also shown that the tongue tends to move in curved paths (Houde, 1968; Perkell, 1969; Kent and Moll, 1972; Schönle, 1988; Munhall *et al.* 1991; Mooshammer *et al.*, 1995; Löfqvist and Gracco, 1994; 2002). In the production of velar stops, the tongue movement trajectory into the stop closure is usually going forward, and the location of the point of contact between the tongue body and the palate at stop closure is influenced by the phonetic context of the stop (cf., Dembowski *et al.* 1998). Less information on movement kinematics is available for stops produced with the tongue tip and tongue blade, /t, d/, although the data published by Kent and Moll (1972) show that the location of the point of contact between the tongue tip and the palate, or alveolar ridge, is influenced by phonetic context. In addition, the pattern of contact between the tongue and the alveolar ridge appears to vary with the shape of the palate (Hiki and Itoh, 1986). Löfqvist and Gracco (2002) recorded tongue movements for stop consonants produced with the tongue tip and the tongue body in different vowel contexts. Their results indicated that both the tongue body and the tongue tip moved during the closure for the consonant. The tongue tip movement was similar to that of the tongue body. In both cases, the movement was upward and forward at consonant onset, and downward and backward at the release. However, both the magnitude and direction of the movement were heavily influenced by the vowel context.

Electropalatographic studies of Italian stops have shown that the amount of tongue palate contact is larger for geminate than for single stops, and also that there is a general

<sup>a)</sup>Electronic-mail: lofqvist@haskins.yale.edu

TABLE I. The linguistic material, text used for presentation, and glosses in English.

/hata/	kanji	“flag”
/hatta/	kanji&hiragana	“(someone) pasted (something)”
/muda/	kanji	“useless”
/budda/	katakana	“Buddha” katakana
/hosa/	kanji	“assistant”
/hossa/	kanji	“attack (medical condition)”
/haja/	kanji	“winner”
/hajja/	kanji	“departure (of a train or a vehicle)”
/saka/	kanji	“slope”
/sakka/	kanji	“author, writer”
/tagu/	katakana	“price tag”
/tagau/	katakana	“tug”

increase in the extent of tongue-palate contact with increasing closure duration (Farnetani, 1990); similar results for American English have been presented by Byrd (1995). An x-ray study of French consonants by Vaxelaire (1995) suggested that the area of tongue palate contact was larger for the long (abuttred) stops than for the short ones.

The purpose of the present study was to extend the findings of Löfqvist and Gracco (2002) to the control of tongue movements in consonants with different durations. As noted above, a speaker is constrained in producing lingual consonants to maintain the contact between the tongue and the palate. Thus, it was hypothesized that the speed of the tongue movement would be slower for a long than for a short consonant. A second hypothesis to be evaluated is that the mag-

nitude of the tongue movement will be larger during a long than during a short consonant. To examine these hypotheses, tongue movements were recorded in native Japanese speakers.

## II. METHOD

### A. Subjects

Five native speakers of Japanese, two male and three female, served as subjects. They reported no speech, language, or hearing problems. They were naive as to the purpose of the study. Before participating in the recording, they read and signed a consent form. (The experimental protocol was approved by the IRB at the Yale University School of Medicine.)

### B. Linguistic material

The linguistic material consisted of Japanese words with an alveolar or velar consonant. These words formed minimal pairs, where the only difference between the pairs was the duration of the consonant. The words are listed in Table I together with English glosses. The linguistic material was organized into randomized lists and presented to the subjects in Japanese writing, with the words occurring in the frame sentence “karewa to itta” (“He said \_”). Fifty repetitions of each word were recorded.

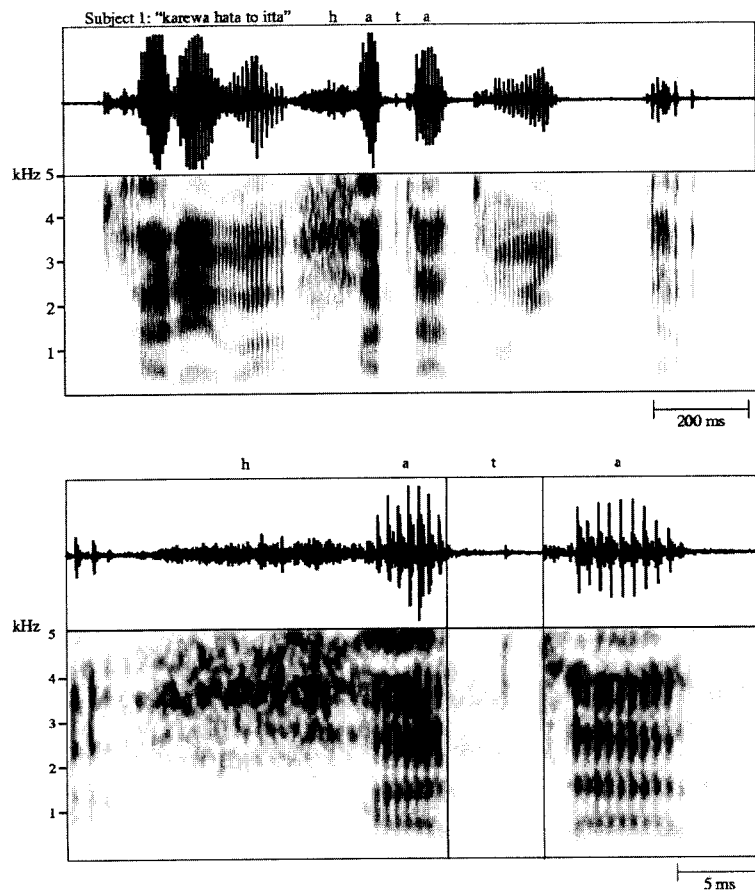


FIG. 1. Waveforms and spectrogram of the utterance “karewa hata to itta” produced by Speaker 1. The top panel shows the whole utterance, and the bottom panel only the word “hata” with the vertical lines showing the onset and release of the voiceless consonant /t/ in “hata.”

### C. Movement recording

The movements of the lips, the tongue, and jaw were recorded using a three-transmitter magnetometer system (Perkell *et al.*, 1992); when proper care is taken during the calibration, the spatial resolution of the system is on the order of 0.5 mm. Receivers were placed on the vermillion border of the upper and lower lip, on three points of the tongue, referred to as tip, blade, and body, and on the lower incisors at the gum line. Two additional receivers placed on the nose and the upper incisors were used for the correction of head movements. The lip and jaw receivers were attached using Isodent, a dental adhesive, while the tongue receivers were attached using Ketac-Bond, another dental adhesive. Care was taken during each receiver placement to ensure that it was positioned at the midline with its long axis perpendicular to the sagittal plane. Two receivers attached to a plate were used to record the occlusal plane by having the subject bite on the plate during the recording. All data were subsequently corrected for head movements and rotated to bring the occlusal plane into coincidence with the  $x$  axis. This rotation was performed to obtain a uniform coordinate system for all subjects (cf. Westbury, 1994). For one speaker, S4, the signal from the tongue body receiver was degraded during the recording session, so no tongue body data are reported for this subject; the same thing happened for subject S3, so only 30 tokens could be used for the words with a velar consonant.

The articulatory movement signals (induced voltages from the receiver coils) were sampled at 500 Hz after low-pass filtering at 200 Hz. The resolution for all signals was 12 bits. After voltage-to-distance conversion, the movement signals were low-pass filtered using a 25-point triangular window with a 3-dB cutoff at 14 Hz; this was done forwards and backwards to maintain phase. To obtain instantaneous velocity of the tongue receivers, the first derivative of the position signals was calculated using a 3-point central difference algorithm. For each tongue receiver, its speed [ $v = \sqrt{(x^2 + y^2)}$ ] was also calculated. The velocity signals were smoothed using the same triangular window. All the signal processing was made using the Haskins Analysis Display and Experiment System (Rubin and Löfqvist, 1996). The acoustic signal was pre-emphasized, low-pass filtered at 4.5 kHz, and sampled at 10 kHz.

The onset and release of the oral closure/constriction for the consonant were identified in waveform and spectrogram displays of the acoustic signal. They were both identified by a change in the amplitude and the spectral properties, cf. Fig. 1.

The magnitude of the tongue movement trajectory during the consonant was obtained by summing the Euclidean distances between successive samples of the tongue tip and tongue body vertical and horizontal receiver positions from the acoustically defined consonant onset to offset. The average speed of the tongue tip and tongue body during the consonant was obtained by adding the speed of all the individual samples between consonant onset and offset and then dividing by the number of samples in the interval. The kinematic signals are expressed in a maxilla-based coordinate system.

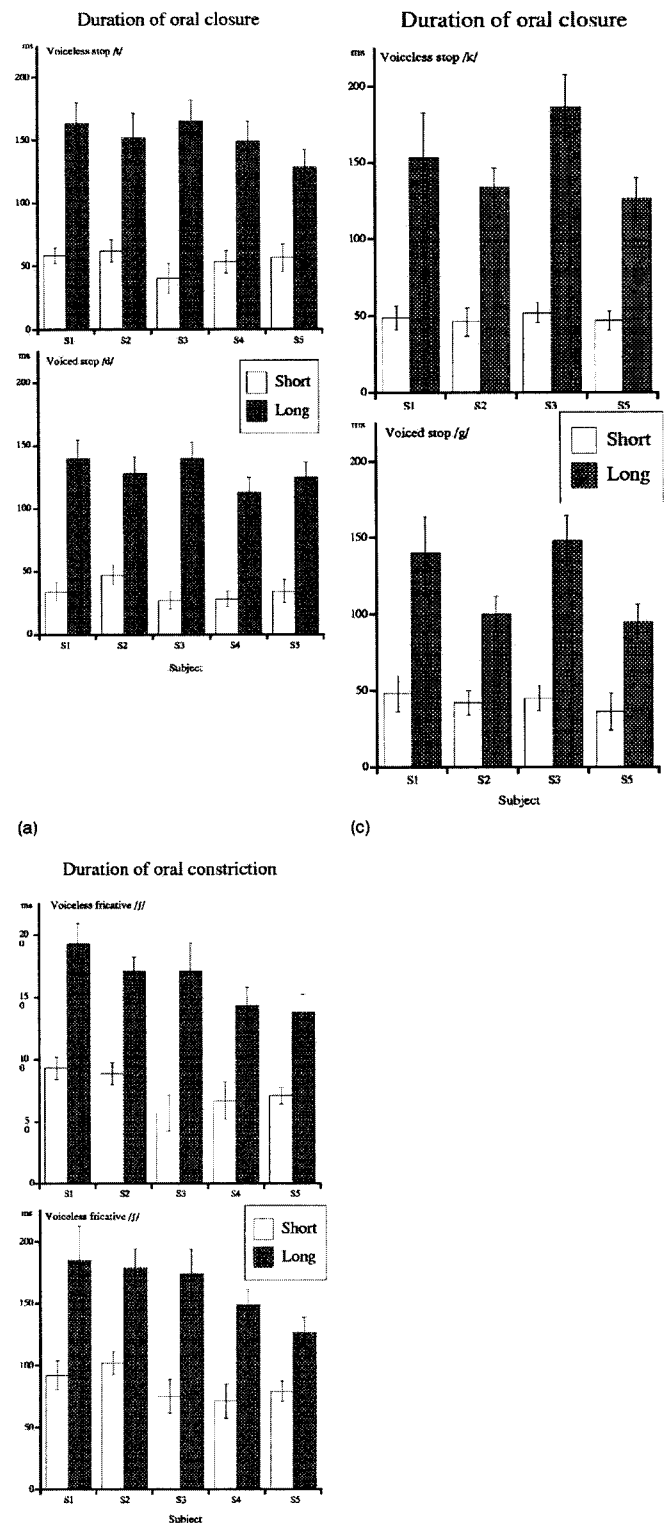


FIG. 2. Closure duration of the consonants (mean and standard deviation). (a) alveolar stops; (b) alveolar fricatives; and (c) velar stops.

Thus, the tongue movement includes the contribution of the jaw, which is appropriate when we are interested in the tongue as the end effector.

$T$ -tests were used to assess differences between the long and short consonants for each subject. Given the large num-

TABLE II. Results of the *t*-tests for the duration of the oral closure/constriction.

	S1	S2	S3	S4	S5
/t, tt/	41.30	29.8	42.92	37.54	28.92
/d, dd/	40.56	36.20	55.93	43.65	42.69
/s, ss/	37.55	39.80	29.51	24.20	30.39
/ʃ, ʃʃ/	22.11	29.85	29.18	29.28	23.94
/k, kk/	35.65	40.87	43.06		36.21
/g, gg/	38.91	28.67	38.90		25.48

ber of comparisons, a conservative  $\alpha$ -level of 0.001 was adopted. Since the variances usually differed between the long and short consonants, as shown by Levene's test, the statistical tests assumed unequal variances, so the degrees of freedom were adjusted (e.g., Winer *et al.*, 1991, p. 67).

### III. RESULTS

#### A. Consonant duration

The duration of the oral closure/constriction for the long and short consonants is summarized in Fig. 2. For all speakers, there is a clear and robust difference between the long and short consonants, with the long ones having twice, or more, the duration of the short ones. The results of the *t*-tests are summarized in Table II. Comparing the durations of the different consonant categories, the voiceless alveolar stop /t/ was always longer than its voiced cognate /d/. Also, the stops were usually shorter than the fricatives.

#### B. Tongue movements

Overall, the tongue movement patterns were similar to those reported previously (e.g., Houde, 1968; Mooshammer *et al.*, 1995; Löfqvist and Gracco, 2002). That is, both the tongue tip and the tongue body were moving up and forward towards the consonant constriction and down and backward after the consonant. There was some variability between and within speakers, but it is beyond the scope of the present paper to present a detailed analysis of these movement patterns.

The first analysis focused on the magnitude of the tongue movement during the consonant, while the second one examined the average speed of the tongue movement during the consonant.

##### 1. Magnitude of tongue movement path

Figure 3 shows the magnitude of the tongue movement during the consonant. The path was generally longer for the long than for the short consonants, irrespective of place and

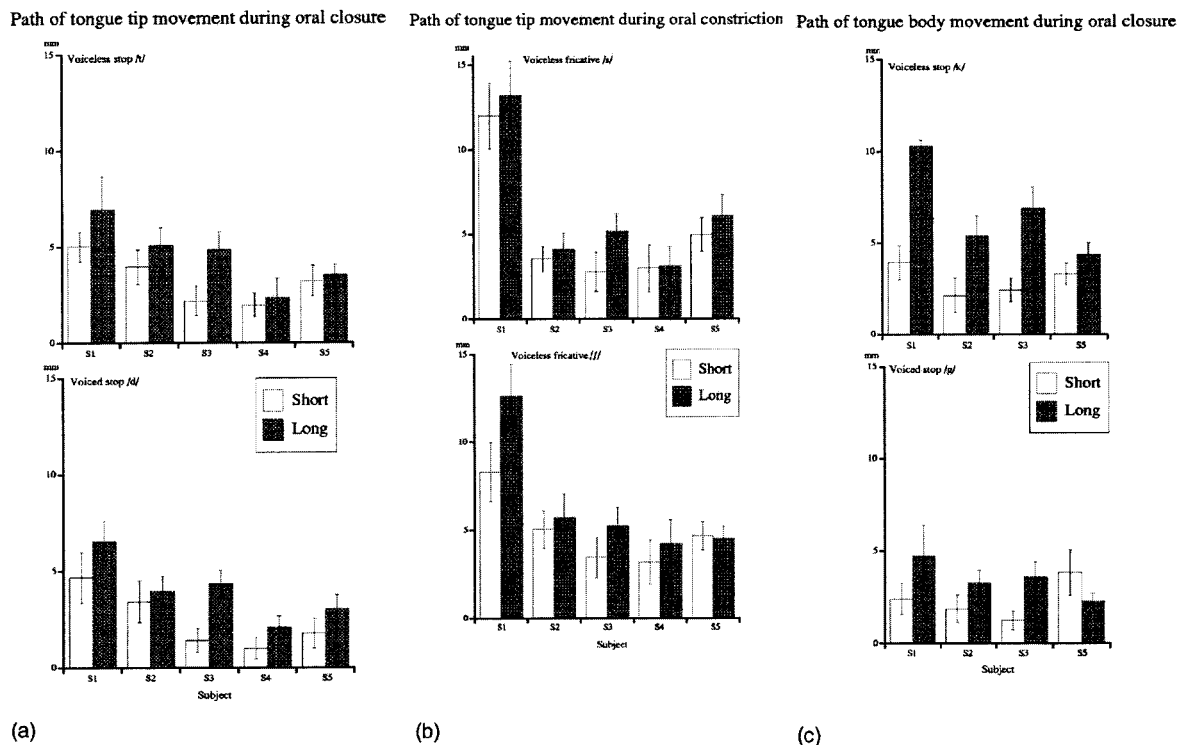
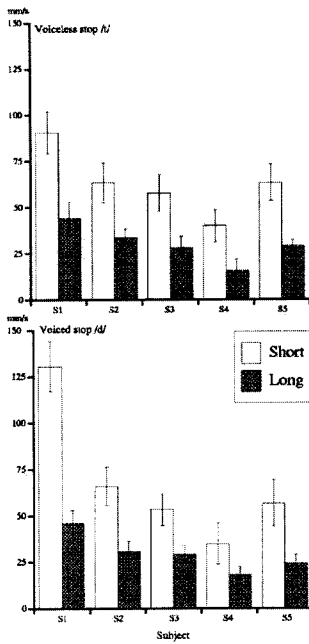


FIG. 3. Path of the tongue movement during the consonant (mean and standard deviation). (a) alveolar stops; (b) alveolar fricatives; and (c) velar stops. For the velar stops, the tongue body movement was measured, while for the other consonants, the tongue tip was used.

TABLE III. Results of the *t*-tests for the path of the tongue during the oral closure/constriction.

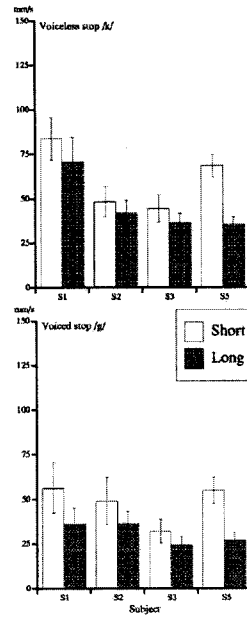
	S1	S2	S3	S4	S5
/t, tt/	6.85	6.35	15.16	2.19	2.42
/d, dd/	6.75	2.54	22.07	9.00	8.46
/s, ss/	3.03	3.33	10.90	0.57	5.15
/ʃ, ʃʃ/	12.44	2.72	8.14	3.95	1.09
/k, kk/	15.0	15.57	17.9		8.25
/g, gg/	15.71	9.92	12.64		3.53

Average speed of tongue tip movement during oral closure



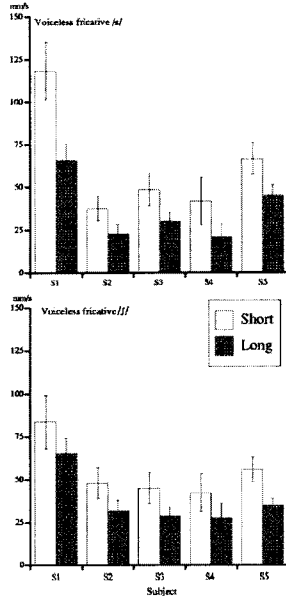
(a)

Average speed of tongue body movement during oral closure



(c)

Average speed of tongue tip movement during oral constriction



(b)

FIG. 4. Average speed of the tongue movement during the consonant (mean and standard deviation). (a) alveolar stops; (b) alveolar fricatives; and (c) velar stops. For the velar stops, the tongue body movement was measured, while for the other consonants, the tongue tip was used.

TABLE IV. Results of the *t*-tests for the average speed of the tongue during the oral closure/constriction.

	S1	S2	S3	S4	S5
/t, tt/	22.24	18.00	17.89	15.42	23.24
/d, dd/	26.68	23.81	17.45	9.53	17.09
/s, ss/	19.19	11.69	12.28	9.04	13.32

less fricative /j/ and the voiced velar stop /g/ for subject S5, where the path was longer for the short consonant. The *t*-tests, summarized in Table III, showed that there was no significant difference between the length of the paths for the long and short consonants in the following cases: /t, tt/ for subjects S4 and S5; /s, ss/ for subjects S1, S2, and S4; /d, dd/ and /ʃ, ʃʃ/ for subjects S2 and S5. Given that the path was longer for the long than for the short consonants, correlations were made between the closure duration and the tongue movement path during the closure. All the correlations were positive, in particular for the short consonants. The patterns for the different consonant categories mirrored those for closure duration mentioned above. That is, the consonants with a longer duration also had a larger tongue movement.

## 2. Average speed of tongue movement

Figure 4 shows the average speed of the tongue body movement from the first to the second vowel. For all subjects and consonants, the average speed was slower for the long than for the short consonant. Table IV provides a summary of the *t*-tests, showing statistically significant differences for all subjects and consonants. There was no clear pattern for the different consonant categories.

## IV. DISCUSSION

This study examined tongue movement kinematics in Japanese long and short consonants. The hypothesis under investigation was based on the idea that speakers are constrained in the production of lingual consonants to maintain the contact between the tongue and the palate. Earlier work has shown that the tongue is moving during these consonants, although the influence of consonant duration on tongue movements has not been explored in any detail. Based on these earlier findings it was hypothesized that a speaker could either stop the tongue movement momentarily during a long consonant or slow down the movement speed. The results show that in all cases the speakers modulated the speed of the tongue, so that it was lower in the long than in the short consonants. There was no instance for any speaker where the tongue movement stopped completely.

If there is a difference in the average speed on the tongue movement between the long and short consonants, one might also predict that speakers modify the speed of the tongue movement within a category, long or short, as a function of consonant length. In particular, one would expect a negative relationship between consonant duration and average tongue movement speed. An examination of this rela-

tion of consonants, but the correlations were not very strong. One reason that such a negative relationship was not found for the short consonants might simply be that their durations spanned a smaller range than those of the long consonants. Overall, the long consonants were produced with a longer tongue movement trajectory during the closure/constriction. Correlations were made between the closure duration and the tongue movement path during the closure to examine this relationship within the long and short consonants separately. All the correlations were positive, in particular for the short consonants.

Earlier work has shown a strong correlation between movement displacement and peak velocity in both speech and nonspeech movements (e.g., Cooke, 1980; Ostry *et al.*, 1983; Kelso *et al.*, 1985; Vatikiotis-Bateson and Kelso, 1993; Hertrich and Ackermann, 1997; Löfqvist and Gracco, 1997). The present study measured the average speed of the tongue movement during the acoustically defined closure/constriction for the consonant. There was, still, a very strong correlation between the magnitude of the tongue movement path and the average speed during the consonant, with almost all of the correlations higher than 0.7.

The present results show that speakers modify the speed of tongue movements in consonant production when the duration of the consonant is varied for linguistic purposes. They thus provide additional evidence for active, task-dependent control of tongue movement speed in speech, since the same was found for the movement from the first to the second vowel in a vowel-bilabial consonant-vowel sequence in Japanese, when the duration of the labial consonant was either long or short (Löfqvist, 2006). In this case, the speed of the tongue movement was lower during a long than during a short consonant. All speakers consistently showed this pattern, thus maintaining a similar, but not identical, pattern of coordination between the lip and tongue movements during long and short consonants.

These studies of the relationship between consonant duration and movement kinematics have used native adult speakers of Japanese. An interesting question is the emergence of the length contrast and its associated movement kinematics during speech development. Although there are methodological issues in the recording of tongue movements in children, lip and jaw movements can be readily analyzed (e.g., Green *et al.*, 2000; 2002). Another issue is how a learner of Japanese as a second language masters the length contrast and if the kinematic differences observed here in native speakers between long and short consonants are also

seen in second-language learners, in particular if there is a gradual change over time towards the kinematic patterns found here.

## ACKNOWLEDGMENTS

I am grateful to Mariko Yanagawa for help with the Japanese material and running the experiments. This work was supported by Grant No. DC-00865 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

- Beckman, M. (1982). "Segment duration and the mora in Japanese," *Phonetica* 39, 113–135.
- Byrd, D. (1995). "Articulatory characteristics of single and blended lingual gestures," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, edited by K. Elenius and P. Branderud, Vol. 2, pp. 438–441, Stockholm, Royal Institute of Technology and Stockholm University.
- Cooke, J. (1980). "The organization of simple, skilled movements," in *Tutorials in Motor Behavior*, edited by G. Stelmach and J. Requin (North-Holland, Amsterdam), pp. 199–212.
- Dembowski, J., Lindstrom, M. J., and Westbury, J. R. (1998). "Articulator point variability in the production of stop consonants," in *Neuromotor Speech Disorders: Nature, Assessment, and Management*, edited by M. P. Cannito, K. M. Yorkston, and D. R. Beukelman (Brookes, Baltimore), pp. 27–46.
- Fametani, E. (1990). "V-C-V lingual coarticulation and its spatiotemporal domain," in *Speech Production and Speech Modelling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 93–130.
- Green, J., Moore, C., and Reilly, K. (2002). "The sequential development of jaw and lip control for speech," *J. Speech Lang. Hear. Res.* 45, 66–79.
- Green, J., Moore, C., Higashikawa, M., and Steeve, R. (2000). "The physiologic development of speech motor control: Lip and jaw coordination," *J. Speech Lang. Hear. Res.* 43, 239–255.
- Han, M. (1994). "Acoustic manifestations of mora timing in Japanese," *J. Acoust. Soc. Am.* 96, 73–82.
- Hertrich, I., and Ackermann, H. (1997). "Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures," *J. Acoust. Soc. Am.* 102, 523–536.
- Hiki, S., and Itoh, H. (1986). "Influence of palate shape on lingual articulation," *Speech Commun.* 5, 141–158.
- Hirata, Y., and Whiton, J. (2005). "Effects of speaking rate on the single/geminate stop distinction in Japanese," *J. Acoust. Soc. Am.* 118, 1647–1660.
- Houde, R. (1968). *A Study of Tongue Body Motion during Selected Consonant Sounds* (Speech Communications Research Laboratory, Santa Barbara), SCRL Monograph 2.
- Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (1985). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling," *J. Acoust. Soc. Am.* 77, 266–280.
- Kent, R., and Moll, K. (1972). "Cinefluorographic analyses of selected lingual consonants," *J. Speech Hear. Res.* 15, 453–473.
- Löfqvist, A. (2006). "Interarticulator programming, Effect of closure duration on lip and tongue coordination in Japanese," *J. Acoust. Soc. Am.* 120, 2872–2883.
- Löfqvist, A., and Gracco, V. L. (1994). "Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context," *Phonetica* 51, 52–67.
- Löfqvist, A., and Gracco, V. (1997). "Lip and jaw kinematics in bilabial stop consonant production," *J. Speech Lang. Hear. Res.* 40, 877–893.
- Löfqvist, A., and Gracco, V. (2002). "Control of oral closure in lingual stop consonant production," *J. Acoust. Soc. Am.* 105, 1864–1876.
- Mooshammer, C., Hoole, P., and Kühnert, B. (1995). "On loops," *J. Phonetics* 23, 3–21.
- Munhall, K., Ostry, D., and Flanagan, R. (1991). "Coordinate spaces in speech planning," *J. Phonetics* 19, 293–307.
- Ostry, D., Keller, E., and Parush, A. (1983). "Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movements during speech," *J. Exp. Psychol. Hum. Percept. Perform.* 9, 622–636.
- Perkell, J. S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA).
- Perkell, J., M. Cohen, M., Svirsky, M., Matthies, M., Garabeta, I., and Jackson, M. (1992). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," *J. Acoust. Soc. Am.* 92, 3078–3096.
- Rubin, P. E. R., and Löfqvist, A. (1996). "HADES (Haskins Analysis Display and Experiment System)," Haskins Labs. Status Rep. Speech Res. (available at: <http://www.haskins.yale.edu/MISC/DOCS/HADESRLn2.pdf>). Viewed 5/11/07.
- Schönle, P. (1988). *Elektromagnetische Artikulographie* (Electromagnet Articulography, Springer, Berlin).
- Vance, T. (1987). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).
- Vatikiotis-Bateson, E., and Kelso, J. A. S. (1993). "Rhythm type and articulatory dynamics in English, French, and Japanese," *J. Phonetics* 21, 231–265.
- Vaxelaire, B. (1995). "Single vs. double (abutted) consonants across speech rate: X-ray and acoustic data for French," in *Proceedings XIII Int. Conf. Phonetic Sci.*, Vol. 1, 384–387, Stockholm, Royal Institute of Technology and Stockholm University.
- Warner, N., and Arai, A. (2001). "Japanese mora-timing: A review," *Phonetica* 58, 1–25.
- Westbury, J. (1994). "On coordinate systems and the representation of articulatory movements," *J. Acoust. Soc. Am.* 95, 2271–2273.
- Winer, B., Brown, R., and Michels, K. (1991) *Statistical Principles in Experimental Design* (McGraw-Hill, Boston).