Vowel-related tongue movements in speech: Straight or curved paths? (L)

Anders Löfqvist^{a)}

Department of Logopedics, Phoniatrics and Audiology, Clinical Sciences, Lund, Lund University, University Hospital, SE-221 85 Lund, Sweden

(Received 15 July 2010; revised 7 December 2010; accepted 11 December 2010)

This paper examines tongue movements between the two vowels in sequences of vowel-labial consonant-vowel, addressing the question whether the movement is a straight line or a curved path. Native speakers of Japanese and Italian served as subjects. The linguistic material consisted of words where the bilabial consonant was either long or short. The inclusion of words with different consonant lengths was motivated by earlier findings that the tongue movement is often longer when the consonant is long, which may be due to a more curved movement path. Tongue movements were recorded using a three-transmitter magnetometer system. To assess the movement path, the movement magnitude was calculated in two ways, as a straight line, the Euclidean distance, and as the actual path, obtained by summing the individual Euclidean distances between successive samples from movement onset to offset. The ratio between the path and the Euclidean distance is 1 when the movement is a straight line and greater than 1 when the path is curved. Results show that in virtually all 21 cases examined the ratio was very close to 1 and in most cases 1.2 or less. There was no reliable influence of consonant length on the ratio. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3533722]

PACS number(s): 43.70.Aj [DAB]

Pages: 1149–1152

I. INTRODUCTION

This paper examines tongue movements between the two vowels in sequences of vowel-labial consonant-vowel. The specific question being addressed is whether the movement follows a straight line or a curved path. Tongue movements in speech have mostly been studied for consonants in this respect (e.g., Houde, 1968; Perkell, 1969; Kent and Moll, 1972b; Munhall *et al.*, 1991; Mooshammer *et al.*, 1995; Löfqvist and Gracco, 2002). Less work has been done on movements between vowels. For example, there are only a few studies of diphthongs (Kent and Moll, 1972a; Sovijärvi, 1962; Hu, 2005a,b). Engstrand (1988) examined tongue movements in Swedish VCV syllables as a function of stress and speaking rate. He noted (p. 1869) that the movement between the two vowels followed a more curved path in slow and stressed productions but presented no trajectories. Material presented by Löfqvist and Gracco (1999) suggested that the tongue movement between two vowels flanking a bilabial consonant could follow both straight and curved paths. These observations were qualitative, however, and no formal analyses were made in that study.

One prediction, based on minimization of effort, is that the movement will be very close to a straight line. Considerations of effort have been used in speech research to understand and explain movement patterns. For example, Lindblom (1983, see also Lindblom 1990) discussed coarticulation and segment reduction in terms of minimization of effort, as indexed by movement amplitude. A series of studies have examined effort in speech using the model proposed by Nelson (1983), which developed a cost measure for movements based on movement time, amplitude, peak velocity, peak

acceleration, and jerk (the third derivative of position). Nelson *et al.* (1984) recorded jaw movements in syllables produced at increasing speech rate. The results showed that as movement time decreased at the rapid rates, the peak velocity and peak acceleration increased for one out of three subjects. There were also differences between speech and non-speech movements at the slower rates in terms of effort, but these differences tended to disappear at the rapid rates, most likely due to a performance boundary. Perkell *et al.* (2002) examined performance spaces for different speaking conditions, such as normal, clear, fast, and slow speech. The results varied between speakers, with some speakers producing the clear speech with more effort. Perkell and Zandipour (2002) compared cyclical non-speech movements with speech movement but failed to find any clear differences in terms of cost.

Löfqvist (2006) examined tongue movements between vowels flanking a long or short bilabial consonant in Japanese. The results showed that the duration was longer and the average speed of the movement was lower during a long consonant than during a short consonant. The magnitude of the tongue movement path did not consistently differ between long and short consonants. Whenever the movement was larger, one possible reason is that it followed a more curved path, but all the movements in that study were calculated as the path and not the Euclidean distance. Thus, the purpose of this paper is to compare vowel-related movements measured in two different ways. To increase the database, three speakers of Italian have been examined, since Italian has a similar durational contrast as in Japanese.

II. METHOD

A. Subjects

Five native speakers of Japanese (the same subjects as in Löfqvist, 2006) and three native speakers of Italian served

^{a)}Author to whom correspondence should be addressed. Electronic mail: lofqvist@haskins.yale.edu

as subjects. They reported no speech, language, or hearing problems. They were naïve as to the purpose of the experiment. The experimental protocol was approved by the Institutional Review Board (IRB) at the Yale University School of Medicine.

B. Linguistic material

The linguistic material consisted of Japanese and Italian words with a sequence of vowel-labial nasal-vowel. These words formed minimal pairs, where the only difference between the pairs was the duration of the labial consonant, which was either long or short. The following Japanese words were used: /kami, kammi/, /kamee, kammee/, /kema, and kemma/. The Italian material consisted of the following words: /ca'mino, ca'mmino/ /sa'remo, and sa'remmo/. These words were selected because there is a substantial movement between the two vowels. The linguistic material was organized into randomized lists and presented to the subjects in Japanese or European writing, with the words occurring in a short frame sentence. Fifty repetitions of each word were recorded.

C. Movement recording

Articulatory movements 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 in the order of 0.5 mm. Receivers were placed on the vermillion border of the upper and lower lip, on three positions of the tongue, referred to as tip, blade and body, and on the lower incisors at the gum line. 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).

The articulatory movement signals 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 forward and backward 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 ($v = \sqrt{(\dot{x}^2 + \dot{y}^2)}$) was also calculated. The velocity signals were smoothed using the same triangular window. The acoustic signal was pre-emphasized, low-pass filtered at 4.5 kHz, and sampled at 10 kHz.

The analysis of the tongue movements used the receiver on the tongue body, which is appropriate since we are interested in the overall movement of the tongue. Movement onsets and offsets were defined algorithmically in the tongue body speed signal as minima during the first and second vowels, see Fig 1. The magnitude of the tongue movement trajectory from the first to the second vowel was measured in two ways. The first one was obtained by summing the Euclidean distances between successive samples of the tongue

FIG. 1. The top panel shows receiver trajectories between the two vowels in the Japanese word "kami." The bottom panel shows audio and tongue body signals for the same word, indicating the points used in the speed signal for defining the onset and offset of the tongue movement between the two vowels. The baseline in the bottom panel represents zero speed. The vertical shaded area is the oral closure for the consonant.

body vertical and horizontal receiver positions from movement onset to movement offset. The second one was the Euclidean distance between tongue movement onset and offset, i.e., a straight line. The ratio between the path and the Euclidean distance is 1 when the movement is a straight line and greater than 1 when the path is curved.

The kinematic signals are expressed in a maxilla-based coordinate system. Thus, the tongue body movement includes

TABLE I. The ratio between the path and the Euclidean distance for Japanese, subjects mean (standard deviation).

Subject					
Word	1	2	3	4	5
kami	1.08 (.07)	1.1 (.03)	1.14 (.1)	1.1 (.05)	1.05 (.02)
kammi	1.03 (.02)	1.09 (.03)	1.19 (.12)	1.1 (.06)	1.03 (.01)
kame	1.03 (.02)	1.11 (.08)	1.38 (.15)	1.06 (.05)	1.03 (.02)
kamee	1.05 (.05)	1.08 (.04)	1.45 (.16)	1.06 (.04)	1.03 (.03)
kema	1.05 (.03)	1.1 (.11)	2.42 (.7)	1.06 (.07)	1.04 (.1)
kemma	1.08 (.05)	1.02 (.01)	2.05 (.5)	1.03 (.02)	1.02 (.01)

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

III. RESULTS

The results for the Japanese subjects are shown in Table I and those for the Italian subjects in Table II. It is evident from these figures that with one single exception, the words /kema and kemma/ of subject No. 3 in Table I, all ratios are close to 1, and in most cases 1.2 or lower. Thus, for the cases examined here, the tongue movement between the two vowels flanking a labial consonant is basically a straight line in extrinsic space. It is also evident from Tables I and II that there is no pattern for the words with long and short consonants within or across subjects. That is, although there were statistically significant differences between the ratios for long and short consonants for the Japanese subjects 1, 2, and 3, they do not follow the same pattern. Only speaker 2 has a consistently greater ratio for the short consonant than for the long consonant. None of the Italian speakers in Table II had any statistically significant difference between the long and short consonants. Correlations between the movement time and the ratio showed no consistent results. Nor did correlations between the movement trajectory and the ratio, It is unclear why one subject showed a curved movement path in one pair of words.

IV. DISCUSSION

The purpose of this study was to examine the tongue movement pattern between two vowels with particular emphasis on the shape of the movement trajectory, i.e., if it followed a curved path or a straight line. This was assessed by measuring the movement as a straight line, the Euclidean distance between movement onset and offset, and as the actual path. The ratio between the path and the Euclidean distance is 1 when the movement is a straight line and

greater than 1 when the path is curved. The results for five speakers of Japanese and three speakers of Italian all suggest that the movement very closely follows a straight line. That is, the ratios were close to 1, and in most cases 1.2 or lower.

The results showed no clear relationship between the ratio and consonant length. That is, speakers did not consistently produce a larger ratio for the short or long consonant. There was also no clear relationship between the ratio and movement time, and path length. Thus, longer paths were not due to a more curved path.

The present study used receiver on the tongue body for measuring tongue movements between two vowels. This is appropriate when the focus is on the overall movement of the tongue. The top panel in Fig. 1 shows all the three receivers on the tongue moving in very similar trajectories.

The tongue is loosely coupled to the jaw, so some of the tongue movement is contributed by the jaw when the movements are expressed in a maxilla-based coordinate system. The magnitude of the jaw's contribution differs between subjects (Noiray *et al.*, 2008). In jaw and arm movements, the end effector can move in a straight or curved path depending on the timing and magnitude of changes in joint angles (e.g., Abend *et al.*, 1982; Munhall *et al.*, 1991). The tongue does not have a bony skeleton (Smith and Kier, 1989), so different trajectories cannot be due to joint angles but rather to different activation patterns of the extrinsic tongue muscles. The present findings of the tongue moving in almost straight lines are most likely due to a minimization of effort (cf. Ohta *et al.*, 2004). In a simple reaching movement, the hand moves in a straight line, whereas if the movement has to pass through successive target regions, the movement is a curved path. This is also true for more complex tongue movements such as those found in a sequence of a vowel-velar consonant-vowel.

ACKNOWLEDGMENTS

The author is grateful to Mariko Yanagawa for help with the linguistic material and running the experiments. This work was supported by Grant DC-00865 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health (A. Lofqvist PI) to Haskins Laboratories, and by the Linneaus Center for Cognition, Communication, and Learning at Lund University, funded by Grant 349-2007-8695 from the Swedish Research Council.

TABLE II. The ratio between the path and the Euclidean distance for Italian subjects, mean (standard deviation).

Subject			
Word	1	2	3
Camino	1.04 (.02)	1.07 (.11)	1.09 (.03)
Cammino	1.03 (.02)	1.06 (.03)	1.09 (.03)
Saremo	1.03 (.02)	1.12 (.09)	1.03 (.04)
Saremmo	1.04 (.08)	1.16 (.14)	1.04 (.04)

- Abend, W., Bizzi, E., and Morasso, P. (1982). "Human arm trajectory formation," *Brain* **105**, 331–348.
- Engstrand, O. (1988). "Articulatory correlates of stress and speaking rate in Swedish VCV utterances," *J. Acoust. Soc. Am.* **83**, 1863–1875.
- Houde, R. (1968). *A Study of Tongue Body Motion during Selected Consonant Sounds* (Speech Communications Research Laboratory, SCRL Monograph 2, Santa Barbara), pp. 1–159.
- Hu, F. (2005a). "Tongue kinematics in diphthong production in Ningbo Chinese," in *Proceedings of the Interspeech*, Lisbon, pp. 1029–1032.
- Hu, F. (2005b). (available at <http://lbms03.cityu.edu.hk/theses/ftt/phd-ctl-b19887395f.pdf>, last visited 100701) "A Phonetic Study of the Vowels of Ningbo Chinese," Ph.D. thesis, City University of Hongkong, Hong Kong.
- Kent, R. D., and Moll, K. L. (1972a). "Tongue body articulation during vowel and diphthong gestures," *Folia Phoniatr Logop* **24**, 278–300.
- Kent, R., and Moll, K. (1972b). "Cinefluorographic analyses of selected lingual consonants," *J. Speech Hear. Res.* **15**, 453–473.

- Lindblom, B. (1983). "Economy of speech gestures," in *The Production of Speech*, edited by P. MacNeilage (Springer, New York), pp. 217–245.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H and H theory," in *Speech Production and Speech Modelling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.
- Löfqvist, A. (2006). "Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese," *J. Acoust. Soc. Am.* **120**, 2872–2883.
- Löfqvist, A., and Gracco, V. (1999). "Interarticulator programming in VCV sequences: Lip and tongue movements," *J. Acoust. Soc. Am.* **105**, 1864–1876.
- Löfqvist, A., and Gracco, V. (2002). "Control of oral closure in lingual stop consonant production," *J. Acoust. Soc. Am.* **111**, 2811–2827.
- 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.
- Nelson, W. (1983). "Physical principles for economies of skilled movements," *Biol. Cybern.* **46**, 135–147.
- Nelson, W., Perkell, J., and Westbury, J. (1984). "Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations," *J. Acoust. Soc. Am.* **75**, 945–951.
- Noiray, A., Iskarous, K., Bolanos, L., and Whalen, D. (2008). (available at <http://issp2008.loria.fr/Proceedings/PDF/issp2008-14.pdf>, last visited 100701) "Tongue-jaw synergy in vowel height production: Evidence from American English," in *Proceedings of the 8th International Seminar on Speech Production*, pp. 81–84.
- Ohta, K., Svinin, M., Luo, Z., Hosoe, S., and Laboissiere, R. (2004). "Optimal trajectory formation of constrained human arm reaching movements," *Biol. Cybern.* **91**, 23–36.
- Perkell, J. S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA), pp. 1–104.
- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabeta, J., and Jackson, M. (1992). "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," *J. Acoust. Soc. Am.* **92**, 3078–3096.
- Perkell, J., Zandipour, M., Matthies, M., and Lane, H. (2002). "Economy of effort in different speaking conditions. I. A preliminary study of inter-subject differences and modeling issues," *J. Acoust. Soc. Am.* **112**, 1627–1641.
- Perkell, J., and Zandipour, M. (2002). "Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements," *J. Acoust. Soc. Am.* **112**, 1642–1651.
- Smith, K., and Kier, W. (1989). "Trunks, tongues and tentacles: Moving with skeletons of muscle," *Am. Sci.* **77**, 29–35.
- Sovijärvi, A. (1962). "Röntgenkinematografisch-akustische Untersuchungen über die Artikulation der Diphthonge," in *Proceedings of the 4th International Congress Phonetic Sciences*, University of Helsinki, The Hague, Mouton, pp. 111–128.
- Westbury, J. (1994). "On coordinate systems and the representation of articulatory movements," *J. Acoust. Soc. Am.* **95**, 2271–2273.