AN MRI-BASED ANALYSIS OF THE ENGLISH /r/ AND /l/ ARTICULATIONS

Shinobu Masaki [1], *Reiko Akahane-Yamada* [1], *Mark K. Tiede* [1],
Yasuhiro Shimada [2] *and Ichiro Fujimoto* [2]

[1] ATR Human Information Processing Research Laboratories
2-2 Hikaridai, Seikacho, Sorakugun, Kyoto 619-02 Japan
[2] Takanohara Chuo Hospital
1-3-3 Ukyo, Narashi, Nara 631 Japan

ABSTRACT

Midsagittal tongue shapes for sustained English /r/ and /l/ sounds between native speakers of American English (AE) and Japanese were compared using the Magnetic Resonance Imaging (MRI) technique. The /r/ sound as produced by AE speakers was characterized by a constriction at the anterior part of the hard palate and the existence of a sublingual cavity, and for /l/ sounds, apical contact to the front teeth and/or alveolar ridge, and the absence of the sublingual cavity. For Japanese speakers, strategies to form the tongue shape contrast between /r/ and /l/ productions were categorized into four types depending on the type of contact/constriction and presence/absence of the sublingual cavity. The first type showed a pattern of tongue shape similar to AE speakers. The second and third types were characterized by /r/ and /l/ oriented production, respectively, for both sounds. In the last type, the distinction between /r/ and /l/ was formed only by the absence or presence of apical contact, while a sublingual cavity was produced for both sounds. These types are discussed in the context of a perceptual evaluation and an acoustical analysis.

1 Introduction

It is well known that native speakers of Japanese have difficulty in distinguishing English /r/ and /l/, both in production and perception. Characteristics of their production have been studied by means of acoustical analysis and perceptual evaluation by English speakers [9, 8]. However, in order to train Japanese speakers to produce these sounds adequately, an improved understanding of the differences in articulation strategies between native American English (AE) and Japanese speakers is necessary.

Several studies have examined the characteristics of articulation for /r/ and /l/ sounds by AE speakers. For /r/ production, observations by cineradiography [3] and x-ray microbeam (XRMB) [7] revealed the existence of several types of tongue shapes. All types however showed two characteristic constrictions: one between either the tip or dorsum of the tongue and the hard palate, and the other between the posterior part of the tongue and the pharyngeal wall. As for /l/ production, apical contact to the alveolar ridge has been observed by electropalatography (EPG) [4, 2], and tongue dorsum retraction and lowering have been observed using XRMB [6]. In contrast, only a few studies have been reported for the /r/ and /l/ articulations of Japanese subjects [4, 5]. Considering the large variation in acoustical characteristics among Japanese speakers in /r/ and /l/ production [8, 9], more articulation data must be accumulated.

With this aim we have used the Magnetic Resonance Imaging (MRI) technique (e.g. [1]). One of the advantages of MRI applied to speech research is its non-invasiveness, which permits accumulation of data from many subjects. Another advantage is that the MRI technique permits three dimensional reconstruction of articulator posture, facilitating analysis of vocal tract shape and its associated transfer function.

In this paper as a first step we have compared midsagittal tongue shapes of sustained English /r/ and /l/ as produced by native speakers of AE and Japanese. The articulatory characteristics are discussed in terms of differences in production strategy in the realization of the /r/-/l/ contrast. Also, we have attempted to probe the effects of these differences on both acoustical characteristics and perceptual evaluation by native speakers of AE.

2 Method

2.1 MRI Data Acquisition

Five native speakers of AE (A1-A5; only A5 was female) and nine native speakers of Japanese (J1-J9, only J4 was female) served as subjects. The AE subjects ranged in age from 26 to 46, and the Japanese subjects from 23 to 50. All of the Japanese subjects were exposed to English instruction at school, starting from about 12 years old until at least 20 years old.

For each subject, we obtained three sagittally oriented images bracketing the mid-line of the vocal tract during the production of /r/ and /l/ sounds. Subjects lay in a supine position and produced sustained /r/ and /l/ by prolonging the articulation of these sounds in "bird" and "pull," respectively. Since the duration of data acquisition (about from one to two minutes) required that subjects interrupt phonation to breathe, they were instructed to do so without changing the position of any articulators.

The MRI system used in this experiment was a Shimadzu SMT-100GUX (static magnetic field = 1T). A spin-echo method (TR = 150, 230 or 300, TE = 15 or 18 ms, number of excitations = 2) was used to obtain images of a 25x25 cm field of view for each slice (thickness = 3 or 5 mm) mapped to 256x256 pixels.

2.2 Practice and Recording of Utterance

Because sustained phonation in a supine position is highly artificial, subjects practiced each utterance under simulated conditions separately before the actual MRI data acquisition. Subjects lay in a supine position, and produced sustained /r/ in "bird" and /l/ in "pull" for about 30s including two interruptions by breathing. They were instructed not to change the posture of articulators during breathing as well as phonation. The productions were recorded and digitized at a sampling frequency of 22.05 kHz with 16-bit resolution. Because of the noise associated with MRI data acquisition, recorded productions from the practice session were used for the later perceptual evaluation and acoustical analysis.

2.3 Perceptual Evaluation by Native Speakers of American English

Productions were evaluated perceptually by native speakers of AE at University of South Florida. The corpus evaluated comprised /r/ and /l/ productions by all nine Japanese plus two AE speakers (A1 and A5) for control. Each stimulus was a 1s stable portion of the second phonation of sustained /r/ or /l/ recorded in the practice session. Five phonetically trained native speakers of AE served as evaluators, and participated in two tests.

In the first test, the evaluators judged the "intelligibility" of the production using a two alternative forced choice (2AFC) from /r/ or /l/. Each of the 22 stimuli were presented five times in random order through headphones. In the second test, they scored the "goodness" of the production as the speaker's intended phoneme (/r/ or /l/) between 1 (worst) and 7 (best). The /r/ and /l/ stimuli were evaluated in separate blocks. In each block, 11 stimuli were each presented five times in random order.

2.4 Acoustical Analysis

Frequencies of the first three formants (F1-F3) were extracted from the same portions used in the perceptual evaluation. Speech waves were down-sampled from 22.05 kHz to 10 kHz, and LPC analysis was used with the following parameters: sampling frequency = 10 kHz, pre-emphasis factor = 0.9, window length = 25 ms (Hanning), window interval = 10 ms, and order of LPC = 14. The values were averaged across all frames in each 1s segment.

3 Results

3.1 MRI Data and Perceptual Evaluation

In describing the characteristics of /r/ and /l/ articulation, it is important to characterize the location of lingual contact or constriction and the presence or absence of a sublingual cavity. In this section, we first describe characteristics of midsagittal tongue shape for /r/ and /l/ as produced by AE subjects. Then we compare differences in production strategy between AE and Japanese speakers.

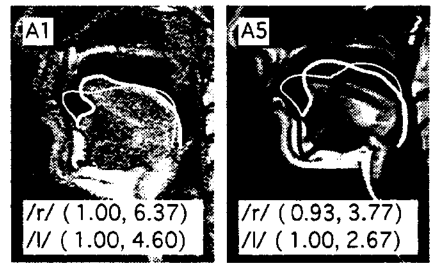

Figure 1: Examples of midsagittal tongue shapes for /r/ and /l/ by native AE speakers. In each panel the original MR image and a thick line represents the tongue shape for /r/, and the superimposed thin line shows that for /l/. Numbers in parentheses at the bottom represent the results of evaluation by native AE speakers; an "intelligibility score" (left; 0 - 1) and a "goodness score" (right; 1 - 7) for sustained utterances of /r/ and /l/.

AE subjects: Figure 1 shows examples of midsagittal tongue shapes for /r/ and /l/ sounds produced by AE subjects (A1 and A5). In this figure, overlaid tracings of the midsagittal tongue shape for /r/ and /l/ are shown by thick and thin white lines, respectively. The numbers in parentheses at the bottom represent the results of the evaluation by native AE speakers; the first is an "intelligibility score" and the second is a "goodness score" for each of the /r/ and /l/ productions. The scores are averaged values from the five evaluators.

All five AE subjects showed the same strategy to effect the tongue shape contrast between /r/ and /l/. The characteristics of the observed tongue shape can be summarized as follows.

/r/ production: the primary constriction was located along the anterior part of the hard palate using either the tongue tip or tongue dorsum. A sublingual cavity was always present.

/l/ production: apical contact was located at the upper front teeth and/or alveolar ridge. Sublingual contact with the lower front teeth was observed along the sagittal midline.

Note that for both /r/ and /l/ production by native AE speakers, a secondary constriction was observed between the posterior part of the tongue and the pharyngeal wall. An extreme example was A5 shown in Figure 1.

In the perceptual evaluation by native AE speakers, these subjects (A1 and A5) received perfect intelligibility scores except for the /r/ of subject A5 (0.93), who had difficulty sustaining a prolonged /r/ and /l/.

Japanese subjects: Based on the existence of lingual contact/constriction and a sublingual cavity, the patterns of tongue shape contrast between /r/ and /l/ production by Japanese speakers were categorized into 4 types, Type-A through Type-D, as shown in Figure 2. Tongue shape contrast and the results of perceptual evaluation are summarized as below. With respect to the perceptual evaluation, since the goodness score is roughly correlated with the intelligibility score, only the results of the intelligibility test will be shown.

Type-A (J1, J3, J5 and J6): These 4 Japanese speakers showed the same tendencies observed for the AE speakers. Out of these subjects, 2 subjects (J1 and J5) received perfect or fairly good intelligibility scores. However, the other 2 subjects (J3 and J6) received good scores only for /l/, and their /r/ production was also perceived as /l/.

Type-B (J2): This subject showed no apical contact to the upper front teeth or hard palate for either /r/ or /l/, and always produced a sublingual cavity. The magnitude of the difference in tongue shape between /r/ and /l/ is smaller than that of the Type-A subjects. The intelligibility score for /r/ was fairly good while /l/ production was completely identified as /r/.

Type-C (J4): This subject also showed only a small variation in tongue shape between /r/ and /l/. To produce both sounds, the tongue tip was inserted between the upper and lower front teeth, and no visible sublingual cavity was produced even for the /r/ production. The intelligibility score for /l/ was fairly good while /r/ production by this subject was identified as /l/.

Type-D (J7, J8 and J9): Like Type-B, these three subjects showed a sublingual cavity for both /r/ and /l/ productions. The difference from Type-B was that the tongue shape contrast between these two sounds was formed by the presence/absence of lingual contact to the alveolar ridge (J7 and J9) or the hard palate (J8). Out of the three subjects, 2 subjects (J7 and J8) received perfect or fairly good intelligibility scores for both /r/ and /l/ productions. However, for the last subject (J9), both the /r/ and /l/ productions were uniformly perceived as /r/.

3.2 Acoustical Analysis

Distribution of F2 and F3 for English /r/ and /l/ production by Japanese speakers has been reported to have a different pattern from native AE speakers[9]. Hence we will concentrate on describing the effects of tongue shape on these acoustical parameters. Figure 3 shows a scatter plot of formant frequencies on the F2-F3 plane for /r/ and /l/, for native speakers of AE and Japanese. Each data point shows the mean value over all analyzed 1s segments. The results can be summarized as follows.

AE subjects: The distribution of data points on the F2-F3 plane for /r/ or /l/ sounds produced by the AE subjects was restricted, indicating that the inter-subject variation of acoustical characteristics is small. This result is consistent with previous reports [3, 7, 9].

Japanese subjects: On the other hand, the relationship of data points between /r/ and /l/ produced by native Japanese speakers was different from subject to subject.

Type-A (J1, J3, J5 and J6): Two subjects (J1 and J5) who received good perceptual scores in the intelligibility test for both /r/ and /l/ showed a similar pattern of F2 - F3 distribution as that seen for AE subjects. However, the other subjects (J3 and J6), who received low scores for /r/ production, showed inconsistent patterns.

Type-B (J2): This subject's value for /r/ was not in the region of

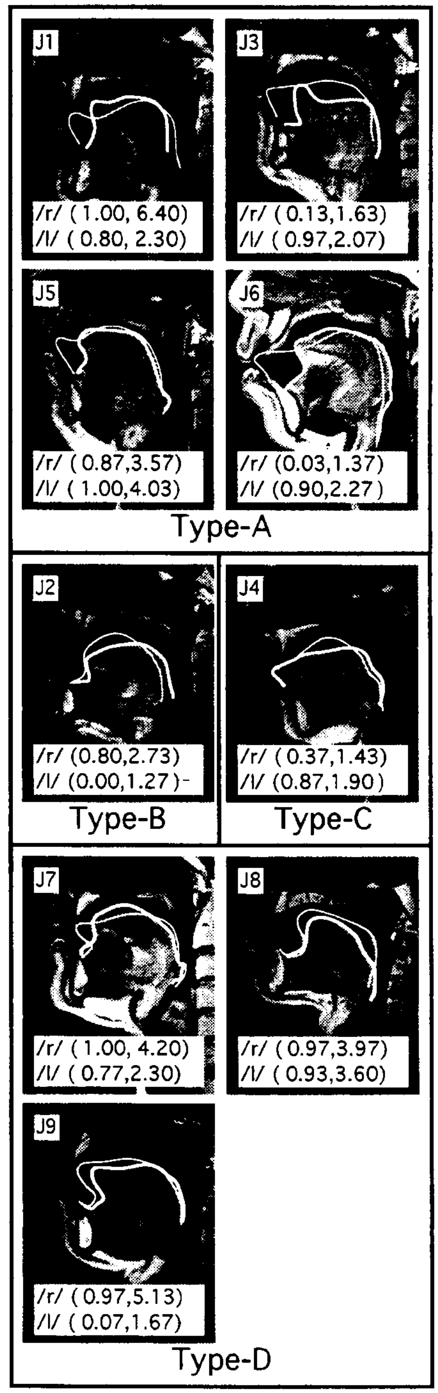

Figure 2: Midsagittal tongue shape for /r/ and /l/ by native speakers of Japanese. The patterns of /r/ - /l/ contrast of tongue shapes were categorized into 4 types (Type-A through Type-D) based on the existence of lingual contact/constriction and a sublingual cavity.

either the /r/ or /l/ value observed for AE subjects, but he showed a value for /l/ consistent with the AE subjects' /r/ value. This corresponded to the result of the perceptual intelligibility test for this subject, in which /l/ production was uniformly perceived as /r/, while the score of /r/ identification for /r/ production was lower.

Type-C (J4): The F3 frequency for both /r/ and /l/ was almost the same as that for /l/ production by AE subjects. This might be the reason for the /l/ oriented intelligibility scores obtained for this subject.

Type-D (J7, J8 and J9): No consistent tendency was observed among these three subjects. J8 received a good intelligibility score for both /r/ and /l/ productions, and showed a pattern of F2 - F3 distribution similar to that of native AE subjects. For J9, whose production of both /r/ and /l/ was consistently identified as /r/, the data points for these sounds were restricted to a region which is near the area of data points for the /r/ production of native AE speakers. For J7, the reason why the /l/ production received a fairly good score might be explained by the location of the data point for /l/ in the F2 - F3 plane which was closer to the points for the /l/ production of AE subjects than that for /r/. However, it is difficult to explain why /r/ received a perfect score in intelligibility test, because the data point for /r/ was also closer to the area for the /l/ production of native AE speakers than that for /r/.

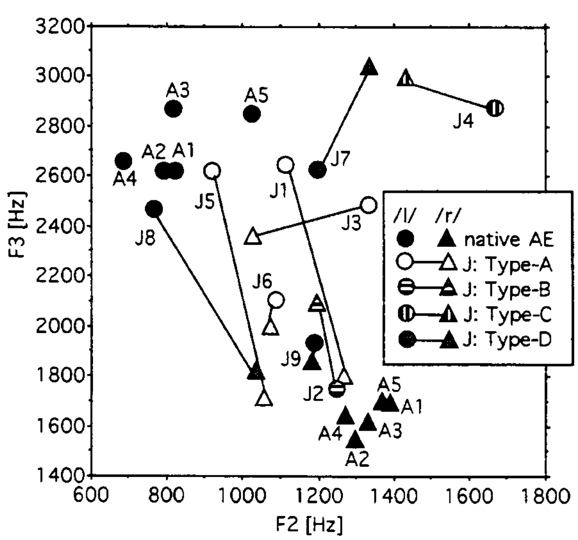

Figure 3: Scatter plot of formant frequencies on F2-F3 plane for /r/ and /l/ by AE and Japanese subjects.

4 Discussion

Strategies for forming the tongue shape contrast for /r/ and /l/ production by Japanese speakers were categorized into four types depending on the existence of lingual contact/constriction and a sublingual cavity. This categorization was often, although not always, coincident with their acoustic properties and/or their qualities as perceptually evaluated by AE listeners. This suggests that the characteristics of lingual contact/constriction and the sublingual cavity observed in the midsagittal vocal tract can provide crucial information for evaluating the validity of Japanese strategies of /r/ and /l/ production. The mismatch between tongue shape and acoustic properties for some productions by Japanese speakers may be due to following reasons. First, the source data for perceptual evaluation and acoustical analysis were recorded separately from the MRI data acquisition. Since /r/ and /l/ articulation by Japanese speakers is potentially inconsistent, the articulatory postures during the MRI scan and the recording session may have differed. Second, the midsagittal image by itself is not sufficient to characterize the articulations. For instance, articulation of /l/ is also characterized by lateral side-channels which cannot be observed from the midsagittal shape alone. In order to address these methodological problems we plan to replicate the experiment using MRI data acquisition of 3D volume in conjunction with simultaneous acoustic recording during scanning.

5 ACKNOWLEDGMENT

We are grateful to Prof. Winifred Strange and Brett H. Fitsgerald at the University of South Florida for their collaboration in the perception experiment. We also thank Rieko Kubo for processing the speech data.

6 REFERENCES

1. T. Bear, J.C. Gore, L.C. Gracco, and P.W. Nye. Analysis of vowel tract shape and dimensions using magnetic resonance imaging: vowels. *Journal of the Acoustical Society of America*, 90(2):799–828, 1991.
2. P.A. Dagenais, L.C. Lorendo, and McCutcheon M.J. A study of voicing and context effects upon consonant linguapalatal contact patterns. *Journal of Phonetics*, 22:225–238, 1994.
3. P. Delattre and D.C. Freeman. A dialect study of American r's by X-ray motion picture. *Linguistics*, 44:29–68, 1968.
4. F.E. Gibbon, W.J. Hardcastle, and H. Suzuki. An electropalatographic study of the /r/, /l/ distinction for Japanese learners of English. *Computer Assisted Language Learning*, 4(3):153–171, 1991.
5. W.R. Naito. English /r/ and /l/ production in American and bilingual Japanese subjects. In *Journal of the Acoustical Society of America; Program of the 130th Meeting*, volume 98, page 2892, 1995.
6. R. Sproat and O. Fujimura. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21:291–311, 1993.
7. J.R. Westbury, M. Hashi, and M.J. Lindstrom. Differences among speakers in articulation of American English /r/: An X-ray microbeam study. In *Proceedings ICPhS 95*, volume 4, pages 50–57, 1995.
8. R.A. Yamada, W. Strange, J.S. Magnuson, J.S. Pruitt, and W.D. Clark III. The intelligibility of Japanese speakers' production of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. In *Proceedings of the 1994 International Conference on Spoken Language Processing*, pages 2023–2026, 1994.
9. R.A. Yamada and Y. Tohkura. Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese. In *Proceedings of the 1990 International Conference on Spoken Language Processing*, pages 757–760, 1990.