An acoustic analysis of American English liquids by adults and children: Native English speakers and native Japanese speakers of English[a)]

Katsura Aoyama,[1,b)] James E. Flege,[2] Reiko Akahane-Yamada,[3] and Tsuneo Yamada[4]
[1]*Department of Audiology and Speech-Language Pathology, University of North Texas, 1155 Union Circle #305010, Denton, Texas 76203-5017, USA*
[2]*Speech and Hearing Sciences, University of Alabama at Birmingham, 1720 2nd Avenue South, Birmingham, Alabama 35294, USA*
[3]*Advanced Telecommunications Research Institute International, 2-2-2 Hikaridai Seika-cho, Sorakugun, Kyoto, 619-0288, Japan*
[4]*Department of Informatics, The Open University of Japan, 2 Chome-11 Wakaba, Mihama Ward, Chiba, 261-8586, Japan*

(Received 7 June 2019; revised 30 September 2019; accepted 4 October 2019; published online 24 October 2019)

This study investigated acoustic characteristics of American English liquids produced by native English (NE) and native Japanese (NJ) speakers reported in Aoyama, Flege, Guion, Akahane-Yamada, and Yamada [(2004). J. Phonetics **32**, 233–250]. For a larger longitudinal study, the data were collected twice to investigate the acquisition of American English by the NJ speakers (Time 1, Time 2). Aoyama, Flege, Guion, Akahane-Yamada, and Yamada [(2004). J. Phonetics **32**, 233–250] evaluated productions of /l/ and /ɹ/ in the NE and NJ adults and children (16 participants each) using NE speakers' perceptual judgments and showed that the NJ children's production of /ɹ/ improved from Time 1 to Time 2. In the current study, four acoustic parameters (duration, $F1$, $F2$, and $F3$) were measured in 256 tokens each of English /l/ and /ɹ/. Results showed that some acoustic parameters, such as $F2$, changed from Time 1 to Time 2 in the NJ speakers' productions, indicating improvements. However, the NJ speakers' productions were different from the NE speakers' productions in almost all acoustic parameters at both Time 1 and Time 2. Results suggest that the improvements in the NJ children's productions of /ɹ/ reported in Aoyama, Flege, Guion, Akahane-Yamada, and Yamada [(2004). J. Phonetics **32**, 233–250] were due to a combination of changes, not due to a change in one acoustic parameter such as $F3$ in /ɹ/.
 https://doi.org/10.1121/1.5130574

[AL] Pages: 2671–2681

I. INTRODUCTION

The perception and production of English /l/ and /ɹ/ by native Japanese (NJ) speakers have been studied extensively (e.g., Aoyama *et al.*, 2004; Bradlow *et al.*, 1997; Flege *et al.*, 1995, 1996; Goto, 1971; Hattori and Iverson, 2009; Larson-Hall, 2006; Yamada, 1995). In general, previous studies indicate that Japanese speakers' perception scores are well below those of native English (NE) speakers for both identification (e.g., Flege *et al.*, 1996; Hattori and Iverson, 2009) and discrimination (e.g., Aoyama *et al.*, 2004; Guion *et al.*, 2000). In production, Japanese speakers' English /l/ is often misidentified as /ɹ/ by NE speakers, and vice versa (Aoyama *et al.*, 2004; Guion *et al.*, 2000). In short, the contrast between English /l/ and /ɹ/ seems to be particularly difficult for NJ speakers, although some previous studies demonstrated that extensive laboratory training can improve Japanese speakers' perception of /l/ and /ɹ/ (Bradlow *et al.*, 1997; Bradlow *et al.*, 1999).

It has been hypothesized that Japanese speakers' difficulty with English /l/ and /ɹ/ is due to perceptual assimilation patterns of English /l/ and /ɹ/ in relation to the Japanese tap /ɾ/. In a perceptual assimilation experiment, Guion *et al.* (2000) demonstrated that Japanese speakers identified both English /l/ and /ɹ/ as the Japanese tap /ɾ/, with equal goodness-of-fit ratings. In Aoyama and Flege (2011), NJ speakers also identified both English /l/ and /ɹ/ as the Japanese tap /ɾ/. However, when the Japanese speakers were asked to rate the goodness-of-fit of English /l/ and /ɹ/ in reference to the Japanese tap /ɾ/, the similarity rating was higher for /l/ than for /ɹ/. In British English, Hattori and Iverson (2009) used identification and best exemplar experiments to determine perceptual assimilation patterns of /l/ and /ɹ/ by Japanese speakers. The results suggested that British English /l/ and /ɹ/ both assimilate to the Japanese tap /ɾ/, but to a different degree. That is, Japanese speakers perceived British English /l/ as a more similar sound to the Japanese tap /ɾ/ than British English /ɹ/ is to the Japanese /ɾ/. Overall, the above studies suggest that English /l/ and /ɹ/ are perceived as equivalent to the same Japanese category, the Japanese tap /ɾ/, by

[a)]This work was presented at the 170th Meeting of the Acoustical Society of America on June 25, 2017 in Boston, MA. An earlier version of this manuscript was published as K. Aoyama, J. E. Flege, R. Akahane-Yamada, and T. Yamada, Proc. Mtgs. Acoust. **30**, 060008 (2017). This work was supported by the National Institutes of Health (DC00257).
[b)]Electronic mail: kat.aoyama@unt.edu

NJ speakers. In terms of the degree of similarity, English /ɹ/ is perceived as being more deviant from the Japanese tap /ɾ/ than English /l/ is.

The Speech Learning Model (SLM; Flege, 1995) hypothesized that the more "dissimilar" a sound in the second language (L2) is, the more likely L2 learners discern the differences. Consequently, the SLM hypothesized that the L2 learners will more likely improve on more dissimilar sounds than less dissimilar sounds. Aoyama *et al.* (2004) investigated NJ adults' and children's perception and production of English /l/ and /ɹ/ to test these hypotheses of the SLM. In Aoyama *et al.* (2004), NE speakers perceptually evaluated Japanese adults' and children's productions of /l/, /ɹ/, and /w/ by choosing a response from the choices that were provided (*l*, *r*, *w*, *d*, *dr*, *br*, *bl*, *ml* on a computer screen). The results showed that the NJ children's production of English /ɹ/, but not /l/, significantly improved over the course of 1 year. However, the acoustic nature of the observed improvement in Aoyama *et al.* (2004) is not known, because the productions of English /l/ and /ɹ/ were evaluated by native American English speakers' perceptual judgments (i.e., intelligibility scores).

The present study aims to investigate the acoustic characteristics of Japanese children's productions of American English liquids to examine the changes observed in Aoyama *et al.* (2004) at a finer level. The secondary aim of the present study is to compare the acoustic nature of English liquids between native speakers and Japanese L2 speakers of English.

A. Acoustic characteristics of English /l/ and /ɹ/

The acoustic differences between American English prevocalic /l/ and /ɹ/ are well established in adult native speakers of English. It has been demonstrated that the third formant ($F3$) is the main acoustic cue to differentiate /ɹ/ from /l/ in adult native speakers of American English (Espy-Wilson, 1992; Ingvalson *et al.*, 2012). The $F3$ is usually below 2000 Hz for prevocalic /ɹ/, whereas the $F3$ for prevocalic /l/ was around 2500 Hz in Espy-Wilson (1992). Due to a lower $F3$ in /ɹ/ than in /l/, the distance between the second formant ($F2$) and $F3$ is also smaller for /ɹ/ than for /l/. Espy-Wilson (1992) suggested that the $F3$-$F2$ distance of 3.5 Bark differentiates prevocalic /l/ from /ɹ/, because the $F3$-$F2$ distance is typically larger than 3.5 for /l/ and smaller than 3.5 for /ɹ/. There are several other acoustic cues for distinguishing /l/ and /ɹ/ in perception including the first formant ($F1$) transition, the onset of $F2$, and duration (Ingvalson *et al.*, 2011; Polka and Strange, 1985; Yamada and Tohkura, 1992). However, the lower $F3$ in /ɹ/ than in /l/ appears to be the primary cue in differentiating /l/ and /ɹ/ in English (Hattori and Iverson, 2009; Ingvalson *et al.*, 2011).

It is also known that American English liquids, especially /ɹ/, are acquired relatively late by children (McGowan *et al.*, 2004; Sander, 1972; Smit *et al.*, 1990). Smit *et al.* (1990) reported that NE-speaking children's productions of /ɹ/ did not reach 90% "acceptance level" (p. 781) until about 7 yr of age. Idemaru and Holt (2013) reported that production accuracy of both /l/ and /ɹ/ was around 90% among 4.5 yr olds. Both Idemaru and Holt (2013) and Smit *et al.* (1990) indicated that NE-speaking children's /l/ and /ɹ/ are almost always perceived as accurate by age 8 and 9 yr.

The development of /l/ and /ɹ/ has also been studied acoustically in NE-speaking children. Dalston (1975) reported that 3- to 4-yr old children's productions of /l/ and /ɹ/ were not as clearly differentiated as adults' productions acoustically, although their /l/ and /ɹ/ productions were perceived as correct and acoustically distinguishable from each other. More recently, Idemaru and Holt (2013) studied the production development of /l/ and /ɹ/ by NE-speaking children between ages 4 and 8 yr. It was found that children's productions of /l/ and /ɹ/ were acoustically differentiated by $F3$ in all groups, including the 4-yr olds. On the other hand, their productions of /l/ and /ɹ/ were not differentiated by the $F3$-$F2$ distance until age 5. In the older groups, productions of English /l/ and /ɹ/ were clearly differentiated by both $F3$ and the $F3$-$F2$ distance. Taken together, the above studies suggest that NE-speaking children's productions of /l/ and /ɹ/ are perceived accurately by adults and acoustically well-distinguished by approximately 8 yr of age.

B. Japanese speakers' productions of English /l/ and /ɹ/

Japanese L2 speakers' productions of English /l/ and /ɹ/ have been studied extensively (e.g., Aoyama *et al.*, 2004; Flege *et al.*, 1995; Ingvalson *et al.*, 2011; Larson-Hall, 2006), although most studies analyzed their productions through perceptual judgments by NE speakers. Previous studies indicated that Japanese speakers' production accuracy of both English /l/ and /ɹ/ is lower than NE speakers (e.g., Aoyama *et al.*, 2004; Ingvalson *et al.*, 2011). Flege *et al.* (1995) demonstrated that Japanese speakers could learn to produce English /l/ and /ɹ/ accurately if they have had extensive experience using English. Two groups of Japanese speakers participated in Flege *et al.* (1995): "experienced" Japanese speakers with over 20 yr of length of residence (LOR) in the U.S. and "inexperienced" Japanese speakers with less than 2 yr of LOR. NE speakers identified the Japanese speakers' productions of /l/ and /ɹ/ by using forced-choice response alternatives ("definitely L," "probably L," "possibly L", "possibly R," "probably R," and "definitely R"). Flege *et al.* (1995) showed that experienced Japanese speakers' productions of /l/ and /ɹ/ were identified as the intended sound more than 95% of the time by NE speakers. In comparison, inexperienced Japanese speakers' productions were identified as the intended sound between 80% and 90% of the time. In Flege *et al.* (1995), NE-speaking judges also rated the degrees of foreign accent in consonant-vowel syllables with an initial /l/ or /ɹ/ (e.g., /lɑ/ and /ɹɑ/). Results showed that experienced Japanese speakers' productions received lower ratings (i.e., stronger accent) than NE speakers' productions, even though their /l/ and /ɹ/ were identified as the intended segment.

Experienced and inexperienced Japanese speakers' productions of /l/ and /ɹ/ were then acoustically analyzed in Flege *et al.* (1995). Acoustically, $F1$ and $F2$ in /ɹ/ did not differ between the NE speakers and the Japanese speakers, but

$F3$ was higher in the Japanese speakers' productions than in the NE speakers' productions. For /l/, the Japanese speakers' productions did not differ from the NE speakers' in $F1$ and $F3$, but the $F2$ was higher in the Japanese speakers' productions than in the NE speakers' productions. These results indicate that, even though the experienced Japanese speakers' productions were almost always identified as the intended sound, their productions of /l/ and /ɹ/ were acoustically different from the native speakers' productions in both /l/ and /ɹ/. Based on these results, Flege *et al.* (1995) suggested that acoustic analysis can provide insights and details that are not observed in native speakers' perceptual judgments.

Saito and Munro (2014) studied the production development of English /ɹ/ by adult Japanese learners of English. Acoustic "benchmarks" (p. 460) were used in the analysis of English /ɹ/ produced by Japanese L2 learners of English with differing LOR in Canada. $F1$, $F2$, $F3$, and transition duration were measured in /ɹ/ produced by L2 Japanese speakers as well as by native Canadian English speakers. Results showed that $F3$ was higher in Japanese speakers' /ɹ/ than in NE speakers' /ɹ/ (14.5 to 15.5 Bark vs 12.5 Bark). Japanese speakers produced target-like $F2$ with less than 1 year of LOR in Canada, indicating some improvements (Saito and Munro, 2014). Saito and Lyster (2012) also studied Japanese L2 learners' production of /ɹ/ acoustically. The results showed that the $F3$ was higher in the NJ speakers' /ɹ/ (around 2500 Hz) than the NE speakers' /ɹ/ (around 1700 Hz). In addition, Saito and Lyster (2012) showed that $F3$ is an important aspect for NE listeners when they judged the goodness of Japanese speakers' production of /ɹ/. That is, the tokens of /ɹ/ with a lower $F3$ (around 2200 Hz) were judged to be "very good" while those with a higher $F3$ (around 2800 Hz) were judged to be "very poor" by NE-speaking judges (p. 616). These studies showed that $F3$ is an important acoustic cue for the production of /ɹ/ for Japanese L2 learners of English. These studies also showed that $F3$ may be more difficult than other cues, such as $F2$, for Japanese speakers to acquire.

The primary aim of the present study is to investigate the acoustic characteristics of the productions of English /l/ and /ɹ/ by NJ-speaking children. Aoyama *et al.* (2004) suggested that the Japanese children's productions of English /ɹ/ improved more than their productions of /l/ over the course of 1 year. The present study analyzes their productions acoustically to examine what contributed to the perceived improvement in the Japanese children's production of /ɹ/. The secondary aim of the present study is to examine overall differences between NE speakers and Japanese L2 speakers of English in the acoustic nature of English /l/ and /ɹ/.

II. METHOD

A. Participants

The participants were 16 NJ-speaking adults and 16 NJ children, as well as 16 NE-speaking adults and 16 NE children (see Table I). All of the NJ participants were born in Japan, and most of them were living in Houston or Dallas, Texas, at the time of testing. All of the NJ adults had studied English for at least 6 years in Japan, whereas only one NJ child had studied English before his arrival in the U.S. The NJ children began attending English-speaking schools upon arriving in the U.S.

TABLE I. Characteristics of the NE and NJ participants. LOR = LOR in the U.S. in years. Age = chronological age in years at Time 1. Standard deviations in parentheses.

			Mean LOR	
	Gender	Mean age	Time 1	Time 2
NE adults	7m/9f	40.3 (4.7)	—	—
NE children	10m/6f	10.6 (2.1)	—	—
NJ adults	8m/8f	39.9 (3.8)	0.5 (0.2)	1.6 (0.3)
NJ children	9m/7f	9.9 (2.4)	0.4 (0.2)	1.6 (0.3)

B. Data collection

Data collection was conducted twice for each participant to study the effect of L2 learning for a larger longitudinal study (Aoyama *et al.*, 2004, Aoyama and Guion, 2007; Aoyama *et al.*, 2008; Oh *et al.*, 2011). The NJ participants' mean LOR in the U.S. was 0.5 yr at the first time of testing (Time 1), and their mean LOR was 1.6 yr at the second time (Time 2). The NE participants were tested in Birmingham, Alabama. They were also tested twice, 1 year apart. In both the NJ and NE groups, the adults were the parents of the children. Eighty-three adults and children originally participated at Time 1. Only 16 each of NJ adults and children were available for testing at Time 2 due to relocations. Sixteen each of NE adults and children were retained in the study in order to have an equal number of participants in the four groups.

C. Elicitation

All participants were tested in a quiet room in their homes, or at the University of Alabama at Birmingham. Twenty-six English words were elicited 3 times from each participant to examine a variety of consonants and vowels for the larger longitudinal study. The participant wore a head-mounted microphone (Shure SM10A, Shure, Niles, IL) connected to a DAT tape recorder (Sony TCD-D8, Sony, Tokyo, Japan). For the first elicitation, an auditory model was played via a loudspeaker. At the same time, a picture and an equivalent word in Japanese were displayed in Japanese orthography on the screen of a laptop computer. For the second and third elicitations, the pictures were displayed without an auditory model. The auditory model for the word was played only when the participant was uncertain about what word to say. The order of display for the 26 pictures was randomized for each participant and for each of the three elicitations.

The participants' productions were digitized at 22.05 kHz with 16-bit amplitude resolution. Two tokens of two words, *light* and *write*, with an initial /l/ and /ɹ/, were analyzed for each participant at each time of testing. In Aoyama *et al.* (2004), tokens of *leaf* and *read* were also included in the production analysis. In the present analysis, only *light* and *write* were analyzed because comparing the vowel context (/a/ vs /i/) was beyond the scope of the study. A set of 26 words was elicited from each participant 3 times.

Two of the three productions (first and third repetitions) were used in the production analysis. The first and third repetitions were originally selected to compare productions with an auditory model and without an auditory model in Aoyama *et al.* (2004). The first and third repetitions were simply treated as two tokens of the same word in Aoyama *et al.* (2004) and the present study because no statistically significant differences were found between them in a preliminary analysis.

D. Data analysis

There were a total of 256 tokens of /l/ and 256 tokens of /ɹ/ for analysis (4 groups × 16 participants × 2 tokens × 2 times of testing per consonant). Acoustic measurements were conducted using *Praat* (Boersma and Weenink, 2019). Consonantal duration, $F1$, $F2$, and $F3$ frequencies were measured because these acoustic parameters differentiate English /l/ and /ɹ/ (Espy-Wilson, 1992) and were measured in previous studies of Japanese speakers' productions of /l/ and /ɹ/ (Flege *et al.*, 1995). For $F1$, $F2$, and $F3$, measurements were taken at a slightly different point in /l/ and /ɹ/ following Flege *et al.* (1995). This decision was made because acoustic characteristics that define /l/ and /ɹ/ are different from each other (Espy-Wilson, 1992). For /l/, the formant frequencies were measured at the consonantal midpoint. Frequencies were measured at the midpoint instead of the onset for /l/, because formants were not always evident for the syllable-initial /l/s and there are abrupt shifts in the formant pattern at the onset and/or offset of the consonant (Dalston, 1975; Espy-Wilson, 1992). For /ɹ/, $F1$, $F2$, and $F3$ frequencies were measured at the point where $F3$ was the lowest. This decision was made because a low $F3$ is the hallmark of American English /ɹ/, following Espy-Wilson (1992), Flege *et al.* (1995), and Saito and Munro (2014).

The duration of the entire initial consonant was measured for both English /l/ and /ɹ/. The onset of the initial consonant was identified by visual inspection of the wideband spectrogram and waveform. The onset of the vowel was identified as the beginning of the vowel steady state. The consonant duration was measured as the interval between the onset of the consonant and the onset of the steady state of the vowel.

Acoustic measurements were first made by trained analysts. Then, all of the 512 tokens were checked, and remeasured if necessary, by K.A. There was one token of /l/ that had an overlapping noise and was excluded from the acoustical analysis. For this NJ child, the values from one token of /l/ were used for statistical analysis at Time 1. Values from two tokens each of /l/ and /ɹ/ were averaged for all other cases. To normalize the formant values, $F1$, $F2$, and $F3$ frequencies were converted from hertz to Bark using the formula provided in Eq. (1) (Traunmüller, 1990)

$$z = [26.81/(1 + 1960/f)] - 0.53. \quad (1)$$

III. RESULTS

A. Duration

Table II summarizes the duration (in ms) and formant frequencies (in Hz) of the productions of /l/. The descriptive statistics indicate that the duration of /l/ was longer in the NJ adults' productions than in the NE adults' productions. The NJ children's productions were comparable to the NE children's at Time 1, but were shorter at Time 2 than at Time 1.

TABLE II. Duration (in ms) and formant frequencies (in Hz) for /l/. Standard deviations in parentheses.

		Duration	$F1$	$F2$	$F3$
NE adults	Time 1	76	427	1116	2914
		(19)	(65)	(211)	(331)
	Time 2	81	417	1111	2878
		(23)	(58)	(192)	(312)
NE children	Time 1	80	468	1245	3128
		(25)	(72)	(180)	(496)
	Time 2	83	507	1238	2999
		(34)	(96)	(222)	(283)
NJ adults	Time 1	108	413	1491	2504
		(37)	(45)	(316)	(426)
	Time 2	106	431	1499	2443
		(31)	(70)	(227)	(409)
NJ children	Time 1	81	517	1854	2907
		(26)	(55)	(277)	(390)
	Time 2	63	563	1464	2934
		(20)	(111)	(209)	(501)

Table III summarizes the duration (in ms) and formant frequencies (in Hz) of the productions of /ɹ/. The descriptive statistics indicate that the duration of /ɹ/ was also longer in the NJ adults' productions than in the NE adults' productions on average. The NJ children's /ɹ/ was comparable to the NE children's /ɹ/ in duration at Time 1. At Time 2, the NJ children's /ɹ/ was shorter than the NE children's /ɹ/.

The duration data were analyzed using Multivariate Analysis of Variance (MANOVA). The dependent variables were durations of /l/ and /ɹ/ (averaged over two tokens for each speaker at each time of testing). Independent variables were Age (2 levels), Language (2 levels) (between subjects), and Time (2 levels) (within subjects). This analysis yielded a significant simple effect of Age, $F(2, 59) = 3.56$, $p = 0.04$; Wilks' Λ

TABLE III. Duration (in ms) and formant frequencies (in Hz) for /ɹ/. Standard deviations in parentheses.

		Duration	$F1$	$F2$	$F3$
NE adults	Time 1	85	476	1109	1641
		(23)	(86)	(136)	(216)
	Time 2	91	505	1161	1679
		(33)	(52	(115)	(159)
NE children	Time 1	90	550	1277	1719
		(33)	(72)	(159)	(130)
	Time 2	101	532	1289	1779
		(39)	(77)	(172)	(219)
NJ adults	Time 1	100	437	1273	2004
		(24)	(78)	(302)	(366)
	Time 2	114	447	1206	1900
		(48)	(74)	(233)	(312)
NJ children	Time 1	92	478	1533	2517
		(36)	(89)	(334)	(451)
	Time 2	70	552	1341	2361
		(23)	(145)	(251)	(325)

= 0.89, partial η^2 = 0.11. The interaction between Language × Age was also significant, $F(2, 59) = 5.60$, $p = 0.006$; Wilks' Λ = 0.84, partial η^2 = 0.16. All other simple effects, two-way interactions, and the three-way interaction were non-significant.

The univariate tests were conducted as a follow-up for the MANOVA. The alpha level was adjusted to $p = 0.025$ for univariate tests because there were two dependent variables (/l/ and /ɹ/) ($p = 0.05$ divided by 2). The interaction between Language × Age was significant for both /l/ and /ɹ/ [for /l/, $F(1, 60)$ = 10.21, $p = 0.002$, partial η^2 = 0.15; for /ɹ/, $F(1, 60)$ = 5.92, $p = 0.02$, partial η^2 = 0.09]. The univariate tests indicated that the NE adults' and children's duration did not differ statistically significantly ($p > 0.05$), but that the NJ adults' productions of both /l/ and /ɹ/ were longer than the NJ children's productions ($p < 0.05$).

B. *F*1

Tables II and III also show formant frequencies (in Hz) for the productions of /l/ and /ɹ/. Figure 1 displays the boxplot of $F1$ (in Bark) of both /l/ and /ɹ/ in the four groups. The NJ adults' mean $F1$ of /l/ was comparable to the NE adults' but the NJ children's mean $F1$ of /l/ was higher than NE children's. For /ɹ/, the NJ adults' mean $F1$ was lower than the NE adults' $F1$. The NJ children's mean $F1$ was lower than the NE children's at Time 1, but was similar to the NE children's at Time 2, indicating improvement in this acoustic parameter.

The $F1$ values in Bark were analyzed using MANOVA. The dependent variables were $F1$ values in Bark for /l/ and /ɹ/ averaged over two tokens for each participant at each time of testing. Independent variables were Age (2 levels), Language (2 levels) (between subjects), and Time (2 levels) (within subjects). The simple effects of Age, Language, and Time were significant, $F(2, 59)$ = 3.54 to 15.00, $p = 0.04$ to 0.001; Wilks' Λ = 0.66 to 0.89, partial η^2 = 0.11 to 0.34. The three-way interaction (Time × Language × Age) was also significant, $F(2, 59)$ = 3.28, $p = 0.045$; Wilks' Λ = 0.90, partial η^2 = 0.10.

The univariate tests showed that the Time × Language × Age interaction was significant for /ɹ/ but not for /l/ [for /l/ $F(1, 60) = 0.31$, $p = 0.58$; for /ɹ/, $F(1, 60) = 5.32$, $p = 0.025$, partial η^2 = 0.08]. The tests indicated that the NJ children's $F1$ for /ɹ/ was higher at Time 2 than at Time 1 ($p < 0.05$), indicating improvement in the production of /ɹ/. No significant difference was found in the other groups.

C. *F*2

Tables II and III summarize $F2$ frequencies in Hz, and Fig. 2 displays the boxplot of $F2$ in Bark for the productions of /l/ and /ɹ/ in the four groups. For /l/, the NJ adults' mean $F2$ was higher than the NE adults' mean $F2$. The NJ children's mean $F2$ was higher than all other groups' mean values at Time 1. Their $F2$ was lower at Time 2 than at Time 1 and was comparable to the NJ adults' values (see Fig. 2). The NE adults' and children's mean $F2$ values were similar for both /l/ and /ɹ/, indicating $F2$ was not an acoustic cue that differentiates /l/ and /ɹ/ in the NE speakers' productions. The NJ adults' mean $F2$ in /ɹ/ was comparable to the NE speakers' $F2$. The NJ children's mean $F2$ in /ɹ/ was higher than

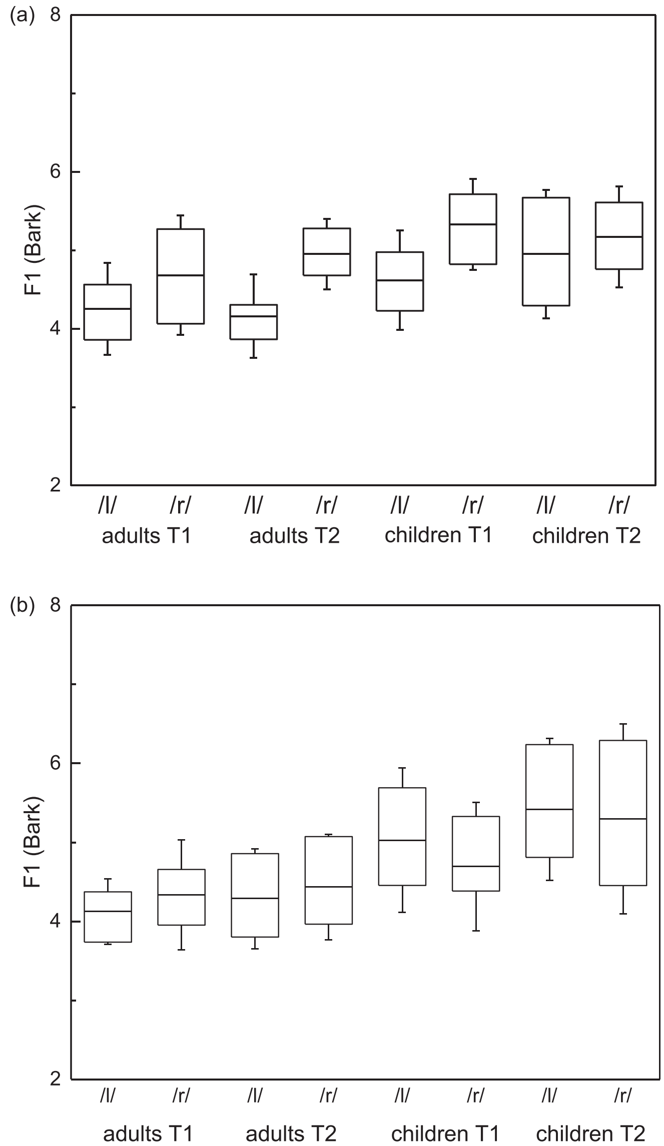

FIG. 1. (a) $F1$ in /l/ and /ɹ/ in the NE groups. (b) $F1$ in /l/ and /ɹ/ in the NJ groups. Note: Box plots indicate 25th and 75th percentile. The line in the box indicates the mean. The whiskers indicate one standard deviation above and below the mean.

other groups' mean $F2$, although it decreased from Time 1 to Time 2, showing improvement in /ɹ/.

The $F2$ values in Bark were analyzed using MANOVA. The dependent variables were $F2$ for /l/ and /ɹ/ in Bark, averaged over two tokens for each participant at each time of testing. Independent variables were Age (2 levels), Language (2 levels) (between subjects), and Time (2 levels) (within subjects). The simple effects of Age, Language, and Time were significant, $F(2, 59)$ = 5.79 to 31.87, $p < 0.005$; Wilks' Λ = 0.48 to 0.83, partial η^2 = 0.16 to 0.52. The three-way interaction (Language × Age × Time) was significant, $F(2, 59)$ = 6.52, $p = 0.003$; Wilks' Λ = 0.82, partial η^2 = 0.18. Two-way interactions (Age × Time, and Language × Time) were also significant, $F(2, 59)$ = 6.04 and 6.91, $p < 0.004$; Wilks' Λ = 0.83 and 0.81, partial η^2 = 0.19 and 0.17.

The univariate tests showed that the Language × Age × Time interaction was significant for /l/ but not for /ɹ/ [for /l/, $F(1, 60)$ = 12.78, $p = 0.001$, partial η^2 = 0.18; for /ɹ/,

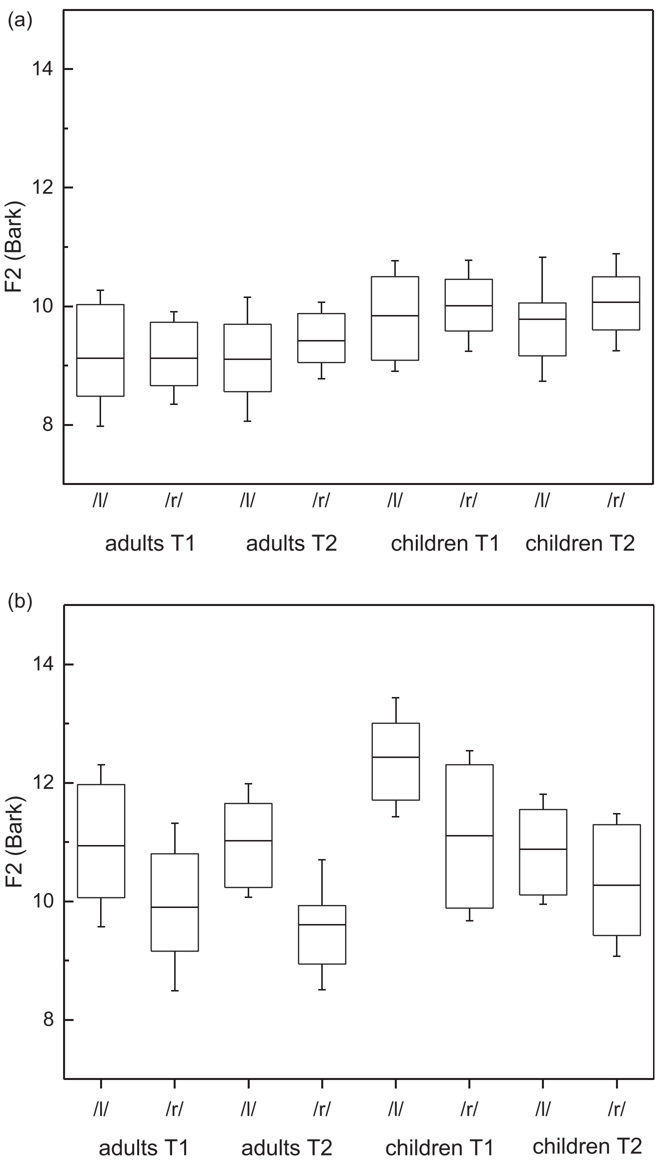

FIG. 2. (a) $F2$ in /l/ and /ɹ/ in the NE groups. (b) $F2$ in /l/ and /ɹ/ in the NJ groups. Note: Box plots indicate 25th and 75th percentile. The line in the box indicates the mean. The whiskers indicate one standard deviation above and below the mean.

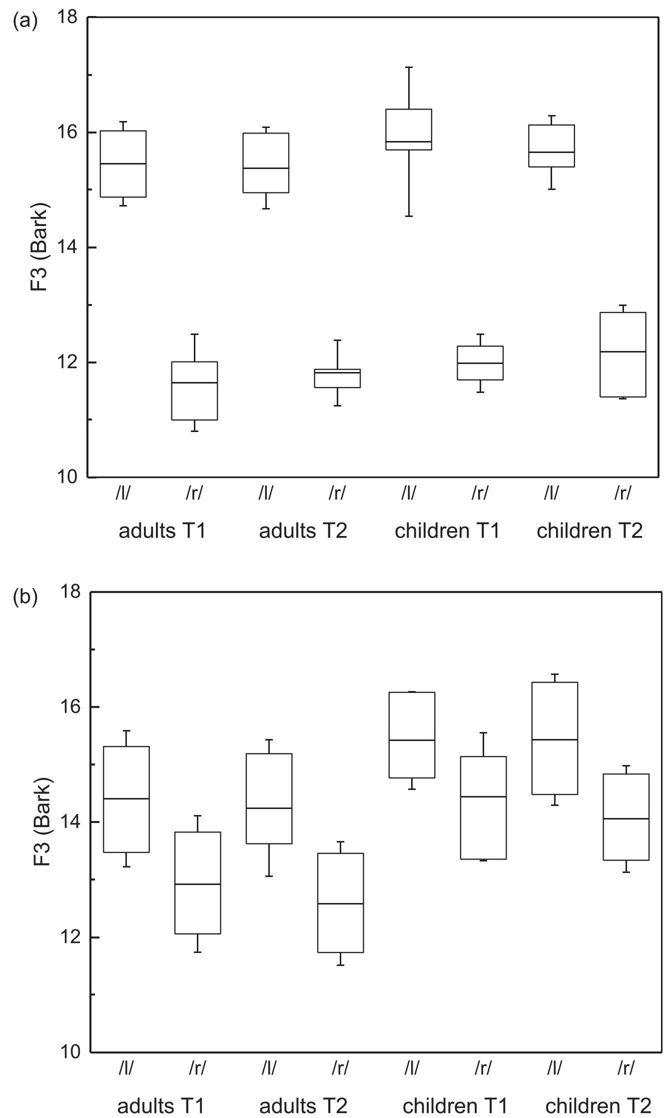

FIG. 3. (a) $F3$ in /l/ and /ɹ/ in the NE groups. (b) $F3$ in /l/ and /ɹ/ in the NJ groups. Note: Box plots indicate 25th and 75th percentile. The line in the box indicates the mean. The whiskers indicate one standard deviation above and below the mean.

$F(1, 60) = 0.27$, $p = 0.60$, partial $\eta^2 = 0.004$]. The *post hoc* tests indicated that the NJ children's $F2$ for /l/ was higher at Time 1 than at Time 2 ($p < 0.05$), but no significant difference was found in the other groups. Overall, the results indicated that the NJ children's $F2$ for /l/ decreased from Time 1 to Time 2, showing improvement, while no change was observed in the other groups.

D. *F*3

Tables II and III summarize $F3$ frequencies in Hz for the productions of /l/ and /ɹ/, and Fig. 3 displays the boxplot of $F3$ in Bark for both /l/ and /ɹ/ in the four groups. For /l/, the NJ adults' mean $F3$ was lower than the NE adults' mean $F3$. The NJ children's mean $F3$ of /l/ was comparable to the NE children's $F3$. For the NE adults' and children's /ɹ/, the values were between 1600 and 1800 Hz, and were mostly consistent with previous studies (Espy-Wilson, 1992). These results confirmed that a low frequency of $F3$ is the primary acoustic parameter that distinguishes /l/ from /ɹ/ in American English. The $F3$ values in the NJ adults' and children's /ɹ/ were higher than in the NE adults' and children's /ɹ/. The NJ adults' and children's mean $F3$ values indicated that their productions of /ɹ/ differed from the NE speakers' /ɹ/ in this crucial acoustic parameter.

The $F3$ values in Bark were analyzed using MANOVA. The dependent variables were $F3$ in Bark for /l/ and /ɹ/, averaged over two tokens for each participant at each time of testing. Independent variables were Age (2 levels), Language (2 levels) (between subjects), and Time (2 levels) (within subjects). The simple effects of Age and Language were significant, $F(2, 59) = 11.57$ and 54.96, $p < 0.001$; Wilks' $\Lambda = 0.72$ and 0.35, partial $\eta^2 = 0.28$ to 0.65. The three-way interaction (Language $\times$ Age $\times$ Time) was non-significant, $F(2, 59) = 0.25$, $p = 0.77$; Wilks' $\Lambda = 0.25$. Two two-way interactions (Language $\times$ Age, and Language $\times$ Time) were significant,

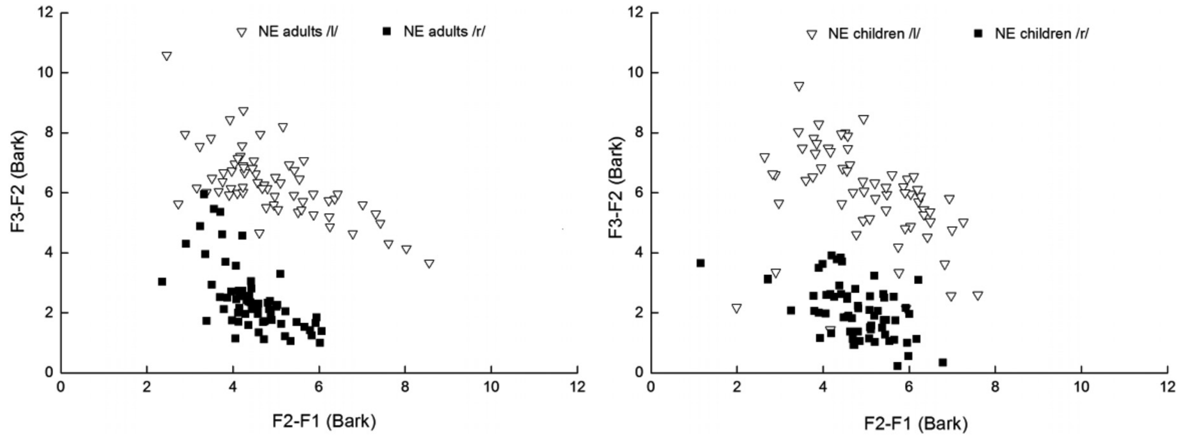

FIG. 4. Scatter plots of the $F2$-$F1$ and $F3$-$F2$ distances (in Bark) of English /l/ and /ɹ/ produced by the NE adults (the left panel) and the NE children (the right panel). Time 1 and Time 2 values were combined.

$F(2, 59) = 4.22$ and 3.55, $p < 0.05$; Wilks' $\Lambda = 0.86$ and 0.89, partial $\eta^2 = 0.13$ and 0.11.

The univariate tests showed that the Language × Age interaction was significant for /ɹ/ but not for /l/ [for /ɹ/, $F(1, 60) = 8.17$, $p = 0.006$, partial $\eta^2 = 0.12$; for /l/, $F(1, 60) = 3.02$, $p = 0.09$]. The tests indicated that the NJ children's $F3$ for /ɹ/ was higher than the NJ adults' $F3$ ($p < 0.05$), but no significant difference was found between the NE adults and NE children. The univariate tests also showed that the Language × Time interaction was significant for /ɹ/ but not for /l/ [for /ɹ/, $F(1, 60) = 6.49$, $p = 0.013$, partial $\eta^2 = 0.10$; for /l/, $F(1, 60) = 0.07$, $p = 0.79$]. The *post hoc* tests indicated that the NJ speakers' $F3$ for /ɹ/ was lower at Time 2 than at Time 1, indicating improvement in this acoustic parameter.

Overall, the NJ speakers' $F3$ for /ɹ/ was higher than the NE speakers' $F3$ (Fig. 3). The NJ speakers' $F3$ (adults and children combined) was lower at Time 2 than at Time 1, indicating a change toward more accurate production of American English /ɹ/. No statistically significant change was found for $F3$ in the NJ adults' and children's productions of /l/.

E. *F*2-*F*1 and *F*3-*F*2 distance

Figure 4 presents scatterplots of the $F2$-$F1$ distance (x axes) and the $F3$-$F2$ distance (y axes) in Bark for the NE adults (left panel) and the NE children (right panel). Espy-Wilson (1992) demonstrated that the $F3$-$F2$ distance is a major acoustic cue that distinguishes /l/ from /ɹ/ in American English, in addition to a low $F3$ in /ɹ/. Figure 4 demonstrates that NE adults' and children's productions of /l/ and /ɹ/ were clearly differentiated by the distance between $F3$ and $F2$.

Figure 5 presents scatterplots of the $F2$-$F1$ distance (x axes) and the $F3$-$F2$ distance (y axes) in Bark for the NJ adults. The left panel in Fig. 5 shows the NJ adults' productions at Time 1, and the right panel shows their productions at Time 2. The NJ adults' /l/ and /ɹ/ were not clearly differentiated from each other for either the $F2$-$F1$ or the $F3$-$F2$ distance. Espy-Wilson (1992) reported that the $F3$-$F2$ difference of 3.5 Bark as the boundary that separates /l/ from /ɹ/, because the $F3$-$F2$ difference is typically larger than 3.5 Bark for /l/ and smaller than 3.5 Bark for /ɹ/. Figure 5 shows that the $F3$-$F2$ difference for *both* /l/ and /ɹ/ was lower than 6.0 Bark in the NJ adult speakers' productions.

Figure 6 presents scatterplots of the $F2$-$F1$ distance (x axes) and the $F3$-$F2$ distance (y axes) in Bark for the NJ children. The left panel in Fig. 6 shows the NJ children's productions at Time 1, and the right panel shows their productions at Time 2. The NJ children's /l/ and /ɹ/ were not clearly differentiated from each other by either the $F2$-$F1$ or

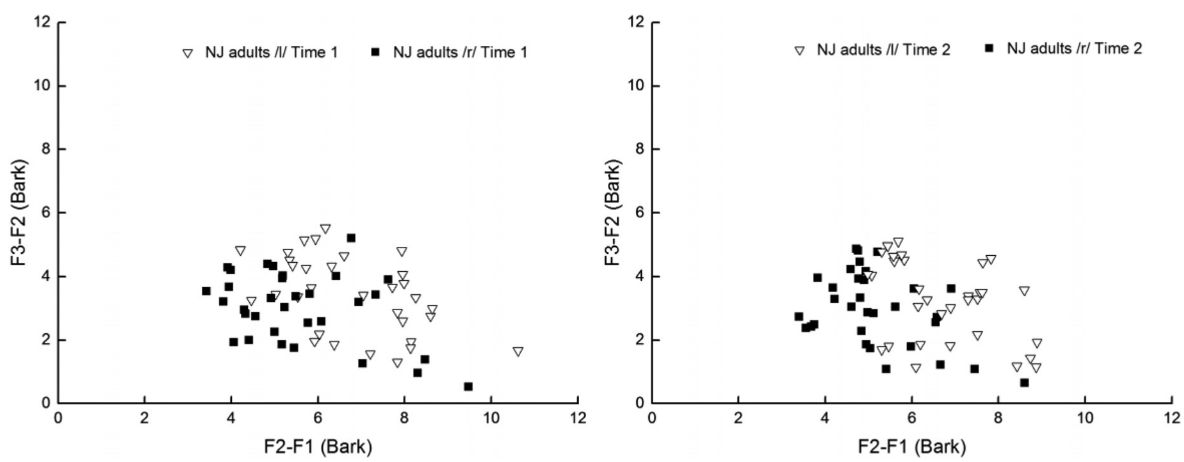

FIG. 5. Scatter plots of the $F2$-$F1$ and $F3$-$F2$ distances (in Bark) of English /l/ and /ɹ/ produced by the NJ adults. The left panel shows the Time 1 data. The right panel shows the Time 2 data.

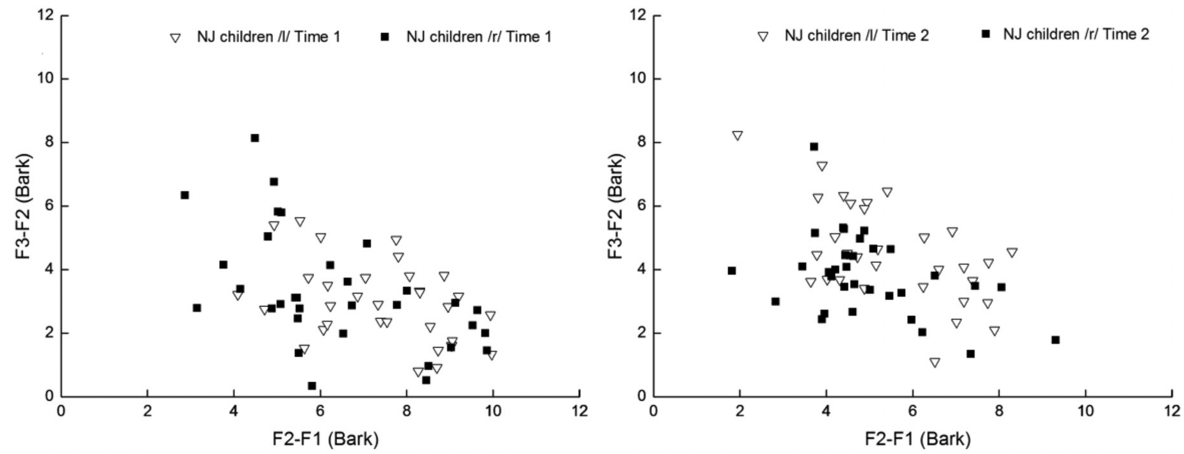

FIG. 6. Scatterplots of the $F2$-$F1$ and $F3$-$F2$ distances (in Bark) of English /l/ and /ɹ/ produced by the NJ children. The left panel shows the Time 1 data. The right panel shows the Time 2 data.

the $F3$-$F2$ distance, and their productions appear to be more widely dispersed than the NJ adults' productions. The $F3$-$F2$ distance for both /l/ and /ɹ/ ranged from 0 to 8 in the NJ children's productions, indicating the two sounds were not differentiated by the $F3$-$F2$ distance, unlike in the NE speakers' productions.

In sum, results showed that the NJ children's productions of liquids changed in some acoustic measures from Time 1 to Time 2, including $F1$ in /ɹ/ and $F2$ in /l/. The NJ adults' and children's $F3$ in /ɹ/ decreased from Time 1 to Time 2, showing improvement in the crucial acoustic parameter for distinguishing English /l/ and /ɹ/. In addition, results showed that the NJ adults' and children's productions of English /l/ and /ɹ/ differed from the NE adults' and children's productions in almost all acoustic parameters both at Time 1 and Time 2. The NJ adults' productions of /l/ and /ɹ/ were longer in duration than the NE adults' productions, while the NJ children's /l/ and /ɹ/ were shorter than the NE children's. The NJ adults' and children's /ɹ/ had a lower $F1$ than the NE adults' and children's. The $F2$ in the NJ groups' productions of /l/ was higher than the NE groups' productions of /l/. The critical acoustic parameter, $F3$ in /ɹ/, was higher in the NJ adults' and children's productions than in the NE adults' and children's productions.

IV. DISCUSSION

The primary aim of the present study was to investigate the acoustic characteristics of American English liquids by NJ-speaking children. Using NE speakers' perceptual judgments, Aoyama *et al.* (2004) reported that the NJ children's productions, especially their production of /ɹ/, improved over the course of 1 year of residence in the U.S. The results of the present study indicated that there were some changes between Time 1 and Time 2 in the NJ adults' and children's productions. The mean $F1$ in the NJ children's /ɹ/ was higher at Time 2 than at Time 1, and their $F2$ in /l/ was lower at Time 2 than at Time 1. In addition, the $F3$ in the NJ speakers' (adults and children combined) /ɹ/ was lower at Time 2 than at Time 1. These changes were in the right direction compared to the NE speakers' productions, indicating improvements in the NJ speakers' productions of American English /l/ and /ɹ/. However, the changes did not occur specifically in the NJ children's production of /ɹ/, as one would expect from the results in Aoyama *et al.* (2004). Changes in acoustic parameters were observed in the NJ adults' productions as well as the NJ children's productions, and they were observed in both /l/ and /ɹ/.

The secondary aim of the present study was to compare the acoustic nature of liquids between native speakers and Japanese L2 speakers of American English. The results showed that the NJ adults' and children's productions of both liquids were different from the NE speakers' productions in almost all acoustic parameters. Most importantly, the $F3$ in /ɹ/, a critical acoustic parameter for the distinction between /l/ and /ɹ/, was higher in the NJ adults' and children's /ɹ/ productions than in the NE adults' and children's productions both at Time 1 and Time 2.

For the durational aspect, the NJ groups' productions showed different characteristics from those of the NE groups. In particular, the NJ children's productions of both liquids were shorter than the NJ adults' liquids at both times of testing (see Tables II and III). In the same group of participants, Aoyama and Guion (2007) showed that the children's overall utterances were longer than the adults' (NE and NJ combined). Oh *et al.* (2011) also showed no consistent differences between the same NJ adults and children in vowel duration. One possibility of the differences in duration is that the NJ children's productions were acoustically similar to the Japanese tap /ɾ/ at least in the durational aspect, because the Japanese tap /ɾ/ is shorter than both English liquids (Hattori and Iverson, 2009; Saito and Munro, 2014). In Saito and Munro (2014), the $F1$ transition duration in /ɹ/ were longer in the Japanese speakers with longer LOR in Canada, showing improvement with increasing experience in the L2. It is possible that, as the NJ children gain more experience, durations of their liquids will become longer. It is puzzling as to why the NJ adults' /l/ and /ɹ/ were the longest compared to all three groups. As Aoyama and Guion (2007) suggested, there may be some prosodic differences between the NE and NJ groups, and possibly between adults and children. We speculate that prosodic differences also contribute to differences in segmental durations observed in the present study.

Some changes were observed in the spectral aspects of the NJ adults' and children's productions. According to Saito and Munro (2014), an acoustic benchmark for $F2$ for English /ɹ/ ranged from 7.9 to 11.0 in Bark. Saito and Munro's (2014) benchmark for $F2$ ranged from 11.8 to 13.2 in Bark for the Japanese tap /ɾ/. In the current study, the mean $F2$ in the NJ adults' /ɹ/ was within the range of English /ɹ/ at both Time 1 and Time 2 (mean 9.9 and 9.6, respectively) (see Fig. 2). The $F2$ in the NJ children's /ɹ/ at Time 1 (mean 11.1) was higher than Saito and Munro's (2014) benchmark. At Time 2, the NJ children's $F2$ in /ɹ/ (mean 10.3) was within the range for English /ɹ/ suggested by Saito and Munro (2014). The NJ children's $F2$ in /l/ at Time 1 (mean 12.4) was within the range for the Japanese tap /ɾ/, suggesting that their productions of /l/ were similar to the Japanese /ɾ/ in $F2$. The $F2$ in /l/ decreased from Time 1 to Time 2 (mean 12.4 vs 10.9) indicating improvement for this acoustic parameter. These results on $F2$ may lend support for Saito and Munro's (2014) claim that $F2$ is a "familiar cue" (p. 456) for Japanese speakers in learning to produce a new category in English. Furthermore, results of the present study suggest that their claim may extend to the production development of English /l/ as well.

The benchmark for $F3$ for English /ɹ/ in Saito and Munro (2014) ranged from 11.4 to 12.6 in Bark. The NE adults' and children's $F3$ values were, as expected, within this range (see Fig. 3). The NJ adults' mean $F3$ values in /ɹ/ were higher than the NE adults' (12.9 at Time 1 and 12.6 at Time 2), and they were slightly above or at the upper end of Saito and Munro's (2014) benchmark. The NJ children's mean $F3$ values (14.4 at Time 1 and 14.1 at Time 2) were higher than the benchmark. In fact, the NJ children's $F3$ values were within the range of benchmark for the Japanese tap /ɾ/ (14.5–15.7 in Bark). Although statistical tests indicated that the $F3$ decreased from Time 1 to Time 2 in the NJ adults' and children's /ɹ/, $F3$ was still higher in the NJ speakers' /ɹ/ than in the NE speakers' /ɹ/ at Time 2.

As mentioned earlier, a low $F3$ is the hallmark of English /ɹ/ (Espy-Wilson, 1992). Saito and Lyster (2012) suggested that tokens of /ɹ/ are considered as "good enough" (p. 622) when the $F3$ was between 2200 and 2300 Hz in Japanese speakers' productions. In the current study, the mean $F3$ in the NJ adults' /ɹ/ tokens were 2004 Hz at Time 1 and 1900 Hz at Time 2, suggesting their /ɹ/ tokens were in the good enough range suggested by Saito and Lyster (2012). The NJ children's mean $F3$ (mean 2517 Hz at Time 1 and 2361 Hz at Time 2), on the other hand, were above the good enough range. This was a somewhat surprising finding, because one would expect the improvement in the production of English /ɹ/ in Aoyama *et al.* (2004) would involve larger changes in the most critical acoustic parameter. Saito and Munro (2014) also suggested that $F3$ may be more difficult for Japanese speakers to acquire than $F2$ and transition duration, because $F3$ is not a primary cue in the Japanese phonetic system.

Perhaps the status of $F3$ in English and Japanese holds the key to Japanese speakers' difficulty with English liquids. It is well documented that a low $F3$ in /ɹ/ is the crucial difference between English /l/ and /ɹ/ (e.g., Espy-Wilson, 1992). In perception, Hattori and Iverson (2009) suggested that "representation of F3" (p. 478) for /l/ and /ɹ/ may be closely related to Japanese speakers' ability to identify them. Yet both Saito and Munro (2014) and the current study suggest that $F3$ may be a more difficult phonetic aspect for NJ speakers to acquire than other cues. The mismatch between the importance of $F3$ in English /ɹ/ and inattention to $F3$ by Japanese speakers may be the source of this well-known difficulty in production and perception of English /l/ and /ɹ/ by Japanese L2 speakers of English.

In addition to $F3$, the distance between $F3$ and $F2$ is an important acoustical cue in distinguishing English /ɹ/ from /l/ (Espy-Wilson, 1992). The $F2$-$F1$ and $F3$-$F2$ distances (Fig. 4) showed that the NE adults' and children's /l/ and /ɹ/ were clearly differentiated by the $F3$-$F2$ distance. Espy-Wilson (1992) demonstrated that English /l/ has a larger $F3$-$F2$ distance than /ɹ/, with a difference of 3.5 Bark as a reference point. In both the NE adults' and children's productions, the $F3$-$F2$ distances were greater than 4.0 Bark for /l/ and smaller than 4.0 Bark for /ɹ/. This finding is consistent with Idemaru and Holt (2013) that NE-speaking children's productions of /l/ and /ɹ/ are differentiated in both $F3$ and $F3$-$F2$ distance by age 8.

In comparison, the $F3$-$F2$ distances were smaller than 6.0 for both /l/ and /ɹ/ in the NJ adults' productions (Fig. 5). The NJ children's /l/ and /ɹ/ productions appear to be overlapping on both the $F3$-$F2$ and $F2$-$F1$ distances (Fig. 6). Figure 5 also suggests that the $F2$-$F1$ distance may be smaller for /ɹ/ and larger for /l/ in the NJ adults' productions. These findings suggest that the NJ adults and children may be differentiating American English /l/ and /ɹ/ with acoustic parameters that are not crucial for native speakers of American English, such as the $F2$-$F1$ distance, rather than the primary cues, a low $F3$ in /ɹ/ and the $F3$-$F2$ distance. Interestingly, this finding coincides with previous studies that Japanese speakers use cues other than $F3$, both in production (Lotto *et al.*, 2004; Saito and Munro, 2014) and in perception (Ingvalson *et al.*, 2012).

The results of the current study are also consistent with Flege *et al.* (1995), who reported that experienced Japanese speakers' /l/ and /ɹ/ productions were acoustically different from the NE speakers', even though the Japanese speakers' productions were almost always identified as the intended segment by the NE speakers. The current study and Flege *et al.* (1995) suggest that the Japanese speakers' /l/ and /ɹ/ productions can be identified as the intended segments in intelligibility judgments, although the acoustic characteristics of the liquids are clearly different from native speakers' productions. It is possible that Japanese L2 speakers of English are differentiating /l/ from /ɹ/ in a different way than native speakers of English, and these differences may not appear in perceptual evaluation by NE speakers. Based on a holistic study of L2 acquisition, Abrahamsson and Hyltenstam (2009) demonstrated that there are subtle linguistic differences between native speakers and L2 speakers, even when L2 speakers are perceived as native-like in overall "nativelikeness" (p. 252). The results of the present study and these previous studies show that there are fine differences between native speakers' and L2 speakers' speech that are not readily observable through perceptual evaluations by native speakers.

Taken together, the changes observed in acoustic characteristics in this study do not appear to be specific to the production of /ɹ/ in the NJ children, as previously suggested by perceptual evaluations of the same /l/ and /ɹ/ productions in Aoyama *et al*. (2004). Why did the NJ children's /ɹ/ productions improve considerably in Aoyama *et al*. (2004), even though the present study showed that the NJ children's productions of both /l/ and /ɹ/ were acoustically quite different from the NE speakers' productions at Time 2? One factor is that the NE speakers' judgments in Aoyama *et al*. (2004) were based on a forced-choice task. That is, the NE-speaking listeners in Aoyama *et al*. (2004) gave a response from the choices that were provided (*l*, *r*, *w*, *d*, *dr*, *br*, *bl*, *ml* on a computer screen). The NJ children's production of /ɹ/ likely changed enough from Time 1 to Time 2 for NE speakers to identify them as "*r*" more often, even though acoustic characteristics of their /ɹ/ production were still different from the NE speakers' /ɹ/. If there was an additional task of goodness rating in Aoyama *et al*. (2004), the differences in the quality of both /l/ and /ɹ/ might have been observed.

In Aoyama *et al*. (2004), improvements were also observed in the NJ children's perception of /l/ and /ɹ/. It cannot be concluded from Aoyama *et al*. (2004) that the NJ children improved specifically on /ɹ/, because their perceptual abilities were assessed by a discrimination task between /l/ and /ɹ/. Likewise, the results of the present study do not directly support or reject the NJ children's perceptual learning of /l/ and /ɹ/. It is not possible to conclude if perceptual learning in the NJ children occurred only for /ɹ/, or for both liquids, from the acoustical analysis of the present study.

In sum, the results of the present study showed that there were acoustic differences in their production of /l/ and /ɹ/ between the two times of data collection in the NJ adults and children. These differences observed between Time 1 and Time 2 are in the direction closer to the NE speakers' values, indicating improvements in production. However, the NJ speakers' productions were different from the NE speakers' productions in almost all acoustic parameters both at Time 1 and Time 2. The results of our current study suggest that, acoustically, the observed improvements in the NJ children's /ɹ/ in Aoyama *et al*. (2004) were likely due to a combination of changes in both /l/ and /ɹ/, rather than a change in a single acoustic parameter, such as $F3$ in /ɹ/. In short, the present study adds to the complexity of the case of Japanese speakers' acquisition of English /l/ and /ɹ/. It also underscores the importance of studying L2 speech production using perceptual judgments as well as acoustic analysis, because the two types of analyses provide different yet complementary perspectives into L2 speech acquisition.

ACKNOWLEDGMENTS

This study was supported by the National Institutes of Health (Grant No. DC00257) awarded to J.E.F. The authors are indebted to Dr. Susan Guion Anderson, who played an active role in designing and implementing the original data collection of this study. The authors thank Linda Legault for her help with data collection, and Christina Borunda and Sarah Troxell for their assistance with data analysis. Lastly, we sincerely thank all the adults and children who participated in this project.

Abrahamsson, N., and Hyltenstam, K. (**2009**). "Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny," Lang. Learn. **59**, 249–306.

Aoyama, K., and Flege, J. E. (**2011**). "Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers," J. Phonetic Soc. Japan **15**, 5–13.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (**2004**). "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/," J. Phonetics **32**, 233–250.

Aoyama, K., and Guion, S. G. (**2007**). "Prosody in second language acquisition: An acoustic analysis on duration and F0 range," in *Language Experience in Second Language Speech Learning*, edited by O. S. Bohn and M. Munro (John Benjamins, Amsterdam), pp. 281–297.

Aoyama, K., Guion, S. G., Flege, J. E., Yamada, T., and Akahane-Yamada, R. (**2008**). "The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English," Int. Rev. Appl. Linguistics Lang. Teach. **46**, 61–90.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer (version 6.1.03) [Computer program]," http://www.praat.org/ (Last viewed September 28, 2019).

Bradlow, A., Akahane-Yamada, R., Pisoni, D., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," Percept. Psychophys. **61**, 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/, IV: Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Dalston, R. M. (**1975**). "Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults," J. Acoust. Soc. Am. **57**, 462–467.

Espy-Wilson, C. Y. (**1992**). "Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English," J. Acoust. Soc. Am. **92**, 736–757.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York Press, Baltimore, MD), pp. 233–277.

Flege, J. E., Takagi, N., and Mann, V. (**1995**). "Japanese adults can learn to produce English /ɹ/ and /l/ accurately," Lang. Speech **38**, 25–55.

Flege, J. E., Takagi, N., and Mann, V. (**1996**). "Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/," J. Acoust. Soc. Am. **99**, 1161–1173.

Goto, H. (**1971**). "Auditory perception by normal Japanese adults of the sounds 'L' and 'R'," Neuropsychologia **9**, 317–323.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–2724.

Hattori, K., and Iverson, P. (**2009**). "English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy," J. Acoust. Soc. Am. **125**, 469–479.

Idemaru, K., and Holt, L. L. (**2013**). "The developmental trajectory of children's perception and production of English /r/-/l/," J. Acoust. Soc. Am. **133**, 4232–4246.

Ingvalson, E. M., Holt, L. L., and McClelland, J. L. (**2012**). "Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency?," Bilingualism: Lang. Cognition **15**, 255–274.

Ingvalson, E. M., McClelland, J. L., and Holt, L. L. (**2011**). "Predicting native English-like performance by native Japanese speakers," J. Phonetics **39**, 571–584.

Larson-Hall, J. (**2006**). "What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /ɹ/ and /l/ by Japanese speakers," Lang. Speech **49**, 521–548.

Lotto, A. J., Sato, M., and Diehl, R. L. (**2004**). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," in *From Sound to Sense*, edited by J. Slifka, S. Manuel, and M. Matthies (MIT Research Laboratory in Electronics, Cambridge, MA), C381–C386.

McGowan, R. S., Nittrouer, S., and Manning, C. J. (**2004**). "Development of [ɹ] in young, midwestern, American children," J. Acoust. Soc. Am. **115**, 871–884.
Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T. (**2011**). "A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting," J. Phonetics **39**, 156–167.
Polka, L., and Strange, W. (**1985**). "Perceptual equivalence of acoustic cues that differentiate /r/ and /l/," J. Acoust. Soc. Am. **78**, 1187–1197.
Saito, K., and Lyster, R. (**2012**). "Effects of Form-Focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English," Lang. Learn. **62**, 595–633.
Saito, K., and Munro, M. J. (**2014**). "The early phase of /ɹ/ production development in adult Japanese learners of English," Lang. Speech **57**, 451–469.
Sander, E. K. (**1972**). "When are speech sounds learned?," J. Speech Hear. Disorders **37**, 55–63.
Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., and Bird, A. (**1990**). "The Iowa articulation norms project and its Nebraska replication," J. Speech Hear. Disorders **55**, 779–798.
Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**, 97–100.
Yamada, R. A. (**1995**). "Age and acquisition of second language speech sounds perception of American English /ɹ/ and /l/ by native speakers of Japanese," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 305–320.
Yamada, R. A., and Tohkura, Y. (**1992**). "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," Percept. Psychophys. **52**, 376–392.