The developmental trajectory of children's perception and production of English /r/-/l/[a]

Kaori Idemaru[b]
Department of East Asian Languages and Literatures, University of Oregon, Eugene, Oregon 97405

Lori L. Holt
Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15203

The English /l-r/ distinction is difficult to learn for some second language learners as well as for native-speaking children. This study examines the use of the second (F2) and third (F3) formants in the production and perception of /l/ and /r/ sounds in 4-, 4.5-, 5.5-, and 8.5-yr-old English-speaking children. The children were tested with elicitation and repetition tasks as well as word recognition tasks. The results indicate that whereas young children's /l/ and /r/ in both production and perception show fairly high accuracy and were well defined along the primary acoustic parameter that differentiates them, F3 frequency, these children were still developing in regard to the integration of the secondary cue, F2 frequency. The pattern of development is consistent with the distribution of these features in the ambient input relative to the /l/ and /r/ category distinction: F3 is robust and reliable, whereas F2 is less reliable in distinguishing /l/ and /r/. With delayed development of F2, cue weighting of F3 and F2 for the English /l-r/ categorization seems to continue to develop beyond 8 or 9 yr of age. These data are consistent with a rather long trajectory of phonetic development whereby native categories are refined and tuned well into childhood.
© 2013 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4802905]

I. INTRODUCTION

Speech perception develops remarkably early. Young infants are language-general listeners, able to distinguish not only sounds that are in their native language but also non-native sounds (Aslin *et al.*, 1981; Best *et al.*, 1988; Polka and Werker, 1994; Streeter, 1976; Trehub, 1976; Werker *et al.*, 1981; Werker and Tees, 1984). This language-general perceptual ability is short lived. Already during their first year, infants' discrimination of non-native sounds declines (Werker and Tees, 1983, 1984; Kuhl, 1998; Best, 1995; Bosch and Sebastián-Gallés, 2003) whereas discrimination between native sounds may increase (Kuhl *et al.*, 2001). For example, English learning children as young as 6 months of age can discriminate sounds such as /ba/ versus /pa/ (a contrast existing in English) and Hindi voiceless retroflex plosive /ʈa/ and non-retroflex /ta/ (a contrast not included in English). However, by the time these children are 1 yr old, they do not discriminate the Hindi contrast, although they continue to discriminate /ba/ and /pa/ (Werker and Tees, 1984). More recently, Bosch and Sebastián-Gallés (2003) showed that whereas both Catalan and Spanish children can discriminate the vowels /e/ and /ɛ/ (a phonemic contrast in Catalan but not in Spanish) at 4 months of age, Spanish babies no longer discriminate these sounds by the time they are 8 months of age, while Catalan babies continue to do so at the same age.

Whereas initial development of speech categories and attunement for native-language processing occurs very early in life, this does not mean that phonetic development is complete in infancy. In fact, there are some aspects of speech perception that continue to develop well into later childhood. Attaining full linguistic competence in a speech sound system includes, but also reaches beyond, learning native speech sound categories. To achieve adult-like speech perception, children also must develop graded internal category responses (Volaitis and Miller, 1992; Miller and Eimas, 1996; Kuhl, 1991), normalization and sensitivity to context (Miller and Liberman, 1979; Kidd, 1989; Newman and Sawusch, 1996), and perceptual weighting of simultaneously available phonetic cues (Diehl and Walsh, 1989; Kluender *et al.*, 1988; Diehl *et al.*, 1991; Kingston and Diehl, 1994; Diehl *et al.*, 2004; Idemaru and Holt, 2011).

Perceptual cue weighting, as an example, appears to have a rather extended developmental course. There have been a number of insightful studies by Nittrouer and colleagues and by others (Nittrouer, 1996, 2002, 2004; Nittrouer and Miller, 1997; Ohde *et al.*, 1995; Ohde and Haley, 1997; Walley and Carrell, 1983; Hazan and Barrett, 2000) that have suggested that children's weighting of multiple acoustic cues to sound categorization differs from that of adults, and it has a rather gradual developmental trajectory. Nittrouer (1992), for example, showed that in categorizing /su/ and /ʃu/, 3-, 5-, and 7-yr-old children give more weight to the dynamic formant transition cue than to the static noise

---

[b]Author to whom correspondence should be addressed. Electronic mail: Idemaru@uoregon.edu

spectra cue and that they give more weight to the dynamic transition cue than adults. This finding has been replicated in a number of studies for this particular contrast (Nittrouer, 1996, 2002; Nittrouer and Miller, 1997) and other contrasts (e.g., Ohde and Haley, 1997; Walley and Carrell, 1983; Ohde et al., 1995 for stop place of articulation; Hazan and Barrett, 2000 for fricative and stop place of articulation and voicing; Nittrouer, 2004 for syllable-final stop voicing). The findings of these studies suggest that it takes years for children to develop adult-like weighting function of acoustic properties that define speech categories in native-language speech productions. Most studies show developmental changes in cue-weights in children between ages 3 and 7 (Nittrouer, 1992, 1996, 2002, 2004; Nittrouer and Miller, 1997; Ohde and Haley, 1997; Walley and Carrell, 1983). In addition, Hazan and Barrett (2000) showed that development continues even among older children. The authors demonstrated that even 12-yr-old children's judgments of ambiguous speech stimuli are not as consistent as those of adults (i.e., their identification functions are not as steep as those of adults), suggesting that fine tuning of speech categories continues well into middle childhood.

There has been considerable attention directed to the English /l/ and /r/ contrast (the more common American English symbol /r/ is used instead of IPA /ɹ/ for simplicity) in the speech perception literature and with respect to perceptual cue weighting. Learning this contrast is particularly difficult for some second language learners (e.g., Yamada and Tohkura, 1992; Iverson et al., 2003, 2005; Ingvalson et al., 2011) as well as English speaking children (e.g., Smit et al., 1990; McGowan et al., 2004; Dalston, 1975). Perhaps due to its difficulty for Japanese learners of American English, it is one of the most studied and best understood contrasts for adult cue weighting. For children, the difficult task is not necessarily separating /l/ and /r/ categories, but rather differentiating them from /w/ (e.g., Dalston, 1975). However, the rich foundation of adult study on /l/ and /r/ presents an opportunity that enables us to examine and interpret children's phonetic development. Here we examine the fine tuning of /l/ and /r/ categories focusing on developmental changes in the acoustic parameters that define these categories.

It is now well understood that in normal adult speech, English /l/ and /r/ in a prevocalic position are distinguished primarily by the onset frequency of the third formant (F3). Whereas F3 frequency for /l/ is relatively high, that for /r/ is considerably lower (e.g., Polka and Strange, 1985). Furthermore, studies have suggested the onset F2 frequency and the rate of the F1 transition also participate in defining /l/ and /r/ categories (Polka and Strange, 1985; Yamada and Tohkura, 1992; Lotto et al., 2004; Ingvalson et al., 2011). Although some have suggested that all of these acoustic correlates are used by listeners in perception (e.g., Polka and Strange, 1985), more recent studies have demonstrated that in categorizing /l/ and /r/, native adult listeners predominantly rely on the onset F3 frequency cue, give a weaker weight to the onset F2 frequency cue, and barely use the F1 transition (Ingvalson, et al., 2011). This mirrors the acoustics in that whereas F3 robustly correlates with /l/ and /r/

categorization, F2 correlates but is less reliable in category prediction (Lotto et al., 2004). The findings of Ingvalson et al. (2011) suggest that children must learn to use F3 and F2 frequencies with more weight on F3 to categorize /l/ and /r/ as adults do.

Prior work on phonetic development, and in particular the development of perceptual cue weighting, has focused primarily on fricatives, affricates and stops (Nittrouer, 1992, 1996, 2002, 2004; Nittrouer and Miller, 1997; Ohde et al., 1995; Ohde and Haley, 1997; Walley and Carrell, 1983; Hazan and Barrett, 2000). Whereas considerable work on /l/ and /r/ distinction in L2 acquisition has helped us to understand its cue weighting for adult English speakers, there has been little work that has examined the development of /l/ and /r/ in children in terms of F2/F3 cue weighting and integration. The well-understood baseline for "adult-like" perceptual cue weighting provides a unique opportunity for a thorough analysis of children's development of perceptual cue weighting for /r/-/l/.

The /l/ and /r/ distinction is known to give some children difficulty: Young children are reported to produce considerable errors in attempting these sounds, often producing something that sounds like /w/ to adult ears (Dalston 1975; Sander, 1972; Smit et al., 1990). Dalston (1975) analyzed F1, F2, and F3 values of word initial /r/, /w/, and /l/ in adults and 3- to 4-yr old children. Both adults' and children's formant values were reported to distinguish these three sounds acoustically (though without statistical analysis): F2 distinguished /w/ from /l/ and /r/ and F3 distinguished /r/ from /w/ and /l/. The mean values of F2 and F3 for children were: F2 for /r/, 1503 Hz, for /l/, 1384 Hz, for /w/, 1020 Hz; F3 for /r/, 2491 Hz, for /l/, 3541 Hz, and for /w/, 3547 Hz. However, the analysis did not include the tokens incorrectly identified by adult raters. Such incorrect production ranged from 2% to 37% across children (mean, 17.4%). Errors were distributed approximately equally in intended /r/ and /l/ productions but never occurred in intended /w/ productions. Dalston also noted that distributions of children's F2 and F3 values for these sounds were more variable and more overlapping than those of adults: Children's /w/ and /l/ showed overlap along the F2 dimension and /l/ and /r/ on the F3 dimension.

More recently, McGowan et al. (2004) longitudinally tracked the development of /r/ production in children from 15 to 32 months old, focusing on the changes in F2 and F3 frequencies over time. In particular, the authors examined the distance between F2 and F3 as an important acoustic cue for /r/. In adult production of /r/, F3 frequency is considerably lower resulting in a small distance between F2 and F3 (Stevens, 1998). The primary interest of McGowan et al. (2004) was the differential development of these formant frequencies in different syllabic positions, i.e., prevocalic (e.g., "right"), postvocalic (e.g., "car"), medial syllabic (e.g., "Burt"), and final syllabic (e.g., "doctor") positions. The overall developmental trend was such that both F2 and F3 decreased with age; however, they decreased at different rates depending on the syllabic position. As a result, whereas the distance between F3 and F2 in prevocalic /r/ remained around 2000 Hz throughout the duration of study, F3-F2 in

postvocalic /r/ decreased to 1000 Hz at the end of observation. These results seem to suggest that the F3-F2 frequency difference, particularly in the prevocalic position, develops further in children older than 3 yr of age. Distance between F2 and F3 frequencies is a relative measure, and as such, it might be important in differentiating /l/ and /r/ categories as a self-normalized parameter.

Together, the findings of Dalston (1975) and McGowan et al. (2004) suggest that the acoustic features that define /r/, distinguishing it from /l/ and /w/ sounds (namely, F2 and F3 frequencies) are still developing in 3-yr-old children and possibly continue to develop beyond age 3, particularly for /r/ in word-initial position. Three- and 4-yr-old children seem to differentiate /r/, /l/, and /w/ in their productions based on F2 and F3 when they are successfully produced. However, there remains rather frequent less-than-adult-like production for these categories. For example, 17% of children's productions were incorrectly perceived by adults (Dalston, 1975). Adult judgments of children's productions support this. Studies report that children start producing intended /l/ and /r/ before age 3, but these early attempts often include variant sounds, such as productions that sound indisputably like /w/ to adult listeners (Sander, 1972; Smit et al., 1990). By age 6 or 7, most of the variant productions disappear and children's productions are identified as the intended productions by adult raters at 90% accuracy (Sander, 1972; Smit et al., 1990). These evaluations of children's /l/ and /r/ productions suggest that important development occurs between 3 and 7 yr old, during which time the proportion of variant productions decreases and the acoustic signatures of /l/ and /r/ develop to approach adult norms.

Whereas these data are significant from a clinical perspective, they assess speech production in a binary manner according to adults' judgments of the "correctness" of pronunciation. From the point of view of understanding the perceptual and learning mechanisms underlying the development of perceptual cue weighting in children's phonetic categories, it is highly desirable to have a finer-grained assessment of children's speech acoustics, their relationship to the children's own developing phonetic perception and perceptual cue weighting, as well as the relationship to adults' ratings of the children's speech. Only this level of detail will provide the data necessary to develop and refine models of perceptual cue weighting and its development. Adult production data (Lotto et al., 2004) suggest that whereas F2 frequency does provide information for /l/-/r/ categorization, it is less reliable than the robust F3 frequency. If children are sensitive to distributional cues in the speech environment, we would expect F3 to be earlier-acquired and F2 to be later-acquired. This may be reflected in both children's production and perception. It is particularly important to investigate both production and perception given that weighting functions of relevant acoustic cues are not always parallel between production and perception (e.g., Idemaru et al., 2012; Shultz et al., 2012) and development or acquisition in one (e.g., production) may not automatically imply the same for the other (e.g., Goto 1971; Sheldon and Strange, 1982).

Given the adult findings that both F3 and F2 frequencies contribute to distinguish /l/ and /r/ in production and perception (Lotto et al., 2004; Ingvalson et al., 2011), the current study examines the development of F3 and F2 frequencies as acoustic and perceptual cues in the /l-r/ categorization at a finer acoustic level in children. The distance between F3 and F2 frequencies is likely to highly correlate with F3 and F2. This measure is examined nonetheless given a suggestion that it might be an important feature that defines the /r/ category (McGowan et al., 2004; Stevens, 1998). Clinical studies have shown that important development occurs between 3 and 7 yr old for /l/ and /r/ productions. The present study, thus, examines 4-, 4.5-, and 5.5-yr-old children to capture early development as well as 8.5-yr-old children as a group who have reached high production accuracy but may not have reached adult-like in terms of perceptual cue weighting (e.g., Hazan and Barrett, 2000).

II. CHILD PRODUCTION MEASUREMENTS

The production study examined the development of F3 and F2 frequencies in /l/ and /r/ productions. The specific aims are to investigate the direction and extent of changes in these formant frequencies across age and the weight of these formant frequencies in categorizing /l/ and /r/ productions.

A. Methods

1. Participants

Forty-eight children (23 girls) were divided into four age groups of approximately equal size:[1] 4-yr-olds (12 children; mean age = 4.16; age range = 3.95-4.37), 4.5-yr-olds (13; 4.71; 4.42-5.04), 5.5-yr-olds (12; 5.49; 5.05-6.13), and 8.5-yr olds (11; 8.45; 7.31-9.54).[2] None of these children had been diagnosed with speech/hearing problems, had six or more ear infections before their second birthday, had an ear infection at the time of testing, had complications at birth, or used a foreign language on a regular basis.[3]

2. Test words and tasks

Productions of eight utterances were elicited: Two lexical words, "light" and "write" (/lait/ and /rait/) and short syllables with initial /l/ and /r/ followed by three point vowels /i, u, a/: /li/, /lu/, /la/, /ri/, /ru/, and /ra/. The elicitation tasks were described as games to children. A native English-speaking research assistant explained to the children that six visitors from outer space, each illustrated on a picture card, wanted to be friends and learn some English words. The six visitors had monosyllabic names, /li/, /lu/, /la/, /ri/, /ru/, and /ra/. The research assistant selected a picture card with a character on it and prompted, "This is Lee. Can you say *This is Lee?*" When the child completed a cycle of naming each of the six characters, the picture cards were shuffled and the procedure was repeated five more times. The last five repetitions were retained as data; the first cycle was discarded as practice.

In the subsequent task, one of the visitors, Roo, appeared on the computer screen with two pictures. The recorded voice of Roo asked the child to teach him the words describing the pictures. One picture showed a hand writing a

K. Idemaru and L. L. Holt: Development of /l/ and /r/

letter, and this picture was associated with the word "write," /rait/. The other picture was that of a lamp and this picture was labeled "light," /lait/. After a few practice trials to encourage children to verbally label the pictures, the pictures were presented one at a time on the monitor. The two pictures were intermixed, and each was presented six times. The last five repetitions of each word were retained for acoustic analysis.

These children were tested individually in a quiet and comfortable room at their school. All the utterances were recorded using a flash digital recorder (Maranz PMD 670) and a light-weight, head-mounted microphone (Shure SM10A). The recording was done at a sampling rate of 22.05 kHz and 16 bit quantization.

3. Transcription and measurement

From each child, we attempted to collect five tokens each of eight test words: /li/, /lu/, /la/, /ri/, /ru/, /ra/, /lait/, and /rait/, a total of 40 tokens per child. For 48 participants, this would be a total of 1920 possible tokens (48 children × 8 words × 5 tokens). Of these, 1857 tokens were collected. Sixty-three tokens (3.3%) were not collected because five children did not complete all five trial cycles and another five children skipped a trial. Among the 1857 tokens collected, 27 tokens (1.5%) from 15 participants were eliminated because of conditions that might affect the accuracy of acoustic measurement, e.g. simultaneous laughter, loud background noise, or the use of an abnormal speaking style. With these tokens excluded, the remaining 1830 tokens were submitted to further analysis.

To examine the accuracy with which /l/ and /r/ were produced, each production was transcribed by a trained phonetician. The accuracy of each production was then determined with regards to the initial consonant: For example, if the transcriber recorded an /l/ for an intended /l/ word, the production received a "correct" score.

For acoustic formant frequency analysis, the F2 and F3 frequencies were measured at the /l/ and /r/ onset using the formant-tracking function of PRAAT 5.0 acoustic analysis software (Boersma and Weenink, 2010). Each sound file containing a test word was down-sampled to 16 kHz and analyzed using a 10-ms Hamming window. F2 and F3 values were extracted from the spectrogram using the burg method (Burg, 1978) at a location where there were clear peaks for the first three formants in the Long-term average spectrum (LTAS), the LPC autocorrelation, and FFT spectra (Fig. 1). If clear peaks were not present at the onset, the measurement location was shifted in time by 10 ms increments until peaks were present. Most measurements were taken within 30 ms from the onset. No measurement location was shifted further than the one-third point of the syllable.

Figure 1 shows that in this production of /la/, formant peaks were found 10 ms following the onset in the spectrogram, where F2 and F3 values were taken. If the output values of the tracker greatly differed from visual inspection of the spectrogram and spectrum, the reference value in the formant tracker was changed and the sound was analyzed again, or the formant was manually estimated by visual
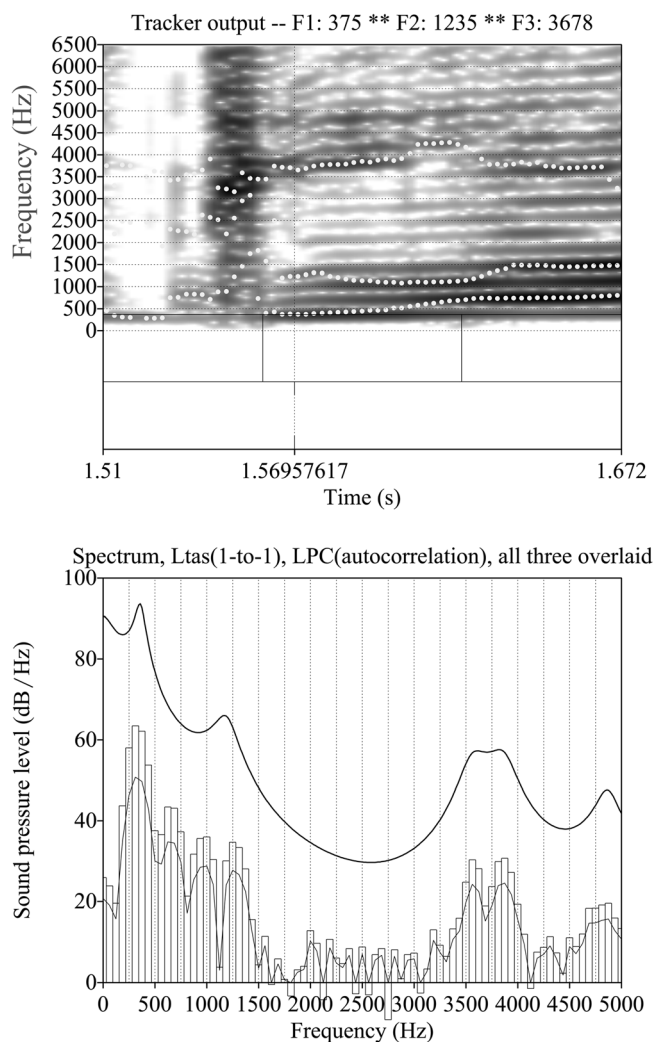

FIG. 1. Spectrogram (top) and LPC and FFT spectra overlaid (bottom) of the word "light" produced by one of the participants. The vertical line at 1.60 ms is showing the measurement location for the formant tracking.

inspection of the corresponding peak in the three spectral displays. A preliminary acoustic analysis was first conducted by a graduate research assistant, followed by a second and a third round of reanalysis by the first author to maximize measurement accuracy. Furthermore, F2 and F3 of 183 randomly selected tokens (10% of the entire database) were remeasured at a later time to examine the consistency of measurements. The mean absolute differences between two measurement points in Hz were 107 (178 SD) for F2 and 130 (106 SD) for F3. The correlation between measurements across the two times was high and significant for both F2 ($r = 0.885$) and F3 ($r = 0.951$) ($p < 0.001$ for both).

B. Results

1. Production accuracy as measured by an adult expert

Production accuracy was measured as a match between transcription by an adult expert phonetician and intended production of the child. Table I reports the percent accuracy of the children's intended /l/ and /r/ in each age group. Qualitatively, whereas transcription of the oldest children's

TABLE I. Mean accuracy and SD of /l/ and /r/ productions by each age group (%) judged by a trained phonetician.

Child Production		/l/	/r/	Overall
			Accuracy (%)	
4-yr-old	mean	94.9	96.0	95.4
	SD	5.3	8.4	5.5
4.5-yr-old	mean	89.8	91.0	90.4
	SD	24.5	15.7	14.7
5.5-yr-old	mean	81.6	92.4	87.0
	SD	31.7	11.5	17.4
8.5-yr-old	mean	99.5	99.5	99.5
	SD	1.5	1.5	1.0

TABLE II. Mean F*2, F*3, and Bark-converted F3-F2 values and SD in /l/ and /r/ for each age group.

Age Group		F*2		F*3		F3-F2	
		/l/	/r/	/l/	/r/	/l/	/r/
4-year-old	Mean	0.23	−0.21	0.75	−0.78	5.97	4.7
	SD	0.44	0.44	0.2	0.21	1.44	1.63
4.5-year-old	Mean	0.12	−0.12	0.67	−0.68	5.6	4.19
	SD	0.54	0.52	0.23	0.22	1.5	1.44
5.5-year-old	Mean	−0.23	0.23	0.68	−0.68	6.58	4.05
	SD	0.29	0.31	0.28	0.27	0.6	1.52
8.5-year-old	Mean	−0.27	0.28	0.92	−0.92	6.52	2.89
	SD	0.39	0.4	0.04	0.04	0.8	0.64

speech was at ceiling (99.5% correct, 1.5 SD for both /l/ and /r/), transcription of younger children's speech yielded lower accuracy and higher variability, particularly for 4.5- and 5.5-yr-olds' /l/ productions. However, a 2×4 (category (/l/, /r/) $\times$ age group) mixed-model analysis of variance (ANOVA) on percent correct found no significant main effects or interaction [category: $F(1, 44) = 0.988$, $p = 326$; age group: $F(3, 44) = 2.484$, $p = 0.073$; category*age group: $F(3, 44) = 0.598$, $p = 0.620$]. It is noted that the effect of age group approaches significance. Tukey's *post hoc* indicates that the difference in transcription accuracy between 5.5- and 8.5-yr-olds approaches significance ($p = 0.073$). The potential lower accuracy for 5.5-yr-olds' speech is likely caused by two children whose speech was transcribed 15.0% and 20.0% correct in the /l/ category, bringing overall transcription accuracy of their speech to 57.7% and 47.5%. The other 10 children in the age group were transcribed with accuracy on average 82.5% or above. These results show that, in general, measured by "correct" and "incorrect" judgment by an expert adult listener, the children's /l-r/ productions do not differ between 4 and 8.5 yr of age. These results are consistent with the findings in prior clinical studies that reported children reach 90% accuracy by age 6 or 7 (Sander, 1972; Smit *et al.*, 1990).

2. Changes in formant frequencies

F2 and F3 frequency values taken at the onset of /l/ and /r/ were first Bark-converted and then normalized using a Lobanov method to eliminate variation caused by physiological differences among children[4] (Thomas and Kendall, 2007). The Bark-converted normalized F2 and F3 values (F*2 and F*3 hereafter) in /l/ and /r/ averaged across four vowel contexts (/i/, /u/, / a/, and /ai/) were compared across the four age groups. The Bark-converted values, but without Lobanov normalization, were used to compute F3-F2 values because subtraction of Bark-converted formant frequencies from each other serves to normalize for physiological differences (Bark difference metric, Thomas and Kendall, 2007). Table II reports the mean values of these measures for each age group.[5] It is important to bear in mind that F3 alone and F3-F2 are highly correlated ($r = 0.898$, $p < 0.001$), and thus it is not feasible to tease apart the effect of F3 versus F3-F2 distance in /l-r/ categorization.

Figure 2 presents box plots illustrating distributions of F*3, F*2, and F3-F2 distance separately for /l/ (gray) and /r/ (black) across age groups. It is clear from Fig. 2(a) that distributions of F*3, the primary cue to the /l-r/ contrast, separate the /l/ and /r/ categories. Furthermore, the variance of this measure decreases considerably among the oldest, 8.5-yr-old, children as indicated by a short length of the box plot whiskers. A 2×4 (category $\times$ age group) mixed-model ANOVA on F*3 found a significant main effect of category, a significant interaction between the two factors, and no main effect of age group [category: $F(1,44) = 636.878$, $p < 0.001$; category*age group: $F(3, 44) = 3.422$, $p = 0.025$; age group: $F(3, 44) = 1.033$, $p = 0.387$]. Given a significant interaction between category and age group, two sets of *post hoc* tests were run, with the first set examining /l-r/ differentiation along the F*3 dimension within each age group and the second set examining development of F*3 across ages separately for /l/ and /r/.

Paired-sample *t*-tests indicated that the F*3 difference between /l/ and /r/ was significant for all age groups: F3 was higher-frequency for /l/ than for /r/ across all age groups ($p < 0.001$ for all, alpha adjusted for 0.013 for 4 comparisons). Thus even the youngest 4-yr-olds acoustically differentiated /l/ and /r/ by F*3 in their speech productions, suggesting that the primary acoustic cue (F3) to this contrast is already robust as young as age 4. In addition, one-way ANOVAs on F*3 in both /l/ and /r/ productions with age group as a factor, found a significant effect of age group for each category [/l/: $F(3, 44) = 3.375$, $p = 0.027$; /r/: $F(3, 44) = 3.417$, $p = 0.025$]. F*3 differed between 4.5- and 8.5-yr-olds ($p = 0.031$ and 0.38 for /l/ and /r/) and between 5.5- and 8.5-yr-olds ($p = 0.047$ and 0.038 for /l/ and /r/), indicating that F*3 changed in the production of both /l/ and /r/ as children grew older. It is important to note the different direction of F*3 development for /l/ and /r/: The mean F*3 for /l/ is 0.67 (4.5-yr-olds) and 0.92 (8.5-yr-olds) and the mean F*3 in /r/ is 0.68 (4.5-yr-olds) and −0.92 (8.5-yr-olds). These results suggest a developmental trend in which F*3 rises in frequency for /l/, whereas it lowers in frequency in /r/ productions, in effect, further differentiating the two categories on this acoustic dimension among older children. This development is observed mostly in differences of 8.5-yr-olds compared to younger children.
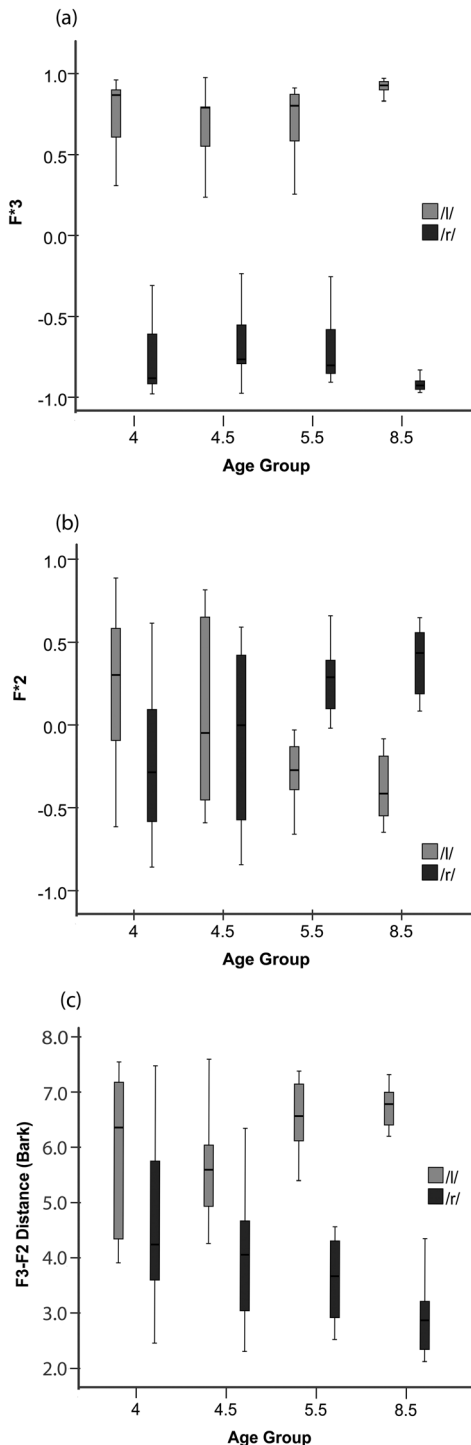
FIG. 2. Box plots illustrating distributions of (a) F*3, (b) F*2, and (c) F3-F2 distance for /l/ (gray) and /r/ (black) for the four age groups.

The distribution of F*2 [Fig. 2(b)] shows more variance than that of F*3 and thus greater overlap between younger children's /l/ and /r/. This variability seems to decrease, thus better separating the two categories in older children's productions. A 2×4 (category $\times$ age group) mixed-model ANOVA on F*2 found no significant main effects [category: F(1, 44) = 0.474, $p = 0.495$; age group: F(3, 44) = 1.051, $p = 0.379$], and a significant interaction between the two factors [category*age group: F(3, 44) = 4.021, $p = 0.13$]. Parallel to the analysis of F*3, two sets of *post hoc* tests

were conducted: The first set examining /l-r/ differentiation along the F*2 dimension within each age group and the second set examining developmental change in F*2 across these young ages.

Paired-sample *t*-tests indicated that the F*2 difference between /l/ and /r/ was not significant in any age groups. Although the *P* level of the test for the two older groups reached 0.022 and 0.041 (5.5- and 8.5-yr-olds), the null hypothesis is not rejected due to Bonferroni's correction to guard for false positives in conducting multiple comparisons (alpha was adjusted to 0.013 for four comparisons). This means that the mean F*2 values did not reliably differentiate /l/ and /r/ categories for any group, but a trend toward differences within age group is implicated. One-way ANOVAs on F*2 in each of /l/ and /r/ productions with age group as a factor found a significant effect of age group in both categories [/l/: F(3, 44) = 4.039, $p = 0.013$; /r/: F(3, 44) = 3.989, $p = 0.013$]. For F*2 in /l/, Tukey's tests found a significant difference between 4- and 8.5-yr-olds ($p = 0.036$) and a marginally significant difference between 4- and 5.5-yr-olds ($p = 0.054$). For F*2 in /r/, there was a significant difference between 4- and 8.5-yr-olds ($p = 0.040$). Again the directionality of F*2 development is noteworthy: F*2 in /l/ decreases across these ages (0.23 to −0.27 from 4- to 8.5-yr-olds), whereas F*2 in /r/ increases (−0.21 to 0.28 from 4- to 8.5-yr-olds).

These results suggest that although F2 has been implicated as a secondary acoustic correlate for the /l-r/ distinction in adult productions (Lotto *et al.*, 2004; Ingvalson *et al.*, 2011), this formant frequency did not correlate with these children's /l/ and /r/ productions. The current data suggest a trend such that F2 may begin to differentiate the categories in older children's speech. The development of F2 as an acoustic cue to the /l-r/ distinction, therefore, may occur in children older than 8.5 yr of age.

It has been suggested that close F3-F2 distance (the small frequency difference between F3 and F2) is related to /r/-ness (Stevens, 1998; McGowan *et al.*, 2004; Dalcher *et al.*, 2008). F3-F2 distance, then, may contribute to distinguish /r/ from /l/. Figure 2(c) illustrates the mean Bark-converted F3-F2 for /l/ and /r/ across age groups. A 2×4 (category $\times$ age group) mixed-model ANOVA on this measure found a significant main effect of category, a significant interaction between category and age group, and no significant main effect of age group [category: F(1,44) = 54.154, $p < 0.000$; category*age group: F(3, 44) = 3.249, $p = 0.031$; age group: F(3, 44) = 2.083, $p = 0.124$].

Paired-sample *t*-tests indicated that F3-F2 difference between /l/ and /r/ was significant in two older groups of children ($p < 0.000$ for 5.5- and 8.5-yr-olds). The F3-F2 distance was greater in /l/ than in /r/, differentiating the two categories in children older than 5 yr of age (5.5- and 8.5-yr-olds). One-way ANOVAs on F3-F2 in each of /l/ and /r/ productions with age group as a factor found a significant effect of age group in /r/ but not in /l/ [/r/: F(3, 44) = 3.480, $P = 0.024$; /l/: F(3, 44) = 1.982, $p = 0.130$]. Tukey's tests for /r/ found a statistically significant difference between 4- and 8.5-yr-olds ($p = 0.015$). These results indicate that the F3-F2 distance develops as a potential acoustic cue in older

TABLE III. Logistic regression models predicting /l/ and /r/ productions with F*2 and F*3 as predictors.

	Percentage classification	Factor	B	s.e.	Wald	df	sig	Exp(B)
4-yr-old	90.9	F*2	−0.525	0.165	10.128	1	0.001	0.591
		F*3	−3.205	0.285	126.083	1	0.000	0.041
4.5-yr-old	84.9	F*2	−0.393	0.135	8.527	1	0.003	0.675
		F*3	−2.235	0.179	155.017	1	0.000	0.107
5.5-yr-old	84.7	F*2	0.322	0.139	5.365	1	0.021	1.380
		F*3	−2.418	0.212	130.220	1	0.000	0.089
8.5-yr-old	99.5	F*2	2.648	1.359	3.796	1	0.051	14.127
		F*3	−19.051	8.724	4.768	1	0.029	0.000

children's speech (between 5.5- and 8.5-yr-olds), and it may further develop in even older children by lowering this feature in /r/ productions.

3. Category classification by formant frequencies

Another method of evaluating formant frequencies as acoustic correlates of category distinctions is to examine how well the formant values predict the category (i.e., /l/ vs /r/) of individual productions as intended by children. To this end, binary logistic regression was used to model the category classification as a function of F*2 and F*3 for each age group. In this analysis, classification accuracy is obtained as a percentage of cases where model's prediction based on F*2 and F*3 match children's intended production. The beta coefficients of the predictors (i.e., F*2 and F*3) in the model indicate the weight of their contribution in predicting the phonetic category.

The prediction model with the two factors was significantly better than the null model for all groups ($p < 0.001$), and the models accurately classified 90.9%, 84.9%, 84.7%, and 99.5% of the productions of 4-, 4.5-, 5.5-, and 8.5-yr-olds, respectively (Table III). The beta coefficients of both F*2 and F*3 were statistically significant for all age groups, indicating a contribution of both frequencies in category classification. This means that even though the comparisons of the central tendency of F*2 showed no difference between /l/ and /r/ [Fig. 1(b)], they may play a role in differentiating the two categories. However, as expected, the coefficient of F*3 is greater than F*2 for every age group, indicating the primary role that F*3 plays in this classification. Another intriguing finding is that whereas the sign of the F*3 coefficient (direction of influence) is negative across the board, the sign of the F*2 coefficient changes with age: It is negative for younger two groups and positive for older two groups. The model's default prediction was set as /r/ with a greater value of F*3 predicting more /l/ (fewer /r/) and a smaller value of F*3 predicting more /r/. The results for F*2, therefore, seem to indicate a change in mapping across age groups: For the younger two groups, the greater the F*2, the more likely the sound is /l/ (fewer /r/), whereas for older two groups, the greater F*2, the more likely it is to be /r/.

Recall that a trend was found that F*2 develops to differentiate the /l/ and /r/ categories in 5.5- and 8.5-yr-old children. The mean of F*2 for /l/ and /r/ crosses over across 4.5- and 5.5-yr-olds (Table II), and the distributional

variance also decreases across these age groups [Fig. 2(b)]. Before age 5, the pattern of F*2 parallels that of F*3: Lower F*2 and F*3 are mapped to /r/, whereas after age 5, this flips for F*2: Lower F*2 is mapped to /l/ while lower F*3 remains mapped to /r/. Thus it appears that F*2 may begin to exert its own pattern for /l/ and /r/ after 5 yr of age.

A one-factor model was also built with Bark-converted F3-F2 as a predictor for each age group. The prediction models ($p < 0.001$ for all cases) showed classification accuracy of 68.0%, 67.9%, 77.9%, and 92.2% for 4-, 4.5-, 5.5-, and 8.5-yr-olds. It is noted that classification accuracy with the F3-F2 distance as the predictor is low for younger children. However, it is also noted that this feature develops to be a reliable acoustic cue for the oldest child group.

4. Adult expert's identification and formant frequencies

How do children's F*3 and F*2 relate to the expert's identification of children's /l/ and /r/? This relationship was examined by binary logistic regression analysis relating the transcriber's identification of each sound (/l/ or /r/) and its F*3 and F*2. Of the 1830 tokens transcribed, the initial consonant of 18 tokens was identified something other than /l/ or /r/: Seven tokens lacked the consonant, five were heard intermediate between /l/ and /r/, four were unintelligible, and two were other consonants (a /f/ and a click). These cases were excluded from the analysis. Only one token was identified as /w/. This case was also excluded because it is not adequate to define a phonetic category with one instance of formant frequencies. The regression model with F*3 and F*2 as predictors was significant ($p < 0.001$), and only F*3 was a significant predictor (beta coefficient for F*3 = −0.873 with /r/ as reference, $p < 0.001$). This means that sounds with higher F*3 were identified as /l/ and lower F*3 as /r/ as expected (e.g., Polka and Strange, 1985).

A one-factor model with Bark-converted F3-F2 also indicated significant prediction of adult expert identification ($p < 0.001$, beta coefficient = −0.219 with /r/ as reference). This indicates that a small distance between F3 and F2 predicts the identification of /r/ and a large distance predicts the identification of /l/, as expected. It is noted that F*2 was not an important factor in this analysis. Recall that the transcription was coded for binary, correct or incorrect identification of the initial consonant. The lack of relationship between F*2 and transcriber's identification may possibly be due to

the subtle influence a secondary cue exerts on speech categorization in the presence of robust primary cue or to the fact that F*2 is not fully developed in these children and the transcriber could not use this information in his judgments.

C. Discussion

Adult native speakers of English distinguish word-initial /l/ and /r/ sounds primarily by the onset F3 cue (Polka and Strange, 1985; Yamada and Tohkura 1992; Iverson and Kuhl, 1996): Low F3 signals /r/, whereas high F3 signals /l/. The results in this study have demonstrated that, in production, this primary acoustic cue develops relatively early in children. The /l/ and /r/ categories are well separated by F3 formant frequency, in terms of their mean values as well as their classification performance as evaluated by binary logistic regression, in the productions of young children at 4 yr of age, the youngest group investigated in this study. Even so, the separation of the two categories as indicated by the F3 values of these categories continues to develop from 4.5-yr-olds to 8.5-yr-olds, through raising F3 for /l/ and lowering it for /r/, further enhancing the category differences on this dimension.

In addition to the onset F3 formant frequency cue, onset F2 frequency has been implicated as a secondary cue to the /l-r/ distinction among adult speech productions (Polka and Strange, 1985; Yamada and Tohkura 1992; Iverson and Kuhl, 1996; Lotto et al., 2004). The current study showed that in contrast to the earlier development of F3, the secondary F2 acoustic cue is only beginning to develop in the speech of the children tested in this study. At age 4, /l/ and /r/ categories overlap considerably along the F2 dimension, but at age 8.5, there is a trend toward separation between the two categories in terms of this dimension. The development of F2 cue thus appears to lag considerably behind that of F3. Given the trajectory of development evidenced in this study, it is possible that F2 matures in children older than 8 or 9 yr of age. This is consistent with the distributional characteristics of F3 and F2 for /l/ and /r/ in the speech environment. In adult speech, F3 is more robust and reliable distinguishing /l/ and /r/ than F2 (Lotto et al., 2004). Therefore children's productions seem to mirror the input statistics in that the robust F3 cue characterizes children's speech earlier than the F2 cue in production. The development of F2 seems to begin around 5 or 6 yr of age when its mapping to /l-r/ categories begins to differ from the mapping pattern of F3 and distributional variance of F2 decreases. The transcriber's categorization of children's /l/ and /r/ tokens seems to reflect this. The F2 cue, still underdeveloped in these children, may not have provided additional information to guide the categorical judgments of the expert adult when the robust, primary F3 cue was available.

F3-F2 distance has been proposed as an important acoustic cue for /r/ in recent years (McGowan et al., 2004; Dalcher et al., 2008). The results here show that the F3-F2 distance develops in older children (after 5 yr of age) to separate the /l/ and /r/ categories. This study is unable to completely tease apart the effect of F3 and F3-F2 distance in /l-r/ categorization: However, it reveals that the acoustic feature of F3-F2 distance develops as a reliable cue to distinguish /l/ and /r/.

III. RATING OF CHILD PRODUCTION BY GENERAL ADULT LISTENERS

A subset of child productions ("light" and "write") were identified and rated for category goodness by adult listeners. The aim of this undertaking was to examine the goodness of children's /l/ and /r/ productions as judged by general adult listeners and relate the adult ratings to the acoustics of children's speech (i.e., F3 and F2).

A. Method

1. Participants

Twelve monolingual English speaking adult listeners (8 females) participated. Their ages ranged from 19 to 24 yr (Mean age = 22). All reported normal hearing.

2. Stimuli

The /lait/ ("light") and /rait/ ("write") tokens produced by children who participated in the production study were identified by the adult listeners. Among the 48 child speakers, there was one (4-yr-old) child who did not produce these words. Seven tokens that were excluded in the acoustic analysis (due to a singing, screaming, or extremely breathy voicing) were also excluded here. An additional nine tokens were excluded due to noise in the signal and a mispronunciation (i.e., "writing").[6] The remaining 454 tokens (231 /lait/ tokens, 223 /rait/ tokens) from 47 children were root-mean-squire (RMS) matched and used for the rating study.

3. Procedure

Seated in front of a computer monitor in a sound attenuated booth, adult participants identified and then rated each of the 454 stimuli twice presented in two random orders. The stimuli were presented diotically over headphones (Beyer DT-150). Each participant was instructed to identify each word as "white," "right," or "light" and then rate the goodness of the production on a 9-point scale from very poor (1) to very good (9). The response choice "right" was used instead of "write" for simplicity. The children's productions sometimes lacked a strong stop release. Participants were instructed to disregard the lack of release for the goodness rating.

A trial began with a simultaneous presentation of the auditory stimulus (e.g., /lait/) through the headphones and a visual stimulus comprised of three words, white, right, and light, on the monitor. The experiment was under the control of E-PRIME experiment software (Psychology Software Tools, Inc.). These words appeared in the same location on the computer screen on every trial: White on the left, right in the middle, and light on the right. Participants were instructed to respond quickly by pressing the "W" key for white, "R" for right, and "L" for light, response key physically matching the relative spatial location of the word on the screen.

The response was followed by a 500-ms interval and then another presentation of the same auditory stimulus and a new visual stimulus on the monitor indicating a 9-point scale and a question asking the degree to which the

J. Acoust. Soc. Am., Vol. 133, No. 6, June 2013

K. Idemaru and L. L. Holt: Development of /l/ and /r/   4239

pronunciation of the word was good/poor. Participants were instructed to respond by pressing the "1" key for 1 (very poor) and so forth. The response triggered the presentation of the next trial. The entire session was completed in approximately 1 hr.

B. Results

A percent accuracy for each child was computed as percent correct identification of the test words by adult listeners and the mean percent accurate scores across four age groups are provided in Fig. 3. Table IV reports the confusion matrix for each of the age groups. A 2×4 (category $\times$ age group) mixed-model ANOVA on percentage accurate score indicated a significant main effect of age group, but no main effect of category or no interaction between the two factors [age group: $F(3, 43) = 4.055$, $p = 0.013$; category: $F(1, 43) = 1.263$, $p = 0.267$; category*age group: $F(3, 43) = 1.483$, $p = 0.233$]. Tukey's *post hoc* tests indicated that the accuracy of 8.5-yr-old production was higher than that of all other groups ($p < 0.05$ for all). These results indicate that children's /l/ and /r/ production develops between age 5.5 (73.5% accurate) and age 8.5 (97.5% accurate) in accuracy as judged by phonetically naive adult listeners' identification of the productions. The accuracy scores observed in this experiment are lower than those observed in the first experiment (Sec. II A), particularly for younger children. It is expected that the lay listeners in this study, who may not be used to child speech, may not able to identify child productions with the same accuracy as a trained phonetician. However, these judgments by lay listeners should reflect more closely how /l/ and /r/ produced by these children would be perceived by general listeners.

Multinomial logistic regression was run to relate adults' identification of children's productions (i.e., /l/, /r/, or /w/) to the acoustics of the production (i.e., F*2 and F*3). The model with F*2 and F*3 as predictors was significant (overall classification of 82.3%, $p < 0.001$), and the beta coefficients of F*2 and F*3 with /l/ as reference (0.537 and -3.038 for /r/, and -0.816 and -1.029 for /w/, $p < 0.001$ for all cases) indicated that /r/ as opposed to /l/ judgments increased as F*2 increased and F*3 decreased, and the judgment was more dependent on F*3 than on F*2. Adults' /w/
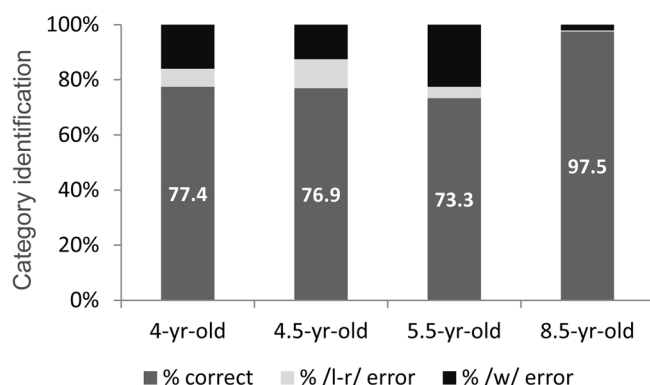
TABLE IV. Confusion matrix showing mean adult identification (%) of child productions.

Child Production		Adult identification			Overall /l-r/ accuracy
		/l/	/r/	/w/	
4-yr-old	/l/	88.0	3.2	8.7	76.8
	/r/	10.3	65.6	24.1	
4.5-yr-old	/l/	85.5	7.7	6.8	77.0
	/r/	13.4	68.4	18.2	
5.5-yr-old	/l/	68.0	3.1	29.0	73.5
	/r/	5.2	78.9	15.9	
8.5-yr-old	/l/	96.6	0.7	2.7	97.5
	/r/	0.2	98.3	1.5	

as opposed to /l/ identification, on the other hand, increased as both F*2 and F*3 decreased. The degree of change in this case was larger on the basis of F*2 (the primary cue for this contrast). These results confirm that adult listeners rely on the F*2 and F*3 cues in categorizing /l/ and /r/, giving more weight to F3 than F2 (Yamada and Tohkura, 1992; Iverson *et al.*, 2005; Ingvalson *et al.*, 2011). Lower F3 and higher F2 signal the English /r/ category, whereas higher F3 and lower F2 signal /l/.

Ratings of children's productions by adult listeners were collected in addition to identification responses. Figure 4 reports the mean rating scores for the correctly identified /l/ and /r/ productions across the age groups. A 2×4 (category $\times$ age group) mixed-model ANOVA on the mean adult ratings for each child speaker[7] indicated a significant main effect of age group, no main effect of category, and no interaction between the two factors [age group: $F(3, 41) = 8.235$, $p = 0.000$; category: $F(1, 41) = 1.543$, $p = 0.221$; category*age group: $F(3, 41) = 0.265$, $p = 0.850$]. Tukey's *post hoc* tests indicated that goodness ratings for the 8.5-yr-old's productions were higher than those of all other age groups ($p < 0.05$ in all cases). These results indicate that children's production of /l/ and /r/ develops between age 5.5 (mean rating $= 7.32$) and age 8.5 (mean rating $= 8.26$) in terms of adults' goodness judgment of their pronunciation. It should be noted, however, that the difference in perceived goodness was rather small. Thus when young children produced accurate /l/ and /r/, they were typically rated as good productions.
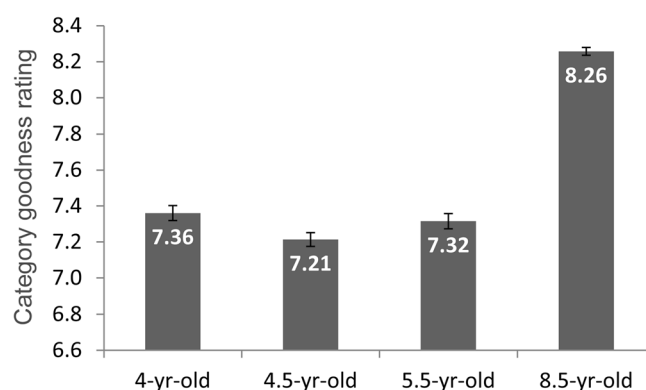

FIG. 3. The mean category identification of children's "light" and "write" productions judged by general adult listeners. The results for "light" and "write" are collapsed.

FIG. 4. The mean goodness rating for the correctly identified child /l/ and /r/ productions. The results for "Light" and "write" are collapsed.

A linear regression analysis was run to relate the adults' ratings of children's /l/ and /r/ productions to the acoustics of the production (i.e., F*2 and F*3). Thus those cases identified as light or right but not white were included in the analysis. The models with F*2 and F*3 as predictors were significant over the null model ($p < 0.001$ for both). F*3 and F*2 were both significant predictors of the goodness ratings of /l/ (beta coefficients: 0.334 and -0.201, $R^2 = 0.019$, $p < 0.001$): Adult listeners rated children's /l/ better when it had high F*3 and low F*2. Only F*3 was a significant predictor for the rating of /r/ (beta: -1.039, $R^2 = 0.162$, $p < 0.001$). Adult listeners rated children's /r/ better when it had lower F*3, and F*2 did not contribute to their goodness ratings. These results inform us that F*3 and F*2 contribute to the category goodness of /l/ (both F*3 and F*2) and /r/ (F*3).

C. Discussion

The acoustic analyses described in the preceding text indicate that 4- to 8-yr-old children's /r/-/l/ consonant acoustics are in the process of development in terms of the fine tuning of F2 and F3 onset frequencies. The current adult rating study provides important support for the findings: There is a perceptual consequence of this development. Adults' identification as well as goodness ratings of children's /l/ and /r/ were better for older children (8.5-yr-olds) than younger children. This presumably reflects developmental acoustic fine tuning of the /l/ and /r/ categories in the older children's productions. Recall that F3 increases for /l/ and decreases for /r/, and F2 decreases for /l/ and increases for /r/ across age. Thus in effect, the development of these formant frequencies results in enhancement of the differences between /l/ and /r/. This fine tuning of acoustic cues (F2 and F3) seems to have payoff in the better intelligibility of child speech judged by general adult listeners. The results also confirmed that both F3 and F2 are used by adults in categorizing /l/ and /r/. We further discovered that the formant frequencies contribute to the category goodness of /l/ and /r/.

IV. CHILD PERCEPTION

Perceptual weighting of F3 and F2 formant frequency by children and adults was investigated using /l-r/ stimuli that varied along the F3 and F2. The aim here was to examine the development of the use of these relevant perceptual cues in children.

A. Method

1. Participants

Of the 48 participants in the production study, 45 children participated in a later perception study, conducted 1–6 days following the production study. Two participants in the 4-yr-old group and one in 4.5-yr-old group (all boys) did not complete the perception task. Thus the child participants in this study were 23 girls and 22 boys: 4-yr-old (10 children, mean age = 4.15; age range = 3.95–4.37), 4.5-yr-old (12; 4.73, 4.42-5.04), 5.5-yr-old (12; 5.49; 5.05-6.13); and 8.5-yr-old (11; 8.45; 7.31-9.54). None of these children had been

diagnosed with speech/hearing problems, had six or more ear infections before their second birthday, had complications at birth, or used a foreign language on a regular basis. Eighteen adults also participated for a small payment as a comparison group. These adult participants were monolingual English undergraduate college students without known speech/hearing problems.

2. Stimuli

The stimuli were a subset of synthesized /lait - rait/ tokens used in Ingvalson *et al.* (2011), which were a close replication of the stimuli of Yamada and Tohkura (1992). Ingvalson *et al.* created these stimuli modeling the F0 and intensity contours of a native English male production while varying F2 and F3 frequencies with the cascade branch of the Klatt synthesizer (Klatt and Klatt, 1990). All stimuli had a total duration of 580 ms with the /rai/-/lai/ portion lasting 360 ms. There were 20 ms of silence between the offset of the diphthong and the onset of the burst; there were 140 ms between the onset of the burst and when the signal amplitude reached 0 dB. The final 60 ms of the stimulus were silence. All stimuli were sampled at 11 025 Hz and RMS amplitude matched in energy. Note that these stimuli were previously used to assess reliance of adult native English listeners on F2 and F3 frequencies. In this study, we use them to investigate the perceptual cue weighting in the same children whose speech productions we measured. The stimuli formed a two-dimensional grid sampling a factorial combination of F2 and F3 onset frequencies. F2 varied in four steps of 200 Hz (i.e., 800, 1000, 1200, and 1400 Hz) while F3 varied in four steps of 400 Hz (i.e., 1600, 2000, 2400, and 2800 Hz), creating 16 unique stimuli.

For all stimuli, F3 onset frequency was held at the initial steady state for 80 ms. It then linearly transitioned to the /a/ portion of the diphthong, 2465 Hz, reached at 180 ms. This value was maintained until 240 ms at which point it linearly transitioned to the /i/ portion of the diphthong, 2735 Hz, reached at 300 ms and held for the remaining 60 ms of the syllable. F2 onset frequency steady state duration varied in conjunction with the onset frequency to keep slope constant from consonant to vowel states in all stimuli. The onset-duration pairings were as follows: 800 Hz and 80 ms, 1000 Hz and 105 ms, 1200 Hz and 130 ms, 1400 Hz and 150 ms. For all stimuli, F2 reached 2350 Hz at 280 ms and maintained this value until the sound's end.

3. Task and procedure

A native English-speaking research assistant (RA) and the first author (KI) tested children individually. The RA explained the task (game) to the child. One of the visitors, Roo, who appeared in the production task, wanted to practice saying the two English words, write and light, which the child taught him in the earlier production task. The recorded voice of Roo then asked the child to tell him and the researcher which word his speech sounded like by pointing to one of the pictures on the computer monitor.

The computer monitor showed a picture of a hand writing a letter (write) and a picture of a lamp (light), the same

pictures used to elicit utterances of those words in the production task. The left/right positioning of the pictures was randomly assigned across trials. A small image of the character, Roo, always appeared in the middle with a sound icon. When KI clicked the sound icon, the auditory stimulus was presented through the headphones worn by the child and KI. The child pointed to one of the pictures to indicate which word they heard Roo say. The RA recorded the response choice. Prior to the presentation of the test stimuli, the RA used her live voice to utter a few practice trials to make sure the children understood the task.

Once the task began, the RA and KI did not praise children for responses so that children were not encouraged to think there was a correct answer. Instead the character Roo appeared on the screen every so often to say how many practices he and the child had finished. The 16 unique stimuli were presented in five random orders to each child. The testing was conducted individually in a quiet and comfortable room at the children's school.

To maintain the children's focus on the tasks, the production (previous section) and perception data collection were conducted 20 min at a time across two sessions. Typically, the first session included the production task and one cycle of presentation of 16 perception stimuli; and the second included a refresher practice of the perception task and four remaining cycles of perception trials. The two sessions took place within 6 days of each other.

B. Results and discussion

The adult listeners identified five random presentations of the stimuli under the control of Eprime. The stimuli were presented diotically over headphones and listeners responded by pressing designated keys on a keyboard to indicate their response choice (i.e., light or right).

1. Perception

Figure 5 illustrates percent /r/ responses as a function of F3 (the x axis) and F2 (lines) for each age group. It is noted that the identification curves did not reach 100% (for low F3) or 0% (for high F3) for younger children (4- and 4.5-yr-olds), making the curves less steep than those for 8.5-yr-olds and adults. It also appears that the adult identification functions were separated when F3 was 2000 Hz, suggesting a possible effect of F2.

Given the apparently strong influence of F3 frequency (the x axis in Fig. 4), listeners' responses to the F3 endpoint stimuli were examined first. The F3 endpoint stimuli, containing robust F3 information, were considered as exemplar /l/ and /r/ stimuli. Perceptual accuracy was calculated for each listener as percent identification of the exemplar /l/ stimuli as /l/ and percent identification of the exemplar /r/ stimuli as /r/ (Table V). This percent accuracy score provides a general sense of how children and adults identified clear /l/s and clear /r/s in the stimulus set.

A 2×4 (category $\times$ age group) mixed-model ANOVA on mean percent accurate scores indicated significant main effect of age group, significant interaction between the factors but no significant main effect of category [age group:
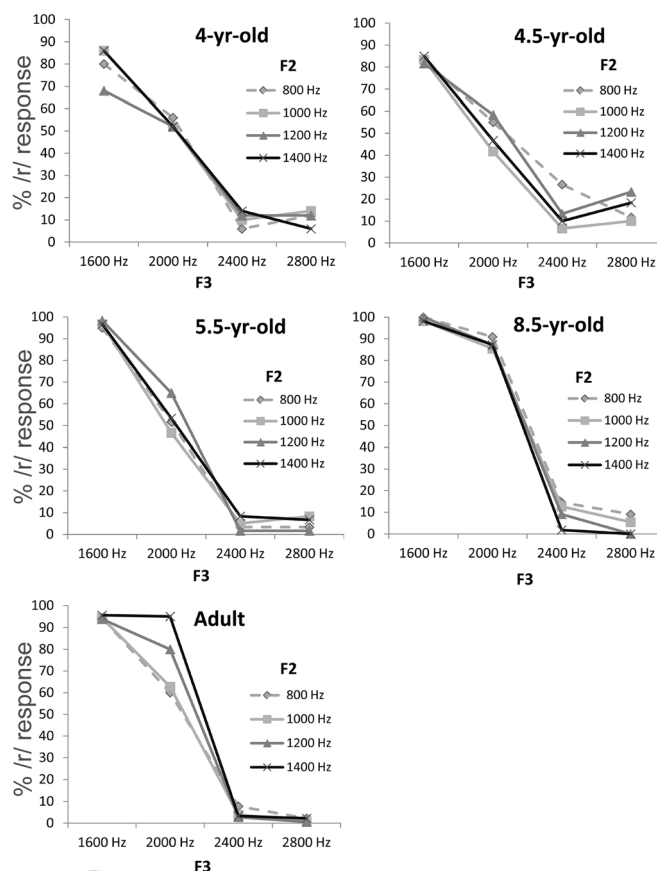

FIG. 5. Mean percent /r/ response as a function of F3 (the x axis) and F2 (lines) for each age group.

$F(4, 58) = 4.483$, $p = 0.003$; category*age group: $F(4, 58) = 2.809$, $p = 0.034$; category: $F(1, 58) = 2.422$, $p = 0.125$]. Tukey's *post hoc* tests indicated that adults (96.5% correct overall) were more accurate than 4.5-yr-olds (83.8% correct overall, $p = 0.028$) in identifying these stimuli, and there was a trend that adults were more accurate than 4-yr-olds (84.5% correct overall, $p = 0.066$) at the task. Older children, 5.5- and 8.5-yr-olds, were not different from adults ($p = 1.00$ and 0.999, respectively). These results indicate that in identifying exemplar /l/ and /r/, perceptual accuracy based on robust and clear F3 frequency information develops around 4–5 yr of age, reaching adult-like accuracy in 5.5 yr of age.

TABLE V. Perception accuracy of identifying F3 endpoint stimuli.

		Perception accuracy (%)		
		/l/	/r/	Overall
4-yr-old	Mean	89.0	80.0	84.5
	SD	*11.7*	*16.5*	*12.7*
4.5-yr-old	Mean	84.2	83.3	83.8
	SD	*20.3*	*22.1*	*20.5*
5.5-yr-old	Mean	95.0	96.7	95.8
	SD	*7.4*	*4.4*	*4.7*
8.5-yr-old	Mean	96.4	99.1	97.7
	SD	*6.4*	*3.0*	*3.3*
Adult	Mean	98.5	94.6	96.5
	SD	*4.2*	*11.0*	*7.4*

TABLE VI. Logistic regression models predicting /l/ and /r/ responses in perception with F*2 and F*3 as predictors.

	Percentage classification	Factor	B	S.E.	Wald	df	sig	Exp(B)
4-yr-old	76.4	F2	−0.010	0.080	0.014	1	0.905	0.991
		F3	−1.368	0.097	197.976	1	0.000	0.255
4.5-yr-old	75.9	F2	−0.022	0.070	0.100	1	0.752	0.978
		F3	−1.228	0.082	224.026	1	0.000	0.293
5.5-yr-old	85.3	F2	0.112	0.091	1.505	1	0.220	1.119
		F3	−2.467	0.153	258.575	1	0.000	0.085
8.5-yr-old	93.4	F2	−0.320	0.118	7.366	1	0.007	0.729
		F3	−3.424	0.227	227.619	1	0.000	0.033
Adult	90.8	F2	0.002	0.000	31.900	1	0.000	1.002
		F3	−0.008	0.000	720.355	1	0.000	0.992

In addition to using the robust primary cue, adult-like speech perception also involves use of secondary acoustic cues, F2 frequency information, in the case of /l-r/ categorization (e.g., Ingvalson *et al.*, 2011). To examine the use of and weighting of F3 and F2 frequencies in categorizing /l/ and /r/, binomial regression analysis was applied to category responses with F3 and F2 as predictors for each age group. All prediction models were significant over a null model ($p < 0.001$ for all cases). As Table VI reports, adults showed reliance on both F2 and F3 in categorizing /l/ and /r/ with more weight given to F3 than to F2 (coefficients of F3 and F2: −0.008 and 0.002). Unlike the adult pattern, F2 is not a significant factor in the categorization for younger children: 4-, 4.5-, and 5.5-yr-olds. These younger groups rely solely on F3 in speech categorization. On the other hand, for the oldest child group, 8.5-yr-olds, both F3 and F2 are significant factors (coefficients of F3 and F2: −3.424 and −0.320): However, the difference in the direction of the F2 effect indicates that 8.5-yr-olds are not entirely adult-like. When F2 rises, adults hear more /r/, whereas 8.5-yr-olds hear more /l/.

As reported in the literature (e.g., Abramson and Lisker, 1985) and demonstrated in Fig. 6, the secondary cue exerts its influence most when the primary cue is ambiguous. When F3 is ambiguous (i.e., 2000 Hz), adult's perception of stimuli changes from 60% to 95% /r/ as F2 increases with the robust 35% difference due to F2's influence. The correlation between the perceptual pattern and F2 is significant for adults at this F3 value ($r = 0.427$, $p < 0.001$). None of child groups showed such a systematic relationship between the perceptual response pattern and F2.
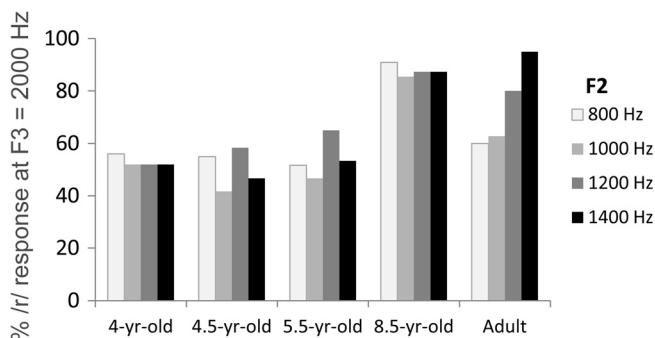
The results indicate that young children are able to use the primary cue (F3) to accurately categorize exemplar tokens of /l/ and /r/, reaching adult-like level of accuracy at 5.5 yr of age. However, parallel to the acoustic patterns characterizing their speech productions, the use of the secondary cue (F2) lags behind in development. Whereas adults show systematic and robust reliance of the secondary cue, children up to 5–6 yr of age do not use the secondary cue, and at 8 or 9 yr of age, children do show sensitivity to the cue but use it in a pattern different from adults'.

2. Perception and production

We obtained a production accuracy score for each child as judged by a trained phonetician in Sec. I, and another as judged by general adult listeners in Sec. II. To examine the relationship between perception and production within each child, the perception accuracy score of the end-point stimuli from this section was correlated to the production accuracy scores for each of /l/ and /r/ category. In general, children's /l/-/r/ perception accuracy was found to correlate only marginally with their production accuracy as judged by adult listeners. Perception accuracy of /l/ marginally correlated with the production accuracy (judged by phonetician) ($r = 0.277$, $p = 0.065$). Perception accuracy of /r/ marginally correlated with the production accuracy (phonetician) of both /l/ and /r/ (/l/: $r = 0.292$, $p = 0.052$; /r/: $r = 0.260$, $p = 0.085$) as well as the production accuracy (general listeners) of /l/ ($r = 0.253$, $p = 0.093$). These results seem to suggest that there is no systematic relationship between children's production and perception of /l/ and /r/.

In both the production study (Sec. I) and the current perception study, we have observed that the development of F2 lags behind that of F3. To investigate the development of F2 in production and perception, the weight of F2 in production was correlated with the weight of F2 in perception. Specifically, the correlation coefficient relating F*2 formant frequency and /l-r/ classification in production and the correlation coefficient relating F*2 formant frequency and perceptual categorization for ambiguous F3 (2000 Hz) was examined for each child. The two were marginally correlated ($r = 0.271$, $p = 0.063$), but the strength of the relationship was weak. The evidence here indicates that there is no robust correlation between production and perception with regards

FIG. 6. Effect of F2 on percent /r/ response for ambiguous F3 (2000 Hz).

J. Acoust. Soc. Am., Vol. 133, No. 6, June 2013

K. Idemaru and L. L. Holt: Development of /l/ and /r/   4243

		4 yr old	4.5 yr old	5.5 yr old	8.5 yr old
I. Production acoustics (formant values)	F3	/l/-/r/ differentiated			
	F2	/l/-/r/ NOT differentiated			(differentiated)
	F3-F2	/l/-/r/ NOT differentiated		/l/-/r/ differentiated	
I. Production cue weighting (beta weight)	F3	primarily weighted			
	F2	secondarily weighted WRONG direction		secondarily weighted right direction	
I. & II. Production accuracy & goodness (adult judgment)	Expert (accuracy)	no difference			(better)
	Non expert (accuracy and rating)	worse			better
III. Perception accuracy	Exemplar stimuli	(worse)	worse		better
III. Perception cue weighting (beta weight)	F3	solely weighted			primarily weighted
	F2	NOT weighted			secondarily weighted WRONG direction

FIG. 7. Summary of findings. Marginally significant findings are presented in parentheses.

V. GENERAL DISCUSSION

Some aspects of speech perception develop early in infancy: By the time infants are 1 yr of age, they have already become attuned to some aspects of the native language (e.g., Werker and Tees, 1984; Kuhl *et al.*, 2001). However, development of speech perception is multi-faceted, and there are aspects that continue to develop well into later childhood. This study examined one such area, namely, the acoustic and perceptual weighting of simultaneously available phonetic cues, focusing on perception and production of /l-r/ categorization in English. Speech categories are typically defined by simultaneously available phonetic cues (Lisker, 1986; Hillenbrand *et al.*, 2000; Polka and Strange, 1985), and weight given to each cue in defining a speech category may be different across languages (Iverson and Kuhl, 1995). Thus an important task for children is to discover what acoustic dimensions characterize a speech category and to give appropriate weight to each dimension.

In learning the /l-r/ distinction in English in an adult-like manner, children must learn to use F3 and F2 formant frequencies, such that higher F3 signals /l/ and higher F2 signals /r/, and to give more weight to F3 (Ingvalson *et al.*, 2011). Examination of production and perception of /l/ and /r/ by 4- to 8-yr-old children in this study revealed that the primary cue to this distinction, F3, is already robust in the youngest 4-yr-old children in production and perception, but the secondary cue, F2, seems only to begin to develop across these ages and development is not yet complete at ages 8 or 9 (see Fig. 7 for the summary of findings). If one considered only F3, one would conclude that all age groups, even the youngest group (4-yr-old), have well-distinguished /l/ and /r/ categories. However, non-expert adults' evaluation of the utterances indicated that there was a cross-sectional improvement in the children's /l/ and /r/ across age, whereby older children's /l/ and /r/ were heard as intended more frequently and were better rated by adult listeners.

Whereas, F2, the secondary cue, did not acoustically distinguish the categories in most child productions, the cross-sectional difference was such that it suggested development in the direction of separating the /l/ and /r/ categories. The perception study also demonstrated that children's categorization of /l/ and /r/ was dependent on the primary cue, F3, and does not incorporate the secondary cue, F2, in an adult-like manner at these ages. F2 as a cue to /l-r/ categorization is likely to continue to develop in children older than 8 yr of age. These findings are consistent with prior work suggesting that learning to use multiple phonetic cues for fricatives, affricates, and stops continues well into 10 or 12 yr of age (i.e., Hazan and Barrett, 2000). This pattern of development is also consistent with the distributional characteristics of /l/ and /r/ acoustics in the speech environment: F3 formant frequency is available in the signal as a robust and reliable acoustic cue, whereas F2 formant frequency is a less reliable acoustic cue (e.g., Lotto *et al.*, 2004). Although F3 is early-acquired and F2 late-acquired consistently across production and perception, production and perception of /l/ and /r/ within each child was not related. Perceptual development did not predict acoustic development or vice versa for /l/ and /r/.

The pattern of cross-sectional differences in F2 for /l/ and /r/ was intriguing. Whereas higher F2 signaled /l/ before age 5, it signaled /r/ after age 5 including adults. F2 formant frequency has been associated with the differences between dark versus clear /l/, resulting from a slight difference in the place of constriction (e.g., van Hofwegen, 2011). Clear /l/ is produced at the alveolar region. Dark /l/, on the other hand,

involves the tongue tip position slightly backward and velarization due to a secondary pharyngeal constriction (van Hofwegen, 2011; Narayanan et al., 1997). For /r/ production, F2 formant frequency has also been associated with a difference in the tongue position or the place of constriction along the alveolar to palatal region (e.g., Espy-Wilson et al., 2000). The acoustic models for /r/ proposed by Espy-Wilson et al. (2000) suggests that retroflex /r/, that produced with the tongue tip raised and the dorsum lowered, would show higher F2 than bunched /r/, that produced with raised tongue dorsum and lowered tongue tip. Given this, children's articulation of /l/ and /r/ may be changing with age such that the place of articulation for /l/ is moving slightly backward, whereas there is more involvement of tongue tip for /r/ production. However, this is highly speculative at this point and further work is needed to understand what changes underlie the pattern of F2 development.

Further work will also be necessary to track the development of F2 formant frequency with regards to the separation of the /l/ and /r/ categories along this dimension and the use of this cue in perception. Given that 8- and 9-yr-old children showed a trend toward more adult-like pattern of F2 in production, this cue may develop fully to separate the two categories in 10-yr-old children's production. However, children's pattern of perception did not suggest a change toward an adult-like use of the secondary cue. Considering the prior finding that even 12-yr-old children do not show adult-like use of multiple acoustic cues in perception (Hazan and Barrett, 2000), it may be necessary to examine children older than 10 yr of age to track this aspect of phonetic development to maturity.

ACKNOWLEDGMENT

[1]Instead of grouping the children with rigid age divides by year, we tested children in the Carnegie Mellon Children's School and grouped children into more-or-less equal groups according to their position in the overall age distribution to collect data from as continuous and large a sample across ages as permitted by our population.

[2]The gender break up of each age group was as follows: 4-yr-old (5 girls and 7 boys); 4.5-yr-old (8 girls and 5 boys); 5.5-yr-old (2 girls and 10 boys); 8.5-yr-old (8 girls and 3 boys).

[3]A total of 61 children took the test. The school where most of these children went to and were tested does not allow researchers to pre-select which children participate. Thus as long as children were in the target age range and were willing, they participated. A questionnaire asking questions related to these screening criteria were later collected from children's care-takers. If any of these criteria were indicated, the child's data were excluded. A total of 13 children were excluded.

[4]The formant values were converted to Bark using the formula, $Z_i = 26.81/(1+1960/F_i) - 0.53$, where F_i is the value for a given formant i (Traunmüller, 1997). Lobanov normalization was performed for individual /l/ and /r/ using the NORM suite (Thomas and Kendall, 2007). The formula used by the NORM suite, which focuses on vowel normalization, $F_{n[V]}^N = (F_{n[V]} - MEAN_n)/S_n$, derives the normalized value for formant

n of vowel V. $MEAN_n$ is the mean value for formant n and S_n is the standard deviation for the speaker's formant n. Using this, we obtained normalized formant values of F2 and F3 for individual productions of /l/ and /r/.

[5]Preliminary analyses found effects of following vowel (i.e., /ɑ/, /i/, /u/, /ai/) on the Bark-converted normalized F2 and F3 frequencies of the preceding /l/ and /r/. These effects were primarily in the direction predicted based on coarticulation effects and more importantly did not affect /l/ and /r/ differentially. F2 in both /l/ and /r/ was higher before /i/ than before any other vowels (/u/, /ɑ/, and /ai/), and it was higher before /u/ than before /ɑ/ and /ai/ [$P < 0.008$]. Whereas F3 in /r/ did not differ as a function of the preceding vowel, F3 in /l/ was slightly higher before /ai/ than before other vowels [$P < 0.004$]. The cause of this single difference is not clear.

[6]These factors did not impede the measurement of formant frequencies at the word onset and thus these tokens were included in the acoustic analysis in the production study described in the preceding text.

[7]None of the /r/ productions of two children in the 4-yr-old group were correctly identified by the adult raters. Thus these children's data were excluded from the analyses here.

Abramson, A. S., and Lisker, L. (1985). "Relative power of cues: F0 shift versus voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (Academic, Orlando, FL), pp. 25–33.

Aslin, R. N., Pisoni, D. B., Hennessy, B. I., and Perey, A. J. (1981). "Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience," Child Dev. 52, 1135–1145.

Best, C. T. (1995). "A direct realist perspective on cross-language speech perception," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 167–200.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. 4, 45–60.

Boersma, P., and Weenink, D. (2010). "PRAAT: Doing phonetics by computer (version 5.0) [Computer program]," http://www.praat.org (Last viewed 9/13/2011).

Bosch, L., and Sebastián-Gallés, N. (2003). "Simultaneous bilingualism and the perception of a language specific vowel contrast in the first year of life," Lang. Speech 46, 217–244.

Burg, J. P. (1978). "Maximum entropy spectral analysis," in *Modern Spectrum Analysis*, edited by D. G. Childers and S. B. Kesler (IEEE Press, New York), Vol. 331, pp. 23–33.

Dalcher, C. V., Knight, R. A., and Jones, M. J. (2008). "Cue switching in the perception of approximants: Evidence from two English dialects," Univ. Pa. Work. Pap. Linguist. 14(2), 63–71.

Dalston, R. M. (1975). "Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults," J. Acoust. Soc. Am. 57, 462–469.

Diehl, R. L., Kluender, K. R., Walsh, M. A., and Parker, E. M. (1991). "Auditory enhancement in speech perception and phonology," in *Cognition and the Symbolic Processes. Applied and Ecological Perspectives*, edited by R. R. Hoffman and D. S. Palermo (Erlbaum, Hillsdale, NJ), Vol. 3, pp. 59–76.

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). "Speech perception," Annu. Rev. Psychol. 55, 149–179.

Diehl, R. L., and Walsh, M. A. (1989). "An auditory basis for the stimulus-length effect in the perception of stops and glides," J. Acoust. Soc. Am. 85, 2154–2164.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., and Alwan, A. (2000). "Acoustic modeling of American English /r/," J. Acoust. Soc. Am. 108(1), 343–356.

Goto, H. (1971). "Auditory perception by normal Japanese adults of the sounds 'l' and 'r,'" Neuropsychologia 9(3), 317–323.

Hazan, V., and Barrett, S. (2000). "The development of phonemic categorization in children aged 6-12," J. Phonetics 28(4), 377–396.

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. 108, 3013–3022.

Idemaru, K., and Holt, L. L. (2011). "Word recognition reflects dimension-based statistical learning," J. Exp. Psychol. Hum. Percept. Perform. 37(6), 1939–1956.

Idemaru, K., Holt, L. L., and Seltman, H. (2012). "Individual differences in cue weights are stable across time: The case of Japanese stop lengths," J. Acoust. Soc. Am 132(6), 3950–3964.

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

Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278.

Iverson, P., and Kuhl, P. K. (**1995**). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," J. Acoust. Soc. Am. **97**, 553–562.

Iverson, P., and Kuhl, P. K. (**1996**). "Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/," J. Acoust. Soc. Am. **92**, 1130–1140.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**(1), 47–57.

Kidd, G. R. (**1989**). "Articulatory-rate context effects in phoneme identification," J. Exp. Psychol. Hum. Percept. Perform. **15**(4), 736–748.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis and perception of voice quality variations among male and female talkers," J. Acoust. Soc. Am. **87**, 820–856.

Kluender, K. R., Diehl, R. L., and Wright, B. A. (**1988**). "Vowel-length differences before voiced and voiceless consonants: An auditory explanation," J. Phonetics **16**, 153–169.

Kuhl, P. K. (**1991**). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories; Monkeys do not," Percept. Psychophys. **50**, 93–107.

Kuhl, P. K. (**1998**). "The development of speech and language," in *Mechanistic Relationships Between Development and Learning*, edited by T. C. Carew, R. Menzel, and C. J. Shatz (Wiley, New York), pp. 53–73.

Kuhl, P. K., Tsao, F.-M., Liu, H.-M., Zhang, Y., and De Boer, B. (**2001**). "Language/culture/mind/brain," Ann. N.Y. Acad. Sci. **935**, 136–174.

Lisker, L. (**1986**). " 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang. Speech **29**(1), 3–11.

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 *Sound to Sense: 50+ Years of Discoveries in Speech Communication*, edited by J. Slifka, S. Manuel, and M. Matthies (Electronic conference proceedings), pp. 181–186.

McGowan, R., Nittrouer, S., and Manning, C. (**2004**). "Development of (r) in young, Midwestern, American children," J. Acous. Soc. Am. **115**, 871–884.

Miller, J. L., and Eimas, P. D. (**1996**). "Internal structure of voicing categories in early infancy," Attention, Percept. Psychophys. **58**(8), 1157–1167.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.

Narayanan, S. S., Alwan, A. A., and Haker, K. (**1997**). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. I. The laterals," J. Acous. Soc. Am. **101**, 1064–1077.

Newman, R. S., and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**, 540–560.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effect of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 1–32.

Nittrouer, S. (**1996**). "The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM," J. Speech Hear. Res. **39**, 1059–1070.

Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.

Nittrouer, S. (**2004**). "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," J. Acoust. Soc. Am. **115**, 1777–1790.

Nittrouer, S., and Miller, M E. (**1997**). "Predicting developmental shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**, 2253–2266.

Ohde, R. N., and Haley, K. L. (**1997**). "Stop-consonant and vowel perception in three- and four-year-old children," J. Acoust. Soc. Am. **102**, 711–722.

Ohde, R. N., Haley, K. L., Vorperian, H. K., and McMahon, C. W. (**1995**). "A developmental study of the perception of onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **97**, 3800–3812.

Polka, L., and Strange, W. (**1985**). "Perceptual equivalence of acoustic cues that differentiate /r/ and /l," J. Acoust. Soc. Am. **78**, 1187–1197.

Polka, L., and Werker, J. F. (**1994**). "Developmental changes in perception of non native vowel contrasts," J. Exp. Psychol. Hum. Percept. Perform. **20**, 421–435.

Sander, E. K. (**1972**). "When are speech sounds learned?" J. Speech Hear. Disord. **37**, 55–63.

Sheldon, A., and Strange, W. (**1982**). "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," Appl. Psycholing. **3**(3), 243–261.

Shultz, A. A., Francis, A. L., and Llanos, F. (**2012**). "Differential cue weighting in perception and production of consonant voicing," J. Acous. Soc. Am, **132**(2), EL95–EL101.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., and Bird, A. (**1990**). "The Iowa articulation norms and its Nebraska replication," J. Speech Hear. Disord. **55**, 779–798.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA and London), p. 535.

Streeter, L. A. (**1976**). "Language perception of 2-month old infants shows effects of both innate mechanisms and experience," Nature **259**, 39–41.

Thomas, E. R., and Kendall. T. (**2007**). "NORM: The vowel normalization and plotting suite," http://ncslaap.lib.ncsu.edu/tools/norm/ (Last viewed 5/15/2012).

Traunmüller, H. (**1997**). "Auditory scales of frequency representation," http://www2.ling.su.se/staff/hartmut/bark.htm (Last viewed 5/15/2012).

Trehub, S. E. (**1976**). "The discrimination of foreign speech contrasts by infants and adults," Child Dev. **47**, 466–472.

van Hofwegen, J. (**2011**). "Apparent time evolution of /l/ in one African American community," Lang. Var. Change **22**, 373–396.

Volaitis, L. E., and Miller, J. L. (**1992**). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," J. Acoust. Soc. Am. **92**, 723–735.

Walley, A. C., and Carrell, T. D. (**1983**). "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 1011–1021.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). "Developmental aspects of crosslanguage speech perception," Child Dev. **52**, 349–355.

Werker, J. F., and Tees, R. C. (**1983**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.

Werker, J. F., and Tees, R. C. (**1984**). "Phonemic and phonetic factors in adult cross language speech perception," J. Acoust. Soc. Am. **75**, 1866–1878.

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.