PRODUCTION AND PERCEPTION OF ENGLISH /l/ AND /r/ BY 4-, 5-, and 8-YEAR-OLD CHILDREN

Kaori Idemaru and Lori L. Holt

University of Oregon and Carnegie Mellon University
idemaru@uoregon.edu and lholt@andrew.cmu.edu

ABSTRACT

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-, 5- and 8-year old native-English-speaking children. The results indicate that whereas the young children's /l/-/r/ production was well-distinguished acoustically by its primary acoustic cue (i.e., F3), they were still developing in regards to how they integrated F3 and F2 in production, and particularly in perception, relative to native adults. These data are consistent with a rather long trajectory of phonetic development whereby native categories are refined and tuned well into childhood.

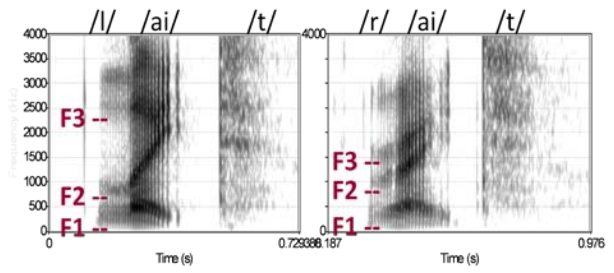
Keywords: speech perception, speech production, English, cue weighting, development

1. INTRODUCTION

In normal adult rhotic varieties of American English, approximants /l/ and /r/ (/r/ will be used throughout this paper instead of /ɹ/) in a prevocalic position are distinguished primarily by the onset frequency of the third formant (F3). Figure 1 shows the production of word initial /l/ and /r/ by an adult male speaker. Whereas F3 for /l/ is high, around 2500 Hz, F3 for /r/ is considerably lower, around 1500 Hz.

In general, speech categories are inherently multidimensional and probabilistic. Typically, no

Figure 1. Spectrograms showing the production of /lɑɪt/ (left) and /rɑɪt/ (right) by an adult male speaker.



single acoustic dimension is necessary or sufficient to define phonetic category membership. The acoustic dimensions thus co-vary in the input and differ in the degree to which they correlate with phonetic categories. Lisker (1986), for example, identified as many as 16 acoustic features that may characterize the voicing distinction in English in syllables such as /ba/ and /pa/. Whereas any of these multiple dimensions may be informative for categorization, their perceptual effectiveness varies. English listeners predominantly use VOT to distinguish /ba/ and /pa/ and use the onset F0 of the vowel as secondary information [1].

Onset F3 frequency, onset F2 frequency, and the rate of the F1 transition have been found to covary with /r/-/l/ category membership, and studies have shown that adult listeners are sensitive to these cues (Polka & Strange, 1985; Yamada & Tohkura 1992). Further, more recent studies have demonstrated that in categorizing /l/ and /r/, native adult listeners systematically use these covarying cues: they predominantly rely on the onset F3 cue, give a slight weight to the onset F2, and barely use the F1 transition. The relative weighting of these three acoustic cues was consistent across adult native listeners [4].

Redundant information conveyed by multiple acoustic dimensions is useful, perhaps necessary, in processing variable and noisy signals and in extracting phonetic information from such signals. The fact that listeners integrate multiple sources of acoustic information with consistent weighting means that children must learn to use appropriate perceptual cues and give appropriate weighting to each cue to achieve adult-like speech perception. Prior work has indicated that phonetic development, including perceptual cue weighting, continues well into childhood, and is not yet adult-like at ages of 6 to 12 [3,7]. These developmental studies have focused on fricatives and affricates. Whereas there has been substantial work on the perception of /l-r/ by adult native listeners and Japanese learners of English, there has been little work that has examined the development of /l/ and /r/, especially in terms of

F2/F3 cue integration. One of a few exceptions is McGowan et al [6], which investigated the development of /r/ production in terms of F2 and F3, reporting that it continued from the age of 15 months- to the age of 32 months. McGowan et al. suggest that in addition to F3 and F2 independently, the distance between F3 and F2 may be an important acoustic cue that characterizes /r/.

Other prior work indicates that before age 3, children start producing intended /l/s and /r/s; however, they may produce variant sounds, such as /w/ as a substitute for /r/ [2]; and F2 and F3 patterns are still developing [6]. By age 7, most of the variant productions of /r/ disappear [9]. This suggests that significant development occurs between 3 and 7 years old. Given this, the current work investigates the perception and production of /l-r/ by children who are 4, 5, and 8 years old.

2. METHODS

2.1. Participants

Participants were 48 American children growing up in a region of the East Coast where rhotic dialect is spoken. The children were divided into four age groups: Youngest (12 children; mean age = 4.16; age range = 3.95-4.37), Middle (13; 4.71; 4.42-5.04), Older (12; 5.49; 5.05-6.13), and Oldest (11; 8.45; 7.31-9.54). None of these children had been diagnosed with speech/hearing problems, had had 6 or more ear infections before their second birthday, had complications at birth, or used a foreign language on a regular basis. All 48 children completed the production tasks; 45 of them completed the perception task (2 in the Youngest and 1 in the Middle group did not). Eighteen adults (Adult) also participated in the perception task, serving as a comparison group.

2.2. Production Tasks and Measurements

Researchers explained to the children that 6 visitors from outer space, each illustrated on a picture card, wanted to be friends with them and learn some English words. The first thing to do was practice saying their names. The 6 visitors had mono-syllabic names, /li/, /lu/, /lɑ/, /ri/, /ru/, and /rɑ/. The children were recorded individually while wearing a light-weight, head-mounted microphone (Shure SM10A) connected to Maranz PMD 670 (22.05 kHz, 16 bit). One of the researchers selected a picture card with a character on it and prompted, “This is Lee. Can you say *This is Lee*?” When the child completed this repetition task one time each for all 6 characters, the picture cards were shuffled and the procedure was repeated 5 times. The last 5 repetitions were retained as data; the first cycle was discarded as practice.

In the subsequent task, children were shown one of two pictures on a monitor and were asked to say a word corresponding to the picture. One picture showed a hand writing a letter, and this picture was associated with the word “write,” /rɑɪt/. The other picture was that of a lamp and this picture was labeled “light,” /lɑɪt/. The two pictures were intermixed and were each presented 6 times. The last 5 repetitions of each word were retained.

Thus, from each child we collected 5 tokens each of /li/, /lu/, /lɑ/, /ri/, /ru/, /rɑ/, /lɑɪt/ and /rɑɪt/. By examining the spectrogram and LPC-smoothed spectrum of each of these test syllables, F1, F2 and F3 were measured at the onset of /l/s and /r/s. If three clear formant peaks were not present at the onset, the measurement location was shifted in time by 10 ms until peaks were present.

2.3. Perception Task

Stimuli used for the perception task were a subset of synthesized /lɑɪt - rɑɪt/ tokens used in [4]. In these stimuli, onset F3 values (1600, 2000, 2400, 2800 Hz) and F2 values (800, 1000, 1200, 1400 Hz) were combined for all combinations creating 16 unique stimuli. These stimuli were presented in 5 randomized orders.

Researchers explained to the children that one of the visitors, Roo, wanted to practice saying two English words, “write” and “light” and that their task was to tell the experimenter which word Roo said by pointing to the picture associated with the word.

On a computer monitor, a picture of a hand writing a letter and a picture of a lamp appeared, with left/right positioning of the pictures randomly assigned. A small image of the character, Roo, always appeared in the middle with a sound icon. When the researcher clicked the sound icon, the auditory stimulus was presented through the headphones worn by experimenter and child. The children pointed to the picture of the word that Roo produced.

3. RESULTS

3.1. Production

The mean F1 values showed no statistically significant variation relative to age-group and no correlation with the chronological age of the participants, indicating that there was no general age effect on the formant frequencies. Thus, raw values of F2 and F3 were used for the analysis. Table 1 reports mean F2, F3 and F3-F2 across age groups. These values were submitted to a separate 2 x 4 (Type: /l/ vs. /r/ x Group: Youngest, Middle, Older, Oldest) ANOVA with repeated measures on Type.

For F2, Group had a marginally significant effect, and there was a significant interaction between Group and Type [Group, $F(3, 44) = 2.666$, $p = .059$; Group*Type, $F(3, 44) = 4.357$, $p = .009$]. Post-hoc tests indicated that the F2 difference due to Type (/l/ vs. /r/) was not significant in all groups, except that for the Older group it was marginally significant [$t(11) = -2.979$, $p = .013$ (alpha adjusted to .013 for 4 comparisons)]. Thus F2 did not consistently differentiate the /l-r/ categories across these age groups.

For F3, both Type and Group were significant, but the interaction was not [Type, $F(1, 44) = 219.252$, $p = .000$; Group, $F(3, 44) = 7.285$, $p=.000$]. The main effect of Type indicates that F3 differentiated /l-r/ in all age groups. As expected, F3 was higher in /l/ than in /r/. Thus, even the youngest 4-year-olds acoustically distinguished the /l-r/ categories using F3, the primary cue for the distinction. Post-hoc tests examining the Group effect indicated that F3 of 8.5-year-olds was different from that of 4-, and 4.5-year-olds [8.5-yr-olds vs. 4-yr-olds, $p = .000$; 8.5-yr-olds vs. 4.5-yr-olds, $p = .004$]. The mean F3 values for both /l/ and /r/ categories decreased across these groups.

Table 1: Mean F2, F3 (F1-normalized) and F3-F2 across four age groups.

	F2		F3		F3-F2	
	/l/	/r/	/l/	/r/	/l/	/r/
Youngest	1711	1448	4129	2994	2418	1546
Middle	1728	1563	3967	2952	2239	1389
Older	1384	1532	3742	2816	2358	1284
Oldest	1369	1497	3673	2303	2304	806

For F3-F2, Type and Group were significant [Type, $F(1, 44) = 97.603$, $p = .000$; Group, $F(3, 44) = 4.938$, $p = .005$]. The F3-F2 value was greater for /l/s than for /r/s as expected; and there was an age difference in that F3-F2 was smaller in the Oldest than in the Youngest group [$p = .002$]. A developmental trend was expected such that F3-F2 for /r/ would be smaller in the older groups, but this was not confirmed.

3.2. Perception

Figure 2 illustrates a sample of the perception data (Youngest, Oldest and Adult): percent /r/ responses are plotted across the test F3 values (the X-axis) and the test F2 values (lines). The identification function appeared steeper in the older groups (Oldest and Adult). It was also noted that for the Adult group the lines were separated when F3 was 2000 Hz, suggesting a possible effect of F2. To examine these observations, mean percent /r/ responses were submitted to separate 4 x 4 (F3 vs. F2) repeated-measures ANOVAs for each age group. The results are reported in Table 2.

Pair-wise post-hoc comparisons found that, for the Youngest, any differences in F3 steps had a reliable effect on percent /r/ responses, except for the steps from 1600 to 2000 Hz and from 2400 to 2800 Hz ($ps < .008$, alpha adjusted to .008 for multiple comparisons). Their categorization of the end stimuli were not robust (80% and 11% /r/ for 1600 Hz and 2800Hz). For the Middle and Older groups, any differences in F3 steps had a reliable effect on percent /r/ responses, except for the step from 2400 to 2800 Hz ($ps < .008$, alpha = .008). Thus, for these groups, F3 changes from 1600 to 2000 Hz and 2000 to 2400 Hz influenced perception of /l/ and /r/. This trend was not observed in the Oldest group. For this group, an F3 step from 1600 to 2000 Hz, and another from 2400 to 2800 Hz did not make a reliable difference ($ps < .008$). Their categorization of the end stimuli was robust (99% and 4% /r/ for 1600 Hz and 2800Hz). Thus, for the oldest group of children, the F3 values were categorized to either /r/ (1600 and 2000 Hz) or /l/ (2400 and 2800 Hz) fairly consistently, suggesting more categorical perception of this continuum. On the other hand, the 3 younger groups were not sure about the /r/-ness of the sounds when F3 was one of the middle values, i.e., 2000 Hz.

The Adult group showed yet a different pattern from the Oldest, reflected by a significant F2 x F3 interaction. First, as in the children, F3 exerted a strong influence on /l-r/ categorization. F3 caused a significant change in perception except for the step from 2400 to 2800 Hz for the lower values of F2 (800 and 1000 Hz). For the 2 higher values of F2 (1200 and 1400 Hz), the peripheral steps of F3 (from 1600 to 2000 Hz, from 2400 to 2800 Hz) did not make a significant difference ($ps < .002$, alpha adjusted to .002 for multiple comparisons). Thus in general, as with the children, F3 had a strong influence on /l-r/ perception, and the response pattern was more categorical in that both 2400 and 2800 Hz were equally associated with the /l/ category.

Most notably, F2 had an effect on adults' perception of /l/ and /r/, whereas it had no effect for children. This F2 effect was localized to a specific F3 value. F2 exerted its influence only when F3 was ambiguous (= 2000 Hz), such that higher values of F2 elicited reliably more /r/ responses (ps < .002).

Figure 2: Mean percent /r/ responses across F3 (the X-axis) and F2 (lines) for Age Groups 1, 4, and Adults. F3 and F2 values are in Hz.

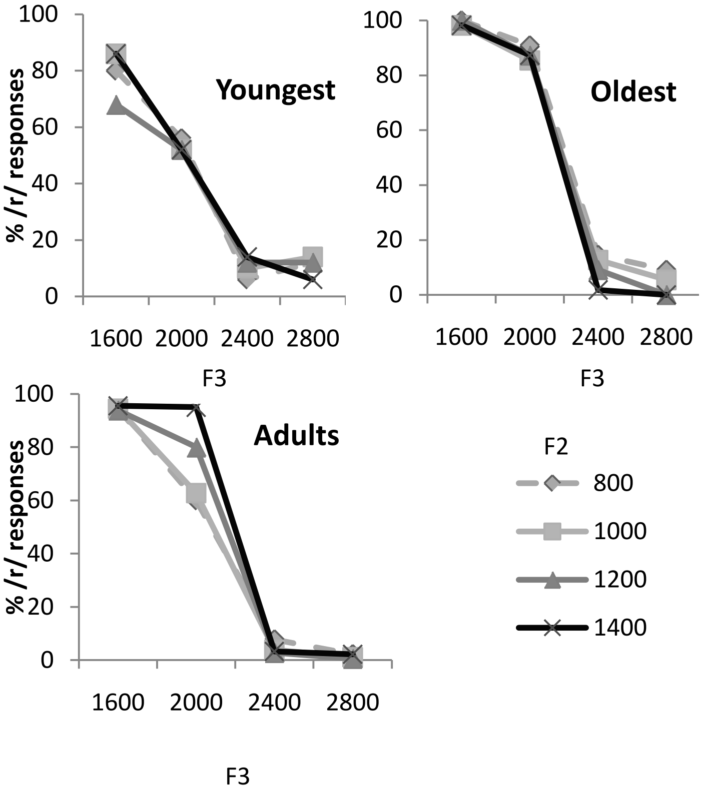


Table 2: Results of ANOVA. Only significant results are reported.

	Source	df 1	df 2	F	p
Age 1	F3	3	27	30.143	.000
Age 2	F3	3	33	20.778	.000
Age 3	F3	3	33	95.262	.000
Age 4	F3	3	30	249.344	.000
Adults	F3	3	51	215.450	.000
	F2	3	51	6.201	.001
	F3*F2	9	153	13.287	.000

4. DISCUSSION

The acoustic examinations reported here have revealed that children's /l/ and /r/ categories are still developing at age 8 and possibly continue to do so in older children. Whereas the primary acoustic cue, F3, distinguished /l/ and /r/ productions in our children's speech, the same was not true for the secondary acoustic cue, F2. Unlike adult speech reported in previous work (Ingvalson et al, in press), F2 did not reliably distinguish /l-r/ categories in the speech of the children tested here. Also, the developmental trend of F3 lowering for /r/ productions was observed between children of 5 and 8 years of age. It is possible that this trend continues even further, in older children. Furthermore, lack of use of F2 by children was evident also in perception. Whereas there was a clear developmental trend of /l/ and /r/ categories being better defined along the F3 dimension, reflected in more categorical-like response patterns in older children, none of the child groups showed the use of F2 as a secondary cue as the adult group did. Our findings here indicate that the development of F2 as a secondary cue to /l-r/ categorization lags behind the development of F3, and establishing adult-like cue weighting of F3 and F2 must occur later in childhood.

5. REFERENCES

[1] Francis, A. L., Kaganovich, N. & Driscoll-Huber, C. J. 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. J. Acoust. Soc. Am. 124, 1234-1251.

[2] Grunwell, P. 1982. *Clinical phonology*. Aspen Publishers.

[3] Hazan, V., & Barrett, **S.** 2000. The development of phonemic categorization in children aged 6-12. *Journal of Phonetics*, 28(4), 377–396.

[4] Ingvalson, E. M., McClelland, J. M., & Holt, L. L. In press. Predicting Native English-Like Performance by Native Japanese Speakers.

[5] Lisker, L. 1986. "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and speech*, 29 (1), 3.

[6] McGowan, R; Nittrouer, S; Manning, C. 2004. Development of (r) in young, Midwestern, American children. *J. Acous. Soc. Am.* 115, 871-884.

[7] Nittrouer, S., Manning, C., & Meyer, G. 1993. The perceptual weighting of acoustic cues changes with linguistic experience. *J. Acous. Soc. Am.* 94, S1865.

[8] Polka, L., & Strange, W. 1985. Perceptual equivalence of acoustic cues that differentiate /r/ and/l/. *J. Acous. Soc. Am.* 78, 1187-1197.

[9] Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., Bird, A. 1990. The Iowa articulation norms and its Nebraska replication. *J. Speech Hear. Disord*. 55, 779-798.

[10] Yamada, R. A., & Tohkura, Y. 1992. The effects of experimental variables on the perception of American