Published in final edited form as:

Clin Linguist Phon. 2019 ; 33(4): 295–315. doi:10.1080/02699206.2018.1477834.

Modeling category goodness judgments in children with residual sound errors

Sarah Hamilton Dugan¹, Noah Silbert¹, Tara McAllister², Jonathan L. Preston³, Carolyn Sotito¹, Suzanne E. Boyce¹

¹University of Cincinnati, USA;

²New York University, USA;

³Syracuse University, USA

Abstract

This study investigates category goodness judgments of /r/ in adults and children with and without residual speech errors (RSE) using natural speech stimuli. Thirty adults, 38 children with RSE (ages 7–16), and 35 age-matched typically-developing (TD) children provided category goodness judgments on whole words, recorded from 27 child speakers, with /r/ in various phonetic environments. The salient acoustic property of /r/—the lowered third formant (F3)—was normalized in two ways. A logistic mixed effects model quantified the relationships between listeners' responses and the third formant frequency, vowel context, and clinical group status. Goodness judgments from the adult group showed a statistically significant interaction with the F3 parameter when compared to both child groups ($p < 0.001$) using both normalization methods. The RSE group did not differ significantly from the TD group in judgments of /r/. All listeners were significantly more likely to judge /r/ as correct in a front vowel context. Our results suggest that normalized /r/ F3 is a statistically significant predictor of category goodness judgments for both adults and children, but children do not appear to make adult-like judgments. Category goodness judgments do not have a clear relationship with /r/ production abilities in children with RSE. These findings may have implications for clinical activities that include category goodness judgments in natural speech, especially for recorded productions.

Keywords

speech perception; speech disorders; speech acoustics

Approaches to Studying Perception and Speech Errors

There are clearly relationships between children's ability to perceive and produce the speech sounds of their language. Accordingly, clinical approaches to intervention in children with speech sound disorders include both auditory/perceptual and articulatory/motor strategies. However, the literature investigating speech perception in children with speech sound

disorders has shown a surprising lack of consensus. Some studies have found that children with speech sound disorders do not have apparent perceptual deficits in phonemic identification or discrimination when compared to typically-developing peers (Broen, Strange, Doyle, & Heller, 1983; Hoffman, Stager, & Daniloff, 1983; Locke, 1980; Saben & Ingham, 1991; Smit & Bernthal, 1983; Tyler, Edwards, & Saxman, 1990). Other experiments using more fine-grained perceptual measures have found that some children with speech errors do appear to have atypical auditory perception (Aungst & Frick, 1964; Monnin & Huntington, 1974; Ohde & Sharf, 1988; Rvachew, 1994; Sharf & Benson, 1983; Shuster, 1998). It is particularly useful to test perception with stimuli from a natural speech distribution of speech sounds varying in category goodness, because this situation allows for better fidelity to the real-world listening environment inside the clinic and elsewhere. In addition, when testing category judgments, children should judge stimuli that reflect their errors (Rvachew, 1994). Children with residual sound errors (RSE) produce errors with variable acoustic characteristics (Shriberg, Flipson, Karlsson, & McSweeney, 2001), suggesting the need to test perception with acoustically variable stimuli. Such distributions may also be a better complement to the articulatory-acoustic characteristics of productions from children with RSE.

Auditory perceptual differences have been observed in children with RSE, whose production errors on late-developing sounds persist past the typical age of phoneme mastery of 8 or 9 years of age (Shriberg, Kwiatkowski, & Gruber, 1994; Shriberg, Austin, Lewis, McSweeney, & Wilson, 1997). Children with RSE show variable impairments in speech perception (Broen et al., 1983; Hoffman, Daniloff, Bengoa, & Schuckers, 1985; Rvachew & Jamieson, 1989) and show neurobiological differences in brain activation during tasks that required auditory perception and judgment (Preston et al., 2012; Cabbage, Hogan, & Carrell, 2016). There is evidence that children with RSE may have increased gray matter in bilateral supratemporal gyrus and supramarginal gyrus (regions involved in integration of auditory and somatosensory information), which may contribute to difficulty with fine-grained phonetic perception and online error detection during phonetic learning (Preston et al., 2014).

Most of the evidence noted above concerning children's perceptual discrimination skills comes from classic categorical perception tests using a continuum between two phonemically distinctive sounds (i.e. "bat" vs. "pat" or "ray" vs. "way"). These have been investigated in different ages and in children with different types of speech sound disorders, including RSE. The evidence from classic studies shows that, even when children are judged to have acquired the ability to perceive a phonemic contrast, refinements in speech perception appear to continue as a function of advancing age (Nitttrouer & Miller, 1997; Walley & Flege, 1999). For instance, Hazan & Barrett (2000) reported that children's responses to a continuum of phonetic contrasts showed increasingly steeper identification functions between 6 and 12 years old. Adults show considerable individual variation in category judgments (Massaro & Cohen, 1983); but Hazan & Barrett (2000) nevertheless found substantial differences in the performance of the adults and the 12 year olds in their study, with adults showing significantly more consistent patterns of phoneme categorization.

Children with speech sound disorders need to be able to identify acceptable and unacceptable productions in order to self-monitor and self-correct their speech (Koegel, Koegel, & Ingham, 1986; Preston, Irwin, & Turcios, 2015). Accordingly, clinical intervention frequently includes activities meant to foster the ability to make *category goodness judgments*; in these, listeners are expected to indicate the accuracy or goodness of the production (i.e. “Was that a good “r” sound in the word?”). Thus, a valid perception task incorporating category goodness would be clinically useful.

Building a Distribution of /r/ “Goodness”

Of the speech sound errors typical of RSE children, the most ubiquitous and frustrating to clinicians is /r/ (Ruscello, 1995). Given the evidence for perception/production links, the developmental trajectory of typically-developing (TD) children’s ability to categorize phonemes, and the possibility of divergence between RSE and TD children, we examined /r/ category goodness judgments in RSE children versus those of adults and TD children.

The most prominent acoustic attribute of /r/ is the lowering of the third formant (F3) (Lehiste, 1964; Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000), which occurs only for /r/ in American English. It has been common in the literature on typical speech to use the narrowing of the difference between F3 and second formant (F2) values as a way to differentiate /r/ from other sonorant consonants and vowels (Stevens, 1999). The F3-F2 narrowing has also been used in the clinical literature for distinguishing accurate from inaccurate productions of /r/. Several studies using this measure have found that /r/ is more likely to be judged as accurate when F3-F2 values are small and to be judged as inaccurate when F3-F2 values are large (Sharf & Ohde, 1983; Flipsen, 2001; Shriberg et al., 2001; Wolfe, Martin, Borton, & Youngblood, 2003; Klein, Grigos, McAllister Byun, & Davidson, 2012; McAllister Byun & Tiede, 2017; Campbell, Harel, Hitchcock, & McAllister Byun, 2017). The difference between F3 and F2 has also been used as a visual marker for spectrographic feedback in clinical interventions (Shuster, Ruscello, & Toth, 1995; McAllister Byun, 2017). For perceptual research using multiple speakers, the F3-F2 value is also important as a form of normalization for individual differences related to vocal tract size (Flipsen et al., 2001). Because both F3 and F2 are influenced by the size of the speaker’s vocal tract, the raw F3-F2 difference cancels out this influence.

An alternative method of normalizing formant measures is a ratio described by Hagiwara (1995): For adult males and females, the F3 of /r/ lowers at least to 80% of a speaker’s average F3 for his/her non-rhotic vowels (although for most speakers, it was 60–70%). There may be advantages to using this ratio measure over the F3-F2 difference measure when predicting goodness judgments. The “near-merging” of F3 and F2 may not be as distinct in children, who have a smaller vocal tract and thus wider-spaced harmonics. In addition, the F3-F2 difference may be rendered less interpretable by the influence of vowel context on the second formant. For example, there is a smaller F3-F2 distance for rhotics in a prevocalic front vowel context (/i/) than a prevocalic back vowel context (/ɑ/), as the second formant is higher (and thus closer to F3) for /i/ (Flipsen, 2001). Vowel context is a largely unexamined factor in goodness judgments. Vowel context can have an effect on rhotic production (Mielke, Baker, & Archangeli, 2010; Stavness, Gick, Derrick, & Fels, 2012), but less is

known about vowel context effects on rhotic perception. It is clinically important to investigate how this and other factors influence listeners' judgments of rhotic errors, especially in populations such as children with RSE, who may make errors with varying acoustic characteristics across vowel contexts.

Purpose.

In this study, we investigate two research questions: (1) can adults' goodness judgments along a distribution of naturally produced /r/ in words be predicted by acoustic measures of rhoticity (F3 ratio or F3-F2) and vowel context and (2) what differences in goodness judgments exist in adults, TD children, and children with RSE?

Study 1: Predicting Goodness Ratings with Acoustic Measures of Rhoticity from a Distribution of Children's Natural /r/ Productions

Method for Study 1

Participants.—Listeners were 30 native speakers of American English between the ages of 18–50 with no self-reported history of speech, language, or hearing problems. Participants were recruited primarily from the university community. Fifteen of the 30 listeners were clinically certified speech-language pathologists. All of the listeners were enrolled in or teaching courses in the Communication Sciences and Disorders department at the University of Cincinnati, where they had their hearing screened as part of instructional activities. One listener was raised as a bilingual Spanish-English speaker, but all other listeners were monolingual.

Stimuli.—Stimuli were selected from recordings of a word probe list repeated by 27 children with RSE (recordings were collected as part of ultrasound biofeedback therapy sessions for /r/ at the University of Cincinnati, Syracuse University, and Haskins Laboratories in New Haven, Connecticut). The age spread of child participants was 8 years (oldest child was 16 years old and youngest was 8 years old). Stimuli were whole words and were selected for acoustic quality (minimal background noise, no interruptions from clinician). Phonological attributes of the stimuli are shown in Table 1 (see Appendix for word list). For descriptive purposes, a clinically-trained listener reviewed all stimuli and, in a four-way category judgment task, determined if the /r/ in each word was best categorized as correct, a substitution, an omission¹, or a distortion.

Trained research assistants measured F3 for the /r/ in every word as well as the point vowels (/i/, /æ/, /u/, and /ɑ/) from each speaker using the acoustical analysis program Praat (Boersma & Weenink, 2016). In order to determine the average F3 of a speaker's vowels, the midpoint of the segment was identified and a 20 ms sample was measured. The sampling frequency was 32000 Hz, and spectrogram frequencies up to 5000 Hz were analysed. For the /r/ F3, a 20 ms sample was measured at the point at which F3 was lowest (Lehiste, 1964)². In some of the stimuli, the F3 of /r/ was not much lower than that of its neighboring

¹Although "omission" was used as a perceptual category, there was acoustic evidence (changes in formant frequencies) for the /r/ segment in all stimuli.

vowel segment. In this case, if the segment was postvocalic, measurement proceeded from the midpoint of the segment, as identified by the beginning and end of the steady-state portion of the segment (excluding transitions). If the segment was prevocalic, the 20 ms sample typically included the transition, and began at the onset of periodicity and ended at the beginning of the steady-state vowel following the /r/.

For presentation to listeners, all stimuli began and ended with 100 ms of silence. Stimuli were equalized to have an average amplitude of 30 (+/-1) dB using the Wavesurfer program (Beskow & Sjölander, 2005).

All stimuli were measured by two different student research assistants. Formant measurements with more than 100Hz discrepancy between researchers were re-measured until consensus was achieved. Approximately 20% of stimuli were re-measured for consensus. Research assistants were blind to the ratings of the clinically trained listener.

Study design.—All adult participants listened to the same set of 159 stimuli over a pair of Sennheiser HD 280 Pro headphones using the Superlab 5.0 stimuli presentation program (Abboud, 2017). Before the experiment, practice trials were administered, which followed the same layout as the experiment, but provided feedback on accuracy of choices in order to familiarize participants with the structure of the experiment. To approximate a natural listening environment situation, stimuli were presented to each listener only once and participants could not replay any stimuli. Stimuli were presented in random order. Participants controlled the pace of experiment by pressing a key when they were ready to hear another stimulus. In a binary forced-choice task, each participant was instructed to press the “y” key if they heard an /r/ in a word that sounded correct to them, and to press “n” if they heard an /r/ that was distorted, substituted, or omitted. Key press instructions remained on the computer screen during each stimulus presentation. Stimulus words were not presented orthographically. Participants heard words in blocks of 50 and were offered a break after each block.

Acoustic characteristics of stimuli.—For each speaker, the F3 of his or her point vowels (/i/, /æ/, /u/, /ɑ/) were used to calculate an average F3. This average was used as input for the normalization procedure (Hagiwara, 1995), calculated as the ratio of the measured F3 (the measured F3 is “F3_m”) in any given production of /r/ to the average F3 of that speaker’s vowels (see Equation 1).

$$\log\left(\frac{F3_m}{F3_{avg}}\right) = F3 \text{ ratio} \quad \text{Equation 1. Equation for normalization}$$

The log of this ratio was taken in order to account for the auditory perceptual response to spectral characteristics of speech signals (Gray & Markel, 1976), and to reduce skewness in the raw ratio values. Because of the log transformation, negative F3 ratios corresponded to

²Due to the shorter duration of most prevocalic /r/ productions, the entire acoustic record was often 20 ms, and this was used as the measurement for prevocalic /r/ tokens.

lower than average F3s and positive ratios to higher than average F3s. The resulting measure (henceforth “F3 ratio”) was used as an independent variable for this investigation. In reference to the acoustic cutoff described by Hagiwara (1995), 50 of the 159 productions were below 80% (“good /r/”) and 109 of the 159 were above 80% (“poor /r/”). It should be stressed that although tokens can be binarized relative to this cutoff, the stimuli actually occupied a range of F3 ratio values. It was anticipated that responses to productions near the 80% “cutoff” value would be more variable across listeners. F3 ratios were transformed to z-scores by computing the mean and standard deviation across each speaker’s data separately.

F3-F2 difference was also calculated so that it could be compared to F3 ratio as a predictor of perceptual ratings. The F3-F2 difference values were also transformed to z-scores by computing the mean and standard deviation across each speaker’s data separately. It was hypothesized that the listeners’ correctness judgments would be influenced by the third formant, with larger F3 ratios or F3-F2 differences corresponding to a greater proportion of stimuli rated as “poor /r/”.

To examine the effect of vowel context on listeners’ responses, vowel context was coded in three categories: front (when /r/ occurred before or after a front vowel, e.g. /ir/ “ear” /rik/ “reek”, /rek/ “rake”, /er/ “air”), mid-central vowel (e.g., /hɜ:/ “her”), or back (e.g., when /r/ occurred before or after a back vowel, e.g. /ɑr/ “are” /rɔk/ “rock”, /ru/ “roo” and /or/ “or”). Mid central vowels were represented by only one phoneme: /ɜ:/ as in “her” /hɜ:/. There were no prevocalic mid-central vowels in the stimuli. Front vowels were /i/, /e/, or /æ/.

Results for Study 1

Statistical analyses were carried out with the statistical software R using the mixed-effects generalized linear models package “lme4” (Bates, Maechler, Bolker, & Walker, 2017). Two multilevel logistic regressions were fit with either the *F3 ratio* or the *F3-F2 difference* as the independent variable (continuous), and listeners’ responses to stimuli (binary) as the dependent variable. Rejections (responses of “no, not a good /r/”) were coded as “1” and acceptance (responses of “yes, a good /r/”) were coded as “0” because we predicted that higher log ratios would correspond to more frequent rejections. Vowel context was included as a fixed effect in the models, with back vowels as the reference level. The model included by-subject random intercepts with random slopes by the acoustic parameter (F3 ratio or F3-F2), as well as by-token random intercepts. “Token” is defined as the production of a given word by a given speaker. Coefficients from two separate logistic regressions are presented in Table 2 below, with one model including the fixed effect of the independent variable F3 ratio and one with F3-F2 difference.

As the results in Table 2 indicate, both acoustic parameters were found to have a significant effect on listeners’ ratings of category goodness. To choose between these parameters, models incorporating one or the other parameter were submitted to a Receiver Operating Characteristic (ROC) Curve analysis using the pROC package in R (Robin et al., 2011). As illustrated in Figure 1, the ROC curve demonstrates the tradeoff between sensitivity (y-axis) and 1-specificity (x-axis) for each model. The overall classification accuracy of a model (with respect to the listeners’ responses) is related to the area under the curve (AUC). The closer the curve is to the 45 degree diagonal, the less accurate the model is. The AUC shows

that the F3-F2 test more accurately predicts subject responses, with area of .79 (95% CI: 0.77–0.80). The F3 ratio has a similar but slightly lower prediction accuracy of .74 (95% CI: 0.72–0.76). The confidence intervals (CI) for F3-F2 AUC and F3 ratio AUC do not overlap, indicating that the F3-F2 AUC is statistically significantly larger than the upper CI limit for the F3 ratio AUC, with a stronger magnitude of effect when compared to the F3 ratio.

There was considerable variation among listener responses³, similar to previous perceptual studies of adults rating children's /r/ productions (Klein et al., 2012). Figure 2 is a violin plot displaying responses from all listeners, where stimuli have been binned into quintiles defined by F3 ratio and F3-F2 differences. Quintiles on the left correspond to tokens with lower F3-F2 or F3 ratios, which were predicted to have fewer rejections from listeners. The y-axis shows proportion of rejections for adult listeners. The width of the violin plots correspond to the kernel density estimates of the response proportions, visually describing the distribution of the data. For example, quintile 1 has a wider base in the plot depicting F3-F2 than in the plot for F3 ratio, indicating that there were more people responding “no” to tokens in the first quintile as defined by the F3-F2 parameter. Each violin plot also contains boxplots within it, where the boxes indicate the interquartile range (IQR), the horizontal lines indicate the median “no” response rate, and the whiskers show 1.5*IQR above and below the IQR. Individual dots indicate outliers (i.e., data points beyond the 1.5*IQR limits).

Figure 2 shows that a lower proportion of “no” responses was observed across all listeners in the quintile with the smallest F3-F2 distance or F3 ratio. As the F3 ratio increases in quintiles 2 and 3, there is an abrupt increase in proportion of /r/ rejections. However, variation across listeners can also be seen to rise in quintiles 2 and 3 for both acoustic parameters.

The results in Table 2 indicate that vowel context also had a statistically significant influence on listener ratings. Listeners were more likely to accept /r/ productions when they occurred in a front vowel context when compared to other vowel contexts (mid-central, back).

Discussion for Study 1

Both acoustic parameters, F3 ratio and F3-F2 difference, were found to significantly predict adult listeners' ratings of category goodness. As predicted, as these two parameters increased, the probability of listeners rejecting a token as an instance of correct /r/ increased.

Based on the results of the ROC analysis (Figure 1), F3-F2 difference was found to be a better model fit than F3 ratio for participants' category goodness judgments. Therefore, F3-F2 will be used as the primary acoustic parameter in the second study, which compares the category goodness judgments of TD children and children with RSE to the adult responses to the same stimuli. In light of the finding that vowel context was a significant predictor of adults' perceptual ratings, the factor is also included in models comparing child listeners.

³There was no significant effect from clinical experience on adult category goodness judgments, as determined in a preliminary analysis that included clinical experience as a predictor variable (Estimate=0.254, Std. Error=0.265, $p>.338$) along with the acoustic parameter. Therefore, all subsequent analyses did not include clinical experience as a factor.

Study 2: What are the Differences in Perceptual Performance Between Adults, Typically-Developing (TD) Children, and Children with RSE?

The goal of the second study was to assess whether children with RSEs respond differently to a distribution of natural /r/ productions when compared to TD children and adults. This aim was motivated in part by evidence that children may pay attention to different information in the speech signal than adults. In a study of fricative identification, Nittrouer, Studdert-Kennedy, and McGowan (1989) found that younger children identified initial fricatives based on syllable information carried by the following vowel, rather than spectral information inherent to the fricative itself, though older children (above age 7) relied on the fricative more than the vowel. It is possible that children with RSE, who are above age 7, will show adult-like sensitivity to frequency information contained within rhotic segments; on the other hand, it is possible that their speech sound disorder will interact with their ability to make adult-like goodness judgments.

There is an exigency for understanding why some children with RSEs make little progress in speech therapy. Of the potential psycholinguistic causes that have been considered, poor auditory perception has particular clinical relevance, as many treatments for speech sound disorders include auditory-based activities, such as “auditory bombardment,” imitating clinician models, or performing category goodness judgments. In addition, the establishment of correct production may ultimately rely on children’s ability to engage in auditory self-monitoring of their speech (Van Riper, 1978). Understanding the extent to which children with RSEs have impaired speech perception may help clinicians develop appropriate intervention plans. We hypothesized that children with RSE and TD children would show distinct perceptual responses to the different acoustic characteristics of the rhotic stimuli tested in Study 1, given that children with RSE have been shown to have phonological processing deficits when compared to TD groups (Preston & Edwards, 2007).

Method for Study 2

Participants.—Participants were children with RSEs (n=38) and age-matched TD children. Child participants with RSE were not the same children who provided the speech tokens under analysis. Children were recruited in three ways: (1) through their participation in a grant to study ultrasound therapy outcomes for children with RSE at Syracuse University, Haskins Laboratories, and University of Cincinnati; (2) through their participation in a university ultrasound clinic for RSE at University of Cincinnati; (3) through word of mouth (for TD children). For inclusion in the experiment, all children passed a hearing screening from 500 to 4000 Hz at 20 dB. An informal articulation assessment was used to determine if the child had typical articulation abilities. One TD child unexpectedly had residual /r/ errors, so her data was assigned to the RSE category in all analyses. Demographics of children with RSE are presented in Table 3.

Children with RSE were diagnosed by a deep articulation assessment in which all phonetic contexts for /r/ are probed. Children lived in a region of the US where a rhotic dialect was spoken. Children’s distortions were qualitatively different in their sound, consistent with observations from Shriberg et al. (2001). For example, some children’s distortions sounded

more like a schwa, where others sounded somewhat like a rounded, dark /ɪ/. TD children were matched within one year to the ages of children with RSE. Children were not matched for sex, as to the best of our knowledge, sex of listener has not been noted to have an effect on children's category goodness judgments.

Stimuli.—To limit the task demands for children, a subset of 98 stimulus words was chosen from the adult experiment (Study 1) and session breaks were included after every 25 stimulus presentations. All acoustic parameters (e.g., duration, amplitude of files, method of z-scoring) were consistent between Study 1 and Study 2. In reference to the Hagiwara range (“good” /r/ occurs when the ratio of the F3 of a production to the speaker's average vowel F3 is less than .80), 40 of the 98 productions were below the 80% threshold (“good”) and 58 of the 98 were above 80% (“incorrect”). Children in Study 2 heard a more balanced representation of “good” and “incorrect” /r/ productions than listeners in Study 1 (who heard 50 out of 159 tokens that fit the acoustic criteria for being correct).

To explore the effects of vowel context, all tokens were coded with a three-way distinction between back vowels, mid central vowels, and front vowels, as in Study 1. Table 4 presents stimuli characteristics for Study 2.

Study design.—Study 2 used the same practice trials as Study 1, and children were instructed prior to beginning the experiment as to its purpose. Children were also asked comprehension questions by the researcher before beginning the experiment, such as “Do you know what a bad “r” could sound like?” and “If you hear an “r” sound that you think is right, what should you do on the computer?” For the RSE group, all children had an understanding of the purpose of category goodness judgments. For TD children, none of whom had attended speech therapy, some instruction about misarticulated speech sounds was needed. Misarticulations were modelled by the experimenter and TD children were given the opportunity to practice judging the accuracy of the experimenter's productions. Stimulus presentation, keyboard responses, and advancement between stimuli during the experiment were identical to Study 1.

Statistical analysis.—As stimuli for Study 2 were a subset of those used in Study 1, stimulus measurement and normalization were identical. Given that the F3-F2 acoustic parameter was determined to be the best fit to the goodness ratings from adults in Study 1, this variable is used as the acoustic independent variable for analyses in Study 2.

Results for Study 2

For the purposes of comparison with children's responses, we reduced the adult rating data sets obtained in Study 1 to the subset of the tokens judged by children. Statistical analysis took the form of a multilevel logistic regression with listeners' binary responses to stimuli as the dependent variable. Independent variables included F3-F2 difference (continuous), vowel context (reference level = back vowel), and group (reference level = TD children). The interaction of F3-F2 distance and group was also examined. The random effects structure was the same as in the models in Study I. Table 5 presents coefficients for all fixed effects in the multilevel logistic regression.

Table 5 shows a statistically significant effect of F3-F2 on listeners' response patterns. As in Study 1, listeners were more likely to reject tokens associated with larger F3-F2 distances. There was also a significant effect of vowel; listeners were significantly more likely to accept productions of /r/ in front vowel contexts as compared to mid-central or back vowel contexts. There was no significant main effect of group, but there was a significant interaction of F3-F2 distance with group. Inspection of the terms of the interaction suggested that the probability of rejecting a token was more strongly associated with F3-F2 distance for adults than for either child group. The interaction did not indicate a difference between RSE and TD children in their response to the acoustic characteristics of /r/. Violin plots in Figure 3 show the distribution of each group's responses to /r/ in stimuli that have again been divided into quintiles on the basis of F3-F2 distance. As in Study 1, the leftmost quintile corresponds to the smallest F3-F2 values, which are expected to be least likely to be rejected. The y-axis corresponds to proportion of rejections from the adult group, the TD group, or the group of children with RSE.

Figure 3 shows an increase in variation in response across all listener groups in quintiles 2 and 3, as the distance between F3 and F2 increases. The medians in each quintile are better-separated in the adult data than in either child group, consistent with the significant interaction between F3-F2 distance and group. Median responses were similar between the two child groups across all quintiles, corresponding with the finding that those two levels of the factor of group were not significantly differentiated in the interaction with F3-F2 distance.

Figure 4 shows a graphical representation of the observed interaction effect of group and the F3-F2 acoustic parameter on the identification contour of responses for adults, children with RSE, and TD children. The x-axis (F3-F2(z)) represents the distribution of r-word stimuli arranged according to the distance of the F3 and F2 (z-scored). The y-axis represents the predicted probability of respondents in that group rejecting stimuli along the /r/ distribution. Steeper slopes indicate sharper transitions from correct to incorrect /r/ for all listeners in the adult (solid line) group. The slopes for children with RSE and TD children overlap to the point that there are no visibly distinct lines between those groups, indicating very similar patterns of response to stimuli as they change in F3-F2 distance. Children in both groups had less sharp boundaries for correct and incorrect /r/ productions when compared to adults.

Discussion for Study 2

We hypothesized that children with RSE would have different category goodness judgments than age-matched TD children. These differences could impact their abilities to detect errors in natural speech. However, there was no significant difference between children with and without RSE. The F3-F2 difference was a statistically significant predictor of category goodness judgments for all listeners, suggesting that this acoustic parameter influences individuals in rating the goodness of /r/ tokens. While it is unexpected that children from RSE and TD groups did not differ in their response to the tokens (or in how the acoustics of the target influenced their responses), it may be the case that auditory perceptual differences, which are common in younger children with speech sound disorder, are uncommon in older children with RSE, who may have a more mature phonological processing system, but

impairments in articulatory precision. Younger children may have speech errors due to being in earlier stages of phonological processing development, and their errors may reflect difficulties in cognitive-linguistic aspects of speech processing (Grunwell, 1988; Smit & Bernthal, 1983). Articulatory distortions consistent with RSEs, however, appear to reflect difficulties in the sensory-motor aspects of articulatory precision (Shriberg, 1993; Redle et al., 2015).

There may be additional variables mediating differences in response pattern across subjects, such as auditory short-term memory. In an intraphonemic identification task, Sharf, Ohde, and Lehman (1988) showed that auditory short-term memory was important for discriminating among similar sounding rhotics. Auditory short-term memory was not probed in this study, but it is possible that further research could show a relationship between this property and rhotic category goodness judgments in children. In addition, there may be task familiarity effects present in the data. Children with RSE in the sample had received speech therapy for at least one year prior to enrolment. Clinicians often encourage children to self-monitor their productions during therapy (Gray & Shelton, 1992) as it has been shown to improve generalization (Shriberg & Kwiatkowski, 1987). Children with RSE in this study likely practiced judging speech productions in previous therapy, but TD children probably had no such explicit practice experience, possibly leading to similarities in performance between the RSE and TD child groups.

The results of this study suggest that perceptual abilities in RSE and TD children, as assessed by category goodness judgments of natural speech words, may be significantly different from adults. Children appear to accept a wider range of /r/ tokens as acceptable than adults, which may be a result of a broader phonetic representation for the sound (Rvachew, 1989; Shuster, 1998). Many previous studies of child perception focus on segments that children produce at very early ages, such as /t/ or /p/ (e.g. Treiman & Breaux, 1982; Zlatin & Koenigsnecht, 1976). As obstruents, these segments have steep identification contours, corresponding to a very small region of acoustic space where listeners perceive these sounds as ambiguous. However, for liquids like /r/, which have shallower identification functions and therefore a wider ambiguous region of acoustic space, the developmental trajectory of perception may be protracted, requiring more listening experience in order to focus on the relevant cues for distinguishing accurate /r/ productions. Perception abilities, in other words, may lag behind production abilities, given that even TD children who produce accurate /r/ sounds were found to differ from adults in their perceptual response to rhotic stimuli.

Vowel context had a statistically significant effect on listener judgments across all groups, with listeners accepting more rhotics that were in a front vowel context. This paper is, to our understanding, the first to consider influences from neighbouring segments on children's goodness judgments. Since F3-F2 distance is so clearly affected by vowel context, we hypothesized that neighbouring segments may contribute an effect on judgments of category goodness. This is an especially intriguing speculation, given that children have been observed to produce /r/ by raising the second formant to approximate the third (McGowan, Nittrouer, and Manning, 2004). This articulatory phenomenon may have an accompanying perceptual effect. While this study did not attempt to determine whether the way child

listeners produced rhotics had a correlation with their perception of them, it would make for an interesting follow-up study.

General Discussion

Category goodness ratings of natural speech stimuli are associated with F3-F2 distance, but children and adults show different degrees of sensitivity to this acoustic parameter. Children, both with and without RSE, had a weaker association between their perceptual ratings and F3-F2 in /r/ when compared to adults. Based on previous evidence that children with RSE tend to have perception and processing deficits, it was hypothesized that TD children and children with RSE would differ in their pattern of response to naturally produced rhotic stimuli, but this prediction was not upheld. However, this study did show that children with and without RSE exhibit a wide range of variability when judging the goodness of /r/ in naturally spoken words.

This study is one of the first to explore how rhotic perception is affected by vowel context, showing that listeners are more likely to accept rhotics as correct when they occur in a front vowel context. This finding must be interpreted with caution, given that other attributes of the natural speech stimuli may also have an effect on category goodness judgments, such as syllable position of /r/ (Klein et al., 2012; Rvachew & Andrews, 2002) or talker identity (Bradlow, Toretta, & Pisoni, 1996). Understanding the effects of contextual factors on error identification, however, is important for clinical training and evaluation. Although vowel context effects on production are well documented, there is very little research on how vowel context affects perception in children, limiting the evidence available to guide the formation of perceptual tests.

Reliably judging category goodness may be a developing skill that only reaches proficiency later in an individual's development. Phonemic discrimination or identification tasks, in contrast with category goodness judgments, may more readily identify group differences in children. Clinicians must be cognizant that children may not have auditory perceptual abilities that are commensurate with their production abilities.

Caveats and Limitations

Some factors limit the conclusions that can be drawn from these results. Children with RSE were not questioned as to the extent of their perceptual training in previous therapy, so it is impossible to assess the extent to which some children had previously practiced rating rhotic stimuli for category goodness. Attention was also not measured, but "focus" is considered an important variable associated with ear training and auditory bombardment (Hodson & Paden, 1991). Thus, future research might extend the results of this study by exploring additional factors that relate to auditory perception in older children with RSE.

Conclusion

This study was conducted to determine (1) if a naturally-variable distribution of natural speech /r/ productions, defined by two normalized measures of F3 height, predicted category goodness judgments of /r/ in adult listeners, and (2) if children with RSEs responded

differently to a distribution of natural speech /r/ productions when compared to TD children and adults. In Study 1, adult listeners showed a significant effect of both acoustic parameters on their judgment of /r/ correctness, but the distance between the third formant and second formant (“F3-F2”) provided a slightly better fit for the data than the F3 ratio measure. The results of Study 2 indicated that children with RSE and TD children were similar to one another but significantly different from adults in their responses to the F3-F2 distribution, with adults showing a greater effect of the acoustic parameter on their responses. These results suggest that perception abilities may have a longer developmental trajectory than production abilities. Further, the study validates F3-F2 difference (and to a lesser extent, the F3 ratio) as useful indicators of category goodness for /r/. Variation in perceptual performance was noted in both of the child participant groups. Larger studies might reveal subgroups of RSE or mediating factors such as auditory short-term memory that may predict individual patterns of performance in this task. Further investigation of these differences will allow us to maximize the clinical utility of perceptual tasks including the present category goodness tasks.

Acknowledgements

The authors wish to thank Jade Clark, Kelsey Douglas, Lucy Durepos, Lindsay Mullins, Kirsten Mosko, Leah Scholl, and Caroline Spencer for their contributions to this research project, as well as all participants in the study.

Statement of interest

The authors report no declarations of interest. This work was partially supported by a NIH Translational R01 Grant “Improving clinical speech remediation with ultrasound technology” #1R01DC013668–01 (Contact PI: Doug Whalen, mPI Suzanne Boyce) and by a Doctoral Research Grant Award from the Ohio Speech Language and Hearing Association.

Appendix 1: Stimulus words for both studies

Study 1 (total n=159)	Study 2 (total n=98)
ear	her
rake	rock
rye	reek
ire	father
row	wrist
roo	deer
door	rug
zipper	car
her	grief
relief	dream
yarn	are
free	roo
or	ray
drop	ry
raw	or
rat	ear

Study 1 (total n=159)	Study 2 (total n=98)
ripe	ire
brag	rat
green	air
prop	row
air	rake
carrot	brag
tire	green
actor	ree
father	door
reek	rye
deer	zipper
rug	free
wrist	/r/ (isolation)
roast	drop
grief	
rock	
dream	
ray	
ree	
are	
car	
/r/ (isolation)	
drop	

References

- Abboud H (2017). Superlab (Version 5.0) [Computer software]. San Pedro, CA: Cedrus Corporation Retrieved from www.cedrus.com
- Aungst LF, & Frick JV (1964). Auditory discrimination ability and consistency of articulation of /r/. *Journal of Speech and Hearing Disorders*, 29(1), 76–85. [PubMed: 14122671]
- Bates D, Maechler M, Bolker B, & Walker S (2017). Package “lme4”: Linear mixed-effects models using “Eigen” and S4 (version 1.1–13) [Computer software]. Retrieved from <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- Beskow J & Sjölander K (2005). Wavesurfer (Version 1.8.5) [Computer software]. Retrieved from <http://www.speech.kth.se/wavesurfer/index2.html>
- Boersma P & Weenink D (2016). Praat: doing phonetics by computer (Version 6.0.21) [Computer software]. Retrieved 25 September 2016 from <http://www.praat.org/>
- Bradlow A, Toretta G, & Pisoni D (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20, 255–272. [PubMed: 21461127]
- Broen PA, Strange W, Doyle SS, & Heller JH (1983). Perception and production of approximant consonants by normal and articulation-delayed preschool children. *Journal of Speech and Hearing Research*, 26(4), 601–608. [PubMed: 6199587]
- Cabbage KL, Hogan T, & Carrell TD (2016). Speech perception differences in children with dyslexia and persistent speech delay. *Speech Communication*, 82, 14–25.

- Campbell H, Harel D, Hitchcock E, & McAllister Byun T (2017). Selecting an acoustic correlate for automated measurement of American English rhotic production in children. *International Journal of Speech-Language Pathology*, 1–9. 10.1080/17549507.2017.1359334
- Espy-Wilson CY, Boyce SE, Jackson M, Narayanan S, & Alwan A (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America*, 108(1), 343–356. [PubMed: 10923897]
- Flipsen P Jr, Shriberg LD, Weismer G, Karlsson H, & McSweeney J (2001). Acoustic phenotypes for speech-genetics studies: reference data for residual /ɜ/ distortions. *Clinical Linguistics & Phonetics*, 15(8), 603–630.
- Gray AH Jr, & Markel JD (1976). Distance measures for speech processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(5), 380–391.
- Gray SI, & Shelton RL (1992). Self-monitoring effects on articulation carryover in school-age children. *Language, Speech, and Hearing Services in Schools*, 23(4), 334–342.
- Grunwell P (1988). Phonological assessment, evaluation and explanation of speech disorders in children. *Clinical Linguistics & Phonetics*, 2(3), 221–252. 10.1080/02699208808985257
- Hagiwara R (1995). Acoustic realizations of /r/ as produced by women and men (Dissertation). UCLA, Los Angeles.
- Hazan V, & Barrett S (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396. 10.1006/jpho.2000.0121
- Hodson BW, & Paden EP (1991). Targeting intelligible speech: A phonological approach to remediation (2nd ed.). Austin, TX: Pro-Ed.
- Hoffman PR, Daniloff RG, Bengoa D, & Schuckers GH (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders*, 50(1), 46–53. [PubMed: 3974212]
- Hoffman PR, Stager S, & Daniloff RG (1983). Perception and production of misarticulated /r/. *Journal of Speech and Hearing Disorders*, 48(2), 210–215. [PubMed: 6621013]
- Koegel LK, Koegel RL, & Ingham JC (1986). Programming rapid generalization of correct articulation through self-monitoring procedures. *Journal of Speech and Hearing Disorders*, 51(1), 24–32. [PubMed: 3945056]
- Klein HB, Grigos MI, McAllister Byun T, & Davidson L (2012). The relationship between inexperienced listeners' perceptions and acoustic correlates of children's /r/ productions. *Clinical Linguistics & Phonetics*, 26(7), 628–645. 10.3109/02699206.2012.682695 [PubMed: 22690718]
- Lehiste I (1964). *Acoustical characteristics of selected English consonants*. Bloomington, IN: Indiana University Press.
- Locke JL (1980). The inference of speech perception in the phonologically disordered child. Part IA Rationale, some criteria, the conventional tests. *Journal of Speech and Hearing Disorders*, 45(4), 431–444. [PubMed: 7442161]
- Massaro DW, & Cohen MM (1983). Categorical or continuous speech perception. *Speech Communication*, 2, 15–35.
- McAllister Byun T (2017). Efficacy of visual–acoustic biofeedback intervention for residual rhotic errors: A single-subject randomization study. *Journal of Speech, Language, and Hearing Research*, 60(5), 1175–1193. 10.1044/2016_JSLHR-S-16-0038
- McAllister Byun T, & Tiede M (2017). Perception-production relations in later development of American English rhotics. *PLOS ONE*, 12(2), 1–22. 10.1371/journal.pone.0172022
- McGowan RS, Nittrouer S, & Manning CJ (2004). Development of [ɹ] in young, Midwestern, American children. *The Journal of the Acoustical Society of America*, 115(2), 871–884. 10.1121/1.1642624 [PubMed: 15000198]
- Mielke J, Baker A, & Archangeli D (2010). Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. *Laboratory Phonology 10*: 699–719.
- Monnin LM, & Huntington DA (1974). Relationship of articulatory defects to speech-sound identification. *Journal of Speech, Language, and Hearing Research*, 17(3), 352–366.
- Nittrouer S, & Miller ME (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, 101, 2253–2266. [PubMed: 9104027]

- Nittrouer S, Studdert-Kennedy M, & McGowan RS (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*, 32(1), 120–132.
- Ohde RN, & Sharf DJ (1988). Perceptual categorization and consistency of synthesized /r-w/ continua by adults, normal children and /r/-misarticulating children. *Journal of Speech and Hearing Research*, 31(4), 556–568. [PubMed: 2976438]
- Preston JL, & Edwards ML (2007). Phonological processing skills of adolescents with residual speech sound errors. *Language, Speech, and Hearing Services in Schools*, 38(4), 297–308. 10.1044/0161-1461(2007/032)
- Preston JL, Felsenfeld S, Frost SJ, Mencl WE, Fulbright RK, Grigorenko EL, ... Pugh KR (2012). Functional brain activation differences in school-age children with speech sound errors: speech and print processing. *Journal of Speech Language and Hearing Research*, 55(4), 1068–1082. 10.1044/1092-4388(2011/11-0056)
- Preston JL, Molfe PJ, Mencl WE, Frost SJ, Hoeft F, Fulbright RK, ... Pugh KR (2014). Structural brain differences in school-age children with residual speech sound errors. *Brain and Language*, 128(1), 25–33. 10.1016/j.bandl.2013.11.001 [PubMed: 24342151]
- Preston J, Irwin J, & Turcios J (2015). Perception of speech sounds in school-aged children with speech sound disorders. *Seminars in Speech and Language*, 36(4), 224–233. 10.1055/s-0035-1562906 [PubMed: 26458198]
- Redle E, Vannest J, Maloney T, Tsevat RK, Eikenberry S, Lewis B, ... Holland SK (2015). Functional MRI evidence for fine motor praxis dysfunction in children with persistent speech disorders. *Brain Research*, 1597, 47–56. 10.1016/j.brainres.2014.11.047 [PubMed: 25481413]
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, & Müller M (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77. 10.1186/1471-2105-12-77 [PubMed: 21414208]
- Ruscello DM (1995). Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders*, 28(4), 279–302. [PubMed: 8576411]
- Rvachew S (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research*, 37(2), 347–357. [PubMed: 8028316]
- Rvachew S, & Jamieson DG (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders*, 54(2), 193–208. [PubMed: 2709838]
- Rvachew S, & Andrews E (2002). The influence of syllable position on children's production of consonants. *Clinical Linguistics & Phonetics*, 16(3), 183–198. 10.1080/02699200110112222 [PubMed: 12064027]
- Saben CB, & Ingham JC (1991). The effects of minimal pairs treatment on the speech-sound production of two children with phonologic disorders. *Journal of Speech, Language, and Hearing Research*, 34(5), 1023–1040.
- Sharf DJ, & Benson PJ (1983). Comparison of speech-language pathologists' and naive subjects' identification of synthesized /r-w/ continua. *Journal of Speech and Hearing Research*, 26(4), 525–530. [PubMed: 6366369]
- Sharf DJ, & Ohde RN (1983). Perception of distorted "R" sounds in the synthesized speech of children and adults. *Journal of Speech and Hearing Research*, 26(4), 516–524. [PubMed: 6366368]
- Sharf DJ, Ohde RN, & Lehman ME (1988). Relationship between the discrimination of /w-r/ and /t-d/ continua and the identification of distorted /r/. *Journal of Speech and Hearing Research*, 31(2), 193–206. [PubMed: 3398493]
- Shriberg LD (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech, Language, and Hearing Research*, 36(1), 105–140.
- Shriberg L, Flipsen P Jr, Karlsson H, & McSweeny JL (2001). Acoustic phenotypes for speech-genetics studies: an acoustic marker for residual/ə/distortions. *Clinical Linguistics & Phonetics*, 15(8), 631–650.
- Shriberg LD, & Kwiatkowski J (1987). A retrospective study of spontaneous generalization in speech-delayed children. *Language, Speech, and Hearing Services in Schools*, 18, 144–157.

- Shriberg LD, Kwiatkowski J, & Gruber FA (1994). Developmental phonological disorders II: Short-term speech-sound normalization. *Journal of Speech, Language, and Hearing Research*, 37(5), 1127–1150.
- Shriberg LD, Austin D, Lewis BA, McSweeney JL, & Wilson DL (1997). The Speech Disorders Classification System (SDCS) extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research*, 40(4), 723–740.
- Shuster LI (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research*, 41(4), 941–950.
- Shuster LI, Ruscello DM, & Toth AR (1995). The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology*, 4(2), 37–44.
- Smit AB, & Bernthal JE (1983). Performance of articulation-disordered children on language and perception measures. *Journal of Speech, Language, and Hearing Research*, 26(1), 124–136.
- Stavness I, Gick B, Derrick D, & Fels S (2012). Biomechanical modeling of English /r/ variants. *The Journal of the Acoustical Society of America*, 131(5), 355–360. 10.1121/1.3695407
- Stevens K (1999). Articulatory-acoustic-auditory relationships In Hardcastle WJ & Laver J (Eds.), *The Handbook of Phonetic Science*. (pp. 462–506). Oxford: Blackwell.
- Treiman R, & Breaux AM (1982). Common phoneme and overall similarity relations among spoken syllables: Their use by children and adults. *Journal of Psycholinguistic Research*, 11(6), 569–598. 10.1007/BF01067613 [PubMed: 7161727]
- Tyler AA, Edwards ML, & Saxman JH (1990). Acoustic validation of phonological knowledge and its relationship to treatment. *Journal of Speech and Hearing Disorders*, 55(2), 251–261. [PubMed: 2329787]
- Van Riper C (1978). *Speech Correction: Principles and Methods* (6th ed.). New York: Prentice Hall.
- Walley AC, & Flege JE (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, 27, 307–332.
- Wolfe V, Martin D, Borton T, & Youngblood HC (2003). The effect of clinical experience on cue trading for the /r-w/ contrast. *American Journal of Speech-Language Pathology*, 12(2), 221–228. 10.1044/1058-0360(2003/068) [PubMed: 12828535]
- Zlatin MA, & Koenigsknecht RA (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech, Language, and Hearing Research*, 19(1), 93–111.

Figure 1.
ROC curves for the acoustic parameters of both models using adult response data.

Figure 2.

Proportion of rejections of /r/ for adult listeners by each acoustic parameter. Pr(NO) indicates the probability that tokens in the quintile are judged incorrect. Left panel shows proportion of rejections using the model with the F3-F2 acoustic parameter and right panel shows the proportion of rejections using the F3 ratio acoustic parameter.

Figure 3.

Proportion of rejections of /r/ for all listeners by F3-F2 acoustic parameter. “TD” refers to typically developing children used as controls. “RSE” refers to children with residual sound errors. Pr (NO) refers to the probability of “no” (“not a correct /r/”) responses to stimuli falling within the quintile.

Figure 4.

Logistic curve depicting interaction effect of status and third formant parameter on the responses of adults (solid line), children with RSE (dashed line), and TD children (dotted line).

Table 1

Stimuli Characteristics for Study 1

Parameter	Sub-categories	N of words in Sub-category
Position	Cluster	21
	Prevocalic	60
	Postvocalic	78
Vowel context	Back	55
	Front	75
	Mid-central	29
Error types	Substitution	49
	Omission	21
	Distortion	56
Correct tokens		33

Note: “Error types and “correct tokens” are based on the perception of one expert listener for descriptive purposes.

Table 2

Logistic Regression Coefficients for Effect of F3 Ratio, F3-F2, and Vowel Context on Listener Responses

	Model with F3 ratio			Model with F3-F2 difference		
	Estimate	Std. Error	<i>p</i>	Estimate	Std. error	<i>p</i>
Intercept	0.279	0.381	0.463	0.805	0.422	0.057
Acoustic parameter	0.809	0.067	<0.001	1.037	0.08	<0.001
Vowel context						
Front	-1.347	0.379	<0.001	-1.933	0.468	<0.001
Mid-central	-0.478	0.627	0.446	-1.171	0.666	0.079

Note: For vowel context, back vowels are the reference level in the logistic regression.

Table 3

Demographic Characteristics of Children with RSE

Sex		Age	
Males	25	Range	7;0–16;8
Females	13	Mean	10;9
		SD	2;5

Table 4

Stimuli Characteristics for Study 2

Parameter	Sub-categories	N of Words in Sub-category
Position	Cluster	9
	Prevocalic	38
	Postvocalic	51
Vowel context	Back	39
	Front	48
	Mid-central	11
Error types	Substitution	27
	Omission	13
	Distortion	38
Correct tokens		20

Table 5

Logistic Regression Coefficients for Effect of F3-F2 Difference on Responses

	Estimate	Std. Error	<i>p</i>
Intercept	0.230	0.212	0.227
F3-F2	0.593	0.059	<.001
Adults ^a	−0.314	0.176	0.07
RSE children ^a	−0.158	0.165	0.33
Front vowel ^b	−0.446	0.139	0.001
Mid-central vowel ^b	−0.025	0.188	0.895
F3-F2 × Adults ^a	0.326	0.088	<.001
F3-F2 × RSE ^a	−0.119	0.079	0.13

Notes:

^aTD children are the reference level.^bBack vowels are the reference level.