DETECTING ERRORS IN AMERICAN ENGLISH /ɹ/ ALONG A NORMALIZED ACOUSTIC THRESHOLD

Sarah M. Hamilton, Keiko Ishikawa, Lindsay Mullins, Suzanne E. Boyce

University of Cincinnati
hamilsm@mail.uc.edu, ishikak@mail.uc.edu, mullinlm@mail.uc.edu, boycese@ucmail.uc.edu

ABSTRACT

Perception of speech sounds has been characterized as largely categorical, such that listeners experience a shift in perception from one sound to a different sound at a certain point on a continuum. Children with speech errors, however, have been found to have atypical category boundaries. Following an observation by Hagiwara [6], who noted that typical speakers show third formant (F3) values for /ɹ/ between 80% and 60% of their average vowel F3, Hamilton et al [7] found that this value replicates a categorical boundary for /ɹ/ for adult listeners: Productions above and below the 80% threshold sounded consistently "incorrect" or "correct", but productions closest to the threshold were given more ambiguous judgments. In this study, we apply this notion of an F3 threshold to investigate whether children with residual sound errors (RSE) respond like expert adult listeners (speech-language pathologists) when presented with natural-speech stimuli along a continuum of F3 distances.

Keywords: rhotics, perceptual categorization, category boundary, English, speech sound disorders

1. INTRODUCTION

In American English, one of the most common residual speech sound errors in children is the phoneme /ɹ/ [13]. Many studies have found that children with residual speech sound errors (RSE) show (1) atypical category boundaries, and (2) difficulty identifying whether their own productions are correct or misarticulated [17], [1], [2], [15]. However, in perceptual experiments for /ɹ/, many ask children to distinguish between /w/ and /ɹ/ and find mixed results, where some children with speech sound disorders are worse at identifying /ɹ/ and /w/ than their typically speaking peers, but other children with speech sound disorders are on par with typical peers [8]. Such mixed results have led researchers to question if a continuum defined by the /ɹ/ and /w/ phonemes is the most representative mental representation to explore when researching disordered perceptual categorization [16]. In addition, most perceptual category discrimination tests have used synthesized speech representing incremental change along an acoustic continuum, but these continua may not represent the true capability of a listener to judge productions that differ along more than one dimension [9]. However, real speech stimuli have been a challenge to use in perceptual tasks, as the acoustic values that define /ɹ/ must be normalized if the speech is to be used as a continuous test variable.

2. ACOUSTIC CHARACTERISTICS OF AMERICAN ENGLISH /ɹ/

The American English rhotic phoneme is characterized as a postalveolar sonorant approximant: /ɹ/. While most vowels and sonorants share similar third formant values, /ɹ/ is distinguished by a third formant (F3) that can be as much as 1000 Hz lower than the F3 of neighboring vowels. For some speakers, in some contexts, F3 may merge with F2. This low F3 is unique among American English speech sounds.

In his dissertation on the acoustics of /ɹ/ as produced by women and men, Hagiwara observed that the third formant for /ɹ/ for individual speakers fell below a certain percentage of the speaker's vowel F3 average [6]. For most speakers, that value was between 60 and 80% of the average vowel F3 value. Any speaker's /ɹ/ production, therefore, could be characterized as having an F3 value below 80% of the speaker's average vowel F3. Hamilton et al [7] hypothesized that the 80% value could function as a sort of perceptual threshold, or categorical boundary, between a production that sounded "/ɹ/-like" and one that did not.

2.1. Defining a perceptual continuum for /ɹ/

Perception of speech sounds has been characterized as largely categorical, where categorical boundaries are a certain point on a continuum where listeners will experience a shift in perception from one sound to a different sound [12]. Categorical boundaries have historically been of interest to speech researchers, as it has been hypothesized that clear categorical boundaries between phonemes are a requirement for accurate speech sound production [13].

Vowel categories are notoriously more fluid than stop consonant boundaries, with a more gradual slope in categorical responses than consonant categories [4]. The slope represents an acoustic region where listeners typically have difficulty identifying one phoneme from another, with a gradual slope representing a wider region of "grey area" or acoustic signatures that fall into both categories for the listener.

A critical component to categorical perception research is that the properties chosen for the continuum must represent acoustic characteristic(s) that are associated with a given speech sound. For example, while many categorical boundary tests have used acoustic characteristics of /ɹ/ and /w/ phonemes as two distinct ends of a perceptual continuum, these continua presume that /ɹ/ and /w/ exist in some kind of intrinsic relationship with each other in a way that other sounds, such as /ɹ/ and /ə/, do not. In this case, /ɹ/ and /w/ may be affected by the tradition of phonemic substitutions in speech-language pathology literature, where a child is perceived as producing a /w/ or /u/ instead of an /ɹ/. However, children are not restricted by their language's phonology when they produce error sounds, and their productions may sound like another phoneme or they may exist outside the expected phonetic structure of English sounds. However, listeners do not always identify these subtle acoustic differences: Listeners' perceptual categorization abilities can actually mask covert contrasts in children's productions, when children make articulatory distinctions that adults fail to notice or transcribe [5].

Instead of investigating /ɹ/ in terms of its acoustic distance to other phonemes, it can be defined in terms of its signature, salient acoustic feature: the third formant. As a continuous variable that lowers dramatically only during /ɹ/ production in American English, the third formant appears to be an excellent candidate to form a continuum along which perception of /ɹ/ can be investigated.

3. A NORMALIZED THRESHOLD FOR /ɹ/ IN AMERICAN ENGLISH

To create a continuum of frequencies that are comparable across speakers, a normalization procedure is required. The particular frequency of the third formant required for perceptually-accurate American English /ɹ/ varies considerably over age and gender [11]. Such variation creates a challenge when trying to define the acoustic characteristics of /ɹ/ across different speakers.

Recalling that Hagiwara [6] found that the third formant for /ɹ/ for individual speakers fell below a certain percentage of the speaker's vowel F3 average, vowel data was collected from all speakers who contributed /ɹ/ productions. For each speaker, the third formants from his or her cardinal vowels (/i/, /æ/, /u/, and /a/) were measured and an average taken. 80% of the resulting value was derived, and used as that speaker's *F3 threshold* value for /ɹ/ (see Equation (1)). The threshold value was unique to that individual speaker and was used as input for a normalization algorithm (Equation (2)).

3.1. Equation for F3 threshold

(1) $$[\frac{(F3_1 + F3_2 + F3_3 + F3_4)}{4}] \bullet 0.8 = F3_t$$

The third formant of each speaker's /ɹ/ production was measured ($F3_m$) and each $F3_m$ was used as the other input variable for the normalization algorithm (Equation (2)).

3.2 Equation for normalization

(2) $$F3_m - F3_t = F3_N$$

$F3_m$ is the third formant measured in the speaker's production, $F3_t$ is the third formant threshold derived from Equation (1), and the resulting value is a normalized F3 value that represents distance from a perceptual threshold, or cut-off, for an accurate /ɹ/ production in American English.

Hamilton et al [7] determined that the normalized F3 threshold constituted a perceptual categorization boundary with expert listeners, where listeners were more likely to indicate that productions falling below the threshold were correct /ɹ/ productions, while productions with third formants that were above the threshold were more likely to be heard as errors. Expert listeners' responses appear to correspond predictably to third formant distances, but it is unknown if the responses of children with difficulty producing /ɹ/ will follow this same pattern.

4. EXPERIMENTAL PROTOCOL

4.1 Listeners

12 children (age 8-15) and 14 expert listeners listened to simple whole word stimuli taken from speech sound therapy sessions. Due to task demands, children listened to a reduced set of stimuli (n=98) while the expert listeners listened to the entire set (n=159). The smaller set for children was balanced to have a distribution of third formant values that was similar to the expert set. In a binary set, forced choice task, participants were instructed to indicate if they heard a perceptually-accurate /ɹ/ production.

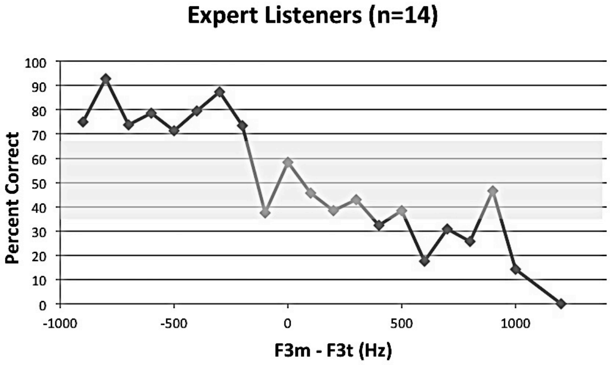
4.2 Stimuli

Stimuli included words with /ɹ/ in both pre- and postvocalic position. 27 speakers' words were used to make the stimulus lists. Stimuli were selected to form a continuum of /ɹ/ productions with third formants at different distances from the normalized threshold. In the interest of investigating categorical boundary effects of third formant distances, the stimulus list included more words with third formants near the threshold.

5. RESULTS

Adult expert listeners' judgments of stimuli are displayed in Fig. 1. A grey region between 35% and 65% is defined, representing a region where strong agreements were not seen across listeners. The "0" point on the X axis defines the normalized threshold for any speaker. Positive values on the X axis represent points on the continuum of measured third formant values (m) that are above a speaker's derived threshold (t) (see Equation (1)). Higher positive values are representative of error /ɹ/ productions. Lower positive values are representative of /ɹ/ productions that are closer to the threshold of perceptually-correct /ɹ/ productions. Negative values are created when the measured third formant values is below the derived threshold for a perceptually correct /ɹ/. Fig. 1 displays that most expert listeners indicated that /ɹ/ tokens with negative F3m-F3t values were instances of correct /ɹ/.

Figure 1: Proportion of "correct /ɹ/" judgments by expert listeners over F3m – F3t measures

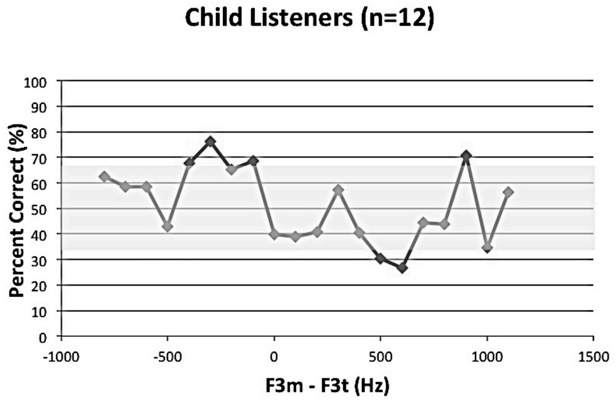


In Fig. 1, there is greater agreement among expert listeners at the extremes of the F3m-F3t continuum. Over 70% of expert listeners rated tokens as "correct" that were -500 to -1000 Hz below the threshold. As is expected in a categorical boundary, there is a steep slope in proportion of judgments of "correct" when the normalized F3 values approximate the 0 threshold. However, for productions of /ɹ/ with normalized values of 0-500 Hz, expert listeners disagreed about whether they were hearing correct or incorrect /ɹ/. It should be noted that the expert listeners' training may have interfered with error judgments: Many experts (speech-language pathologists) stated that their clinical training made it difficult to give "error" scores to productions that they felt were close to correct.

In Fig. 1, there is a precipitous drop in proportion of expert listeners' judgments of "correct" as the normalized third formant moves from below to above the threshold. However, for children with residual sound errors, there is no such overall decrease (Fig. 2) in the proportion of correct /ɹ/ judgments over the F3m-F3t continuum.

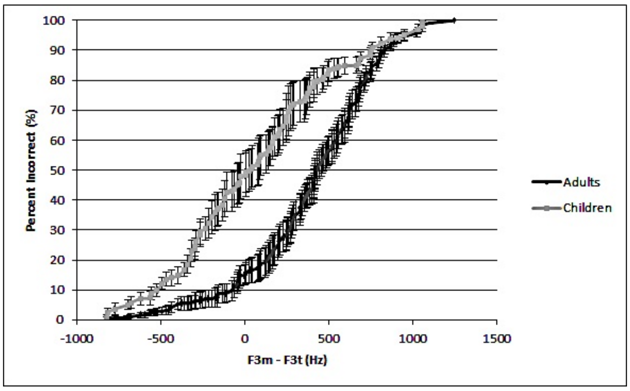
When stimuli with third formant values *near* the threshold are presented to children with residual sound errors, they respond like experts, meaning that their responses fall in the grey region when stimuli are 0-500 Hz from the threshold (Fig. 2). However, unlike the expert listeners, children with RSE do not show greater agreement at the extremes of the continuum. Their responses largely fall in the grey region even when they are given a word with /ɹ/ that has a normalized third formant at a great distance from the threshold.

Figure 2: Proportion of "correct /ɹ/" judgments by child listeners over F3m – F3t measures



In Fig. 3, the proportion of judgments of "incorrect" for stimuli on the continuum is plotted for both children with RSE (in grey) and expert listeners (in black). Curves are based on average and standard deviation of all listeners' performance. Error bars indicate standard deviations of judgments. Greater standard deviations for productions near the normalized threshold for both adults and children suggest that this region defines a perceptual boundary or grey area.

Figure 3: Comparison of average and standard deviation curves for percent incorrect across F3m-F3t measure in adult experts and children with RSE.



As a group, children with RSE have greater standard deviations in their responses, especially near the normalized threshold at 0 Hz. While the two proportion curves converge for the productions that are farthest from the threshold in the positive direction (the predicted "worst" /ɹ/ productions), there children and experts do not converge in their judgments for the predicted "best" productions. Children with RSE are more likely than experts to judge an /ɹ/ as incorrect even when the third formant is well below the "0" threshold.

6. CONCLUSION

When presented with natural speech stimuli along a continuum of normalized F3 values, children with RSE do not respond like adults. While more agreement between adults and children occur in the region closest to the F3 threshold (minimal distances, F3m-F3t), more discrepancy occurs in rating productions that are far from the threshold. Children with RSE, as a group, appear to have wider ranges of acceptability for /ɹ/ than adults. By investigating judgments across a continuum of normalized F3 distances, it appears that children with RSE have a different sensitivity to acoustic cues for /ɹ/ in real speech. The normalization technique for /ɹ/ in speech stimuli allows for an investigation of gradations of quality in productions with reference to an individually-derived threshold value. The results of this study provide further support for Ladefoged's claim that a lower third formant results in a better percept of rhoticity [10].

7. REFERENCES

[1] Aungst, L. F., Frick, J. V. 1964. Auditory discrimination ability and consistency of articulation of /r/. *Journal of Speech and Hearing Disorders,* 29, 76-85.

[2] Byun, T. M., Tiede, M. 2014. Perceptual acuity and production distinctness in child speech: Data from American English /r. *The Journal of the Acoustical Society of America,* 135, 2316-2316.

[3] Espy-Wilson, C. Y., Boyce, S., Jackson, M., Narayanan, S., Alwan, A. 2000. Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America,* 108, 343-356.

[4] Fry, D. B., Abramson, A. S., Eimas, P. D., Liberman, A. M. 1962. The identification and discrimination of synthetic vowels. *Language and Speech,* 5, 171-189.

[5] Gibbon, F. E. 1999. Undifferentiated Lingual Gestures in Children With Articulation/Phonological Disorders. *J Speech Lang Hear Res*, 42, 382–397.

[6] Hagiwara, R. 1995. *WPP No. 90: Acoustic Realizations of American /r/ as Produced by Women and Men.* (PhD Dissertation), UCLA, UC Los Angeles.

[7] Hamilton, S. M., Ishikawa, K., Boyce, S. E., Mullins, L. 2014. Perceptual categorization of /r/ for children with residual sound errors. *Journal of the Acoustical Society of America,* 136, 2261-2261.

[8] Hoffman, P. R., Daniloff, R. G., Bengoa, D., Schuckers, G. H. 1985. Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders,* 50, 46-53.

[9] Kent, R. D. 1996. Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech Language Pathology,* 5, 7-23.

[10] Ladefoged, P., Maddieson, I. 1996. *The sounds of the world's languages*. Oxford: Blackwell Publishers.

[11] Lee, S., Potamianos, A., Narayanan, S. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105, 1455–1468.

[12] Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review,* 74, 431-461.

[13] Locke, J. L. 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, 431-444.

[14] Ruscello, D. M. 1995. Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders,* 28, 279-302.

[15] Sharf, D. J., Benson, P. J. 1982. Identification of synthesized /r–w/ continua for adult and child speakers. *The Journal of the Acoustical Society of America,* 71, 1008-1015.

[16] Sharf, D. J., Benson, P. J. 1983. Comparison of Speech-Language Pathologists' and Naive Subjects' Identification of Synthesized /r-w/ Continua. *Journal of Speech and Hearing Research,* 26, 525-530.

[17] Shuster, L. I. 1998. The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language and Hearing Research,* 41, 941-950.

[18] Strange, W., Broen, P. A. 1981. The relationship between perception and production of /w/,/r/, and/l/ by three-year-old children. *Journal of Experimental Child Psychology,* 31, 81-102.