Published in final edited form as:

Clin Linguist Phon. 2018 ; 32(1): 70–87. doi:10.1080/02699206.2017.1330898.

Deriving individualised /r/ targets from the acoustics of children's non-rhotic vowels

Heather Campbell and Tara McAllister Byun

NYU Steinhardt School of Culture, Education, and Human Development, New York, NY, USA

Abstract

In visual-acoustic biofeedback for rhotic errors, learners are guided to match the third formant (F3) location to a visual target on a real-time acoustic spectrum. As the acoustic properties of correct English /r/ differ across speakers, this study aimed to improve target selection by investigating the validity of individualised targets derived from children's non-rhotic vowels. A previously proposed prediction formula was adjusted using data from a child normative sample and tested in two groups of children. Study 1 found that predicted values were unexpectedly higher than actual F3 values in children whose /r/ errors had been remediated. To understand this discrepancy, Study 2 applied the formula to typically developing children and found that predicted values were also higher than actual F3 values, suggesting that different normative data might better represent the current samples. An updated formula is proposed, which can be used to generate individualised targets within acoustic biofeedback applications.

Keywords

Articulation disorders; biofeedback; intervention; residual speech errors; speech sound disorders

Introduction

Some young children with speech sound disorders (SSD) are able to resolve their errors spontaneously or through traditional speech therapy, but others continue to exhibit residual errors in later school age or young adulthood (Flipsen, 2015; Lewis & Shriberg, 1994). Errors producing US English rhotics¹ are considered among the most challenging to remediate (Shuster, Ruscello, & Toth, 1995), a fact that is likely related to its unique articulatory and acoustic properties. Articulatorily, production of /r/ features lip rounding (Shriberg, 1980) and requires near-simultaneous anterior and posterior lingual constrictions (Flipsen, 2015), which can be achieved with tongue configurations characterised as 'retroflex' or 'bunched,' as well as many intermediate variants (Tiede, Boyce, Holland, & Choe, 2004). Despite articulatory variability, perceptually accurate /r/ has stable acoustic

CONTACT Heather Campbell : heather.campbell@nyu.edu, Department of Communicative Sciences and Disorders, New York University, 665 Broadway, Floor 9, New York, NY 10012, USA.

ORCID

Heather Campbell <http://orcid.org/0000-0002-6072-5850>

¹Following a common clinical convention, the class of English rhotics will collectively be referred to as 'r/' in the current analysis, and 'ɹ/' will be used to refer to the stressed syllabic variant in particular.

properties: it is distinguished from other sonorants by the low frequency of the third formant (F3) (e.g. Boyce, Hamilton, Scholl, & Schmidlin, 2013; Hagiwara, 1995), which comes in close proximity to the relatively high second formant frequency (F2) (Delattre & Freeman, 1968). Treatment of /r/ is further complicated by the existence of different positional allophones (e.g. in the onset position versus as the nucleus of a syllable). However, the current study focuses exclusively on stressed syllabic /ɹ/ (as in 'fur'), since normative data for this variant are readily available for both child and adult speakers.

Treatment for /r/ errors

In traditional therapies for /r/ misarticulation, a speech-language pathologist (SLP) typically provides a model of an auditory target for the child to imitate, along with articulatory placement cues and feedback on perceptual accuracy (Van Riper, 1978). Some children respond well to this treatment approach, while for others, errors persist despite extensive treatment (Ruscello, 1995). In addition, error patterns within and across individuals are inconsistent (Curtis & Hardy, 1959; Magloughlin, 2016), such that treatment approaches must be tailored to each child while remaining flexible. Furthermore, /r/ errors may be associated with inadequate internal representations of the auditory-perceptual target for /r/ (Shuster et al., 1995), leading to difficulty classifying one's own productions as correct or incorrect.

In light of these challenges, SLPs have called for more effective treatment strategies for individuals with residual /r/ errors (Ruscello, 1995). Visual-acoustic biofeedback is a promising approach that shifts the focus beyond the articulatory and auditory domains that dominate traditional treatment (Maas et al., 2008). The form of acoustic biofeedback in the current analysis makes use of linear predictive coding (LPC) to display the resonant frequencies (formants) as peaks along a wave-like shape that updates in real-time. Recent research suggests that this approach can facilitate acquisition of perceptually accurate /r/ (McAllister Byun & Hitchcock, 2012), and these effects may be particularly prominent when biofeedback is used to establish an accurate motor plan prior to a traditional treatment phase (McAllister Byun & Campbell, 2016).

While viewing the LPC spectrum, learners modify their speech in an attempt to align their formant peak locations with a target representing a correct production (Figure 1). As the height of F3 is understood to be the primary acoustic marker distinguishing /r/ from other sonorants (Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000), visual-acoustic biofeedback for /r/ typically emphasises an appropriate F3 target for the learner to match during sustained /r/ productions. It is known that vocal tract length is negatively related to formant heights (González, 2004). Since F3 heights differ across speakers, it is necessary to find a way to predict individualised F3 targets for use in therapy.

Although research evidence has accumulated to support acoustic biofeedback as an effective clinical option for /r/ treatment, its adoption in many clinical settings may be limited due to financial restrictions or a lack of familiarity with the technology. To facilitate clinical adoption of visual-acoustic biofeedback, efforts are ongoing to develop a low-cost, user-friendly iOS application that SLPs can use to provide visual-acoustic biofeedback for clients

with residual rhotic errors (McAllister Byun, Campbell, Liang, Park, & Svirsky, in press). Since it can be challenging to determine the optimal acoustic target for an individual speaker, one way to maximise the user-friendliness of the app is to include a routine that will automatically compute such a target. The aim of the current study was to investigate whether individualised F3 targets for /r/ can reliably be calculated from measurements of children's non-rhotic vowels, with the ultimate goal of applying this knowledge to reduce barriers to clinical adoption of app-based acoustic biofeedback intervention.

Derivation of an F3 target formula

Hagiwara (1995) investigated the relationship between formant frequencies of non-rhotic vowels and /r/ in adults. As non-rhotic vowels are relatively uniform in F3 height within individuals, he used the average F3 of the four corner vowels (/i/, /æ/, /ɑ/, and /u/) to approximate the height of F3 across non-rhotic vowels ('neutral vowel F3'). Hagiwara observed that F3 values of various perceptually accurate initial, syllabic, and final /r/ productions typically fell within a range between 60% and 80% of the neutral vowel F3 for a given speaker.

Boyce et al. (2013) followed up on Hagiwara's observation by positing that 80% of an individual's corner vowel average (the upper limit of Hagiwara's proposed range) would serve as a 'critical frequency', defined in (1):

$$\text{critical frequency} = \left[\frac{(F3_i + F3_{\text{æ}} + F3_a + F3_u)}{4} \right] * 0.80 \quad (1)$$

They hypothesised that perceptually incorrect rhotics would have F3 values that fall above this critical frequency for an individual, and that productions would begin to sound correct as they approached the critical frequency. In a sample of children with /r/ misarticulation, they found that all F3 values for tokens of /r/ rated perceptually incorrect fell above the critical frequency, while the few /r/ productions judged perceptually correct fell below the critical frequency. However, only nine out of 84 tokens in that sample were correct, providing insufficient evidence to draw a general inference about whether all rhotic tokens rated correct would show F3 values below the critical frequency.

In sum, Hagiwara's (1995) study included mostly perceptually accurate rhotics whose F3 values fell below the 80% critical frequency, whereas the study by Boyce et al. (2013) included mostly incorrect tokens falling above this frequency. This study tests whether the same cutoff accurately classifies both correct and incorrect rhotics within individual speakers. This value represents the edge of the range of acoustic variation across typical speakers and is important to rhotic perception. However, for the purpose of defining treatment targets, it may be optimal to provide an F3 value representing a canonical F3 for /r/ drawn from the centre rather than the edge of the range of acoustic variation. Estimating this target frequency represents another major goal of the present study.

To estimate this canonical F3 target, we drew on normative acoustic measures from Lee, Potamianos, and Narayanan (1999), who reported fundamental frequency and the first three formants in tokens of stressed vocalic /ɜ/ and nine vowels (/i/, /ɪ/, /e/, /æ/, /ɑ/, /ʌ/, /ɔ/, /ʊ/, /u/) elicited in a carrier phrase context. Productions were obtained from groups of males and females aged 5 through adult (19+), with 9–25 speakers representing each gender and age combination.

Figure 2 compares the average F3 of the corner vowels with the average F3 of /ɜ/ for each age group for females (top panel) and males (bottom panel) in the normative sample. The higher dark bars represent the mean F3 across the corner vowels, while the lower light bars represent the mean F3 of /ɜ/ in each sample. To calculate the actual F3 of /ɜ/ values across age groups, we lowered the scale factor, or the proportion of the vowel average, from 0.8 in increments of 0.01. Based on visual inspection, the scale factors that best matched the actual F3 of /ɜ/ values across ages in the normative group data were 0.70 for female speakers and 0.67 for male speakers,² represented by the dot superimposed over each bar in Figure 2. Henceforth, these scale factors applied to the average F3 of the corner vowels will be referred to as the norm-derived ‘predicted F3 of /ɜ/’, abbreviated F3_{pr}, and shown in (2).

$$\begin{aligned} \text{a. } F3_{pr}(\text{males}) &= \left[\frac{(F3_i + F3_{\text{æ}} + F3_a + F3_u)}{4} \right] * 0.67 \quad (2) \\ \text{b. } F3_{pr}(\text{females}) &= \left[\frac{(F3_i + F3_{\text{æ}} + F3_a + F3_u)}{4} \right] * 0.70 \end{aligned}$$

The goal of this paper is to address whether the F3_{pr} formula can correctly predict the mean F3 measured across perceptually accurate rhotics within individuals, and to compare inaccurate and accurate tokens against the critical frequency. The ultimate aim is to generate F3 targets to be used in acoustic biofeedback applications. In Study 1, we compared data from children with remediated rhotic errors to both the proposed 80% critical frequency and norm-based F3_{pr}. In Study 2, we did the same with a sample of typically developing (TD) children.

Study 1 – remediated sample

Study 1 investigated data from two children with /r/ misarticulation who reached ceiling-level accuracy over the course of an ultrasound treatment study (McAllister Byun, Hitchcock, & Swartz, 2014). It was hypothesised that those tokens that were rated as perceptually incorrect, which were mainly elicited prior to and early in treatment, would fall above the critical frequency. Conversely, tokens rated correct, which were primarily elicited later in treatment, would fall well below the critical frequency, near the calculated F3_{pr} value. If remediated tokens did differ from F3_{pr} values, it was predicted that observed F3 values would be slightly higher than F3_{pr} values derived from normative data. The rationale

²Separate scale factors were used for male and female speakers as a result of differences observed between the two sexes following visual inspection. As dividing by sex yields accurate predictions across most ages in the plots of both sexes in Figure 2, scale factors were not further subdivided by age.

for this prediction came from research by Shriberg, Flipsen, Karlsson, and McSweeny (2001), who observed a non-significant difference in correct rhotic productions between children with and without a history of rhotic misarticulation. Their data suggested that children with a history of rhotic misarticulation may produce /r/ with a less extreme degree of rhoticity than children with no history of misarticulation.

Method

Participants—The children selected for inclusion in this post-hoc analysis were 7 year-old ‘Autumn’ and 11 year-old ‘Lilianne’ (pseudonyms from McAllister Byun et al., 2014). For both participants, blinded expert listeners’ (certified speech-language pathologists’) perceptual ratings of /ʒ/ in words showed improvement from nearly 0% accuracy during pre-treatment baseline sessions to nearly 100% accuracy during post-treatment maintenance sessions.

Stimuli and measurement—In order to test the critical frequency value and to compute an $F3_{pr}$ value for both Autumn and Lilianne, F3 frequencies were measured in the four corner vowels and in /ʒ/ produced in selected syllable/word probes elicited at different treatment stages. The vowel measurements were also drawn from probes that had been assembled to elicit /r/ (see syllable/word column in Table 1) and thus were taken from words that contained /r/ in onset position. There is a strong possibility that perseverative coarticulatory effects could have lowered the F3 values for vowels in this context (Tunley, 1999), and results will be interpreted in consideration of this possibility.

Acoustic measurements of recordings from McAllister Byun et al. (2014) were analysed using the Praat acoustic software (Boersma & Weenink, 2014). Recordings had been collected in a sound-shielded room using the Computerised Speech Lab (KayPentax, Model 4150B) with 16-bit encoding and a sampling frequency of 48,000 Hz with the microphone approximately five inches from each participant’s mouth. After an optimal formant setting was determined by inspection for each participant, a point was manually selected in the steady state portion of each vowel or at the minimum F3 value for rhotics, and a Praat script was used to extract the formant frequencies at that point.

Analyses— $F3_{pr}$ was calculated for each child by applying the 0.7 scale factor for females to the average across corner vowels. The primary analyses were non-parametric one-sample Wilcoxon signed-rank tests evaluating whether each child’s $F3_{pr}$ value differed significantly from the actual distribution of F3 values observed in /ʒ/. For comparison purposes, the actual scale factor needed to reach the observed mean F3 of /ʒ/ was then calculated for each child. One-sample Wilcoxon signed-rank tests were additionally used to compare observed F3 of /ʒ/ values with the mean normative value for each girl’s age group. To follow up on the hypothesis that the critical frequency would divide incorrect and correct productions, 80% of the corner vowel average was calculated for both participants. The number of perceptually correct and incorrect tokens that fell on either side of the critical frequency is reported.

Results

The mean F3 values for corner vowels and for both perceptually correct and incorrect /ʒ/ productions for both participants are reported in Table 1. Autumn's calculated $F3_{pr}$ of 2293.9 Hz was higher than the mean across the F3 values observed in her correct /ʒ/ tokens (mean = 2070.6 Hz). This difference of 223.3 Hz was statistically significant ($V = 6$, $p = .0009$). Lilianne's calculated $F3_{pr}$ of 2232.09 Hz was also found to be significantly higher than her F3 values in correct /ʒ/ tokens ($V = 26$, $p = .02$), although the difference was smaller in this case (86.3 Hz). In sum, lower scale factors would be needed to map from the non-rhotic vowel mean to the average F3 of /ʒ/ for both participants. For Autumn, the scale factor that could be applied to her corner vowel average to obtain her actual mean F3 of correct /ʒ/ was calculated to be 0.63. For Lilianne, the corresponding scale factor was calculated to be 0.67.

The finding that scale factors derived for our remediated participants differed significantly from the norm-derived scale factors raises a question as to whether the rhotics they produced after remediation were consistent with normative values. For Lilianne, the observed mean F3 did not differ significantly from that reported for her age-matched peers in the normative sample ($V = 46$, $p = 0.2$), whereas Autumn's mean F3 of /ʒ/ fell significantly below the mean for her age ($V = 0$, $p < .0001$).

Figure 3 shows F3 values for Autumn and Lilianne over the course of the treatment study, including baseline (BS), treatment (TX) and maintenance (MN) sessions. The figure depicts how the F3 values measured for each perceptually correct and incorrect /ʒ/ token fell in relation to the 80% critical frequency (dashed line) and the $F3_{pr}$ value (solid line) for both children. For Autumn, 20/21 tokens that had been judged perceptually incorrect by blinded expert listeners fell above the critical frequency, while 15/15 tokens perceived as correct fell below the critical frequency. Similarly for Lilianne, 14/15 tokens judged as incorrect fell above the critical frequency, while 16/17 tokens judged correct fell below it. In total, the F3 values of most incorrect tokens (34/36) fell above the critical frequency and the F3 values of most correct tokens (31/32) fell below the critical frequency, yielding a 95.6% correct classification rate. This finding supports previous literature positing that perceptually correct and incorrect /ʒ/ can be separated by a critical frequency representing 80% of the mean across non-rhotic vowels (Boyce et al., 2013; Hagiwara, 1995).

Discussion

Study 1 was undertaken with the hypothesis that if there was a difference between the derived $F3_{pr}$ and the actual F3 observed in perceptually accurate rhotics produced by children with remediated /ʒ/, that F3 would be higher than $F3_{pr}$. Contrary to this expectation, observed F3 values of correct /ʒ/ were significantly lower than $F3_{pr}$ values for both children. Three possible explanations are considered for this unexpected result.

The first possible explanation is that there was a task effect stemming from the different manners in which target sounds were elicited in Lee et al. (1999) and McAllister Byun et al. (2014). The normative study elicited words in a carrier phrase context, whereas the treatment study elicited isolated syllables and words. It has been shown that an individual's vowel

space tends to be more expanded in slow than in regular connected speech (Tjaden & Wilding, 2004), and in isolated words than in phrases (Van Bergem, 1993). Another possible source of task-related differences is the fact that the normative study elicited all speech samples in a single visit with no specific target sound, whereas McAllister Byun et al. (2014) was a longitudinal study focused on rhotic production. Either or both of these task effects could have occurred, which may explain why F3 values were found to be lower in the treatment study than in the varied vowel phrase task.

A second possible explanation is that children who undergo remediation for rhotic misarticulation and TD children produce /ɜ/ differently. Shriberg et al. (2001) suggested that F3 of /ɜ/ values would be lower for TD children than for children with SSD. However, low F3 of /ɜ/ values in the remediated sample in this study suggests that these children might instead have learned an extreme, hyperarticulated version. The finding that the normative F3 value of /ɜ/ for Autumn's age was higher than her actual F3 values is consistent with this hypothesis.

A third possible explanation is that the normative data from Lee et al. (1999) were not representative of the children in the present sample. One way in which such a discrepancy could occur is if the relationship between vowels and /ɜ/ taken from the group data in Lee et al. (1999) did not apply at the individual speaker level. This could be true if speakers showed a high degree of variability such that scale factors differed across children. Another way in which such a discrepancy could arise would be if the F3 values of vowels differed between the children in the normative sample and those in the current sample, possibly as a consequence of dialectal variation. Both Autumn and Lilianne had F3 values of corner vowels more than one standard deviation (SD) from the normative mean. The mean F3 of /æ/ was below one SD for both Autumn and Lilianne, whereas the mean F3 of /u/ was above one SD for Lilianne. This suggests that vowel averages may have differed between the remediated and normative samples.

Study 2 – typically developing sample

Study 2 was designed to investigate why F3 values predicted from the norm-derived $F3_{pr}$ formula were higher than actual F3 values observed in study 1, with the ultimate goal of updating the formula to represent a better fit for children similar to the treatment recipients in study 1. To disambiguate the three potential explanations laid out above, $F3_{pr}$ was calculated from F3 values of vowels in a sample of TD children in two tasks. The first was a replication of the phrase task used to collect the normative data in Lee et al. (1999). The second task replicated a probe used to elicit /r/ in both syllables and words in the McAllister Byun et al. (2014) treatment study. Observed F3 values were compared with calculated $F3_{pr}$ and critical frequency values for each individual in each task. Three possible outcomes for these comparisons were anticipated, each compatible with one of the three previously proposed explanations.

The explanation most supported in previous literature was the hypothesis that different outcomes could be attributed to a task effect. Under this scenario, the norm-derived $F3_{pr}$ formula would accurately predict F3 of /ɜ/ elicited as one of many targets in a carrier phrase

context, but would be higher than F3 values of /ʒ/ elicited in isolated syllables/words in a task focused on rhotics. A second explanation considered was that the children in study 1 were hyperarticulating as a consequence of having just received treatment targeting rhotic misarticulation. This conclusion would be supported if predicted F3 values were not found to differ from actual F3 values in either the word or the phrase task for the sample of TD children in study 2. A third possible explanation was that the $F3_{pr}$ formula was fit to normative group averages that were in some sense not representative of the current sample. This conclusion would be supported if predicted F3 values were found to be greater than actual F3 values in both tasks in study 2.

Method

Participants—Thirteen TD children were recruited from a sample of recent participants in a similar study (McAllister Byun & Tiede, 2017). In both the original and current studies, inclusionary criteria included parent report of no major neurobehavioural, speech-language, or hearing impairment and passing a hearing screening at 500, 1000, 2000, and 4000 Hz at 20 dB HL. The 13 participants initially enrolled included eight females and five males with a mean age of 13;1 (SD = 22.5 months), ranging from 9;4–15;10. After two exclusions, which will be described in the next section, the final sample of 11 children (seven females and four males) had a mean age of 12;11 (SD = 22.2 months), ranging from 9;4 to 15;3.

Stimuli and measurement—Target vowel and /ʒ/ productions were collected in two different tasks for each participant. Recording parameters and formant measurement procedures were identical to those reported for study 1, with the exception that recordings were elicited in a soundproof booth at a sampling frequency of 44,100 Hz.

The first task involved production of 10 CVC target words five times each within the carrier phrase ‘I say a __ again’, following the protocol in Lee et al. (1999). In this condition, henceforth referred to as the ‘phrase task’, participants were instructed to speak in a normal voice as if talking to a friend. Although only a subset of words (/i/ ‘bead’, /æ/ ‘bat’, /ɑ/ ‘pot’, /u/ ‘boot’, and /ʒ/ ‘bird’) were measured for use with the $F3_{pr}$ formula, the full set of words was elicited (/ʌ/ ‘but’, /ɔ/ ‘ball’, /ɛ/ ‘bet’, /ɪ/ ‘bit’, and /ʊ/ ‘put’) to ensure fidelity of the replication.

The second task involved production of syllables and words, collected in a fashion similar to the routine /r/ probes in McAllister Byun et al. (2014). Stressed syllabic /ʒ/ was elicited 20 times within words (five productions each of ‘her’, ‘purr’, ‘fur’, ‘sir’), whereas the corner vowels were elicited within syllables (three productions each of /ri/, /ræ/, /rɑ/, /ru/).³ Together, these tasks will be referred to as the ‘syllable/word task’. To elicit a careful speech condition, participants were instructed to speak in a clear voice as if they were talking to someone who has difficulty understanding English (Ghosh et al., 2010). This condition was intended to induce hyperarticulated production in the TD children, as a careful condition is associated with more extreme formant values than a casual condition (Ferguson & Kewley-

³All word/syllable stimuli were presented orthographically for each participant to produce. For the syllables, each participant was familiarised with correct pronunciation in a practice block in which they were given feedback. If they made an error in syllable production during the trial, they were given a reminder.

Port, 2007). Due to the possibility of coarticulatory influence from preceding rhotics on vowels in the syllable/word task, all comparisons were within-task (i.e. phrase vowels versus phrase rhotics; syllable vowels versus word rhotics).

For data cleaning purposes, means and SD values of F3 and F3-F2 distances for all repetitions of each target word were plotted for all participants for both tasks. Plotted values that stood out on visual inspection were re-examined to detect any formant tracking errors. After re-measurement, two participants with means for /i/, /æ/, /ɑ/, /u/, or /ɜ/ that fell outside two standard deviations of the mean F3 or mean F3-F2 distance for all participants were excluded from further analysis. The excluded participants were a bidialectal child from a family speaking British English (B5), who exhibited outlier F3 and F3-F2 values for /u/ in both tasks, and a bidialectal child from a family speaking African-American Vernacular English (G5), who exhibited outlier F3 and F3-F2 values for /ɜ/ in both tasks, as well as outlier F3 of /æ/ and /ɑ/ and F3-F2 of /u/ in the phrase task only. However, there is no indication that a non-rhotic dialect substrate was influencing the Midwestern English speakers in Lee et al. (1999), so these exclusions were considered appropriate.

Analyses—For each participant, $F3_{pr}$ was calculated by applying the appropriate scale factor (0.70 for females and 0.67 for males) to the average of the corner vowels, calculated separately for the phrase task and the syllable/word task. Paired Wilcoxon signed-rank tests were used to determine whether $F3_{pr}$ values differed from the average F3 of /ɜ/ in the corresponding task across children. The scale factor that would need to be applied to the corner vowel average in order to obtain the mean F3 of /ɜ/ value was then calculated for each child in each task. Finally, observed F3 values were compared against the 80% critical frequency value. To investigate the role of dialectal difference in the phrase task, paired Wilcoxon signed-rank tests were conducted to compare the mean F3 for each vowel across all participants with corresponding normative values.

Results

Table 2 presents a summary of the measurements that formed the basis of the $F3_{pr}$ comparison, including vowel averages, $F3_{pr}$ values, and actual F3 of /ɜ/ values by participant for both tasks. The mean of actual F3 values was lower than the $F3_{pr}$ value in 9 out of 11 participants for the phrase task (exceptions were subjects B3 and G7), while the mean of actual F3 values was lower than the predicted value for all participants in the syllable/word task. Across participants, the predicted $F3_{pr}$ values for the phrase task exceeded actual F3 of /ɜ/ values by an average of 159.2 Hz, while $F3_{pr}$ values for the syllable/word task exceeded actual F3 of /ɜ/ values by an average of 207.62 Hz; both differences were significant ($V = 63$, $p = 0.005$; $V = 66$, $p = 0.001$). As expected, $F3_{pr}$ was generally lower for the syllable/word task than the phrase task, which was attributed to the fact that /r/preceded vowels in the syllable/word task, potentially exerting a coarticulatory lowering influence on F3 in adjacent segments (Tunley, 1999).

Scale factors that could be applied to the average F3 across corner vowels to obtain the observed mean F3 of /ɜ/ for each child are reported in Table 3. Averaged across all children, scale factors of 0.64 and 0.62 were calculated for the phrase task and the syllable/ word task,

respectively. When participants were subdivided by sex, the scale factors for females had an average of 0.65 for the phrase task (range 0.60–0.72) and 0.63 for the syllable/word task (range 0.56–0.68), while the scale factors for males were 0.61 for the phrase task (range 0.57–0.67) and 0.60 for the syllable/word task (range 0.55–0.64).

The comparison of individuals' F3 of /ʒ/ values (red dots) against the critical frequency (green triangles) is depicted in Figure 4. Across participants, no token fell above the 80% upper limit for either task, in support of the 80% critical frequency. The large difference noted between the critical frequency and actual F3 values highlights the need for a lowered scale factor. For comparison, the F3_{pr} values are also plotted (blue diamonds), and indicate that an overwhelming majority of F3 values are at or below the predicted values.

Finally, at a Bonferroni corrected alpha level of 0.005, the Lee F3 norms were found to be significantly greater than actual F3 values for /ʒ/ ($V = 65$, $p = 0.002$) and /ʌ/ ($V = 63$, $p = 0.0049$), while differences that could be considered marginally significant were found for /æ/ ($V = 61$, $p = 0.0098$), /e/ ($V = 59$, $p = 0.019$), and /ɪ/ ($V = 56$, $p = 0.042$).

Discussion

Results from study 2 show that F3_{pr} values calculated using the formula in (2) were significantly greater than actual F3 of /ʒ/ values in both the phrase and syllable/word tasks, suggesting that the scale factors derived from the normative data were not an optimal fit for the children in the current sample. Competing hypotheses that there had been a task effect or that the children with remediated rhotic errors exhibited a particularly high level of hyperarticulation were not supported.

The finding that norm-derived F3_{pr} values did not fit the children in the current sample corresponded with various potential explanations. One possibility was that the relationship between vowels and /ʒ/ observed in the current reanalysis of Lee et al. (1999) was a consequence of group averaging, and individual child speakers could show a distribution of F3_{pr} values both above and below the group mean value. The present data did not support this possibility, as most participants' F3_{pr} values exceeded the mean observed F3 of /ʒ/, while no participant's F3_{pr} value was substantially lower than the observed F3.⁴ It is thus unlikely that a different sample of children from the same population would yield F3_{pr} values falling significantly below their actual F3 values. Although findings from the current sample are sufficiently consistent across individual children to appear robust, future research examining a larger sample of children would allow stronger conclusions to be drawn.

Another possibility is that regional dialectal factors could explain the observed differences between the normative sample, collected in the Midwestern USA, and study 2, collected in the New York City area. The observed differences between the two samples in F3 of /ʒ/ and of non-rhotic vowels support the hypothesis that the source of the error in fitting the formula might be attributable to dialectal differences in acoustics of /ʒ/ and vowels. On the other hand, F3 is not generally considered an acoustic target in vowel production (Ladefoged,

⁴The two children who were excluded may have had F3 of /ʒ/ values that were greater than F3_{pr}, contrasting with the children retained in the analysis. As noted above, exclusion of these individuals exposed to a non-rhotic dialect was considered appropriate from the standpoint of replicating Lee et al. (1999).

1993), and previous dialectal comparisons of vowels have not reported significant F3 differences (Hagiwara, 1997). Thus, more research should investigate whether the observed cross-dialectal differences in F3 of /ɜ/ and non-rhotic vowels is robust.

Alternatively, it could be that the differences between the Lee norms and the actual F3 values were driven by group differences not related to regional dialect, such as longitudinal changes in children's heights. Height and vocal tract length are negatively related to formant heights (González, 2004) and population height has increased over time in the US (Komlos & Lauderdale, 2007). It is thus possible that the higher F3 values found in the normative sample could be attributed to generally shorter heights and correspondingly smaller vocal tract lengths in the Lee et al. (1999) sample. Whether the discrepancy between predicted and actual F3 values in the current study was due to dialect or physical size differences, we reach the same conclusion that the scale factors need to be updated for use with the population represented by the current sample. Further research is needed to know whether the same would be true for other regions.

General discussion

Studies 1 and 2 showed that the formula derived from normative data yielded $F3_{pr}$ values that were higher than actual F3 of /ɜ/ values in children with remediated rhotic errors and in TD children. After ruling out alternative explanations, it was judged that intrinsic differences between the normative sample and the current sample represented the most likely explanation for this discrepancy. The corresponding next step is to lower the scale factors to represent F3 of /ɜ/ values within phrases or words/syllables for TD children. The 80% critical frequency formula (1) represents a strong upper bound under which all F3 of /ɜ/ tokens are expected to fall. In practice, clinicians may wish to emphasise a considerably lower value when targeting typical acoustics for /ɜ/, especially in the context of biofeedback treatment. Based on the data from this study, the following revised scale factors are proposed for therapeutic use with children from the northeast US in either a phrase or word/syllable task⁵:

$$\begin{aligned} \text{a. revised } F3_{pr}(\text{males}) &= \left\lfloor \frac{(F3_i + F3_{æ} + F3_a + F3_u)}{4} \right\rfloor * 0.605 \quad (3) \\ \text{b. revised } F3_{pr}(\text{females}) &= \left\lfloor \frac{(F3_i + F3_{æ} + F3_a + F3_u)}{4} \right\rfloor * 0.64 \end{aligned}$$

Limitations

Although adequate for the population of study, this research may be limited in the extent to which it can be applied across diverse populations. First, the revised formula was fitted to TD children but was intended to be used with children with SSD. Revisiting the two remediated children from study 1, the two participants' scale factors (0.63 and 0.67) fit

⁵Scale factors pooled across gender but separated by task were 0.64 for phrases and 0.62 for syllables/ words. Since a larger difference was found between scale factors pooled across tasks but separated by gender, gender was chosen as the partition variable.

within the range observed for TD females in the rhotic-focused syllable/word task (0.56 to 0.68). Shriberg et al. (2001) had hypothesised that there were different subtypes of /r/ misarticulation, but a larger sample size would be needed to state with confidence whether these two children were heterogeneous with respect to the subtype of SSD.

Second, the current formula was fitted solely to stressed syllabic /ɹ/ instead of the variety of /r/ contexts that were included in Hagiwara's study. There is disagreement in the literature about whether there are acoustic or articulatory differences in /r/ across allophonic contexts in US English (Curtis & Hardy, 1959; McGowan, Nitttrouer, & Manning, 2004; Lockenvitz, Kuecker, & Ball, 2015). However, some research has suggested that F3 of /r/ does vary in a meaningful way between contexts such as onset versus nuclear /r/. If this is the case, further research to calculate refined scale factors for different allophones of /r/ may be useful.

Third, scale factors for males were consistently lower than for females in the present sample, suggesting that males may display lower F3 in rhotics relative to non-rhotic vowels than females. Visual inspection showed no clear age-related differences in the relationship between vowel averages and /ɹ/ in the present sample. However, further research with a larger sample size is unquestionably needed before we can conclude that the observed gender difference is robust and determine whether age or other factors play a significant role.

Clinical implications

Each of the presented caveats to the $F3_{pr}$ formula should be considered in future attempts to precisely predict the acoustic properties of children's perceptually correct /r/ on the basis of non-rhotic vowel measurements. However, the magnitude of imprecision is unlikely to affect the formula's usefulness as a treatment target. The average difference between predicted and actual F3 values across all children in both tasks in study 2 was +183.4 Hz, which was less than the average SD (228.2 Hz) across age/gender groups in the Lee et al. (1999) normative study. As an illustrative example, the SD of F3 for /r/ from 10 year-olds in the normative sample was 258 Hz for females and 211 Hz for males. Thus, the observed differences between predicted and actual values have raw acoustic magnitudes that would fall within the range of normal variation found among typical children of the same age and gender. This observation suggests that further refinement may be targeting a finer level of precision than is necessary for clinical purposes.

A final consideration is that the $F3_{pr}$ formula relies on the premise that F3 height is sufficient as the primary acoustic target to be emphasised in intervention. Since the F2 height contributes as a secondary cue in the perception of rhoticity (Idemaru & Holt, 2013), F3-F2 distance may be a more robust indicator of perceptual accuracy than F3 alone (Flipsen, Shriberg, Weismer, Karlsson, & McSweeney, 2001; McAllister Byun & Tiede, 2017). Furthermore, z-scores of F3-F2 and F3/F2 have been shown to control for age/sex differences across individuals (Flipsen et al., 2001). However, these z-scores are calculated relative to age and gender normative values, so the impact of F2 in target selection is unknown. Future biofeedback research should investigate whether visual-acoustic targets that incorporate both F2 and F3 would be more effective than F3 targets alone.

A formula to generate individualised rhotic F3 targets addresses the need for efficient and accurate ways to select appropriate acoustic goals for individuals enrolled in biofeedback therapy. The individualised F3 target generation formula can immediately be incorporated into acoustic biofeedback therapy apps (e.g. McAllister Byun et al., in press), which could help increase uptake of biofeedback and fidelity in implementation. Other than providing a target for therapy, the formula has the potential to support automated scoring of /r/ accuracy by defining a confidence interval around an individual's target F3. Both target generation and automated scoring will be valuable features to SLPs who use acoustic biofeedback applications for clients with /r/ misarticulation. SLPs may also be able to use formula-derived targets when working with second-language learners who wish to reduce their accent by matching their formants for /r/ (or other sonorants visible on an acoustic spectrum) to those of native speakers.

Conclusion

The primary focus of the current study was on refining a formula to predict F3 heights in rhotic targets based on acoustic measurements of non-rhotic vowels at the individual speaker level. Previous formulas to estimate F3 of /r/ were adapted to child samples using normative data from Lee et al. (1999). Study 1 showed that the rhotic F3 values predicted by the norm-derived formula were significantly higher, i.e. less rhotic, than actual F3 values measured in perceptually accurate /ʒ/ produced by two children with remediated /ʒ/. Study 2 showed that predicted F3 values derived with the same formula were also significantly higher than actual F3 of /ʒ/ values measured from a sample of typically developing children. Further investigation suggested that this discrepancy most likely arose because the Lee et al. (1999) normative sample was not optimally representative of the individual children in either of the two current samples. In response, the scale factors used in the formula for predicting F3 of /ʒ/ were lowered from the norm-derived values, yielding scale factors of 0.64 for females and 0.61 for males. The proposed formula contributes to an empirical understanding of the acoustic properties of rhotics, and can immediately be applied in the context of target-setting for acoustic biofeedback therapy. This formula to predict /r/ targets addresses a current need for a simple, straightforward way to identify individualised acoustic targets for use in visual-acoustic biofeedback intervention for rhotics, which may in turn increase uptake of this evidence-supported intervention.

Acknowledgments

The authors gratefully acknowledge the contributions of numerous student assistants at NYU. Many thanks as well to all participants and their families for their cooperation throughout the study.

Funding

This work was supported by the National Institutes of Health (NIH) under grant #R01DC013668. The authors are involved in the development of staRt, an app to provide visual-acoustic biofeedback treatment. They do not receive financial compensation for their role in developing the app.

References

- Boersma, P., Weenink, D. Praat: Doing phonetics by computer (version 5.3.84) [computer program] 2014 Retrieved from <http://www.fon.hum.uva.nl/praat/>

- Boyce, SE, , Hamilton, SM, , Scholl, LM, , Schmidlin, S. Paper presented at the American Speech-Language Hearing Association Convention Atlanta, GA: 2013 Defining the acoustic targets for correct production of /r/.
- Curtis JF, Hardy JC. A phonetic study of misarticulation of /r/. *Journal of Speech, Language, and Hearing Research*. 1959; 2(3):244–257.
- Delattre P, Freeman DC. A dialect study of American r's by x-ray motion picture. *Linguistics*. 1968; 6(44):29–68.
- Espy-Wilson CY, Boyce SE, Jackson M, Narayanan S, Alwan A. Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America*. 2000; 108(1):343–356. [PubMed: 10923897]
- Ferguson SH, Kewley-Port D. Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*. 2007; 50(5):1241–1255.
- Flipsen P. Emergence and prevalence of persistent and residual speech errors. *Seminars in Speech and Language*. 2015; 36(4):217–223. [PubMed: 26458197]
- Flipsen P, Shriberg LD, Weismer G, Karlsson HB, McSweeney JL. Acoustic phenotypes for speech-genetics studies: reference data for residual /s/ distortions. *Clinical Linguistics & Phonetics*. 2001; 15(8):603–630.
- Ghosh SS, Matthies ML, Maas E, Hanson A, Tiede M, Ménard L, Perkell JS. An investigation of the relation between sibilant production and somatosensory and auditory acuity. *Journal of the Acoustical Society of America*. 2010; 128(5):3079–3087. [PubMed: 21110603]
- González J. Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics*. 2004; 32(2):277–287.
- Hagiwara, R. (Ph D dissertation) UCLA; Los Angeles, CA: 1995 Acoustic realizations of American /r/ as produced by women and men.
- Hagiwara R. Dialect variation and formant frequency: the American English vowels revisited. *Journal of the Acoustical Society of America*. 1997; 102(1):655–658.
- Idemaru K, Holt LL. The developmental trajectory of children's perception and production of English /r-/l/. *Journal of the Acoustical Society of America*. 2013; 133(6):4232–4246. [PubMed: 23742374]
- Komlos J, Lauderdale BE. The mysterious trend in American heights in the 20th century. *Annals of Human Biology*. 2007; 34(2):206–215. [PubMed: 17558591]
- Ladefoged, P. A course in phonetics 3rd. Fort Worth, TX: Harcourt Brace & Company; 1993
- Lee S, Potamianos A, Narayanan S. Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*. 1999; 105(3): 1455–1468. [PubMed: 10089598]
- Lewis, BA, , Shriberg, LD. Talk presented at the Miniseminar at the American Speech-Language Hearing Association Convention New Orleans, LA: Nov, 1994 Life span interrelationships among speech, prosody-voice, and nontraditional phonological measures.
- Lockenvitz S, Kuecker K, Ball MJ. Evidence for the distinction between “consonantal-/r/” and “vocalic-/r/” in American English. *Clinical Linguistics & Phonetics*. 2015; 29(8–10):613–622. [PubMed: 26172586]
- Maas E, Robin DA, Hula SNA, Freedman SE, Wulf G, Ballard KJ, Schmidt RA. Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*. 2008; 17(3):277–298. [PubMed: 18663111]
- Magloughlin L. Accounting for variability in North American English /ɹ/: Evidence from children's articulation. *Journal of Phonetics*. 2016; 54:51–67.
- McAllister Byun T, Campbell H. Differential effects of visual-acoustic biofeedback intervention for residual speech errors. *Frontiers in Human Neuroscience*. 2016; 10(567):1–17. [PubMed: 26858619]
- McAllister Byun T, Campbell H, Liang W, Park TH, Svirsky M. Enhancing intervention for residual rhotic errors via app-delivered biofeedback: a case study. *Journal of Speech, Language, and Hearing Research*. (in press).

- McAllister Byun T, Hitchcock ER. Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology*. 2012; 21(3):207–221. [PubMed: 22442281]
- McAllister Byun T, Hitchcock ER, Swartz MT. Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research*. 2014; 57(6):2116–2130.
- McAllister Byun T, Tiede M. Perception-production relations in later development of American English rhotics. *Plos One*. 2017; 12(2):e0172022. [PubMed: 28207800]
- McGowan RS, Nittrouer S, Manning CJ. Development of [ɹ] in young, Midwestern, American children. *Journal of the Acoustical Society of America*. 2004; 115(2):871. [PubMed: 15000198]
- Ruscello DM. Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders*. 1995; 28(4):279–302. [PubMed: 8576411]
- Shriberg LD. An intervention procedure for children with persistent /r/ errors. *Language, Speech, and Hearing Services in Schools*. 1980; 11(2):102–110.
- Shriberg LD, Flipsen P, Karlsson HB, McSweeney JL. Acoustic phenotypes for speech-genetics studies: an acoustic marker for residual /ʁ/ distortions. *Clinical Linguistics & Phonetics*. 2001; 15(8):631–650.
- Shuster LI, Ruscello DM, Toth AR. The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology*. 1995; 4(2):37–44.
- Tiede MK, Boyce SE, Holland CK, Choe KA. A new taxonomy of American English /r/ using MRI and ultrasound. *Journal of the Acoustical Society of America*. 2004; 115(5):2633–2634.
- Tjaden K, Wilding GE. Rate and loudness manipulations in dysarthria: acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research*. 2004; 47(4):766–783.
- Tunley, A. (Ph.D. dissertation) University of Cambridge; Cambridge, England: 1999 Coarticulatory influences of liquids on vowels in English.
- Van Bergem DR. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*. 1993; 12(1):1–23.
- Van Riper, C. *Speech correction: Principles and methods* 6th. Englewood Cliffs, NJ: Prentice-Hall; 1978

Figure 1.

A screen shot of a live LPC spectrum with a superimposed target F3 of correct /r/ (line with starfish), adapted from 'staRt' the Speech Therapists' App for /r/ Treatment from McAllister Byun, Campbell, Liang, Park, and Svirsky (in press).

Figure 2.

For each age group for females (top) and males (bottom) in Lee et al. (1999) normative data, dark bars represent average of corner vowels /i,æ,ɑ,u/ while light bars represent actual F3 of /ɜ/ values. To the nearest hundredth, predicted F3 of /ɜ/ values (dots) correspond most accurately with actual F3 of /ɜ/ values at 70% of vowel average for females and at 67% of vowel average for males.

Figure 3.

Plotted F3 of vocalic /ə/ tokens for Autumn and Lilianne across the course of ultrasound biofeedback therapy in McAllister Byun et al. (2014). Vocalic /ə/ production in words improved from baseline (BS) to maintenance (MN) sessions, consistent with the decreasing F3 across treatment (TX) sessions. The grey dashed line is the 80% critical frequency. The solid black line is the $F3_{pr}$ value, which was 70% of the vowel average for females.

Figure 4.

The 80% critical frequency (green triangles), predicted F3 of /s/ (blue diamonds), and actual F3 of /s/ values (dots) for each participant and each task.

Table 1

Token frequency of syllables/words for each vowel and /ʒ/, selected from probes administered over the course of the McAllister Byun et al. (2014) treatment study. Average automatically extracted third formant frequencies for each vowel, the vowel average, and all /ʒ/, all correct /ʒ/, and all incorrect /ʒ/ are reported for both participants.

CHILD	SYLLABLE/WORD	SOUND	AVERAGE F3 (Hz)
AUTUMN	<i>read</i> (9); <i>ri</i> (6)	/i/	3568.1
	<i>crab</i> (8); <i>crack</i> (8)	/æ/	3050.1
	<i>ra</i> (6); <i>rock</i> (9)	/ɑ/	3269.3
	<i>ru</i> (6); <i>root</i> (9)	/u/	3220.4
		AVERAGE /i,æ,ɑ,u/	3277.0
	<i>her, purr, fur, sir</i>	ALL /ʒ/	3034.3
	(18 each)	CORRECT /ʒ/	2070.6
		INCORRECT /ʒ/	3722.6
LILIANNE	<i>read</i> (8); <i>ri</i> (6)	/i/	3406.9
	<i>crab</i> (7); <i>crack</i> (7)	/æ/	3019.4
	<i>ra</i> (6); <i>rock</i> (8)	/ɑ/	3147.5
	<i>ru</i> (6); <i>root</i> (8)	/u/	3181.0
		AVERAGE /i,æ,ɑ,u/	3188.7
	<i>her, purr, fur, sir</i>	ALL /ʒ/	2581.9
	(16 each)	CORRECT /ʒ/	2145.8
		INCORRECT /ʒ/	3076.2

Table 2

For each participant in each task, vowel averages were multiplied by a scale factor (by sex) to determine predicted F3 of /s/ value, which is viewed next to actual measured F3 of /s/. Actual mean /s/ F3s that are lower than predicted, and thus driving the significant finding, are marked with an asterisk.

	Subject	Scale factor	Phrase task				Syllable/Word task			
			Average vowels	Predicted /s/ F3	Actual /s/ F3		Average vowels	Predicted /s/ F3	Actual /s/ F3	
Males	B1	.67	3273.8	2193.5	1908.7*		3181.8	2131.8	1736.1*	
	B2		3140.4	2104.1	1796.6*		3042.2	2038.3	1734.9*	
	B3		2604.6	1745.1	1746.0		2455.6	1645.3	1560.7*	
	B4		2621.5	1756.4	1651.2*		2449.1	1640.9	1540.9*	
Females	G1	.70	2854.0	1997.8	1877.4*		2806.4	1964.5	1897.3*	
	G2		3110.3	2177.2	2009.5*		3092.0	2164.4	1821.6*	
	G3		3075.0	2152.5	2069.9*		3026.5	2118.6	1952.3*	
	G4		3177.7	2224.4	1906.4*		3117.5	2182.3	1747.3*	
	G6		2886.8	2020.7	1748.6*		2773.3	1941.3	1826.5*	
	G7		3042.0	2129.4	2203.2		2974.9	2082.4	1983.6*	
	G8		2689.5	1882.7	1715.2*		2689.9	1882.9	1707.6*	
All	Mean		2952.3	2034.9	1875.7		2873.6	1981.2	1773.5	

Table 3

Scale factors needed to predict F3 of /s/ accurately from average of measured vowels for each participant within each task in study 2.

Subject	Age	Sex	Scale factor (Phrase Task)	Mean scale factor	Scale factor (Syllable/Word Task)	Mean scale factor
B1	9;4	M	0.58	.61	0.55	.60
B2	11;8	M	0.57		0.57	
B3	13;11	M	0.67		0.64	
B4	15;1	M	0.63		0.63	
G1	11;4	F	0.66	.65	0.68	.63
G2	12;1	F	0.65		0.59	
G3	12;3	F	0.67		0.65	
G4	12;7	F	0.60		0.56	
G6	13;11	F	0.61		0.66	
G7	14;11	F	0.72		0.67	
G8	15;3	F	0.64		0.63	