Enhancing Intervention for Residual Rhotic Errors Via App-Delivered Biofeedback: A Case Study

Tara McAllister Byun,[a] Heather Campbell,[a] Helen Carey,[b]
Wendy Liang,[a] Tae Hong Park,[a] and Mario Svirsky[c]

Purpose: Recent research suggests that visual-acoustic biofeedback can be an effective treatment for residual speech errors, but adoption remains limited due to barriers including high cost and lack of familiarity with the technology. This case study reports results from the first participant to complete a course of visual-acoustic biofeedback using a not-for-profit iOS app, Speech Therapist's App for /r/ Treatment.
Method: App-based biofeedback treatment for rhotic misarticulation was provided in weekly 30-min sessions for 20 weeks. Within-treatment progress was documented using clinician perceptual ratings and acoustic measures. Generalization gains were assessed using acoustic measures of word probes elicited during baseline, treatment, and maintenance sessions.

Results: Both clinician ratings and acoustic measures indicated that the participant significantly improved her rhotic production accuracy in trials elicited during treatment sessions. However, these gains did not transfer to generalization probes.
Conclusions: This study provides a proof-of-concept demonstration that app-based biofeedback is a viable alternative to costlier dedicated systems. Generalization of gains to contexts without biofeedback remains a challenge that requires further study. App-delivered biofeedback could enable clinician–research partnerships that would strengthen the evidence base while providing enhanced treatment for children with residual rhotic errors.
Supplemental Material: https://doi.org/10.23641/asha. 5116318

This special issue contains selected papers from the March 2016 Conference on Motor Speech held in Newport Beach, CA.

Deviations in speech sound production can impact intelligibility and pose a barrier to participation across academic and social domains, with potentially negative consequences for educational and occupational outcomes (Hitchcock, Harel, & McAllister Byun, 2015; McCormack, McLeod, McAllister, & Harrison, 2009). While speech production normalizes in most children by 8–9 years of age, a subset of older speakers continue to exhibit distorted production of certain phonemes, described as residual speech errors (Shriberg, Gruber, & Kwiatkowski,

1994). In American English, distortions of the rhotic sounds /ɝ/ and /ɹ/ are among the most common residual speech errors. Clinicians report finding rhotic errors particularly challenging to remediate, and many clients are discharged with these errors still unresolved despite months or years of therapy (Ruscello, 1995).

Recent evidence suggests that visual biofeedback technologies may enhance the efficacy of intervention for residual rhotic errors (McAllister Byun & Hitchcock, 2012; McAllister Byun, Swartz, Halpin, Szeredi, & Maas, 2016; Preston et al., 2014). In visual biofeedback, instrumentation is used to provide real-time information about aspects of speech that are typically outside the speaker's conscious awareness or control (Davis & Drichta, 1980). Learners view a model representing the target speech behavior, often side-by-side with or superimposed on the real-time feedback display, and are encouraged to explore different production strategies to make their output match the model.

The rationale for biofeedback is rooted in the literature investigating principles of motor learning, which hypothesizes that different conditions of practice and feedback can impact the acquisition and generalization of new motor

[a]Steinhardt School of Culture, Education, and Human Development, New York University
[b]Tandon School of Engineering, New York University
[c]Langone Medical Center, New York University

Correspondence to Tara McAllister Byun: tara.byun@nyu.edu

skills (Bislick, Weir, Spencer, Kendall, & Yorkston, 2012; Maas et al., 2008). Biofeedback is a form of detailed *knowledge of performance* feedback (Volin, 1998). As such, it is predicted to facilitate the acquisition of new motor skills (Maas et al., 2008; Preston, Brick, & Landi, 2013), although its impact on long-term retention and generalization may be neutral or even detrimental (Hodges & Franks, 2001; Maas et al., 2008). It has also been suggested that biofeedback may have its effect by encouraging speakers to adopt an external direction of attentional focus (McAllister Byun et al., 2016), which has been found to speed the acquisition of motor skills in nonspeech contexts (see discussion in Maas et al., 2008).

Various technologies, such as ultrasound and electropalatography, have been used to provide real-time information about the location and movements of the articulators during speech (e.g., Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007; Gibbon, Stewart, Hardcastle, & Crampin, 1999; McAllister Byun, Hitchcock, & Swartz, 2014; Preston et al., 2013, 2014). Another alternative is *visual-acoustic biofeedback,* such as a dynamic visual display of the formants or resonant frequencies of the vocal tract. Figure 1 displays a linear predictive coding (LPC) spectrum with frequency on the *x*-axis and amplitude on the *y*-axis; the peaks of the spectrum represent formants. Rhotics are distinguished acoustically from other sonorants by the low height of the third formant (F3), which closely approximates the second formant (F2). Both case studies (Shuster, Ruscello, & Smith, 1992; Shuster, Ruscello, & Toth, 1995) and single-subject experimental studies (McAllister Byun, 2017; McAllister Byun & Campbell, 2016; McAllister Byun, Halpin, & Szeredi, 2015; McAllister Byun & Hitchcock, 2012) have reported that visual-acoustic biofeedback featuring a lowered F3 target can improve rhotic production in speakers who have not responded to other forms of intervention. One caution that has been raised in previous studies of various types of biofeedback (e.g., Gibbon & Paterson, 2006; McAllister Byun & Hitchcock, 2012; Preston et al., 2014) is that gains made in the treatment setting do not automatically generalize to contexts in which enhanced feedback is not available. If generalization
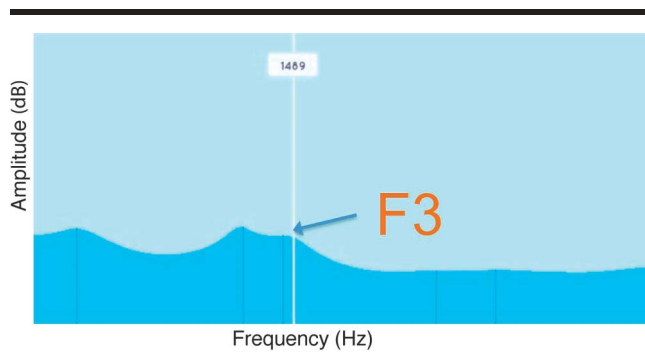
is not immediately achieved, it may be encouraged through a period of follow-up treatment in which task difficulty is adjusted in an adaptive fashion, including expanded opportunities for practice without biofeedback (Hitchcock & McAllister Byun, 2014).

Despite evidence that visual-acoustic biofeedback can be an effective supplement to treatment for residual rhotic errors, few clinicians have adopted the method to date. In this case, the high cost and complexity of the required equipment may represent a major limiting factor. From the point of view of *implementation science* (e.g., Olswang & Prelock, 2015), intervention researchers should not only collect evidence regarding the efficacy of different treatment approaches, but also work to overcome barriers to widespread uptake of evidence-supported methods. Giving practitioners access to a low-cost tool for visual-acoustic biofeedback could help bridge the research–practice gap in the treatment of residual speech errors. Making the tool user-friendly and easy to navigate could further lower barriers to adoption of biofeedback (Muñoz, Hoffman, & Brimo, 2013). To address these aims, we undertook the development of an inexpensive, accessible iOS app for clinical use with individuals with residual rhotic errors.

The staRt App

In app-based biofeedback, mobile technology is used to generate a visual representation of the speech signal and present a model for learners to match. *Speech Therapist's App for /r/ Treatment* (staRt) is an iOS app currently in development at New York University (NYU) that aims to increase the number of speech-language pathologists (SLPs) using visual-acoustic biofeedback to augment treatment for residual rhotic errors. The core mechanism of the app is a real-time LPC spectrum on which a visual target is superimposed in the form of an adjustable line (see Figure 1). The target is positioned to encourage a low height of F3 for rhotic targets; different settings are suggested based on the user's age and sex. Following instructions provided by the app, the treating clinician can cue a client to produce a rhotic sound and make articulatory adjustments until their F3 peak approximates the target line. The wavelike real-time LPC display was selected because it is the form of visual-acoustic feedback that has been most extensively tested in previous literature (McAllister Byun, 2017; McAllister Byun & Campbell, 2016; McAllister Byun & Hitchcock, 2012; McAllister Byun et al., 2016). As an early-stage investigation of the functionality of app-based biofeedback, the present study avoided major deviations from previous research with respect to both the nature of biofeedback and the protocol for treatment delivery. However, minor design modifications have been incorporated with the goal of enhancing the user experience for both client and clinician.

Design decisions in the development of the staRt app were made with the goal of maximizing appeal across different age groups and genders. The biofeedback "wave" is presented in a beach-themed environment with a cheerful

Figure 1. Formant frequencies represented as peaks of a linear predictive coding spectral display in the Speech Therapist's App for /r/ Treatment (staRt) app. The third formant (F3), which is targeted in rhotic treatment, is labeled.

gender-neutral palette. A starfish character guides users through a tutorial explaining how to use the biofeedback display (see Figure 2). Manipulating the wave is framed as the primary challenge of the game, while the starfish and the clinician act as supportive mentors in an environment that encourages vocal experimentation and experiential learning. The user interface draws on principles from Google's material design language (https://design.google.com), which promotes ease of navigation by evoking the surfaces and edges of physical objects.

In accordance with best practices from the fields of engineering and design, the staRt interface is being developed through an iterative process informed by frequent informal user testing. The present study aimed to test a minimum viable product of staRt in a controlled laboratory setting. To focus attention on the core components of the app, the tested version included only the biofeedback display and those settings needed to control it. Future modifications will expand the functionality of the app while incorporating user feedback from the present study and future pilot studies. A video of the functioning of the staRt app at the time of writing is provided in online Supplemental Material S1.

Method

Participant

This case study reports results from the first individual to complete a course of staRt-based biofeedback treatment. The participant, pseudonym "Hannah," was a 13-year-old girl with residual rhotic errors who exhibited otherwise typical language, hearing, and cognitive development. Before enrolling in the study, she passed a hearing screening, structural and functional examination of the oral mechanism, and a standardized test of receptive language (Auditory Comprehension subtest of the Test of Auditory Processing Skills–Third Edition; Martin & Brownell, 2005). Hannah had received early intervention services for concerns about speech development and had seen a school SLP for 6 months and a private SLP for 6 months in the year prior to her involvement in biofeedback treatment studies at NYU. At the time she was initially evaluated, she was receiving speech treatment in a group at school every other week for 30 min, but this was suspended while lab-based intervention was in progress.

Immediately preceding the present study of app-based biofeedback treatment, Hannah had completed another research study at NYU in which she received ten 30-min sessions of traditional articulatory treatment and ten 30-min sessions of visual-acoustic biofeedback treatment using the Sona-Match module of the KayPentax Computerized Speech Lab (CSL; KayPentax, Lincoln Park, NJ). The treating clinician reported that Hannah could produce perceptually accurate rhotics at the syllable level while viewing the biofeedback display; the clinician's ratings also suggested a small increase in Hannah's ability to produce /r/ at the word level during treatment trials. However, Hannah showed no long-term improvement in response to either treatment; blinded listeners' ratings indicated nearly 0% accuracy in rhotic production at the word level across baseline, midpoint, and maintenance probes. Because Hannah had shown some

Figure 2. Still frame from the Speech Therapist's App for /r/ Treatment (staRt) introductory tutorial. Notable features include the side navigation panel, the biofeedback display, the starfish character, and sand dollars representing the user's progress through the tutorial.
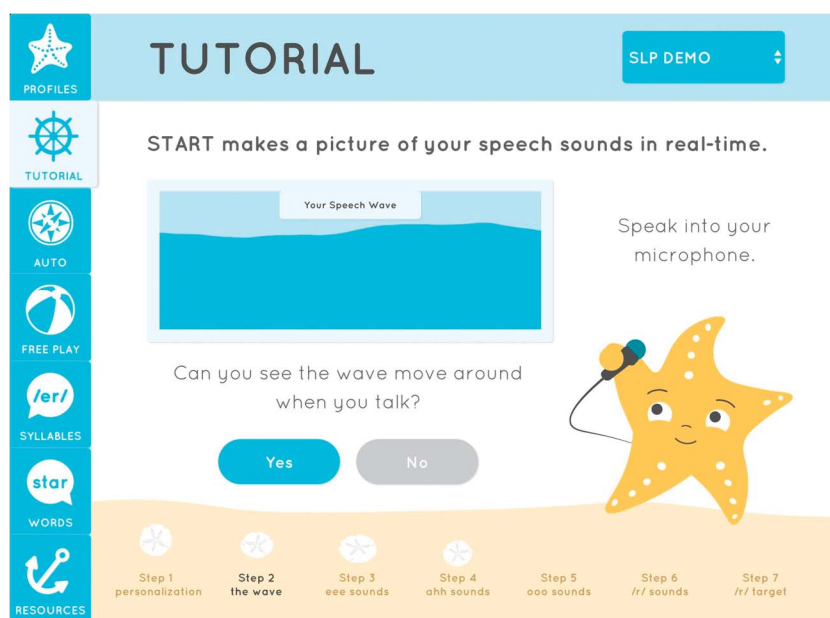
ability to benefit from enhanced visual feedback but her gains had not generalized beyond the context of treatment, she was judged to be a candidate for a follow-up phase of treatment in which she would receive additional opportunities to practice producing accurate rhotics both with and without biofeedback. It is important to note that the present study was not structured in a way that permits systematic comparison of the efficacy of biofeedback provided with the staRt app versus CSL Sona-Match. Rather, this preliminary study aimed only to test whether an individual who had shown some degree of therapeutic response to visual-acoustic biofeedback provided with standard CSL technology would continue to make gains when biofeedback was instead provided using the newly developed staRt app. Future research is needed to document the effects of app-based biofeedback provided without a preceding period of CSL-based treatment, as well as to compare staRt intervention against other types of biofeedback.

Study Design

This study was carried out over 28 weeks, including three baseline sessions, twenty 30-min treatment sessions, and three maintenance sessions provided on a roughly weekly basis. Baseline and maintenance sessions elicited a standard probe featuring 50 words containing rhotics in various phonetic contexts, presented in random order. All treatment sessions began with randomly ordered elicitation of a fixed 25-word subset of the standard probe. Treatment sessions were individually delivered by the second author, a certified SLP. User experience feedback was solicited through brief surveys that were administered at the start of the first session and at nine additional points over the course of treatment. Some of the survey questions required the user to rate her experience with various aspects of the app on a Likert scale, while other questions provided opportunities for open-ended commentary.

Visual-acoustic biofeedback with the staRt app, running on a model A1458 iPad operating iOS version 8.1.3 (Apple, Cupertino, CA), was provided as part of each treatment session. Hannah's F3 target was set to roughly 2000 Hz, an appropriate value for her age and gender (Lee, Potamianos, & Narayanan, 1999). Sessions began with 5 min of unstructured practice in which the participant, with cueing from the clinician, could explore different strategies to produce more accurate rhotics. This was followed by 60 structured trials in blocks of five. Each block began with one verbal focusing cue, which could be articulatory in nature or could refer to the app display. For example, the clinician could point out the client's third formant and remind her to focus on moving that peak to meet the vertical line representing her F3 target. The participant then produced five syllable or word trials from orthographic stimuli displayed on a screen. Biofeedback was visible throughout most trials, with exceptions as described below. After each block, the clinician provided summary feedback by indicating which trial in the preceding block she had judged to be most accurate.

A custom-designed software, Challenge-R (McAllister Byun, Hitchcock, & Ortiz, 2014) was used to implement a system of adaptive treatment difficulty in accordance with principles of motor learning (e.g., Maas et al., 2008). At the time of testing, this software was operated on a separate computer, but in future versions it will be incorporated directly into the staRt app. Based on the participant's accuracy over the preceding 10 trials, Challenge-R adjusts treatment parameters to make practice more difficult (e.g., reduce the number of trials in which biofeedback is available) or easier (e.g., restore a higher level of clinician scaffolding). Adjustments were determined automatically based on the clinician's online judgments of perceptual accuracy, which she entered into the software after each trial. Finally, treatment began with a limited subset of rhotic variants (/ɝ, ɑɚ, ɔɚ/), selected because these are considered relatively facilitative, early-emerging contexts for accurate rhotic production (Klein, McAllister Byun, Davidson, & Grigos, 2013). Although advancing to other rhotic variants was possible in principle, Hannah did not reach the requisite level of accuracy in this study.

Measurement

Both probe measures and treatment sessions were audio-recorded to a CSL Model 4150B (KayPentax) with a 44.1 kHz sampling rate and 16-bit encoding. Using Praat software (Boersma & Weenink, 2010), trained graduate assistants measured formants in each rhotic production from baseline, within-treatment, and maintenance probes, as well as a subset of treatment trials. Students followed a semiautomated protocol as described in McAllister Byun and Hitchcock (2012). The distance between the second and third formants (F3–F2 distance), which provides partial adjustment for individual differences in vocal tract length (Flipsen, Shriberg, Weismer, Karlsson, & McSweeny, 2001), is used as the primary measure of rhoticity in this study. Because accurate rhotics are characterized by a low F3 frequency, improvement in rhotic production accuracy should correspond with a reduction in F3–F2 distance.

To assess the reliability of acoustic measurement, 15% of files (3/26 probe files and 4/20 within-treatment files) were remeasured by a different student assistant. F3–F2 distances obtained from the original and remeasured files were compared using intraclass correlation with single random raters. The calculated intraclass correlation of .94 indicates strong agreement between measurements carried out by different individuals.

Results
Progress Within Treatment

Based on the clinician's ratings assigned in real time, Hannah's accuracy during biofeedback practice increased steadily from 1% correct in the first session to a maximum of 73% correct in session 10. At that point, Hannah began to reach higher levels of complexity in the Challenge-R hierarchy. Modifications triggered by this process included

reductions in biofeedback frequency (from 100% to 50% to 0%), fading of clinician models, and increased stimulus complexity (from the syllable to the word level). Hannah maintained a steady level of accuracy throughout these increases in complexity, which is evident in the plateau in Figure 3. By the final session, Hannah was judged to be 65% accurate in producing /r/ at the word level with no clinician model and no biofeedback.

Because ratings assigned in real time by the treating clinician are vulnerable to bias, acoustic measures of treatment trials were obtained as a corroborating data source. All 60 treatment trials were measured from every other treatment session, starting with the second session and ending with session 20. Figure 4 shows a steady decrease in mean F3–F2 distance in the first half of treatment, followed by a stable low level in sessions 10–20. The mean F3–F2 in sessions 10–20 hovered around 500 Hz, comparable to the mean F3–F2 distance of 493 Hz reported for rhotics produced by typical 13-year-old girls in Lee et al. (1999). In short, acoustic measures supported the treating clinician's ratings in showing that Hannah's rhotic production within the treatment setting improved in the first half of the study, then held steady at a relatively high level of accuracy in the second half.

As noted previously, Hannah's attitudes toward the staRt app were probed in short surveys administered at the beginning of roughly every other treatment session. Overall, she consistently expressed a positive attitude toward the app (4 on a scale of 5). Her ratings of how much the app helped her speech increased from 2/5 at the beginning of the study to 4/5 at the end, in keeping with the increased accuracy she exhibited within the treatment setting.

Figure 4. Mean F3–F2 distance in rhotic syllables/words in treatment trials. Bars represent 95% confidence interval around the mean.
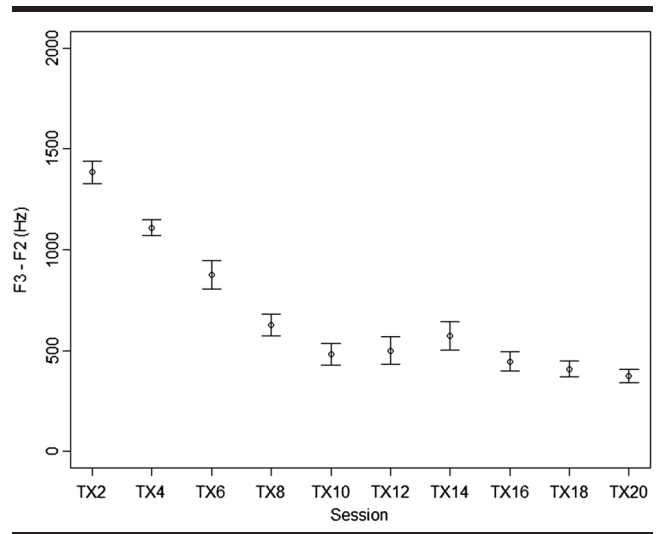
Generalization Word Probes

Although Hannah showed improved rhotic production during practice trials in the treatment setting, it is essential to consider how much these gains generalized to a non-treatment context, as measured through word probes administered in baseline and maintenance sessions and before each treatment session. Figure 5 depicts F3–F2 distances measured from rhotic word probes during baseline, treatment, and maintenance intervals. Because only a subset of

Figure 3. Points represent percentages of rhotic syllables/words rated perceptually accurate in treatment trials, reflecting scores assigned in real time by treating clinician. Shaded area represents 95% confidence interval around best-fit loess curve (line).
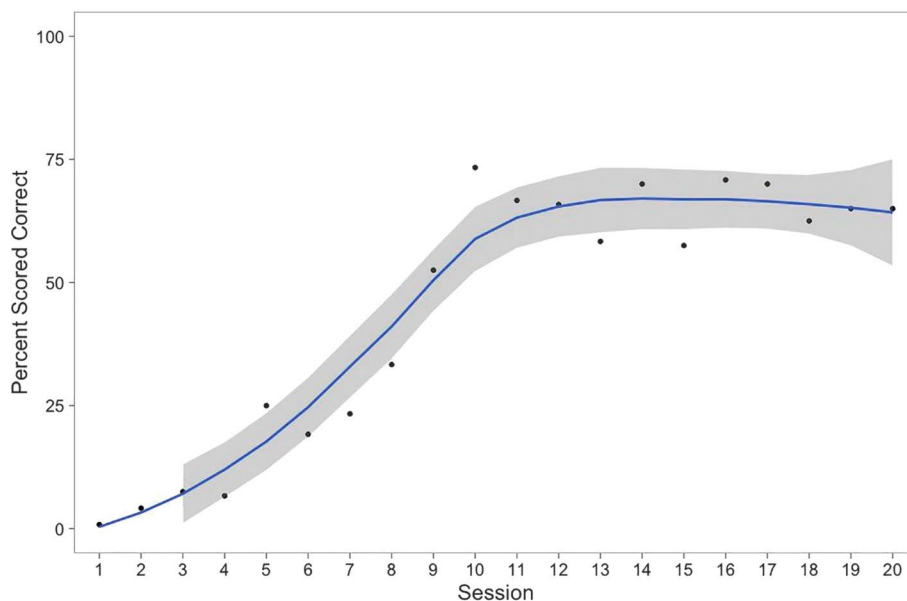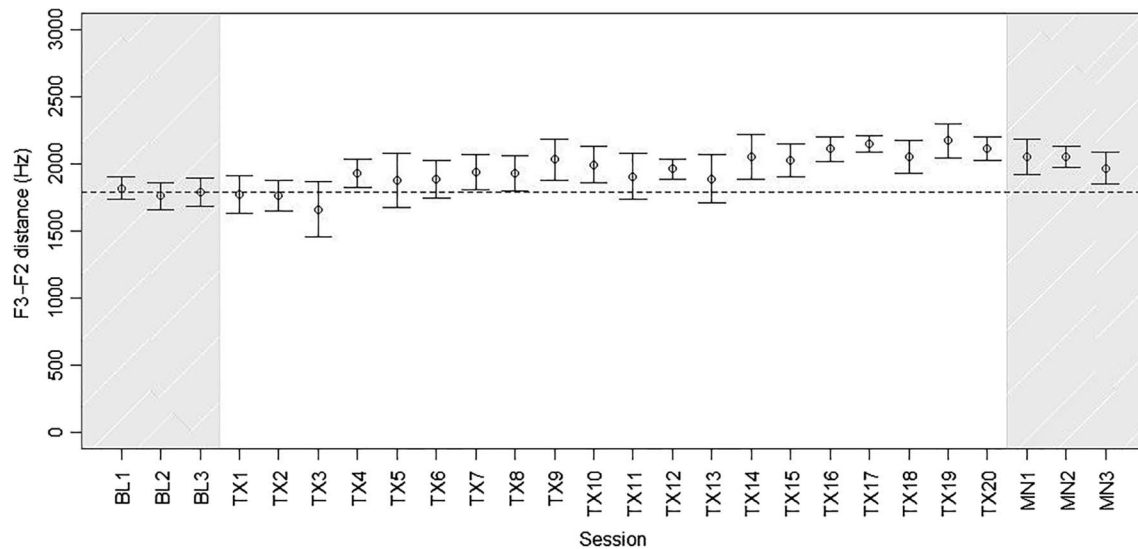
Figure 5. Mean F3–F2 distance in /ɝ, ɑɚ, ɔɚ/ words produced in baseline, within-treatment, and maintenance probe measures. Bars represent 95% confidence intervals. Dotted line represents mean F3–F2 distance across the three baseline sessions.

variants of /r/ were treated, and there was minimal generalization to untreated variants, the plots and analyses below focus on words representing the treated categories /ɝ, ɑɚ, ɔɚ/.[1] Even for these treated categories, Hannah showed no improvement on generalization probes over the course of treatment. In fact, mean F3–F2 distance was higher across the three maintenance sessions than it had been at baseline (i.e., less accurate). Although the difference in mean F3–F2 between baseline and maintenance phases was modest in magnitude (156 Hz), it was statistically significant ($t = -3.18$, $df = 134.7$, $p < .01$). The treating clinician's perceptual judgments were stable in indicating near-zero accuracy throughout baseline, treatment, and maintenance phases.

Discussion and Conclusion

This case study represents a proof-of-concept test of the functionality of a new app to provide visual-acoustic biofeedback for residual rhotic errors. Prior to her enrollment in this study, Hannah received biofeedback treatment using CSL Sona-Match, a commercially available software that has been tested in multiple published studies (McAllister Byun, 2017; McAllister Byun & Campbell, 2016; McAllister Byun & Hitchcock, 2012; McAllister Byun et al., 2016). Although she showed some improvement within the treatment setting in the previous study, generalization was minimal, and Hannah was judged to be a candidate for further treatment incorporating biofeedback. The present study was conducted to determine whether she would continue to make progress when biofeedback was

instead provided with the staRt app, and whether generalization to contexts in which biofeedback was not available would occur. In this study, both the treating clinician's ratings and acoustic measures showed that Hannah made strong gains within the context of treatment while using staRt biofeedback; accurate production within the treatment setting continued as the frequency of biofeedback was faded from 100% to 50% to 0%. However, there was still no generalization of within-treatment gains to word-level probe measures elicited without feedback. In fact, acoustic measures showed a small but significant *increase* in F3–F2 distance, indicating less strongly rhotic productions, from the baseline to the maintenance period.[2]

As a case study reporting a single participant's response to treatment, the present investigation is highly limited in the scope and strength of conclusions that can be drawn. We can answer our most basic question—is an app-generated LPC a viable option for the provision of visual-acoustic biofeedback?—in the affirmative. However, any estimates of the absolute or relative efficacy of app-based biofeedback must be deferred until follow-up studies can be conducted with larger sample sizes and controlled experimental designs.

This case study also highlighted a challenge that has been acknowledged in previous biofeedback treatment research, whereby even strong gains within the treatment

---

[1]Each probe measure elicited 10 tokens representing these treated rhotic variants. Therefore, the means and confidence intervals plotted in Figure 5 are based on calculations with $n = 10$.

---

[2]It remains unclear why F3–F2 distances grew slightly larger at the end of the treatment interval. Possible explanations include a loss of motivation or a maladaptive articulatory strategy. However, neither of these explanations can satisfactorily account for the observation that F3–F2 distance increased in generalization probes but not in treatment trials in the same time period, even though some treatment blocks in the final sessions were elicited entirely without biofeedback.

setting do not automatically generalize to other contexts. To date, it remains unknown why the extent of generalization varies so widely between subjects and studies. This may be a question of individual characteristics (e.g., children with normal auditory perception might be more likely to generalize than children with perceptual deficits) or the structure of treatment (e.g., earlier fading of biofeedback followed by an extended period of no-biofeedback practice might enhance generalization). Larger-scale studies will be essential to answer these questions.

In the introduction, we described how additional features would be added to the app in response to feedback received over the piloting process. Several potential improvements are suggested by Hannah's responses to survey questions and the authors' observations over the course of her treatment. Because frequent practice across varying contexts may enhance generalization gains (e.g., Maas et al., 2008), we plan to develop a module for home practice. To ensure that home practice reinforces the child's new and more accurate production patterns, users will be able to record their home practice sessions and transmit them to the clinician for scoring. We aim to implement automated scoring based on the computer's measurements of the acoustic properties of the user's productions. In addition, to enhance motivation, future versions will include progress-tracking features that will allow users and clinicians to monitor changes in rhotic production accuracy over time. To further increase user engagement, we intend to incorporate gamified rewards. For example, participants who have completed a practice goal or achieved a higher level of accuracy could be awarded "sand dollars," which could be exchanged in-app for embellishments that can be added to personalize the starfish character.

The introduction to this article also discussed the field of implementation science, which encourages researchers not only to document the efficacy of a particular approach, but also to make active efforts to ensure that effective methods are made widely available and easily accessible to practitioners. As stated above, a long-term aim of this line of research is to disseminate the staRt app broadly on a not-for-profit basis. In turn, we intend to ask clients and participating children/families to consider supporting our efforts by allowing deidentified data to be shared securely and privately for research purposes. By creating an "ecosystem" in which researchers and practitioners play mutually beneficial roles (Heffernan & Heffernan, 2014), we can potentially conduct higher-quality research studies, including well-powered randomized controlled trials. Harnessing next-generation technologies, especially in the form of easy-to-use, easy-to-disseminate apps, will play a crucial role in expanding these researcher–clinician partnerships.

Acknowledgments

References

Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology, 16*(2), 128–139.

Bislick, L. P., Weir, P. C., Spencer, K., Kendall, D., & Yorkston, K. M. (2012). Do principles of motor learning enhance retention and transfer of speech skills? A systematic review. *Aphasiology, 26*(5), 709–728.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.25) [Computer program]. Retrieved from http://www.praat.org/

Davis, S. M., & Drichta, C. E. (1980). Biofeedback: Theory and application to speech pathology. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice,* Volume 3 (pp. 283–286). New York, NY: Academic Press.

Flipsen, P., Shriberg, L. D., Weismer, G., Karlsson, H. B., & McSweeny, J. L. (2001). Acoustic phenotypes for speech-genetics studies: Reference data for residual /r/ distortions. *Clinical Linguistics & Phonetics, 15*(8), 603–630.

Gibbon, F., Stewart, F., Hardcastle, W. J., & Crampin, L. (1999). Widening access to electropalatography for children with persistent sound system disorders. *American Journal of Speech-Language Pathology, 8*(4), 319–334.

Gibbon, F. E., & Paterson, L. (2006). A survey of speech and language therapists' views on electropalatography therapy outcomes in Scotland. *Child Language Teaching and Therapy, 22*(3), 275–292.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments Ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education, 24*(4), 470–497.

Hitchcock, E., Harel, D., & McAllister Byun, T. (2015). Social, emotional, and academic impact of residual speech errors in school-aged children: A survey study. *Seminars in Speech and Language, 36*(4), 283–293.

Hitchcock, E. R., & McAllister Byun, T. (2014). Enhancing generalization in biofeedback intervention using the challenge point framework: A case study. *Clinical Linguistics & Phonetics, 29*(1), 59–75.

Hodges, N. J., & Franks, I. M. (2001). Learning a coordination skill: Interactive effects of instruction and feedback. *Research Quarterly for Exercise and Sport, 72*(2), 132–142.

Klein, H. B., McAllister Byun, T., Davidson, L., & Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech-Language Pathology, 22*(3), 540–553.

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*(3), 1455–1468.

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277–298.

Martin, N., & Brownell, R. (2005). *Test of Auditory Processing Skills–Third Edition.* (TAPS-3) [Measurement instrument]. Novato, CA: Academic Therapy.

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, 7,* 1–19.

McAllister Byun, T., & Campbell, H. (2016). Differential effects of visual-acoustic biofeedback intervention for residual speech errors. *Frontiers in Human Neuroscience, 10,* 567.

McAllister Byun, T., Halpin, P. F., & Szeredi, D. (2015). Online crowdsourcing for efficient rating of speech: A validation study. *Journal of Communication Disorders, 53,* 70–83.

McAllister Byun, T., & Hitchcock, E. R. (2012). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology, 21*(3), 207–221.

McAllister Byun, T., Hitchcock, E., & Ortiz, J. (2014). Challenge-R: Computerized challenge point treatment for /r/ misarticulation. Paper presented at ASHA 2014, Orlando, FL.

McAllister Byun, T., Hitchcock, E. R., & Swartz, M. T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research, 57*(6), 2116–2130.

McAllister Byun, T., Swartz, M. T., Halpin, P. F., Szeredi, D., & Maas, E. (2016). Direction of attentional focus in biofeedback treatment for /r/ misarticulation. *International Journal of Language and Communication Disorders, 51,* 384–401.

McCormack, J., McLeod, S., McAllister, L., & Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology, 11*(2), 155–170.

Muñoz, M. L., Hoffman, L. M., & Brimo, D. (2013). Be smarter than your phone: A framework for using apps in clinical practice. *Contemporary Issues in Communication Science & Disorders, 40,* 138–150.

Olswang, L. B., & Prelock, P. A. (2015). Bridging the gap between research and practice: Implementation science. *Journal of Speech, Language, and Hearing Research, 58*(6), S1818–S1826.

Preston, J. L., Brick, N., & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology, 22*(4), 627–643.

Preston, J. L., McCabe, P., Rivera-Campos, A., Whittle, J. L., Landry, E., & Maas, E. (2014). Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language, and Hearing Research, 57*(6), 2102–2115.

Ruscello, D. M. (1995). Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders, 28*(4), 279–302.

Shriberg, L. D., Gruber, F. A., & Kwiatkowski, J. (1994). Developmental phonological disorders III: Long-term speech-sound normalization. *Journal of Speech, Language, and Hearing Research, 37*(5), 1151–1177.

Shuster, L. I., Ruscello, D. M., & Smith, K. D. (1992). Evoking [r] using visual feedback. *American Journal of Speech-Language Pathology, 1*(3), 29–34.

Shuster, L. I., Ruscello, D. M., & Toth, A. R. (1995). The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology, 4*(2), 37–44.

Volin, R. A. (1998). A relationship between stimulability and the efficacy of visual biofeedback in the training of a respiratory control task. *American Journal of Speech-Language Pathology, 7*(1), 81–90.