Can perceptual training alter the effect of visual biofeedback in speech-motor learning?

Adam Klaus,¹ Daniel R. Lametti,² Douglas M. Shiller,³ and Tara McAllister^{4,a)}

¹Gallatin School of Individualized Study, New York University, 1 Washington Place, New York, New York 10003, USA

²Department of Psychology, Acadia University, Horton Hall, 18 University Avenue, Wolfville, Nova Scotia, B4P 2R6, Canada

³École d'orthophonie et d'audiologie, Université de Montréal, C.P. 6128, succursale Centre-ville, Montreal, Quebec, H3C 3J7, Canada

⁴Department of Communicative Sciences and Disorders, New York University, 665 Broadway, Suite 900, New York, New York 10012, USA

(Received 11 May 2018; revised 2 January 2019; accepted 15 January 2019; published online 13 February 2019)

Recent work showing that a period of perceptual training can modulate the magnitude of speech-motor learning in a perturbed auditory feedback task could inform clinical interventions or second-language training strategies. The present study investigated the influence of perceptual training on a clinically and pedagogically relevant task of vocally matching a visually presented speech target using visual-acoustic biofeedback. Forty female adults aged 18–35 yr received perceptual training targeting the English /æ-ɛ/ contrast, randomly assigned to a condition that shifted the perceptual boundary toward either /æ/ or /ɛ/. Participants were then asked to produce the word *head* while modifying their output to match a visually presented acoustic target corresponding with a slightly higher first formant (*F*₁, closer to /æ/). By analogy to findings from previous research, it was predicted that individuals whose boundary was shifted toward /æ/ would also show a greater magnitude of change in the visual biofeedback task. After perceptual training, the groups showed the predicted difference in perceptual boundary location, but they did not differ in their performance on the biofeedback matching task. It is proposed that the explicit versus implicit nature of the tasks used might account for the difference between this study and previous findings.

© 2019 Acoustical Society of America. <https://doi.org/10.1121/1.5089218>

[BVT]

Pages: 805–817

I. INTRODUCTION

Studies investigating how speakers respond to perturbed auditory feedback have played a major role in illuminating the mechanisms that underlie speech production and speech-motor learning. The perception-production links revealed through this research paradigm have implications for clinical efforts to remediate speech disorders as well as efforts to assist second-language (*L*₂) learners in achieving native-like pronunciation of *L*₂. However, direct applications of this line of investigation to clinical or *L*₂ populations have remained limited, in large part because the equipment and technical expertise required to successfully implement auditory feedback perturbation are rarely available outside the research setting. The present study is part of a broader line of research aiming to translate insights from the literature on perturbed auditory feedback into methods that might be more amenable to application in clinical practice or *L*₂ pedagogy. Specifically, this study investigated whether properties of the relationship between perceptual training and speech-motor learning, as revealed through tasks involving perturbed auditory feedback (e.g., Lametti *et al.*, 2014a; Shiller and Rochon, 2014), could be applied to enhance the effect of

speech training incorporating visual-acoustic biofeedback (e.g., McAllister Byun, 2017).

A. Perception-production links

Speech production and perception are distinct abilities that are often handled separately in the study and treatment of speech disorders. However, previous studies have shown that there is a close relationship between a person's ability to perceptually distinguish phonemes and the precision with which they produce that same speech sound. For example, Perkell *et al.* (2004) and Ghosh *et al.* (2010) demonstrated that participants who showed greater sensitivity in perceptually discriminating between /s/ and /ʃ/ also produced a larger acoustic difference between the two phonemes. More recently, McAllister Byun and Tiede (2017) reported that adolescents who were less consistent in classifying sounds along a continuum as instances of /ɹ/ or /w/ also produced /ɹ/ sounds with a less extreme acoustic signature of rhoticity, and Franken *et al.* (2017) found that individuals who exhibited more acute vowel discrimination produced vowels with less token-to-token variability. Another form of perception-production linkage is evident in the now sizable literature documenting the phenomenon of phonetic accommodation or convergence, in which speakers who are perceptually exposed to speech with a particular characteristic (e.g.,

^{a)}Electronic mail: tkm214@nyu.edu

lengthened voice onset time) tend to reflect back this property in their own subsequent productions (e.g., [Nielsen, 2011](#); [Pardo, 2013](#)).

The existence of perception-production links also gives rise to the prediction that alterations in auditory feedback should result in compensatory adjustments in production. These predictions have been explored in the experimental paradigm of perturbed auditory feedback, where a speaker's voice is altered using digital signaling processing software and played back to the speaker in near real time. Compensation for perturbed auditory feedback, in which speakers adjust their default pattern of speech production to offset these externally imposed alterations, has been observed across multiple dimensions of the speech signal, including vocal pitch/ F_0 (e.g., [Elman, 1981](#); [Kawahara, 1995](#); [Jones and Munhall, 2000, 2005](#)), vowel formant frequencies ([Houde and Jordan 1998, 2002](#); [Purcell and Munhall, 2006a, 2006b](#); [Villacorta et al., 2007](#)), and the spectral characteristics of consonants ([Shiller et al., 2009](#)). A smaller number of studies looking at child participants have shown that children also perceive and respond to perturbed auditory feedback ([Ménard et al., 2008](#); [Shiller et al., 2010](#); [MacDonald et al., 2012](#); [Terband et al., 2014](#)), although a lack of compensatory response has been documented in some subgroups, including toddlers ([MacDonald et al., 2012](#)) and children with speech sound disorder ([Terband et al., 2014](#)).

B. Perceptual influences on speech-motor learning

The above-described perception-production links raise a theoretically and clinically important question of whether perceptual training can have a lasting influence on speech production. Although it is straightforward to induce short-term changes in listeners' perceptual classification of speech sounds (e.g., [Guediche et al., 2014](#); [Callan et al., 2003](#)), studies examining the longer-term effects of perceptual reorganization on speech production have yielded mixed results. For instance, [Bradlow et al. \(1997\)](#) found that perceptual training of a non-native phonemic contrast led to improvements in production of the same contrast, whereas [Baese-Berk \(2010\)](#) found that perceptual training was not reliably associated with changes in the production domain. Although the preconditions for transfer from perception to production remain incompletely understood, studies using precisely calibrated conditions have succeeded in inducing small but reliable effects of perceptual training on production; the present investigation was modeled after one such experiment ([Lametti et al., 2014a](#)).

[Lametti et al. \(2014a\)](#) engaged listeners in a period of perceptual training using a synthetic ten-step continuum starting at the vowel / ϵ / in the context of the word *head* and ending at the vowel / æ / in the context of the word *had*. Participants listened to each step of the continuum multiple times in random order and were asked to identify each token as *head* or *had*. These responses were used to identify each participant's perceptual boundary, operationalized as the point along the continuum at which the participant perceives the word as *head* 50% of the time and *had* 50% of the time.

After their boundary location was identified, participants completed perceptual training tasks with feedback intended to shift the location of this boundary (Fig. 1). Participants were randomly assigned to a *head-shift* condition, in which the location of the boundary was moved closer to the *head* end of the continuum, or a *had-shift* condition, which moved the boundary in the opposite direction. The *head-shift* condition tended to encourage participants to classify perceptually ambiguous stimuli as instances of *had*, while the *had-shift* condition encouraged participants to classify ambiguous tokens as instances of *head*.

Participants were subsequently engaged in a perturbed auditory feedback task. They were asked to produce the word *head* repeatedly while listening to their own speech through headphones. They were not informed that they were hearing an altered version of their voice in which the first formant (F_1) of the vowel in *head* was raised, making it sound slightly more like the word *had*. All participants heard the same magnitude and direction of perturbation, but it was hypothesized that this feedback would have a different impact on the *head-shift* versus the *had-shift* group. This prediction was grounded in previous research ([Niziolek and Guenther, 2013](#)) indicating that listeners exhibit a larger magnitude of compensation for altered auditory feedback when the perturbation applied causes a token to cross a boundary from one phoneme category into another. Since participants in the *head-shift* group had been trained to classify ambiguous vowels with a slightly raised F_1 as instances of *had*, they should be more likely to perceive their perturbed feedback as boundary-crossing and thus take larger compensating action. The *had-shift* group, by contrast, was conditioned to identify ambiguous tokens as still belonging to the *head* category, and they should correspondingly be less likely to undertake an articulatory adjustment in response to the perturbed feedback. Consistent with their hypothesis, [Lametti et al. \(2014a\)](#) found that the *head-shift* group showed a significantly greater magnitude of compensation than the *had-shift* group. The difference between the groups remained significant in a post-test after the auditory perturbation was withdrawn.

FIG. 1. (Color online) Schematic representation of perceptual training paradigm from [Lametti et al. \(2014a\)](#), replicated in the present study. Feedback was provided to encourage listeners to shift their psychometric identification function toward one endpoint (*head* or *had*), with individuals randomly assigned to the *head-shift* or *had-shift* direction.

C. Visual-acoustic biofeedback training for speech

Compensation for perturbed auditory feedback is a type of sensorimotor learning: it involves adjustments to the mapping between motor commands and sensory consequences associated with a particular speech target. Many forms of speech production training, whether in the context of clinical rehabilitation or *L2* pronunciation pedagogy, also involve sensorimotor learning tasks that aim to alter the motor routine that the speaker uses for a given target sound. Given this parallel, the results from Lametti *et al.* (2014a) raise the possibility that the effects of speech intervention or pronunciation training might also be enhanced if treatment sessions consisted of an initial period of perceptual training to sensitize the speaker to the target contrast, followed by a period of intensive speech-motor practice. The present study tests the hypothesis that perceptual training can modulate the magnitude of change in a task of matching a visually presented acoustic target (visual-acoustic biofeedback).

Visual biofeedback involves the use of instrumentation to capture measurements of some aspect of physiology or behavior, which can then be displayed in real time to a learner. This visual display provides a novel source of insight intended to allow the learner to exercise a higher level of conscious control over the behavior in question (Davis and Drichta, 1980; Volin, 1998; Bliss *et al.*, 2018). In the specific context of visual-acoustic biofeedback for speech, speakers view a dynamic image of the acoustic signal of their speech output, typically in the form of a linear predictive coding (LPC) spectrum (e.g., McAllister Byun and Hitchcock, 2012) or a spectrogram (e.g., Shuster *et al.*, 1992). Learners are also provided with a visual model representing correct production of the target sound and cued to alter their speech output so that the resonant frequencies of their speech, visible as peaks in the LPC spectrum or horizontal bars in the spectrogram, align more closely with their locations in the model. In the context of *L2* learning, significant positive effects of biofeedback training have been reported in the production of fricatives (Olson, 2014), vowels (Kartushina *et al.*, 2014), and vowel length contrasts (Okuno and Hardison, 2016). In the context of clinical research, evidence that visual-acoustic biofeedback can be effective in improving speech errors that have not responded to other forms of intervention has been reported in case studies (Shuster *et al.*, 1992; Shuster *et al.*, 1995) and single-case experimental studies (McAllister Byun and Hitchcock, 2012; McAllister Byun and Campbell, 2016; McAllister Byun *et al.*, 2016; McAllister Byun, 2017). However, these clinical studies have been uniform in reporting that gains made within the treatment setting do not automatically generalize to a context in which biofeedback is not available. Thus, there is a need for further research to identify methods that could amplify the effects of biofeedback training and promote long-term retention and generalization of gains made in the clinic setting.

D. Does perceptual training influence performance in visual-acoustic biofeedback tasks?

The present study investigated perceptual training as one factor that could potentially modulate the effects of visual-acoustic biofeedback. This study followed Lametti

et al. (2014a) in separating participants into *head-shift* and *had-shift* groups for perceptual training. Participants were then engaged in a task of producing the word *head* while viewing visual-acoustic biofeedback and trying to match a visual target that was raised relative to their typical *F1* for “head.” We hypothesized that participants in the *had-shift* group would be more successful in matching the visual-acoustic target than participants in the *head-shift* group. This is based on the notion that the change in *F1* needed to match the visual target is more likely to cross the perceptual boundary between / ϵ / and / $\ae$ / for participants in the *head-shift* group, and this conflicting auditory feedback could inhibit them from undertaking a change in production of this magnitude.

It is important to acknowledge that the mechanisms involved in perturbed auditory feedback and visual-acoustic biofeedback are not identical. Previous literature examining limb movements has recognized that “motor learning” is not a monolithic process but can be broken down into at least two subtypes (Krakauer and Mazzoni, 2011; Taylor *et al.*, 2014; McDougle *et al.*, 2016). One subtype is an implicit learning route in which prediction error (i.e., the difference between the expected and perceived sensory consequences of movement) is used to update stored mappings between motor and sensory representations. This form of learning has been characterized as slow but lasting in its effects, and it is associated with low variability in performance outcomes (Benson *et al.*, 2011). Another subtype is an explicit learning mechanism in which strategies or heuristics are used to make adjustments to the motor plan in an effort to match an external target. In this case, the learner receives error feedback relative to an external target, but there should be no mismatch between the predicted and actual sensory consequences of the planned action, and therefore no updates to the stored sensory-motor mapping (Krakauer and Mazzoni, 2011). This type of learning is associated with faster increases in accuracy in the short term, but it is characterized by high variability and may also yield a lesser degree of long-term carryover learning (Benson *et al.*, 2011).

It is highly likely that these distinct types of motor learning are invoked differentially by the task in the present study versus that studied by Lametti *et al.* (2014a). In perturbed auditory feedback, participants attempt to achieve their habitual speech target, but they receive feedback indicating a deviation from that target, requiring adjustments to the preexisting production routine. These speakers are typically not aware of the perturbation or their own compensatory response. In our visual biofeedback task, participants are instructed to match an external target and provided with explicit strategies for doing so. Because they hear their own output without modification, they should not receive feedback reflecting prediction error or sensory mismatch.

Given that both *L2* production training and interventions for speech disorders involve explicit learning, it is important to evaluate whether the influences of perception on motor learning described by Lametti *et al.* (2014a) also hold in the context of an explicit task like biofeedback matching. More generally, the differences between explicit and implicit motor learning processes have been understudied in speech compared to other motor behaviors (e.g., reaching).

Although explicit and implicit learning are distinct mechanisms, there is evidence that they can interact (e.g., [Benson et al., 2011](#)). In fact, in [Lametti et al. \(2014a\)](#), an explicit training task was used to shift the location of speakers' perceptual boundary, which subsequently influenced the magnitude of their compensation for perturbed auditory feedback. If an explicitly shifted perceptual boundary can influence implicit learning, it is reasonable to ask whether it could also modulate performance in an explicit task of matching an external target. The present study specifically investigated whether the magnitude of participants' adjustments to match a visual-acoustic target would reflect the degree of divergence between their actual output and their internally specified auditory target, manipulated through explicit perceptual training.

II. METHODS

A. Participants

This study was approved by the ethical review board at New York University. Participants in the experiment were 40 females ranging in age from 18 to 33 yr [mean = 22.25 yr, standard deviation (SD) = 3.56 yr]. Three additional participants completed the experiment but were excluded from data analysis due to errors in the administration of the study protocol. Participants were required to be native speakers of American English with no history of speech or hearing impairments, per self-report on a screening questionnaire. A connected speech sample was elicited and screened for any major deficits in articulation, voice, resonance, or fluency. Participants were additionally required to pass a hearing screening [20 dB hearing level (HL) at 1000, 2000, and 4000 Hz]. Once eligibility was determined, participants were randomly assigned to equal-sized *had*-shift and *head*-shift groups.

B. Protocol

1. General

The protocol can be divided into two main types of tasks: perception and production. For simplicity, they will be described separately here, but they were presented in an interleaved order in the actual protocol. The complete sequence of tasks carried out over the course of the experiment is schematized in Fig. 2. Participants completed a baseline task of producing the words *head* and *had*, followed by a baseline perceptual task requiring identification of tokens

along a synthetic *head-had* continuum. They then completed two blocks of perceptual training intended to shift the location of their perceptual boundary in a specified direction, randomly assigned for each participant. This was followed by a task of producing the word *head* while attempting to match a visually presented formant target, subdivided into three phases. In the first phase, the external biofeedback target was available for all trials; in the second phase, it was available for half of all trials; and in the final phase, it was withdrawn altogether. The withdrawal of biofeedback across phases was included to provide maximal opportunity to measure changes that carry over into a no-feedback phase, which are considered indicative of implicit rather than explicit learning ([Krakauer and Mazzoni, 2011](#)). The experiment concluded with repetition of the production and perceptual discrimination tasks administered at baseline. The complete experimental protocol is available as a supplement to this paper (Appendix A).¹

2. Perception

The perceptual component of the study consists of four tasks designed to measure and then shift the location of the participant's perceptual boundary between /*ɛ*/ and /*æ*/ in the context of *head* and *had*. A synthetic continuum was created using tokens of *head* and *had* elicited from an adult female native speaker of English (northeastern United States dialect region). To create the intermediate steps, the first and second formants associated with the vowel in *head* were shifted in 41 approximately equal steps toward the formant values measured from the speaker's naturally produced *had* token. Formant shifting was carried out using the STRAIGHT speech algorithm ([Kawahara et al., 2013](#)). *F*₁ and *F*₂ values from each of the resulting tokens were measured and plotted, and a subset of 11 tokens was selected that most closely approximated a linear sequence in *F*₁ and *F*₂ space (in Hertz). To oversample the probable boundary region, the stimuli numbered 3–9 were selected to be more tightly spaced (mean distance = 26.4 Hz, SD = 10.3 Hz) than the outermost pairs of stimuli (mean distance = 40.2 Hz, SD = 15.0 Hz). The stimuli used are available in [Mm. \(1\)](#); formant measurements for each are reported in [Table I](#).

[Mm. 1](#). Eleven-step continuum from *head* to *had*. This is a file of type “wav” (817 KB).

All perceptual stimuli were presented on a laptop computer over Sennheiser HD429 headphones (Wedemark, Germany) in a soundproof booth. In each perceptual task, participants heard each continuum step 20 times in random order for a total of 210 tokens per task. Participants were instructed to click either *head* or *had* to classify each token they heard. The first perceptual task (pre-test) served as a baseline measurement of participants' perceptual boundary between *head* and *had*. Immediately after administration of the perceptual pre-test, an *R* script (adapted from [McAllister Byun and Tiede, 2017](#)) was used to fit a logistic function over the number of *head* responses at each step in the continuum. The script also computed each participant's boundary location, defined as the point on the continuum where the

FIG. 2. Schematic representation of components of the experiment. Unaltered production of *head* and *had* (pre-probe) was followed by one perceptual pre-test and two blocks of perceptual training (PT1 and PT2). Participants were then familiarized with biofeedback and trained to produce the word *head* while matching a visually presented formant target representing a raised *F*₁ (closer to *had*). Biofeedback was faded from 100% to 0% over the course of 120 trials. The experiment concluded with a perceptual post-test and a final elicitation of unaltered *head* and *had* (post-probe).

TABLE I. Formant frequencies (Hertz) measured from stimuli in the synthetically generated *head-had* continuum.

Step	<i>F</i> 1	<i>F</i> 2	<i>F</i> 3
1 (<i>had</i>)	767.1	2083.4	3211.5
2	785.2	2049.7	3134.1
3	830.9	2029.5	3108.1
4	859.1	2011.6	3106.3
5	874.3	1991.6	3089.7
6	913.9	2003.6	3071.3
7	939.5	1956.8	3042.6
8	954.0	1950.5	3014.0
9	989.4	1929.6	2990.8
10	1039.9	1890.4	2925.4
11 (<i>head</i>)	1086.6	1891.6	2924.2

fitted logistic function reached its 50% probability point; this value was rounded to the nearest integer to correspond with an actual step on the continuum. Participants then completed two perceptual training tasks (PT1 and PT2) aimed at shifting their perceptual boundary by two continuum steps in the direction of either *had* or *head* (randomly assigned), starting from the participant's individually determined rounded boundary location. The number of trials presented was the same as at pre-test, but in this case participants received feedback ("correct" or "incorrect") on their responses. The stimulus presentation script was manually updated to reflect the shifted perceptual boundary calculated for each participant. Tokens below the shifted perceptual boundary were classified as *head*, while tokens at or above the shifted boundary were classified as *had*. Participants could see a tally of incorrect responses at the bottom of their computer screen, and they were encouraged to minimize their error. Finally, participants completed a perceptual post-test that was identical to the pre-test in that participants did not receive accuracy feedback. The purpose of the post-test was to assess whether any effects of perceptual training would be maintained at the end of the experiment after the biofeedback production task.

3. Production

Recordings for all production tasks were made in the same soundproof booth using a Shure SM48 microphone (Niles, IL). Recordings were captured with a KayPENTAX

Computerized Speech Lab model 4500 (Lincoln Park, NJ) and digitized with a 44.1 kHz sampling rate and 16-bit encoding. Prior to any experimental manipulations, participants performed a baseline task of producing the words *head* and *had* 45 times each in random order (henceforth, pre-probe). Immediately after this baseline recording, while the participant was completing perceptual tasks, the first author measured formant frequencies from a subset of baseline productions for use in setting an individualized acoustic target. Ten productions each of *head* and *had* were randomly selected for measurement from among the 90 total trials. Different candidate LPC filter orders for each participant were visually inspected (see more detailed description in Sec. II C), and formant measurements were taken from the steady-state region at the approximate vowel midpoint. Average *F*1 (Hertz) was calculated for each vowel, and the midpoint between these two values was then used as the participant's target for visual-acoustic biofeedback. The visual target was represented by placing a vertical marker over the real-time LPC spectrum at the calculated frequency (Fig. 3).

The production training component of the session began with an introduction to biofeedback provided by the experimenter. The Sona-Match module of the Computerized Speech Lab software suite (KayPENTAX, Lincoln Park, NJ) was used to generate the biofeedback display. Verbal descriptions and demonstrations were used to familiarize participants with the peaks of the LPC spectrum and explain how they changed in connection with the production of different speech sounds. Participants were informed that the location of the first peak reflected the height of their jaw and degree of mouth opening. They were then engaged in a guided pre-practice period, beginning with an instruction to stretch out the vowel in the word *head* and note the location of the first peak in the spectrum, followed by suggestions to adjust jaw height and mouth opening while observing the influence of these manipulations on the location of the first peak.

After the pre-practice period ended, participants were encouraged to produce the word *head* with a normal duration while attempting to make their first peak line up as closely as possible to the visual target. This speech motor task was broken up into three phases of increasing duration. Because pilot data suggested that some individuals quickly attained

FIG. 3. (Color online) Screenshot of visual-acoustic biofeedback display generated with Sona-Match software (KayPENTAX, Lincoln Park, NJ). A vertical marker superimposed over the dynamic LPC spectrum served as an *F*1 target for participants to match.

ceiling-level performance in the relatively straightforward task of matching a visual formant target, the percentage of trials in which biofeedback was available was reduced over successive phases. (Recall that prompt withdrawal of biofeedback was implemented in an effort to maximize opportunities to observe changes that carry over in the absence of an external target, potentially reflecting a more implicit type of motor learning.) In production phase 1, participants produced 20 trials of *head* with the biofeedback display visible during 100% of trials. Production phase 2 elicited 40 trials of *head*. The first ten were produced with biofeedback, while for the next ten, the biofeedback was withdrawn by minimizing the display window, and participants were instructed to continue producing the word *head* in the same way they had while they were trying to reach the visual target. This procedure was repeated for a second set of 20 trials. Production phase 3 elicited 60 trials of *head*. No biofeedback was available during this phase, but participants were instructed to continue to produce the vowel just as they had while viewing the visual target.

The last step of the protocol (post-probe) elicited 45 trials each of *head* and *had* in random order. No biofeedback was provided during this task. This post-test elicitation was used to assess whether perceptual training and/or production practice would have any lasting effect on participants' vowel formants relative to their productions during pre-probe elicitation.

C. Measurement

Praat acoustic analysis software (Boersma and Weenink, 2016) was used to obtain acoustic measures of each utterance of *head* across all five production tasks (pre-probe, phases 1–3 of production training, and post-probe). Files were annotated at the word and phoneme level using the DARLA forced aligner (Reddy and Stanford, 2015), which in turn makes use of FAVE-Extract (Rosenfelder *et al.*, 2014). Prior to forced alignment, trained student assistants reviewed the acoustic record of each file and trimmed out any false starts or other speech that did not correspond with the target utterances of *head*. Formant measurements were obtained at the midpoint of each vowel interval. To select an appropriate LPC filter order for each participant, spectrograms with superimposed LPC formant tracks were visually inspected with different formant settings (e.g., five formants in a range up to 5000 Hz, 5500 Hz, 6000 Hz, or 6500 Hz). The setting judged to produce the best match between the LPC formant tracks and visible areas of energy concentration was selected for use in all formant measurements for a given individual. The measurement process was repeated for all target utterances in all five production tasks. Finally, outliers were cleaned on a speaker-by-speaker basis by identifying and discarding all tokens that fell more than two median absolute deviations (MADs) away from the median *F1* frequency for a given individual; 214 out of approximately 8400 tokens (2.5%) were eliminated in this way.

To assess measurement reliability, all data used in analyses reported here were re-measured by hand. Student assistants were told which filter order settings had been selected for each participant. They visually inspected the spectrogram for each word and marked a point representing a steady state

of *F1* in the vowel interval. Formant measurements were then extracted using the same Praat script noted above. Formant values determined through hand annotation versus forced alignment were significantly correlated (Pearson's $r = 0.88$, $p < 0.001$).

D. Analyses

The first experimental question asked if participants' perceptual boundaries between the vowels /*ɛ*/ and /*æ*/ were successfully shifted by means of perceptual training. This was investigated with a series of nonparametric Wilcoxon tests comparing the distribution of boundary locations measured in the pretest phase against the boundary locations measured in subsequent phases. Because four comparisons were conducted, the threshold for significance was Bonferroni-corrected to 0.0125. The second experimental question looked for an association between performance on the perceptual task and performance on the production task of matching a visually presented *F1* target. This relationship was examined using a linear mixed model with a normalized measure of *F1* frequency as the dependent variable. Independent variables included the phase of assessment (pre- versus post-training, described in more detail below) and the direction of perceptual training (*head*-shift or *had*-shift), as well as the interaction between them. A random intercept for participant was included to account for speaker-level variation not captured by the fixed effects. Additionally, a by-phase random slope was examined for inclusion as permitted by model convergence.

III. RESULTS

Complete data and code used to generate the results reported here are available as supplementary materials (Appendix B: perceptual data; Appendix C: production data; Appendix D: reliability data; Appendix E: analysis code).¹

A. Perception

We begin by examining whether perceptual boundary location differed significantly between participants in the *had*-shift and *head*-shift groups at four points in the experiment. Figure 4 shows the distribution of perceptual boundary locations measured for each group over time. At pre-test, the median bias was very similar between the two groups (*had*-shift = 5.75, *head*-shift = 5.9). After one block of perceptual training (PT1), the groups showed a difference in median bias (*had*-shift = 6.7, *head*-shift = 4.45) with further divergence after an additional block of training (PT2; *had*-shift = 6.85, *head*-shift = 3.65). At post-test, after the speech motor task, the *had*-shift group showed a median bias of 7.2, while the *head*-shift group exhibited a median bias of 4.45. The Bonferroni-corrected difference between groups was not significant at pre-test ($W = 192$, $p = 0.84$), but it was significant at all subsequent time points (after PT 1, $W = 400$, $p < 0.001$; after PT 2, $W = 400$, $p < 0.001$; at post-test, $W = 381.5$, $p < 0.001$).

B. Production

Following Lametti *et al.* (2014a), all production values were normalized by dividing the observed *F1* value by the

FIG. 4. Perceptual boundaries of the two groups at pre-test, after one block of perceptual training (PT1), after two blocks of perceptual training (PT2), and at post-test. Middle bars represent medians, boxes represent the interquartile range (IQR; 25th–75th percentile), whiskers extend to the most extreme non-outlier data point, and outliers (more than 1.5IQR outside the IQR) are represented as single points.

speaker's mean $F1$ for head during the baseline phase. Visual inspection revealed a great deal of variability in how individual participants' $F1$ values changed over the phases of the study. As an illustration of this diversity, Fig. 5 shows changes in normalized $F1$ frequency across the five phases for four individuals hand-selected from the *had-shift* group. The first participant shown, A03 (upper left), gradually increased the normalized $F1$ of her productions from pre-test through post-test. By contrast, participant A08 (upper right) showed a large increase in $F1$ in phase 1, which featured 100% biofeedback, but showed decreasing $F1$ values as biofeedback frequency was decreased across subsequent phases. Participant A16 (lower left) showed a robust increase in $F1$ throughout all training phases but then fell back to baseline levels at post-test. Finally, participant A05 (lower right) illustrates a case of minimal response; none of the measurements during and after training showed a significant increase relative to $F1$ at pre-test. Figure 6 shows the same data for selected participants from the *head-shift* group. Note that each of the patterns described in Fig. 5 was also attested in the *head-shift* group: participant B21 (upper left) showed gradually increasing $F1$ across successive phases, B17 (upper right) showed dramatically increased $F1$ in phase 1 with values falling off thereafter, B19 (lower left) showed increased $F1$ in all phases except post-test, and B22 (lower right) showed minimal change in production across the phases of the study. These examples are provided merely to illustrate the diversity of individual responses across phases in both *had-shift* and *head-shift* groups; we take up the question of average patterns of response below.

Figure 7 shows boxplots representing changes in normalized $F1$ frequency across the five points at which productions were elicited, pooled across individuals in each group. In the pre-probe administered prior to any training, the median $F1$ was very similar between the two groups (*had-shift*: 731.6 Hz, *head-shift*: 735.5 Hz). In phase 1 of production training, where the visual formant target was presented in all 20 trials, both groups produced slightly higher median $F1$ values but remained similar to one another (*had-shift*: 809.3 Hz, *head-shift*: 785.21 Hz). This pattern remained mostly stable in phases 2 and 3 when biofeedback was provided less frequently and then withdrawn. In the post-probe, which elicited 45 trials of *head* with no visual formant target, both groups returned to a slightly lower median $F1$ (*had-shift*: 758.1 Hz, *head-shift*: 762.0 Hz).

A generalized linear mixed-effects model was used to test for differences in normalized $F1$ between phases of the study and between groups. Despite the heterogeneity in individual response patterns described above, participants tended to show the greatest magnitude of shift in normalized $F1$ during phase 1, when biofeedback was available in 100% of trials. To maximize the potential to observe between-phase differences, only data from the baseline pre-probe and phase 1 (100% biofeedback) were included in the mixed-effects model. As indicated above, the hypothesis-driven model structure included fixed effects of condition (*had-shift* versus *head-shift*) and phase, as well as the interaction between them. Model comparison using likelihood ratio tests was used to determine the random-effects structure of the final model; chi-square comparisons indicated that the best-fitting model included a random intercept for participant with by-

FIG. 5. Normalized $F1$ frequencies produced by individuals in the *had-shift* group across five phases of production. Middle bars represent medians, boxes represent the IQR (25th–75th percentile), whiskers extend to the most extreme non-outlier data point, and outliers (more than 1.5IQR outside the IQR) are represented as single points.

phase random slopes. There was a significant main effect of phase [$\beta = -0.10$, standard error (SE) = 0.015, $p < 0.001$], indicating that normalized $F1$ values were significantly lower at pre-probe than in phase 1 (100% biofeedback). There was no main effect of condition (*had-shift* versus *head-shift*; $\beta = -0.01$, SE = 0.02, $p = 0.68$) and no significant interaction between phase and condition ($\beta = 0.01$, SE = 0.02, $p = 0.69$).

An additional generalized linear mixed-effects model was used to test for differences in normalized $F1$ when comparing pre-probe versus post-probe productions. As previously, the model included fixed effects of condition (*had-shift* versus *head-shift*) and phase (pre-probe versus post-probe) and the interaction between them, as well as a random intercept for speaker with a by-phase random slope. As previously, there was a significant main effect of phase ($\beta = -0.02$, SE = 0.011, $p = 0.04$), although this result would not survive correction for multiple comparisons. There was no main effect of condition ($\beta = 0.02$, SE = 0.02, $p = 0.33$) and no significant interaction between phase and condition ($\beta = -0.02$, SE = 0.02, $p = 0.33$).

C. Perception-production relationship

Pearson's correlations between measures of perception and production were examined across subjects. There was no significant correlation between magnitude of change in

production from pre-probe to phase 1 and magnitude of shift in perceptual boundary from pre-test to PT2 [$r(35) = -0.16$, $p = 0.34$]. Magnitude of change in production also showed no significant association with perceptual boundary location, as measured either at pre-test [$r(35) = -0.16$, $p = 0.34$] or at PT2 [$r(35) = -0.21$, $p = 0.20$].

IV. DISCUSSION

A. Perception

In the present sample of typical female speakers aged 18 to 35 yr, the two perceptual training groups differed significantly in the location of the / ϵ /-/ $\ae$ / phoneme boundary in all perceptual assessments after the initial perceptual training (PT1, PT2, and post-test). No significant between-group difference was expected in the pre-test, which was simply a baseline measurement of perceptual boundary location in two randomly assigned groups. As soon as perceptual training was introduced, perceptual boundary location moved up for participants in the *had-shift* group and down for participants in the *head-shift* group. Thus, the perceptual training was successful in shifting participants' boundary location in the intended direction. This is consistent with the perceptual findings from previous work (Lametti *et al.*, 2014a; Lametti *et al.*, 2016), and was expected because the protocol used for perceptual training in the current study was based on that used in Lametti *et al.* (2014a).

FIG. 6. Normalized $F1$ frequencies produced by individuals in the *head-shift* group across five phases of production. Middle bars represent medians, boxes represent the IQR (25th–75th percentile), whiskers extend to the most extreme non-outlier data point, and outliers (more than 1.5IQR outside the IQR) are represented as single points.

FIG. 7. Normalized $F1$ frequencies of the two shift condition groups at pre-probe, in three phases of training (phase 1 = formant target presented in 100% of trials; phase 2 = formant target presented in 50% of trials; phase 3 = formant target provided in 0% of trials), and at post-probe. Middle bars represent medians, boxes represent the IQR (25th–75th percentile), whiskers extend to the most extreme non-outlier data point, and outliers (more than 1.5IQR outside the IQR) are represented as single points.

Perceptual boundaries between phonemes are influenced by a variety of factors in addition to the explicit training examined in the present study. Previous studies investigating speech motor adaptation to alterations in auditory feedback have noted changes in perceptual representations following motor learning tasks. For example, Shiller *et al.* (2009, 2010) observed changes in perceptual categorization of the /s-/ʃ/ contrast following speech adaptation to a real-time alteration of fricative spectra during production of words containing /s/. Similarly, Lametti *et al.* (2014b) observed changes in the location of the vowel category boundary along an /ε-/ɪ/ continuum following motor adaptation to altered *F1* frequency during /ε/ production.

In the present study, the final perceptual test of the /ε-/æ/ continuum (post-test) was carried out only *after* the speech production task in order to minimize the delay between the end of perceptual training and the beginning of speech adaptation (following the protocol used by Lametti *et al.*, 2014a). Given the order of these tasks, one cannot rule out the possibility that changes in speech motor representations may have influenced the subsequent perceptual boundary. The direction and magnitude of any such effect is difficult to predict, as the current protocol differs in potentially important ways from prior studies (e.g., the present study did not involve altering auditory feedback). Critically, in the present study, two perceptual training groups were included, involving boundary shifts toward opposite ends of the phonetic continuum. The perceptual post-test showed a significant difference between the two groups, confirming that the perceptual training was successful at inducing the expected perceptual changes. If the motor learning task did contribute to the perceptual shift, it was clearly not sufficient to outweigh the effect attributable to perceptual training.

B. Production

A linear mixed-effects model revealed that *F1* values in production phase 1 (100% biofeedback) were significantly higher than *F1* values in the pre-probe. Thus, training with a visually presented formant target did succeed in creating an increase in participants' mean *F1* values within the first 20 trials. This finding is consistent with previous results from the *L2* learning literature, where it has been reported that biofeedback-enhanced training can produce measurable changes in production patterns even when the duration of practice is brief (Kartushina *et al.*, 2015; Gick *et al.*, 2008). Kartushina *et al.* (2015) documented greater gains in a group of participants who practiced with visual-acoustic biofeedback versus auditory feedback only. Unlike that study, the present design did not include a no-biofeedback condition, which limits the strength of the conclusions that can be drawn about the influence of biofeedback on learning outcomes. Nevertheless, the idea of using a visual feedback display to present an altered formant target may in itself be of broad interest to the speech research community. This paradigm makes it possible to evaluate the ability of talkers to modify their speech production patterns without introducing an *L2* target sound (which poses confounds of perceptibility and distance to the nearest *L1* category) or using perturbed

auditory feedback (which requires specialized equipment and can be a challenge to configure). Additionally, if an altered visual formant target is paired with auditory masking, it could represent a unique opportunity to assess speech-motor skill independent of auditory acuity.

A second linear mixed-effects model compared *F1* in productions elicited at pre-probe and post-probe. Although this difference was also statistically significant, it was much smaller in magnitude than the effect of training in phase 1, reflecting the fact that *F1* tended to fall off as biofeedback was withdrawn over the course of the study. Thus, while visual biofeedback was successful in inducing short-term changes in participants' production patterns, generalization was limited. This finding is in keeping with prior biofeedback studies targeting formant alterations in individuals with rhotic misarticulation (e.g., McAllister Byun, 2017; McAllister Byun and Hitchcock, 2012; McAllister Byun *et al.*, 2016), where it has been reported that gains made through biofeedback training do not automatically generalize to contexts where biofeedback is not available. For the purpose of future research, as well as clinical and pedagogical applications, it would be beneficial to identify modifications to the protocol for practice that could result in greater long-term learning, as discussed in the next paragraph.

The broader literature on principles of motor learning (e.g., Maas *et al.*, 2008) suggests numerous other parameters whose modification could represent possible avenues to increase generalization learning. Perhaps the most obvious modification would be an increase in the number of practice attempts, which here was limited to 120 trials. A larger number of trials could feasibly be carried out by typical adult participants, and could potentially increase retention and transfer of speech-motor skills (e.g., Edeal and Gildersleeve-Neumann, 2011). The schedule of biofeedback availability is also a potential target for manipulation in future research. Previous research suggests that generalization gains may be maximized when enhanced feedback is provided in the early phases of practice, but later withdrawn (McAllister Byun and Campbell, 2016). Because we expected to see many participants perform at ceiling in the relatively straightforward task of matching a visual formant target, fading of biofeedback was initiated after only 20 trials. In actuality, no ceiling effect was observed, and relatively few participants showed changes that extended through the post-probe used to evaluate generalization learning. In short, the task of matching a visually presented formant target was more difficult for our typical adult participants than we anticipated. This suggests that some individuals did not achieve their full potential to match a visually presented formant target before fading of the biofeedback was initiated, and larger or more lasting changes might be observed in future research if biofeedback is provided with a higher intensity over a longer period of time.

C. Perception-production relationship

Based on the results of Lametti *et al.* (2014a), we hypothesized that the magnitude of change to match a visual target would differ across perceptual training groups.

However, the interaction between phase and perceptual training condition (*had-shift* versus *head-shift*) did not reach significance in either of the two mixed-effects models reported above, indicating that the two groups did not differ significantly in the magnitude of change from pre-probe to phase 1 (100% biofeedback) or from pre-probe to post-probe. The present study was matched to Lametti *et al.* (2014a) with respect to sample size and used the same perceptual training paradigm. However, matching the sample size of a previous study is not a guarantee of sufficient power in a replication (Button *et al.*, 2013). Moreover, numerous other factors may have contributed to the discrepancy in results between the two studies.

Perhaps the most salient difference is our use of an explicit target-matching task, in contrast with the implicit adaptation task from Lametti *et al.* (2014a). While participants in the present study were instructed to modify their speech in order to match a visually presented formant target, subjects in Lametti *et al.* (2014a) were not given any external target; their changes in production are presumed to reflect the involuntary response of a feedback control mechanism to altered auditory feedback. Given the existence of well-documented differences between the properties of implicit and explicit learning, it is perhaps unsurprising that the present study did not reproduce the effects seen in Lametti *et al.* (2014a). In particular, explicit motor learning is characterized as having a large and fast but not durable effect, with a high level of variability across individuals (e.g., Benson *et al.*, 2011). All of these characteristics can accurately be applied to the results of the present study. It is possible that the large, highly variable changes associated with strategy-based motor learning could swamp any subtle influences of perceptual training on performance.

This line of reasoning suggests that longer-term motor learning (which is slow to accrue but more lasting and lower in variability) could provide a more appropriate point of comparison with the implicit learning studied by Lametti *et al.* (2014a). We did look at long-term learning in our second mixed-effects model comparing *F1* in pre-probe versus post-probe productions, and we found no significant interaction between phase and condition (*had-shift* versus *head-shift*). However, in Sec. IV B we discussed the fact that the relatively small number of trials elicited and rapid fading of biofeedback may not have been conducive to a large magnitude of long-term learning in the present study. It is possible that a difference between groups would emerge if practice conditions were modified to encourage a larger magnitude of change at the post-probe time point. However, there are no trends in the present data that explicitly point toward such an outcome; in fact, the present data show a trend for a lesser degree of long-term learning in the *had-shift* group relative to the *head-shift* group, although there was no statistically significant interaction.

It is also important to note that a high level of individual variability was observed both in the present study and in the results of the perturbed auditory feedback task in Lametti *et al.* (2014a). At the present time, we have an incomplete understanding of individual factors that might predict degree of success in matching an altered formant target or

compensating for perturbed auditory feedback. In the present study, magnitude of change in production was not significantly correlated with either perceptual boundary location (measured either at pre-test or immediately before the production task) or magnitude of change in perceptual location. Further research is needed to understand what factors are predictive of this individual variability, such as auditory or somatosensory acuity or trial-to-trial variability in speech production (*compactness*; Kartushina and Frauenfelder, 2014). Understanding these sources of variability is important because it might ultimately be useful in predicting strong versus poor responders to biofeedback training.

V. CONCLUSION

The present study examined whether perceptual training would alter performance in a task of matching a visually presented formant target that tended to make the vowel in *head* sound more similar to *had*. Based on findings from Lametti *et al.* (2014a), which served as the conceptual and methodological model for this study, it was hypothesized that individuals in the *had-shift* group, who were trained to accept acoustically intermediate tokens as acceptable instances of *head*, would exhibit a larger change in production than individuals in the *head-shift* group. This study was successful in replicating the perceptual training component of Lametti *et al.* (2014a), since *had-shift* and *head-shift* groups differed significantly in perceptual boundary location at all points after baseline. Both groups also made a significant change in normalized *F1* height in production, indicating that visual biofeedback was successful in producing at least short-term changes in speech production. Contrary to our hypothesis, though, the *had-shift* and *head-shift* groups did not differ in the magnitude of change in production. We suggest that the discrepancy between the present findings and those of Lametti *et al.* (2014a) may be attributable to the explicit versus implicit nature of the speech-motor learning tasks used in the two studies. Specifically, subtle effects of perceptual training may not be detectable in the large-magnitude, highly variable changes in production that are typical of explicit learning outcomes. It is still of interest to ask whether an effect of perceptual training could be detected if we assess generalization learning over a more extended period of training. Follow-up research along these lines would additionally be advantageous because, relative to compensation for perturbed auditory feedback, it would provide a more accurate reflection of the type of activities that typically take place during speech training in a clinical or pedagogical setting.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant No. NIDCD- R01DC013668).

¹See supplementary material at <https://doi.org/10.1121/1.5089218> for the complete experimental protocol and complete data and code used to generate the results reported here.

Baese-Berk, M. M. (2010). "An examination of the relationship between speech perception and production," Doctoral dissertation, Northwestern University, Chicago, IL.

- Benson, B. L., Anguera, J. A., and Seidler, R. D. (2011). "A spatial explicit strategy reduces error but interferes with sensorimotor adaptation," *J. Neurophysiol.* **105**(6), 2843–2851.
- Bliss, H., Abel, J., and Gick, B. (2018). "Computer-assisted visual articulation feedback in L2 pronunciation instruction," *J. Second Lang. Pronunciation* **4**(1), 129–153.
- Boersma, P., and Weenink, D. (2016). "Praat: Doing phonetics by computer (version 6.0.16) [computer program]," <http://www.praat.org/> (Last viewed February 4, 2019).
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," *J. Acoust. Soc. Am.* **101**(4), 2299–2310.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., and Munafò, M. R. (2013). "Power failure: Why small sample size undermines the reliability of neuroscience," *Nat. Rev. Neurosci.* **14**(5), 365–376.
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., and Akahane-Yamada, R. (2003). "Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast," *Neuroimage* **19**(1), 113–124.
- Davis, S. M., and Drichta, C. E. (1980). "Biofeedback theory and application in allied health," *Biofeedback Self-Regul.* **5**(2), 159–174.
- Edeal, D. M., and Gildersleeve-Neumann, C. E. (2011). "The importance of production frequency in therapy for childhood apraxia of speech," *Am. J. Speech-Lang. Pathol.* **20**(2), 95–110.
- Elman, J. L. (1981). "Effects of frequency-shifted feedback on the pitch of vocal productions," *J. Acoust. Soc. Am.* **70**(1), 45–50.
- Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., and Hagoort, P. (2017). "Individual variability as a window on production-perception interactions in speech motor control," *J. Acoust. Soc. Am.* **142**(4), 2007–2018.
- Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F. H., Lane, H., and Perkell, J. S. (2010). "An investigation of the relation between sibilant production and somatosensory and auditory acuity," *J. Acoust. Soc. Am.* **128**(5), 3079–3087.
- Gick, B., Bernhardt, B., Bacsfalvi, P., and Wilson, I. (2008). "Ultrasound imaging applications in second language acquisition," in *Phonology and Second Language Acquisition*, edited by J. G. Hansen Edwards and M. L. Zampini (John Benjamins, Amsterdam), pp. 309–322.
- Guediche, S., Holt, L. L., Laurent, P., Lim, S.-J., and Fiez, J. A. (2014). "Evidence for cerebellar contributions to adaptive plasticity in speech perception," *Cereb. Cortex* **25**(7), 1867–1877.
- Houde, J. F., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* **279**(5354), 1213–1216.
- Houde, J. F., and Jordan, M. I. (2002). "Sensorimotor adaptation of speech I: Compensation and adaptation," *J. Speech Lang. Hear. Res.* **45**(2), 295–310.
- Jones, J. A., and Munhall, K. G. (2000). "Perceptual calibration of F0 production: Evidence from feedback perturbation," *J. Acoust. Soc. Am.* **108**(3), 1246–1251.
- Jones, J. A., and Munhall, K. (2005). "Remapping auditory-motor representations in voice production," *Curr. Biol.* **15**(19), 1768–1772.
- Kartushina, N., and Frauenfelder, U. H. (2014). "On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation," *Front. Psychol.* **5**, 1246.
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2015). "The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds," *J. Acoust. Soc. Am.* **138**(2), 817–832.
- Kawahara, H. (1995). "Transformed auditory feedback: The collection of data from 1993.1 to 1994.12 by a new set of analysis procedures," TRH-120, ATR Human Information Processing Research Laboratories, Kyoto, pp. 1–52.
- Kawahara, H., Morise, M., Banno, H., and Skuk, V. G. (2013). "Temporally variable multi-aspect N-way morphing based on interference-free speech representations," in *The Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2013 Asia-Pacific, pp. 1–10.
- Krakauer, J. W., and Mazzoni, P. (2011). "Human sensorimotor learning: Adaptation, skill, and beyond," *Curr. Opin. Neurobiol.* **21**(4), 636–644.
- Lametti, D. R., Krol, S. A., Shiller, D. M., and Ostry, D. J. (2014a). "Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning," *Psychol. Sci.* **25**(7), 1325–1336.
- Lametti, D. R., Oostwoud Wijdenes, L., Bonaiuto, J., Bestmann, S., and Rothwell, J. C. (2016). "Cerebellar tDCS dissociates the timing of perceptual decisions from perceptual change in speech," *J. Neurophysiol.* **116**(5), 2023–2032.
- Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., and Ostry, D. J. (2014b). "Plasticity in the human speech motor system drives changes in speech perception," *J. Neurosci.* **34**(31), 10339–10346.
- Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., and Schmidt, R. A. (2008). "Principles of motor learning in treatment of motor speech disorders," *Am. J. Speech-Lang. Pathol.* **17**(3), 277–298.
- MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., and Munhall, K. G. (2012). "Children's development of self-regulation in speech production," *Curr. Biol.* **22**(2), 113–117.
- McAllister Byun, T. (2017). "Efficacy of visual-acoustic biofeedback intervention for residual rhotic errors: A single-subject randomization study," *J. Speech Lang. Hear. Res.* **60**, 1175–1193.
- McAllister Byun, T., and Campbell, H. (2016). "Differential effects of visual-acoustic biofeedback intervention for residual speech errors," *Front. Human Neurosci.* **10**, 567.
- McAllister Byun, T., and Hitchcock, E. R. (2012). "Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation," *Am. J. Speech-Lang. Pathol.* **21**(3), 207–221.
- McAllister Byun, T., Swartz, M. T., Halpin, P. F., Szeredi, D., and Maas, E. (2016). "Direction of attentional focus in biofeedback treatment for /r/ misarticulation," *Int. J. Lang. Commun. Disord.* **51**(4), 384–401.
- McAllister Byun, T., and Tiede, M. (2017). "Perception-production relations in later development of American English rhotics," *PLoS One* **12**(2), e0170222.
- McDougle, S. D., Ivry, R. B., and Taylor, J. A. (2016). "Taking aim at the cognitive side of learning in sensorimotor adaptation tasks," *Trends Cognit. Sci.* **20**(7), 535–544.
- Ménard, L., Perrier, P., Aubin, J., Savariaux, C., and Thibeault, M. (2008). "Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children," *J. Acoust. Soc. Am.* **124**(2), 1192–1206.
- Nielsen, K. (2011). "Specificity and abstractness of VOT imitation," *J. Phonetics* **39**(2), 132–142.
- Niziolek, C. A., and Guenther, F. H. (2013). "Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations," *J. Neurosci.* **33**(29), 12090–12098.
- Okuno, T., and Hardison, D. M. (2016). "Perception-production link in L2 Japanese vowel duration: Training with technology," *Lang. Learn. Technol.* **20**(2), 61–80.
- Olson, D. J. (2014). "Benefits of visual feedback on segmental production in the L2 classroom," *Lang. Learn. Technol.* **18**(3), 173–192.
- Pardo, J. (2013). "Measuring phonetic convergence in speech production," *Front. Psychol.* **4**, 559.
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (2004). "The distinctness of speakers' /s/—/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," *J. Speech Lang. Hear. Res.* **47**(6), 1259–1269.
- Purcell, D. W., and Munhall, K. G. (2006a). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," *J. Acoust. Soc. Am.* **120**(2), 966–977.
- Purcell, D. W., and Munhall, K. G. (2006b). "Compensation following real-time manipulation of formants in isolated vowels," *J. Acoust. Soc. Am.* **119**(4), 2288–2297.
- Reddy, S., and Stanford, J. (2015). "A web application for automated dialect analysis," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp. 71–75.
- Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., and Yuan, J. (2014). "FAVE (Forced Alignment and Vowel Extraction) Program Suite (version 1.2.2) [computer program]," <http://doi.org/10.5281/zenodo.9846> (Last viewed February 4, 2019).
- Shiller, D. M., Gracco, V. L., and Rvachew, S. (2010). "Auditory-motor learning during speech production in 9–11-year-old children," *PLoS One* **5**(9), e12975.
- Shiller, D. M., and Rochon, M.-L. (2014). "Auditory-perceptual learning improves speech motor adaptation in children," *J. Exp. Psychol.: Human Percept. Perform.* **40**(4), 1308.
- Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (2009). "Perceptual recalibration of speech sounds following speech motor learning," *J. Acoust. Soc. Am.* **125**(2), 1103–1113.

- Shuster, L. I., Ruscello, D. M., and Smith, K. D. (1992). "Evoking using visual feedback," *Am J. Speech-Lang. Pathol.* 1(3), 29–34.
- Shuster, L. I., Ruscello, D. M., and Toth, A. R. (1995). "The use of visual feedback to elicit correct /r/," *Am J. Speech-Lang. Pathol.* 4(2), 37–44.
- Taylor, J. A., Krakauer, J. W., and Ivry, R. B. (2014). "Explicit and implicit contributions to learning in a sensorimotor adaptation task," *J. Neurosci.* 34(8), 3023–3032.
- Terband, H., Maassen, B., Guenther, F., and Brumberg, J. (2014). "Auditory–motor interactions in pediatric motor speech disorders: Neurocomputational modeling of disordered development," *J. Commun. Disord.* 47, 17–33.
- Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (2007). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," *J. Acoust. Soc. Am.* 122(4), 2306–2319.
- Volin, R. A. (1998). "A relationship between stimulability and the efficacy of visual biofeedback in the training of a respiratory control task," *Am J. Speech-Lang. Pathol.* 7(1), 81–90.