Sensorimotor adaptation affects perceptual compensation for coarticulation

William L. Schuerman[a)]
Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Srikantan Nagarajan
Department of Radiology, University of California–San Francisco School of Medicine, San Francisco, California 94143, USA

James M. McQueen[b)]
Radboud University, Nijmegen, the Netherlands

John Houde
Department of Otolaryngology Head and Neck Surgery, University of California–San Francisco School of Medicine, San Francisco, California 94143, USA

(Received 26 October 2016; revised 22 March 2017; accepted 23 March 2017; published online 14 April 2017)

A given speech sound will be realized differently depending on the context in which it is produced. Listeners have been found to compensate perceptually for these coarticulatory effects, yet it is unclear to what extent this effect depends on actual production experience. In this study, whether changes in motor-to-sound mappings induced by adaptation to altered auditory feedback can affect perceptual compensation for coarticulation is investigated. Specifically, whether altering how the vowel [i] is produced can affect the categorization of a stimulus continuum between an alveolar and a palatal fricative whose interpretation is dependent on vocalic context is tested. It was found that participants could be sorted into three groups based on whether they tended to oppose the direction of the shifted auditory feedback, to follow it, or a mixture of the two, and that these articulatory responses, not the shifted feedback the participants heard, correlated with changes in perception. These results indicate that sensorimotor adaptation to altered feedback can affect the perception of unaltered yet coarticulatorily-dependent speech sounds, suggesting a modulatory role of sensorimotor experience on speech perception. © *2017 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4979791]

[ZZ] Pages: 2693–2704

I. INTRODUCTION

The drive to find parity between production and perception reflects the fact that humans are both producers and perceivers of speech (Liberman and Whalen, 2000; Casserly and Pisoni, 2010). One of the central questions in speech perception is whether and to what extent our perception of an acoustic speech signal maps onto the physical mechanisms utilized to produce that sound (Liberman and Mattingly, 1989). While some theories claim that representations accessed during speech perception can be described succinctly in terms of acoustics (e.g., Blumstein and Stevens, 1981), others posit that perception involves accessing representations more directly related to the articulatory gestures that generated the speech (e.g., Liberman and Mattingly, 1985; Fowler, 1986; Poeppel and Monahan, 2011). This study sought to contribute to this debate by asking whether altering speakers' articulation-to-sound mapping for the production of a vowel has consequences for their perception of coarticulated consonants whose interpretation is dependent on vowel context.

While many phoneticians and psycholinguists describe the representations utilized for perception in articulatory terms, researchers in the field of speech motor control have, somewhat paradoxically, as Hickok *et al.* (2011) has noted, more consistently characterized speech production not as implementing rigid articulatory programs but as attempting to hit acoustic or somatosensory targets (Houde and Nagarajan, 2011). Such models suggest that the articulatory sequences themselves are flexible and can be changed in order to generate a particular acoustic pattern. Therefore, the stability of an articulatory motor program rests only on its ability to consistently generate intended sensory targets. If speech perception involves mapping from acoustics to articulation, altering this mapping in a production task should alter speech perception as well.

Substantial evidence for a sensory-centric view of speech production stems from experiments utilizing Altered Auditory Feedback (AAF) devices (Houde and Jordan, 1998, 2002) that enable researchers to manipulate spectral and temporal properties of a speaker's voice in real time. In response to repeated and consistent perturbations of auditory feedback, speakers alter their productions to more closely

[a)]Electronic mail: Will.Schuerman@mpi.nl
[b)]Also at: Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.

approximate their intended sensory outcomes, and this adaptation effect persists even when altered feedback is replaced by masking noise or removed (Purcell and Munhall, 2006). Similar experiments using altered somatosensory feedback suggest that sensory target need not necessarily be acoustic (Lametti *et al.*, 2012).

In addition to investigating speech motor control, this technique has also probed the relationships between production and perception. For example, participants who show more acute discrimination of first formant differences are also found to compensate more in response to perturbations of that formant (Villacorta *et al.*, 2007). Other experiments suggest that adaptation to AAF may alter the way speech is perceived, possibly due to a "restructuring" of the motor-to-acoustic mappings.

In a study by Shiller *et al.* (2009), for example, participants were asked to produce [s]-initial CV or CVC words under conditions of altered (AF) or unaltered (UF) auditory feedback. Prior to and following speech training, both groups identified stimuli along a continuum between "a said" and "a shed." Compared to the pre-test, the AF group reported more instances of [s], while the UF group reported fewer instances of [s]. This suggests that the changes in the representations accessed during phoneme categorization had been altered by the participants' experiences during the production task. This claim was further supported by the results of a passive listening group which listened to an average participant from the AF group, but showed no difference between pre- and post-exposure phoneme categorization.

Adaptation to AAF can also affect subsequent vowel perception. Lametti *et al.* (2014b) performed two experiments in which participants were tested on categorization of vowel stimuli between "head" ([ɛ]) and "had" ([æ]) [experiment (exp.) 1] or "head" and "hid" ([ɪ]) (exp. 2). The purpose of the experiment was to determine which of the influences, auditory or articulatory, would lead to a change in the categorization of the test stimuli. Participants were separated into two groups, differing in the direction of the feedback perturbation. All participants were tasked with producing the word "head," containing the vowel [ɛ]. In one group, the frequency shifted the vowel in the direction of [æ], in which case the participants would have to articulate a more [ɪ]-like vowel in order to compensate for the shift; the direction of the perturbation was reversed in the other group. In both experiments, it was found that only the group that *articulated* into the test continuum region (as a result of adaptation) showed significant changes in vowel categorization. In exp. 1, only participants who had compensated for the shifted feedback by articulating a more [ɪ]-like vowel during the production of the word "head" demonstrated a change in the perception of a continuum between [ɛ] and [ɪ], while in exp. 2, it was found that only participants who articulated a more [æ]-like vowel demonstrated a shift in perception. The authors conclude from this that the shift in perception follows the direction of the articulation rather than the acoustic input. As in Shiller *et al.* (2009), the possibility that this effect was due to simple auditory exposure to shifted feedback was ruled out by inclusion of a passive listening control group.

These two studies, Shiller *et al.* (2009) and Lametti *et al.* (2014b), suggest that altering motor-to-auditory mappings can alter the perception of speech sounds. In both studies, the authors suggest that their findings support the idea that production and perception are closely linked and that the motor system plays an active role during perception. However, in both studies, the speech segment utilized in the adaptation phase was the target in the perception phase. Therefore, it is unclear to what extent such production-induced shifts are simply the result of a bias induced by the altered feedback procedure or represent true perceptual changes induced by auditory-motor remapping.

In the current experiment, we capitalize on a well-studied linguistic phenomenon, "perceptual compensation for coarticulation" (*CFC*) to eliminate the possibility that such effects may be due to response bias. It has long been observed that the articulation of a speech sound is, as a rule, extremely influenced by its surrounding context, and the same acoustic signal can be widely interpreted based on the context in which it is produced (Liberman *et al.*, 1952). One example of the coarticulated nature of speech can be found in vowel-consonant coarticulation; in English, a vowel preceding a nasalized consonant will also tend to be nasalized (Bell-Berti and Krakow, 1991). Both behavioral (Fowler and Brown, 2000) and neuroimaging experiments (Flagg *et al.*, 2006) reveal that English listeners are sensitive to this nasalization as an indicator of an upcoming nasal consonant. The ability of listeners to recognize segments as the same under such varying conditions, and to utilize such cues, was a primary factor leading to the characterization of representations involved in speech perception as ultimately articulatory in nature (Liberman and Mattingly, 1985).

In addition to being sensitive to acoustic cues for coarticulation, listeners have been found to "undo" common coarticulatory effects. For example, in the study by Fowler and Brown (2000) on the perception of vowel-nasal consonant sequences, listeners were found to perceive a nasalized vowel as *less* nasal if it was followed by a nasal consonant compared to when it was followed by an oral consonant. Perceivers perceptually "compensated" for the coarticulatory effects of the nasal consonant on the preceding vowel by attributing acoustic information from one segment to the following segment. Research has identified a range of contexts in which *CFC* effects occur (e.g., Mann and Repp, 1980; Repp and Mann, 1981; Mann and Soli, 1991; Lotto and Kluender, 1998; Elman and McClelland, 1988; Mitterer and Blomert, 2003; Fowler, 2006), including synthesized, natural, and even non-speech contexts.

Purely perceptual experiments have attempted to disentangle articulatory and acoustic accounts for *CFC* by demonstrating that non-linguistic stimuli may also affect speech categorization (Holt, 2005) or by utilizing stimuli that make different predictions based on articulatory or acoustic information (Viswanathan *et al.*, 2010). In this experiment, however, we attempt to more directly probe the interactions between production and perception by altering the relationship between articulation and acoustics for a particular speech sound and then examining whether this remapping may affect the *CFC* response. If adaptation to AAF involves

sensorimotor remapping, then generalization of the remapping to the unadapted speech sound would suggest that the perception of continuous speech involves active utilization of articulatory knowledge to classify speech sounds. Furthermore, testing *CFC* responses on unadapted segments rules out the possibility that any observed effects can be attributed to response bias.

However, this does not exclude the possibility that any observed changes in phonetic categorization may be caused simply by altered sensory experience (exposure to a non-standard vowel), rather than changes in the relationship between articulatory and acoustic information. Therefore, we conducted a control experiment in which participants were exposed to either altered (i.e., what was heard) or unaltered (what was said) recordings from exp. 1. If such sensory exposure is sufficient to induce shifts, then differences in categorization should be observed between groups.

II. EXPERIMENT 1

A clear *CFC* effect that has been demonstrated in the literature is the effect of vowel quality on the categorization of a preceding fricative (Kunisaki and Fujisaki, 1977; Mann and Repp, 1980; Whalen, 1981; Nittrouer and Studdert-Kennedy, 1987). For example, when producing the word "sheep," the tongue position for [i] is already being prepared during the articulation of the fricative. This has the effect of raising or lowering the centroid frequency of the coarticulated fricative. The centroid frequency will tend to be higher before [i] and lower before [u]. Therefore, in fricative vowel sequences, a certain portion of the "lowness/highness" of the centroid frequency of the preceding fricative can be attributed to the speaker's preparation to articulate the following vowel (in the same manner that nasality on a vowel can be attributed to a following nasal consonant). Early research by Kunisaki and Fujisaki (1977) found different perceptual responses to fricative stimuli along a continuum between [s] and [ʃ] dependent on the quality of the following vowel. These responses ran counter to the articulatory effects; listeners were more likely to categorize an ambiguous fricative between [s] and [ʃ] as [s] in the context [-u], and [ʃ] in the context [-i]. The results suggest that listeners perceptually compensated for the effects of coarticulation on the acoustic realization of the intended phones.

In the first experiment, we investigated whether changes in the articulation of the vowel [i] due to exposure to AAF may change the perception of an unshifted yet contextually dependent fricative continuum between "see" and "she." In two sessions, we asked participants to categorize fricative-vowel stimuli after short periods of production training under conditions of altered (AF) and unaltered (UF) auditory feedback. This within-subjects design enabled us to compare the same participants under conditions of unaltered and altered feedback.

In AF sessions, participants' auditory feedback in response to their productions of words containing [i] was shifted to more closely approximate their average productions of the vowel [u]. While for certain participants this shift involved some change to the first formant, corresponding to vowel height, the majority of the shifted feedback was confined to altering the second formant, which generally corresponds to vowel frontness/backness. To oppose this shift, participants would need to hyper-articulate their productions of the vowel [i] in order to increase the frequency of the second formant.

We propose that articulatory and acoustic accounts make opposing predictions about how the response to AAF should affect behavior in identification tasks. There are two features of this response that distinguish these predictions. First, while total compensation has been found for very small shifts, in response to larger shifts (as utilized here), participants' compensatory articulations have been found to only counteract shifts in auditory feedback by approximately 20% (Katseff *et al.,* 2012; MacDonald *et al.,* 2011). Therefore, while participants may oppose the shift induced by AAF, they are still being exposed to stimuli that are shifted compared to an average production. Second, there exists the possibility that participants may fail to produce opposing articulatory responses upon exposure to altered feedback. A meta-analysis conducted by MacDonald *et al.* (2011) of seven AAF experiments (116 participants) found a range of inter-speaker variability in vocal responses to the auditory feedback. In all data analyzed, the target word to be produced was "head," and the shift consisted of an increase in F1 by 200 Hz and a decrease in F2 by 250 Hz. Measuring the difference between the last 15 utterances of the baseline phase and the last 15 utterances of the full shift phase, the authors found that 14 of the 116 talkers exhibited a following response in F2, with an additional 4 exhibiting a following response in both F1 and F2 (15% of participants).

These two features, that compensatory adaptation only counteracts a small percentage of the shifted feedback and that participants may not necessarily oppose, but rather follow, the shift in feedback, enable us to dissociate the predictions of articulatory and acoustic accounts. Some participants may oppose the shift in feedback while others follow it, resulting in divergent articulatory responses (more or less [u]-like) compared to baseline. However, due to the fact that compensatory adaptation is only partial, regardless of articulatory behavior all participants will hear a vowel that is more [u]-like than normal. Therefore, the acoustic and articulatory accounts make opposite predictions as to the direction of the perceptual change. In terms of the articulation of the vowel, opposing the shift in feedback leads the articulated vowel to become more [i]-like, in which case we would expect an increase in the amount of stimuli subsequently perceived as "she." A following response would lead the articulated vowel to become more [u]-like, in which case we would expect a decrease. If, however, it is not articulation that matters but auditory experience, then we should see no differences between the two groups in the direction of the change in fricative identification. Overall, there could be increases or decreases in the proportion of "she" responses (because exposure to the vowel during production training could have either assimilative or contrastive effects on vowel perception), but the direction of this effect should be the same in opposers as in followers.

A. Participants

Twenty-four North American native speakers of English (14 = female) took part in the experiment. Average age at time of testing was 27.42 years (min = 21, max = 36), and all were residents of the Northern California Bay Area at time of testing.

B. Ethics declaration

Ethical approval for this study was obtained from the Institutional Review Board of the University of California, San Francisco (exp. 1), as well as the Ethics Committee of the Social Sciences Faculty of Radboud University (exp. 2). Written consent was obtained from each participant on the first day of the study. Participants were informed that their participation was voluntary and that they were free to withdraw from the study at any time without any negative repercussions and without needing to specify any reason for withdrawal. All were reimbursed for their participation.

C. Design

The experiment comprised two sessions, an unaltered feedback (UF) and an altered feedback (AF) session. All participants completed both sessions. The UF session always preceded the AF session, to prevent any possible carryover effects from the altered feedback. Each session was separated by a minimum of two weeks.

Within each session, participants performed two tasks separated into blocks: a speaking task and an identification task. In the speaking task, participants read aloud simple CVC words. In the perception task, participants listened to synthesized stimuli ranging along a continuum between clear "see" and clear "she," and reported via button press which word they thought they had heard. Both sessions consisted of seven identification blocks (IB) and six speaking blocks (SB), presented in alternation (Fig. 1). A session began with an IB, followed by a SB, and continued in this manner until all seven IBs and SBs had been completed. All participants completed both sessions.

D. Speaking task

1. Equipment and AAF signal processing

For both sessions, participants were seated in front of a laptop, and fitted with Beyerdynamic DT 770 PRO noise-isolating headphones. Speech was recorded utilizing a Micromic.C 520 VOCAL head-mounted microphone. Microphone input was routed through an RDL HR-mP2 Dual Microphone Preamplifier to an M-Audio Delta 1010 external sound card.

In order to determine the optimal audio processing settings for altered feedback, each participant was first recorded producing the words "heed," "who'd," and "had," corresponding to the point vowels [i], [u], and [æ], respectively. Spectral measurements using varying levels of LPC coefficients and frequency cutoff levels were then generated, and the number of coefficients and the frequency cutoff level giving the best formant tracking were selected by visual

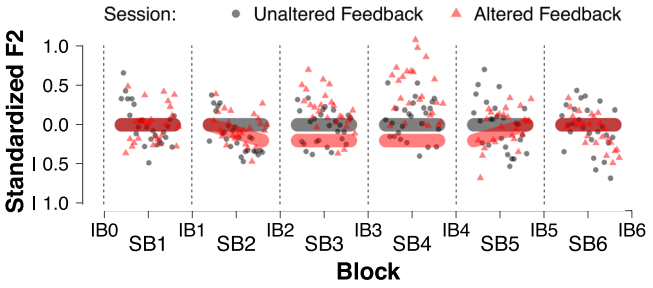

FIG. 1. (Color online) Overview of experiment design. Each session consisted of alternating identification blocks (IB) and speaking blocks (SB). Grey and red bars indicate auditory feedback in the unaltered (UF) and altered (AF) sessions, respectively. In the AF session, auditory feedback was shifted 20% of the acoustic distance towards each participant's average [u] values, with ramp-up, hold, and ramp-down phases. Grey circles (UF session) and red triangles (AF session) denote grand average standardized F2 by trial.

inspection of the results. Following calibration, participants were recorded producing the same words three times each. The averages of the first and second resonant frequencies for the point vowels [i] and [u] were then utilized as the basis for the frequency shifted feedback (see Sec. II D 3).

The input signal from the microphone was analyzed by a frequency alteration device (FAD) as described by Katseff *et al.* (2012), based on the method of sinusoidal synthesis (Quatieri and McAulay, 1986). The input signal was recorded with a 32-bit sampling depth at a rate of 11 025 Hz, generating a frame size of 3 ms. This frame was then ported into a 400 sample, 36 ms buffer for spectral analysis. The acoustic envelope was converted into a narrow-band magnitude frequency spectrum in order to obtain the spectral envelope from the signal. This spectral envelope was utilized to estimate, and modify, the fundamental and resonant frequencies present in the recorded input. The new narrow band magnitude frequency spectrum was then used as the basis for sinusoidal synthesis, in which each harmonic of the spectrum is represented as a sinusoid. The acoustic signal was then generated by sinusoidal addition. In order to maintain continuity between frames, the preceding 3 ms frame was also used as input to the estimation of the current 3 ms frame, totaling a 6 ms analysis window. Additional processing delays between microphone and headphones led to an overall delay of approximately 12 ms, regardless of whether or not feedback had been altered.

2. Speaking task stimuli

Stimuli for the speaking task consisted of 13 monosyllabic English words in orthographic form ("peep," "beep," "deep," "keep," "peat," "beet," "bead," "deed," "keyed," "peak," "beak," "teak," "geek"). Words were presented in white 30-point font against a black background. All words began with a voiced or voiceless stop consonant, followed by the vowel [i], and ended with a voiced or voiceless stop consonant. Crucially, no words contained a fricative consonant.

3. Procedure

In a SB, participants were instructed to read aloud the words presented on screen. The 13 stimulus words were

presented in random order. Each word was presented twice, with all words being presented at least once before repetition began, totaling 26 presentations per block.

In the UF session, there was no altered feedback except for the downsampling of the signal and the processing delay of 12 ms. In the AF session, there was no altered feedback in the first (baseline) SB. Starting in the second SB ("on-ramp"), the participants' auditory feedback was shifted towards their average [u] production. In their first experiment, Purcell and Munhall (2006) applied a fixed frequency shift of ±200 Hz to each participant's F1, which they argue may have led to differing compensatory responses given the vocal tract parameters of the individual participants. Furthermore, because our study utilized both male and female participants (with differing vocal tract sizes), a fixed frequency shift may not create similar perceptual consequences for each participant. Therefore, following Niziolek and Guenther (2013), we defined custom frequency shifts for each participant, based on per-participant differences in average F1 and F2 for the vowels [i] and [u]. For each participant, maximum feedback was 20% of the distance in F1-F2 space from [i] to [u]. Feedback alteration began at 0% perturbation at trial 1 of SB 2, and reached a maximum of 20% perturbation of both F1 and F2 by trial 26 ("on-ramp"). This averaged −263.29 Hz for F2, and 6.67 Hz for F1, indicating that the primary dimension of the shift was along F2. Perturbation remained at 20% over blocks three and four. In block five ("off-ramp"), perturbation decreased from 20% at trial 1 to 0% at trial 26. Feedback was unaltered in block six.

4. Acoustic analysis

Semi-automated formant measurements were conducted with the same software utilized for the feedback alteration, and MATLAB (Mathworks, 2012). An algorithm measuring the periodicity of the acoustic signal was first utilized to identify the start and end time of the vowel segment of each recording. Formant tracking results were visually inspected to determine that measurements were taken from vowel midpoint or the closest suitable point.

While the altered feedback was defined in terms of change in F1 and F2, the primary direction of change occurred along the F2 dimension. We therefore excluded F1 measurements from analysis. The acoustic measurements for F2 were converted from Hz to Mel, a logarithmic frequency scale that more closely approximates human hearing, using the formula $2595 \times (\log(1 + (F2\_\mathrm{Hz}/700)))$.

In order to compare production responses across participants, Mel frequency measurements of F2 were standardized according to the following procedure: For each participant, the first SB of each experimental session was designated as the "baseline" block for that session. The Mel value of each trial was subtracted from the average Mel value of the baseline, and then standardized by dividing by the standard deviation of the baseline block,

$$F2_{standardized} = (F2 - \mathrm{mean}(F2_{baseline}))/sd(F2)_{baseline}. \quad (1)$$

Therefore, these standardized values represented changes from baseline production values in each session with respect to baseline formant variability. All statistical tests were performed utilizing these standardized measurements.

For comparison with other studies, we also calculated an alternative compensation index according to the following formula:

$$F2_{compensation} = 1 - \frac{(F2 - F2_{baseline}) - shiftF2}{shiftF2}, \quad (2)$$

where for a given trial, compensation was defined as the percentage return towards baseline from the shifted formant value. Subtracting the resulting proportion from 1 separates responses opposing the direction of the shift (positive values) from those following it (negative values).

E. Identification task

1. Stimuli

Stimuli for the identification task were created according to the following: A female native speaker of North American English produced the sentence "say the word 'see' " three times. After selecting the most natural sounding version, the word "see" was extracted. The duration values for each phone segment and the prosodic contour of the word was extracted utilizing PRAAT (Boersma and Weenink, 2013) and converted into a text-based format readable by Mbrola (Dutoit *et al.*, 1996), a text-based diphone synthesizer. This method enabled the creation of two endpoint stimuli with identical phone durations and prosodic contours. Sample-by-sample interpolation (McQueen, 1991) was then utilized to create a 100-step continuum between unambiguous "see" and unambiguous "she."

2. Pre-test

Before the first session, participants completed a pre-test, utilizing the same materials as the test phase, in order to determine the stimulus-step at which participants switched from hearing "see" to "she," by means of an adaptive staircase procedure. During pilot testing it was found that consecutive presentation of ambiguous stimuli led to shifts in the categorical perception boundary. Therefore, a random "filler" endpoint stimulus (0 or 100) was presented in odd-numbered trials. These filler trials did not count towards the staircase procedure results. Even numbered trials first began with presentation of the endpoint stimuli (0 or 100, corresponding to unambiguous "see" and unambiguous "she"). Initial step size was set at 100; after trial four, step size decreased by half after each reversal until either 12 reversals had occurred or step size remained at one for three consecutive trials.

3. Procedure

In an identification block (IB), participants were instructed to listen to an auditorily presented stimulus and indicate via keyboard whether the stimulus sounded more like "see" (button 1) or "she" (button 2). Participants were

instructed to wait until the sound file had finished before responding, and to respond with whichever hand felt most comfortable. As it was essential to the experiment that participants listen to both the fricative and the following vowel in order to assess the *CFC* effect, responses made prior to the end of stimulus presentation were not accepted. In such cases participants were verbally reminded to wait until stimulus presentation had finished before responding. While this eliminated the possibility of comparing reaction times, this method ensured that participants listened to both the fricative and the vowel before responding.

Presented stimuli consisted of ± 1, 3, 5, 7, 9, 11, 13, 15, 17 steps above and below each participant's pre-test boundary as well as the endpoint stimuli (step 0 and step 100), for a total of 20 stimuli. Stimulus presentation was pseudo-randomized into sets of four stimuli: One stimulus fewer than 10 steps above pre-test boundary, one stimulus fewer than 10 steps below, one stimulus greater than 10 steps above, and one stimulus greater than 10 steps below. In an IB, each stimulus was presented twice for a total of 40 presentations per IB, with all stimuli presented once before repetition. Responses in the identification task were coded as 0 (for "see") and 1 (for "she").

F. Data exclusions

Due to software failure, two participants were unable to complete IBs five and six and SB six of the UF session. When possible, all remaining data from these participants were included in the analyses.

Out of 7436 total trials, 297 trials (4% of total data) were excluded prior to acoustic analysis due to either recording error (e.g., wrong word spoken or spoken outside of recording window) or failure of the formant tracker to locate a stable point for formant measurement. In the remaining trials, each participant's formant measurements were visually inspected for formant tracking errors. Five trials were removed in which formant values for F1 were exceedingly large (greater than 5 standard deviations from centered mean of all participants). We then excluded trials in which F2 was greater than 3 standard deviations from the mean of all participants of that gender (15 trials). Finally, values for F1 and F2 were centered and scaled on a by-participant basis, and based on visual inspection of the data, a conservative value of $\pm 4sd$ was set as the cutoff at which extreme values would more likely have arisen due to tracker error. In total, 41 trials were excluded from analysis in this manner, comprising 0.5% of the data.

G. Results

1. Production results

Inspection of individual results in the AF session revealed that, while many participants appeared to oppose the AAF, a number of participants exhibited a decrease in F2 compared to baseline. This suggests that in contrast to an opposing response, certain participants responded to the altered feedback by producing a "following" response in the same direction as the shifted feedback (MacDonald *et al.*, 2011).

We therefore conducted a simple *post hoc* division of participants into three groups based on whether standardized F2 averaged over the third and fourth blocks of the AF session, during which the altered auditory feedback was held constant at its maximum value, was greater than baseline for both blocks, less than baseline for both blocks, or above baseline in one block but below baseline in another. Standardized F2 was above baseline for 11 participants (mean of blocks 3 and $4 = 0.917$, standard deviation (sd) $= 1.353$) and below baseline for 7 participants (mean of blocks 3 and $4 = -0.508$, sd $= 1.128$), and mixed for 6 participants (mean of blocks 3 and $4 = -0.003$, sd $= 0.867$). We classified these groups as "opposers," "followers," and "mixed," respectively. Average standardized F2 for the three groups in the UF and AF sessions are displayed in Fig. 2 [panels (A)–(C)].

Standardized F2 values were analyzed with linear mixed effects regression in R (R Development Core Team, 2013) using the LME4 package (Bates *et al.*, 2015). This technique is robust to missing data and allowed us to include two participants who failed to complete the last few blocks of the UF session. Due to the standardization procedure, block 1 contained essentially no variance and therefore was excluded from the analyses. Model fitting began with a maximal model containing fixed effects for "type" (opposer, follower, mixed), session (AF or UF), and block (1–6; as categorical variables), as well as all two-way and three-way interaction terms. We utilized a maximum random effects structure (Barr *et al.*, 2013) including a random intercept for participant as well as random slopes for block and session. Significance of predictors was assessed by conducting likelihood ratio tests between nested models with and without the candidate fixed effect term. Removal of the three-way interaction term for type, block and session was found to significantly decrease model fit [$\chi^2(8) = 63.116, p < 0.001$]. To simplify model interpretation, we subsetted the data and fit separate models for the AF and UF sessions.

For each by-session model, we began with a maximal random effect structure including random intercepts for participant and a random slope for block, and a maximal fixed effect structure including main effects for type and block as well as the interaction term. Removal of this interaction term significantly reduced model fit in the AF session [$\chi^2(8) = 19.377, p < 0.013$]. Model estimates (with p-values obtained by Satterthwaite approximation) are reported in contrast to block 2 of the AF session for the opposer group. For this group, standardized F2 was found to increase significantly in block 3 [estimate (est.) $= 0.610$, standard error (SE) $= 0.158$, $p < 0.001$] and block 4 (est. $= 1.038$, SE $= 0.324$, $p = 0.005$). In block 2, the opposers differ significantly from followers (est. $= -0.516$, SE $= 0.236$, $p < 0.05$) but not from the mixed group. This difference between opposers and followers from block 2 onwards exhibits a significant decrease in block 3 (est. $= -0.761$, SE $= 0.253$, $p < 0.007$) and a marginally significant decrease in block 4 (est. $= -1.083$, SE $= 0.520$, $p = 0.05$). The difference between opposers and mixed becomes

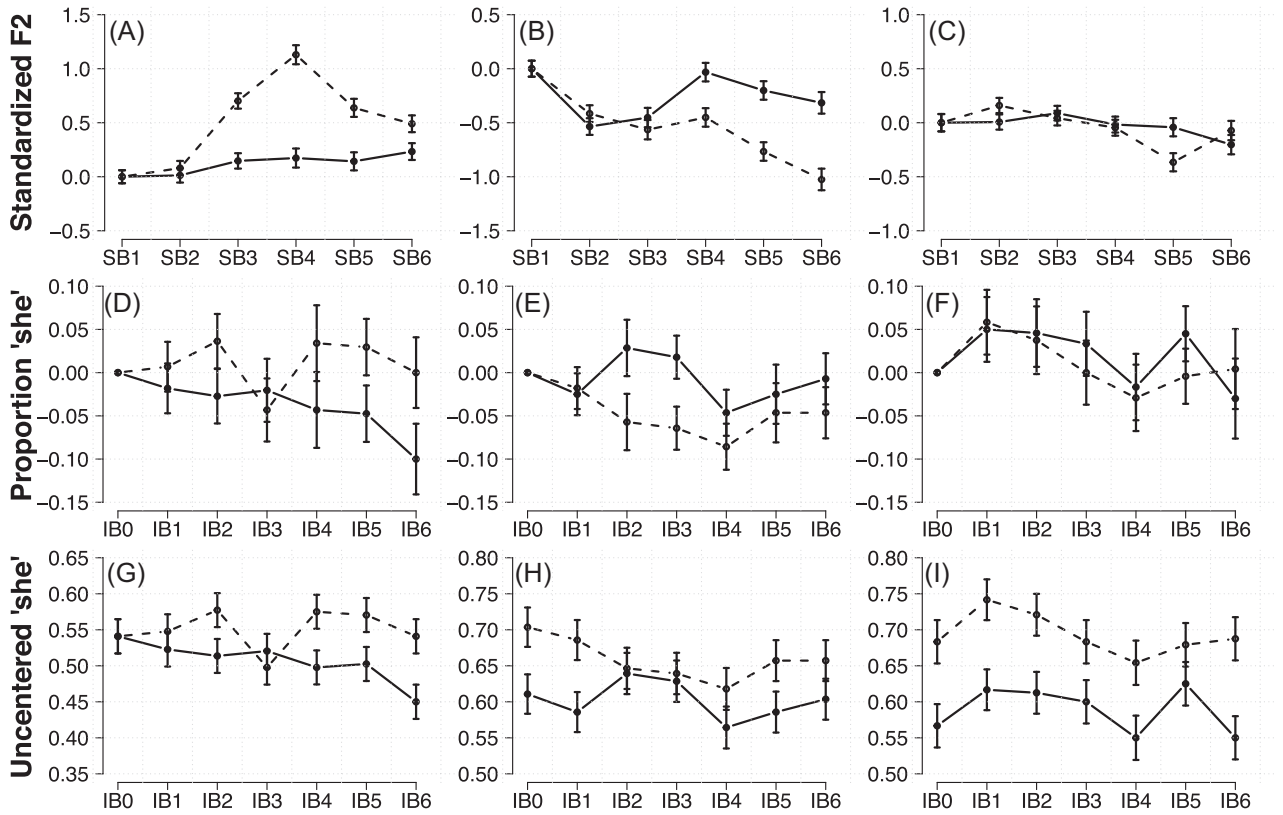

FIG. 2. Average standardized F2 and proportion "she" responses by type of participant. Solid lines refer to the UF session, dashed lines refer to the AF session. Panels (A)–(C) represent average standardized F2 for opposers (A), followers (B), and mixed (C) participants. Panels (D)–(F) represent group-averages of baseline-centered proportions of "she" responses for opposers (D), followers (E), and mixed (F). Panels (G)–(I) represent group-averages of uncentered proportions of "she" responses for opposers (G), followers (H), and mixed (I).

significant in block 3 and remains significant until block 6 (all $p < 0.05$).

In contrast, in the UF session, removing the interaction term [$\chi^2(8) = 5.004$, $p = 0.75$], the main effects of type [$\chi^2(2) = 2.596$, $p = 0.27$], and the main effect of block [$\chi^2(4) = 2.756$, $p = 0.6$] all failed to significantly impact model fit. These combined results indicate that the opposer, follower, and mixed groups do not reliably differ in standardized F2 when feedback is unaltered, but produce differing response patterns when exposed to altered auditory feedback.

As has been found in previous experiments, adaptation was not sufficient to completely counteract the altered feedback (Katseff *et al.*, 2012; Purcell and Munhall, 2006; Houde and Jordan, 1998). In contrast to the average F2 compensation of 23.2% reported for perturbations of [ɛ] (MacDonald *et al.*, 2011), in this experiment opposers altered their productions by an average of only 5.6% in the third and fourth blocks of the AF session (min = 0.5%, max = 18.1%). This is likely attributable to the choice of target vowel in the current experiment, which occupies a relative endpoint in articulatory space. Followers were found to alter their F2 by only −2.9% (min = −5.8%, max = −0.3%).

2. Identification results

Our key question concerned whether participants' perceptual responses differed with respect to conditions of altered and unaltered production feedback. However, the production results indicate that participants' vocal motor behavior differed in response to AAF, justifying their division into three separate groups. Figures 2(D)–2(F) displays the baseline-centered proportion of "she" responses in each block and session for the opposers, followers, and mixed. Uncentered proportions are displayed in Fig. 2, panels (G)–(I).

These graphs indicate that, despite the staircase procedure, the proportion of "she" responses in the first block appears to have shifted between the UF and AF sessions for participants in the follower and mixed groups (mean difference AF-UF: opposers = 0.000, followers = 0.093, mixed = 0.117). Rather than averaging responses in each block over stimuli, binary coded identification responses were analyzed using mixed-effects logistic regression (Breslow and Clayton, 1993; Jaeger, 2008).

The graphs in Figs. 2(D)–2(F) suggest that with regards to overall proportion of responses, participants in all three groups tended to perceive more stimuli as "she" in the AF session compared to the UF session. However, because this increase was present from the first block in the follower and mixed groups, this overall increase in "she" responses cannot be attributed to the AAF. To assess the effect of altered auditory feedback on identification, we conducted separate regression analyses examining the effects of block and session within each group. As with the production data, model fitting proceeded by comparison of backwards-fitted nested models.

For opposers, the initial model contained a random effect structure containing random intercepts for participant and item, as well as random slopes for block and session

over participant and session over stimulus. Removing the interaction between session and block significantly decreased model fit [$\chi^2(6) = 29.372$, $p < 0.001$]. Parameter estimates [treatment coded with the baseline block (block 0) of the UF session as reference] indicate that in the UF session, the proportion of "she" responses was significantly less than baseline in the sixth block (est. $= -1.11$, $Z = -3.37$, $p < 0.001$). The estimates of the interaction indicate that in the AF session participants tended to perceive more stimuli as "she" in the second (est. $= 0.717$, $Z = 2.204$, $p < 0.028$), fourth (est. $= 0.856$, $Z = 2.641$, $p < 0.009$), fifth (est. $= 0.845$, $Z = 2.564$, $p < 0.011$), and sixth blocks (est. $= 1.097$, $Z = 3.322$, $p < 0.001$) compared to the corresponding blocks in the UF session [with respect to their change from baseline; Fig. 2(D)].

The model for the follower group included fixed effects of block, session, and the interaction term, as well as random intercepts for participant and stimulus with random slopes for session. Removing the interaction between session and block also significantly decreased model fit ($\chi^2(6) = 15.941$, $p < 0.02$). Inspection of model estimates indicates that in the UF session, the probability of perceiving stimuli as "she" in block 4 was significantly lower than in the baseline block (est. $= -0.6838$, $Z = -2.132$, $p < 0.04$). In contrast to the opposers, model estimates indicate that participants were already more likely to report stimuli as being "she" in the baseline of the AF session than in the baseline of the UF session (est. $= 1.2496$, $Z = 2.747$, $p < 0.007$). This increased tendency to report stimuli as "she" decreased significantly in the second (est. $= -1.2236$, $Z = -2.717$, $p < 0.007$) and third (est. $= -1.1650$, $Z = -2.587$, $p < 0.01$) blocks of the AF session.

For the mixed group, a model containing the same fixed effect structure was fit with random intercepts for subject and stimulus with random slopes for session and block over participant and session over stimulus. As in the production results, removal of the interaction and main effects did not significantly impact model fit (all $p > 0.05$).

3. Interactions between production and identification

As in other experiments (Shiller *et al.*, 2009; Lametti *et al.*, 2014b), no within-session significant correlations were found between production (standardized F2) and perception (centered response). The results seem to suggest that the effect of the altered feedback on perception only emerges when considering the differences in production and identification across the two sessions. We computed differences scores for both the production and the identification data: For the production results, we utilized the standardized F2 scores; for the identification results, we baseline-centered each participant's results to the average of block 0 of each session. This allowed us to compare both the identification and the production results with respect to changes from baseline in a given session. Each participant's average standardized F2 for production and average baseline-centered proportion of "she" responses for identification in the AF session were subtracted from the corresponding results in the UF session. These difference scores were found to have a

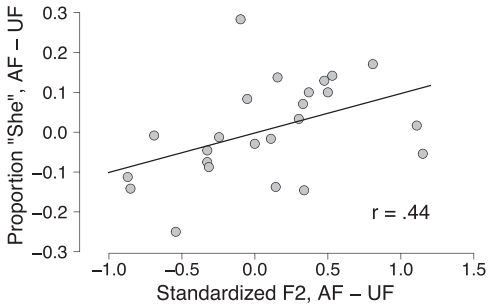

FIG. 3. Correlation of difference scores. X-axis denotes, for each participant, the average standardized F2 in the UF session subtracted from average standardized F2 in the AF session. Y-axis denotes average baseline-centered proportion of "she" responses in the UF session subtracted from those in the AF session.

positive correlation (Fig. 3; $r(22) = 0.44$, $p < 0.031$), indicating that a higher standardized F2 value in the AF session compared to the UF session correlates with an increased likelihood of "she" responses in the AF session compared to the UF session. The fact that this correlation is found with regard to by-session and by-participant baseline-centered proportions of "she" removes the possibility that this pattern is simply due to an overall difference in the number of "she" responses in one session compared to another. For blocks three and four, the two blocks corresponding to maximum perturbation in the AF session, differences scores were found to correlate only found for the fourth block [$r(22) = 0.47$, $p = 0.016$], and not for the third [$r(22) = 0.32$, $p = 0.13$]. This accords with what is seen in the group results, in which a sharp drop is seen in the proportion of "she" responses in the opposer group [Fig. 2(D)], even though standardized F2 is above baseline in this block [Fig. 2(A)].

Inspection of individual by-trial results revealed that some participants exhibited a strong opposing or following response in the initial half of a block, but then returned to baseline in the second half (or vice versa; see Fig. 4, SB3). While the reason of these within-block changes is unclear, we explored whether these may have had an effect on identification responses. If vocal behavior in the second half of an SB differed from vocal behavior in the first, this may have "cancelled out" the effect of adaptation on previous trials. If this is the case, then we surmised that stronger correlations between the difference scores may be found if standardized F2 measurements were limited to the second half of each SB

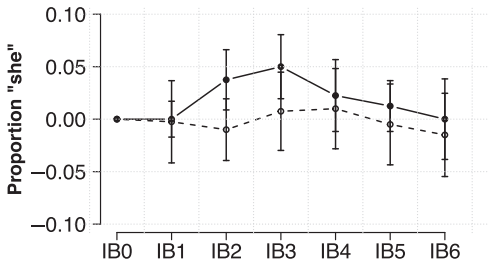

FIG. 4. Passive listening results. Filled circles and solid lines indicate average "she" responses for participants exposed to the unshifted recordings of a model participant from exp. 1. This model participant exhibited an increase in standardized F2 in response to the altered feedback that was typical of the opposer group. Empty circles and dashed lines denote participants exposed to the shifted feedback that the model participant heard during the AF session.

(trial > 13). Limiting production data to the last half of the block was found to slightly increase the correlation coefficient [$r(22) = 0.47$, $p < 0.019$]. In addition, we found that the previously non-significant correlation in block three became marginally significant if we limited our measurements to the last 10 trials of the block [$r(22) = 0.41$, $p = 0.05$]. These within-block changes in F2 may explain the sudden drop in identification scores in the third IB despite an increase in average standardized F2 in the corresponding SB.

H. Discussion

The combined results from the Speaking and Identification tasks suggest that a change in produced F2 correlated with a change in the proportion of stimuli perceived as "she." Even though opposers counteracted only 5.6% of the shifted feedback, they nevertheless exhibited an increase in the proportion of stimuli reported as "she." The direction of the correlation accords with results from similar *CFC* experiments in which a more fronted vowel is associated with more [ʃ] responses [Kunisaki and Fujisaki (1977); Mann and Repp (1980)]. This provides evidence against a purely acoustic account of the results, because each participant (both opposers and followers) heard themselves producing the vowel [i] with a much lower F2 than normal. Instead, the shift in perception correlated with changes in the motoric behavior, supporting an articulatory basis for the perceptual shift. In line with the results of Lametti *et al*. (2014b), changes in perception were variable and depended on the direction of the articulatory change, rather than universally according with the direction of the shifted auditory feedback.

The relatively large amount of following responses found in this experiment compared to others may possibly be due to the location of the target vowel ([i]). Acoustically, this vowel occupies an endpoint in terms of both F1 and F2, while articulatorily, the degree of lingual contact is greater for [i] than for non-closed vowels, such as [ɛ], which may increase the importance of somatosensory feedback relative to acoustic feedback (Mitsuya *et al*., 2015). As in the meta-analysis conducted by MacDonald *et al*. (2011), we found no significant correlation between a participant's average standardized F2 in the maximum shift blocks and standard deviation for F2 in baseline blocks in either mean centered F2 [$r(22) = 0.031$, $p = 0.88$] or standardized F2 [$r(22) = -0.036$, $p = 0.8654$].

III. EXPERIMENT 2

The results of exp. 1 suggest that sensorimotor adaptation in response to altered auditory feedback can affect the perception of an unadapted, yet contextually dependent, fricative. Furthermore, the direction of the observed effects suggests that the change in perceptual function corresponds to motoric changes rather than auditory feedback. Previous experiments have found that passive exposure to the recorded speech of a participant compensating in response to AAF fails to change perceptual function, regardless of whether the recorded speech is made by an average compensating speaker (Shiller *et al*., 2009) or consists of a random selection of stimuli take from several compensating speakers (Lametti *et al*., 2014b). However, these experiments differed from the present study in that they did not examine how passive listening to such stimuli may affect *CFC*. Therefore, in order to determine whether passive listening may also affect phonetic categorization in CFC, we recruited an additional 20 participants to perform a passive listening version of the same task. If we observe no perceptual changes similar to those observed in exp. 1, this likely indicates that sensorimotor remapping requires experiencing error between an intended sensory target and the articulatory movement enacted to produce that target (Houde and Nagarajan, 2011).

Based on the behavioral division of participants into opposers and followers in exp. 1, we decided to directly test whether two different types of auditory stimuli may elicit differential perceptual changes. Half of the participants were exposed to the recordings of one participant from exp. 1. The recordings consisted of this participant's unshifted input to the AAF device. If the passive listeners in this "unaltered" group perform like the opposers in exp. 1, we would expect an increase in the number of "she" responses.

The remaining participants were exposed to the shifted auditory feedback that this participant heard during the first experiment (in which the [i] vowel was altered to sound more like [u]). According to an auditory account (Kunisaki and Fujisaki, 1977; Mann and Repp, 1980), the passive listeners in this "altered" group should, if they exhibit a change in perception, report fewer stimuli as "she" compared to baseline.

A. Participants

Twenty-two North American native speakers of English (all residing in the Netherlands at the time of testing) took part in exp. 2. Of these participants, two were excluded due to errors in the pre-test boundary finding procedure, in which they responded "see" to almost all stimuli during the test phase. This left twenty participants (9 = M, average age = 27.55). Three of the twenty received eight euros for completing the task, while the remainder declined payment.

B. Materials

Materials for the identification task were identical to those used in exp. 1. Materials for the passive listening task consisted of either the recorded input (32-bit, 11 050 Hz sampling rate) from one male speaker from exp. 1 during the AF session (unaltered-group), or the altered auditory feedback which this speaker heard during the experiment (altered-group). As in Shiller *et al*. (2009), this speaker was selected from amongst the above-baseline participants (the opposers) for exhibiting an average but not extreme increase in standardized F2 in response to the AAF.

C. Procedure

Participants were seated in a sound-proof booth in front of a computer monitor. The identification boundary pre-test and identification task were identical to that reported in exp. 1, except that the experimental software used to deliver the stimuli was Presentation (Version 0.70, www.neurobs.com). In contrast to the previous experiment, the speaking task was replaced with a passive listening task (PLT), in which

participants were instructed to silently read the words that appeared on the screen and listen to the voice as it read each word aloud.

D. Results and discussion

As in exp. 1, the average proportion of [ʃ] responses were centered to each participant's average in the baseline block (block 0). Average proportion of "she" responses in the input group was 0.017 (sd = 0.085), while average proportion of "she" responses in the output group was −0.002 (sd = 0.104). As in exp. 1, results were analyzed using mixed-effects logistic regression. Models with random slopes for block failed to converge, therefore the model contained random intercepts for participant and centered stimulus. Backwards-fitting began with a maximal fixed effects structure including main effects for block and group (input/output) as well as the interaction term. Model comparison revealed no significant interactions, and neither the effect of group [$AIC = 2945.5$, $BIC = 2972$, $\log Lik = -1468.8$, $\chi^2(1) = 0.55, p = 0.46$] nor block [$AIC = 2945.9$, $BIC = 3005.6$, $\log Lik = -1464$, $\chi^2(6) = 10.15$, $p = 0.11$] was found to improve the null model.

While this accords with previous results (Shiller *et al.*, 2009), the lack of any change in the perceptual boundary may seem rather counterintuitive given research demonstrating that exposure to ambiguous auditory stimuli in the context of biasing lexical information can drive perceptual retuning (Samuel and Kraljic, 2009). In such experiments, participants are exposed to an ambiguous sound that falls between two phonemic categories, such as [f] and [s], in a lexical context that leads the listener to categorize the ambiguous sound as belonging to one of the two categories (Norris *et al.*, 2003). This lexical bias has been found to have a strong effect on how an ambiguous phone is perceived. The resulting effect is that if the lexical context biases listeners to categorize the ambiguous sound as [f], they are subsequently more likely to classify stimuli along an [f]-[s] continuum as [f], while the inverse is found if participants are led to categorize the ambiguous sound as [s]. Further experiments with this paradigm have found that this effect can transfer to unexposed words (McQueen *et al.*, 2006) and can remain stable for as long as 12 h (Eisner and McQueen, 2006), comparable to the durable effects seen in sensorimotor adaptation (Ostry *et al.*, 2010; Nourouzpour *et al.*, 2015). The apparent contradiction between the lexically-guided retuning results and the results of exp. 2 may be due to the specificity of the adaptation; Kraljic and Samuel (2005) found that perceptually guided lexical retuning for fricatives along an [s]-[ʃ] continuum transferred from a female training voice to a male test voice, but not in the opposite direction. The authors attribute this asymmetry to the fact that the female training stimuli were close to the frequency of the male test items, while the male training stimuli were far from the female test stimuli, suggesting that transfer may depend on acoustic similarity. This provides a reasonable explanation for the fact that we see no effects in exp. 2, as the silent listening task utilized a male voice while the identification stimuli were based on a female voice.

IV. GENERAL DISCUSSION

The results of the first experiment indicate that participants changed their vocal behavior, as reflected in acoustic measurements of their baseline standardized second formant (F2), in response to the altered auditory feedback. Our results differ from previous experiments in that, while 11 participants adapted to the feedback by opposing it (i.e., their standardized F2 increased), a relatively large number (13 of 24 participants) failed to oppose the shifted feedback. Instead, in 7 of these participants standardized F2 was found to decrease relative to baseline, following the direction of the shifted feedback.

Critically, the 11 opposers differed from the 7 followers with regard to their behavior in the identification task. The opposing group reported more instances of "she" in the altered feedback blocks while the following group exhibited a decrease in the number of stimuli identified as "she." Differences in the speaking task between altered and unaltered feedback sessions were found to correlate with corresponding differences in the identification task, suggesting that responses in the identification blocks were influenced by vocal behavior in the preceding speaking blocks. This suggests that the perceptual processes involved in compensation for coarticulation can be modulated by the observer's own sensorimotor experience.

While exposure to nonlinguistic acoustic stimuli is known affect *CFC* (Holt, 2005), the pattern of results observed in this experiment conflict with a purely auditory explanation. In both groups (opposers and followers), the effect of the altered feedback was to decrease the F2 that participants heard by a substantial amount. As adaptation in articulation only counteracted a small portion of the shift, all participants heard themselves producing the vowel [i] with a lower than average F2. We predicted that if perception is influenced by articulatory behavior rather than acoustic feedback (Lametti *et al.*, 2014b), then participants who opposed the shift in auditory feedback should exhibit an increase in the proportion of stimuli identified as "she," while those who followed the shift in feedback should exhibit a decrease. However, if the *CFC* effect is instead modulated by auditory experience, we predicted that the direction of the change in identification should be the same in opposers as in followers. The articulatory account appears to have been borne out in the results, as exhibited by the divergent shifts in identification responses between participants who opposed vs followed the direction of the altered auditory feedback.

Furthermore, shifts in perception were found to correlate with motoric adaptation, not auditory exposure, as found by Nasir and Ostry (2009) and Mattar *et al.* (2011), and no consistent changes in perceptual function were found in two groups of passive listeners (exp. 2). However, it should be noted that the correlation between articulatory and perceptual responses was only found when taking into account differences in behavior between unaltered and altered sessions. Similar studies have failed to find correlations between adaptation and perceptual change (Lametti *et al.*, 2014b), leading some to posit that motoric and sensory adaptation proceed somewhat independently (Nourouzpour *et al.*, 2015).

As has been suggested in previous research on adaptation to altered feedback in production (Jones and Munhall, 2005), we propose that adaptation results in a remapping between an articulatory/somatosensory representation of a phonetic target and its acoustic consequences. The results of the experiment suggest that during the identification task, the listener's remapping between the articulation or phonetic category onto the acoustics of the context vowel results in a different phonetic categorization of the same fricative.

Prior to exposure to altered auditory feedback, participants have a relatively stable mapping between a certain vocal tract state and a certain sensory target, such as a vowel (Mitsuya *et al.,* 2011; Niziolek and Guenther, 2013; Reilly and Dougherty, 2013). When attempting to produce the specific vowel, the participant initiates a motor sequence and compares their expectations of the intended sound to auditory feedback (Houde and Nagarajan, 2011), which may lead to swift articulatory adjustments if the vowel is off-target (Niziolek *et al.,* 2013). During the identification task, the participant is exposed to repeated ambiguous and unambiguous stimuli produced by the same voice and must map the incoming speech sounds onto abstract phonetic categories in order to respond. This may involve comparing the exposure voice to representations stored in memory, and adjusting to the idiosyncrasies of the exposure voice (Liu and Holt, 2015).

Acoustic-to-phonetic category mapping proceeds normally during the unaltered feedback session, producing a certain proportion of "she" responses in response to changes in the frequency of the fricative while the context remains relatively constant (though repeated presentation or production of a sound may also alter perception; Eimas and Corbit, 1973; Shiller *et al.,* 2009). Introducing the altered feedback during the production task alters the relationship between a given articulatory trajectory and its acoustic outcome, and adaptation reflects a stabilization of this shifted mapping (Purcell and Munhall, 2006). The change in the proportion of "she" responses indicates that shifting this mapping alters perception as well, suggesting that the listeners actively use knowledge of acoustic-to-motor mappings to classify contextually dependent speech sounds.

This experiment contributes to the growing body of research demonstrating that sensory adaptation in speech can influence motoric learning (Bradlow *et al.,* 1997; Lametti *et al.,* 2014a), and that motoric adaptation can affect perception of other speakers (Shiller *et al.,* 2009; Lametti *et al.,* 2014b). It is interesting that the remapping during production generalized to the CFC processes during perception, given that perceptual retuning is notoriously specific (Kraljic and Samuel, 2005; Reinisch *et al.,* 2014). This is somewhat true as well for adaptation to altered feedback. Thus far, generalization of articulatory adaptations has only been examined in with regard to generalization in production changes; such studies have found that changes in articulation are not limited to the adapted segment, but can also generalize to other tone categories (Jones and Munhall, 2005), or other vowels (Cai *et al.,* 2010). Our results demonstrate that perceptual changes can also occur even if the adapted sound is only contextually related to the target sound, as in the case of *CFC*. Testing the effects of adaptation on perceptual compensation for coarticulation eliminated the possibility that the adaptation-induced reported here and in previous experiments (Shiller *et al.,* 2009; Lametti *et al.,* 2014b) can be attributed to response biases. Finally, the experiment capitalized on the fact that when participants oppose a shift in auditory feedback, the opposing response may not be complete (Katseff *et al.,* 2012; MacDonald *et al.,* 2011). Thus, participants with differing articulatory responses heard very similar auditory feedback, enabling adjudication between articulatory and acoustic accounts for the observed effects.

While perception may not depend on production, it is clear from this study and others that sensorimotor processes can and do affect perception. Such a view accords with models of speech perception that suggest that the motor system, rather than playing a crucial role in the online decoding of speech sounds, plays a more modulatory but still important role in sensorimotor integration during speech perception (Hickok *et al.,* 2011). Accurately identifying rapidly coarticulated segments is a common problem listeners must face in decoding a continuous speech signal. Listeners have a wealth of production experience available to them, and our results, as well as others, suggest that this experience assists in decoding speech (Poeppel *et al.,* 2008). Sensorimotor integration processes may serve to increase the intelligibility of accented speakers (Adank *et al.,* 2010; Adank *et al.,* 2013) or support phonetic alignment between interlocutors (Pardo, 2006). Exploring modulatory relationships such as these may help to reconcile disparate bodies of research that focus on production and perception in isolation.

Adank, P., Hagoort, P., and Bekkering, H. (**2010**). "Imitation improves language comprehension," Psychol. Sci. **21**(12), 1903–1909.

Adank, P., Rueschemeyer, S.-A., and Bekkering, H. (**2013**). "The role of accent imitation in sensorimotor integration during processing of intelligible speech," Front. Hum. Neurosci. **7**, 634.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Mem. Lang. **68**(3), 255–278.

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. (**2015**). "lme4: Linear mixed-effects models using Eigen and S4," J. Stat. Softw. **67**(1), 1–48.

Bell-Berti, F., and Krakow, R. A. (**1991**). "Anticipatory velar lowering: A coproduction account," J. Acoust. Soc. Am. **90**(1), 112–123.

Blumstein, S. E., and Stevens, K. N. (**1981**). "Phonetic features and acoustic invariance in speech," Cognition **10**(1-3), 25–32.

Boersma, P., and Weenink, D. (**2013**). "Praat: Doing phonetics by computer" [computer program], version 5.0.1, http://www.praat.org/ (Last viewed 9/12/2016).

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.

Breslow, N. E., and Clayton, D. G. (**1993**). "Approximate inference in generalized linear mixed models," J. Am. Stat. Assoc. **88**(421), 9–25.

Cai, S., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (**2010**). "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization," J. Acoust. Soc. Am. **128**(4), 2033–2048.

Casserly, E. D., and Pisoni, D. B. (**2010**). "Speech perception and production," WIREs Cogn. Sci. **1**(5), 629–647.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and van der Vrecken, O. (**1996**). "The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes," in *Proceedings of the Fourth International Conference on Spoken Language Processing. ICSLP'96*, IEEE, Philadelphia, PA, Vol. 3, pp. 1393–1396.

Eimas, P. D., and Corbit, J. D. (**1973**). "Selective adaptation of linguistic feature detectors," Cogn. Psychol. **4**(1), 99–109.

Eisner, F., and McQueen, J. M. (**2006**). "Perceptual learning in speech: Stability over time," J. Acoust. Soc. Am. **119**(4), 1950–1953.
Elman, J. L., and McClelland, J. L. (**1988**). "Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes," J. Mem. Lang. **27**(2), 143–165.
Flagg, E. J., Oram Cardy, J. E., and Roberts, T. P. (**2006**). "MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs," Neurosci. Lett. **397**(3), 263–268.
Fowler, C. A. (**1986**). "An event approach to the study of speech perception from a direct-realist perspective," J. Phon. **14**, 3–28.
Fowler, C. A. (**2006**). "Compensation for coarticulation reflects gesture perception, not spectral contrast," Percept. Psychophys. **68**(2), 161–177.
Fowler, C. A., and Brown, J. M. (**2000**). "Perceptual parsing of acoustic consequences of velum lowering from information for vowels," Percept. Psychophys. **62**(1), 21–32.
Hickok, G., Houde, J., and Rong, F. (**2011**). "Sensorimotor integration in speech processing: Computational basis and neural organization," Neuron **69**(3), 407–422.
Holt, L. L. (**2005**). "Temporally nonadjacent nonlinguistic sounds affect speech categorization," Psychol. Sci. **16**(4), 305–312.
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.
Houde, J. F., and Nagarajan, S. S. (**2011**). "Speech production as state feedback control," Front. Hum. Neurosci. **5**, 82.
Jaeger, T. F. (**2008**). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," J. Mem. Lang. **59**(4), 434–446.
Jones, J. A., and Munhall, K. G. (**2005**). "Remapping auditory-motor representations in voice production," Curr. Biol. **15**(19), 1768–1772.
Katseff, S., Houde, J., and Johnson, K. (**2012**). "Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback?," Lang. Speech **55**(2), 295–308.
Kraljic, T., and Samuel, A. G. (**2005**). "Perceptual learning for speech: Is there a return to normal?," Cogn. Psychol. **51**(2), 141–178.
Kunisaki, O., and Fujisaki, H. (**1977**). "On the influence of context upon perception of voiceless fricative consonants," Ann. Bull. Res. Inst. Logoped. Phoniatr., Univ. Tokyo **11**, 85–91.
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**, 1325–1336.
Lametti, D. R., Nasir, S. M., and Ostry, D. J. (**2012**). "Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback," J. Neurosci. **32**(27), 9351–9358.
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**, 10339–10346.
Liberman, A., Delattre, P., and Cooper, F. (**1952**). "The role of selected stimulus-variables in the perception of the unvoiced stop consonants," Am. J. Psychol. **65**(4), 497–516.
Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**(1), 1–36.
Liberman, M., and Mattingly, I. G. (**1989**). "A specialization for speech perception," Science **243**(4890), 489–494.
Liberman, M., and Whalen, D. H. (**2000**). "On the relation of speech to language," Trends Cogn. Sci. **4**, 187–196.
Liu, R., and Holt, L. L. (**2015**). "Dimension-based statistical learning of vowels," J. Exp. Psychol.: Human Percept. Perform. **41**(6), 1783–1798.
Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**(4), 602–619.
MacDonald, E. N., Purcell, D. W., and Munhall, K. G. (**2011**). "Probing the independence of formant control using altered auditory feedback," J. Acoust. Soc. Am. **129**(2), 955–965.
Mann, V., and Soli, S. D. (**1991**). "Perceptual order and the effect of vocalic context of fricative perception," Percept. Psychophys. **49**(5), 399–411.
Mann, V. A., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [ʃ]-[s] distinction," Percept. Psychophys. **28**(3), 213–228.
Mathworks (**2012**). "MATLAB and statistics toolbox" [computer program].
Mattar, A. A. G., Nasir, S. M., Darainy, M., and Ostry, D. J. (**2011**). *Enhancing Performance for Action and Perception—Multisensory Integration, Neuroplasticity and Neuroprosthetics, Part I*, Vol. 191 of *Progress in Brain Research* (Elsevier, Amsterdam).
McQueen, J. M. (**1991**). "The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity," J. Exp. Psychol. Hum. Percept. Perform. **17**(2), 433–443.
McQueen, J. M., Cutler, A., and Norris, D. (**2006**). "Phonological abstraction in the mental lexicon," Cogn. Sci. **30**(6), 1113–1126.
Mitsuya, T., MacDonald, E. N., Munhall, K. G., and Purcell, D. W. (**2015**). "Formant compensation for auditory feedback with English vowels," J. Acoust. Soc. Am. **138**(1), 413–424.
Mitsuya, T., MacDonald, E. N., Purcell, D. W., and Munhall, K. G. (**2011**). "A cross-language study of compensation in response to real-time formant perturbation," J. Acoust. Soc. Am. **130**, 2978–2986.
Mitterer, H., and Blomert, L. (**2003**). "Coping with phonological assimilation in speech perception: Evidence for early compensation," Percept. Psychophys. **65**(6), 956–969.
Nasir, S. M., and Ostry, D. J. (**2009**). "Auditory plasticity and speech motor learning," Proc. Natl. Acad. Sci. U.S.A. **106**(48), 20470–20475.
Nittrouer, S., and Studdert-Kennedy, M. (**1987**). "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Lang. Hear. Res. **30**(3), 319–329.
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.
Niziolek, C. A., Nagarajan, S. S., and Houde, J. F. (**2013**). "What does motor efference copy represent? Evidence from speech production," J. Neurosci. **33**(41), 16110–16116.
Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cogn. Psychol. **47**(2), 204–238.
Nourouzpour, N., Salomonczyk, D., Cressman, E. K., and Henriques, D. Y. P. (**2015**). "Retention of proprioceptive recalibration following visuomotor adaptation," Exp. Brain Res. **233**(3), 1019–1029.
Ostry, D. J., Darainy, M., Mattar, A. A. G., Wong, J., and Gribble, P. L. (**2010**). "Somatosensory plasticity and motor learning," J. Neurosci. **30**(15), 5384–5393.
Pardo, J. S. (**2006**). "On phonetic convergence during conversational interaction," J. Acoust. Soc. Am. **119**(4), 2382–2393.
Poeppel, D., Idsardi, W. J., and Van Wassenhove, V. (**2008**). "Speech perception at the interface of neurobiology and linguistics," Philos. Trans. R. Soc. London B: Biol. Sci. **363**(1493), 1071–1086.
Poeppel, D., and Monahan, P. J. (**2011**). "Feedforward and feedback in speech perception: Revisiting analysis by synthesis," Lang. Cogn. Process. **26**(7), 935–951.
Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**(2), 966–977.
Quatieri, T., and McAulay, R. J. (**1986**). "Speech transformations based on a sinusoidal representation," IEEE Trans. (ASSP) **34**(6), 1449–1464.
R Development Core Team (**2013**). R [computer program].
Reilly, K. J., and Dougherty, K. E. (**2013**). "The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback," J. Acoust. Soc. Am. **134**, 1314–1323.
Reinisch, E., Wozny, D. R., Mitterer, H., and Holt, L. L. (**2014**). "Phonetic category recalibration: What are the categories?," J. Phon. **45**, 91–105.
Repp, B. H., and Mann, V. A. (**1981**). "Perceptual assessment of fricative–stop coarticulation," J. Acoust. Soc. Am. **69**(4), 1154–1163.
Samuel, A. G., and Kraljic, T. (**2009**). "Perceptual learning for speech," Atten. Percept. Psychophys. **71**(6), 1207–1218.
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**, 1103–1113.
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.
Viswanathan, N., Magnuson, J. S., and Fowler, C. A. (**2010**). "Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech," J. Exp. Psychol.: Human Percept. Perform. **36**(4), 1005–1015.
Whalen, D. H. (**1981**). "Effects of vocalic formant transitions and vowel quality on the English [s]–[ʃs] boundary," J. Acoust. Soc. Am. **69**(1), 275–282.