

Psychological Science

<http://pss.sagepub.com/>

Brief Periods of Auditory Perceptual Training Can Determine the Sensory Targets of Speech Motor Learning

Daniel R. Lametti, Sonia A. Krol, Douglas M. Shiller and David J. Ostry
Psychological Science 2014 25: 1325 originally published online 8 May 2014
DOI: 10.1177/0956797614529978

The online version of this article can be found at:
<http://pss.sagepub.com/content/25/7/1325>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Association for Psychological Science](http://www.sagepublications.com)

Additional services and information for *Psychological Science* can be found at:

Email Alerts: <http://pss.sagepub.com/cgi/alerts>

Subscriptions: <http://pss.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Jul 14, 2014

[OnlineFirst Version of Record](#) - May 22, 2014

[OnlineFirst Version of Record](#) - May 8, 2014

[What is This?](#)

Brief Periods of Auditory Perceptual Training Can Determine the Sensory Targets of Speech Motor Learning

Daniel R. Lametti^{1,2}, Sonia A. Krol¹, Douglas M. Shiller^{3,4,5}, and David J. Ostry^{1,6}

¹Department of Psychology, McGill University; ²Institute of Neurology, University College London;

³Research Center, Sainte-Justine Hospital, Université de Montréal; ⁴School of Speech Pathology and Audiology, Université de Montréal; ⁵Centre for Research on Brain, Language and Music, McGill University; and

⁶Haskins Laboratories, New Haven, Connecticut

Psychological Science
2014, Vol. 25(7) 1325–1336
© The Author(s) 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797614529978
pss.sagepub.com


Abstract

The perception of speech is notably malleable in adults, yet alterations in perception seem to have little impact on speech production. However, we hypothesized that speech perceptual training might immediately influence speech motor learning. To test this, we paired a speech perceptual-training task with a speech motor-learning task. Subjects performed a series of perceptual tests designed to measure and then manipulate the perceptual distinction between the words *head* and *had*. Subjects then produced *head* with the sound of the vowel altered in real time so that they heard themselves through headphones producing a word that sounded more like *had*. In support of our hypothesis, the amount of motor learning in response to the voice alterations depended on the perceptual boundary acquired through perceptual training. The studies show that plasticity in adults' speech perception can have immediate consequences for speech production in the context of speech learning.

Keywords

cognitive neuroscience, speech perception, motor processes, learning

Received 10/18/13; Revision accepted 3/7/14

The perception of speech is remarkably plastic, yet alterations in speech perception seem to have little immediate impact on speech production. People quickly come to understand English spoken with a foreign accent, for instance, but this perceptual change does not cause them to suddenly adopt that foreign accent. This phenomenon contrasts with other behaviors like reaching, where increased visual acuity from, say, a new pair of glasses would immediately be used by the brain to make more accurate movements. Here, we provide initial evidence that alterations in speech perception do, in fact, have immediate consequences for speech production in the context of speech motor learning.

The perceptual goals of speech movements are typically identified by their acoustic properties. Different vowels, for instance, are contrasted mainly on the basis of peaks in the acoustic spectrum, or *formants* (Ladefoged, 1975). These frequency peaks are a major

perceptual target in speech motor control, just as visual or somatosensory targets guide limb movement. The perception of speech sounds has been shown to be highly flexible. It is apparent both anecdotally and experimentally that people adapt their speech perception to the unfamiliar acoustic properties of foreign accents (Clarke & Garrett, 2004; Maye, Aslin, & Tanenhaus, 2008) and computer-altered speech (Dupoux & Green, 1997). However, within a speaker's first language, changes in

Corresponding Authors:

Daniel R. Lametti, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, United Kingdom
E-mail: danielrlametti@gmail.com

David J. Ostry, Department of Psychology, McGill University, 1205 Dr. Penfield Ave., Montreal, Quebec H3A1B1, Canada
E-mail: david.ostry@mcgill.ca

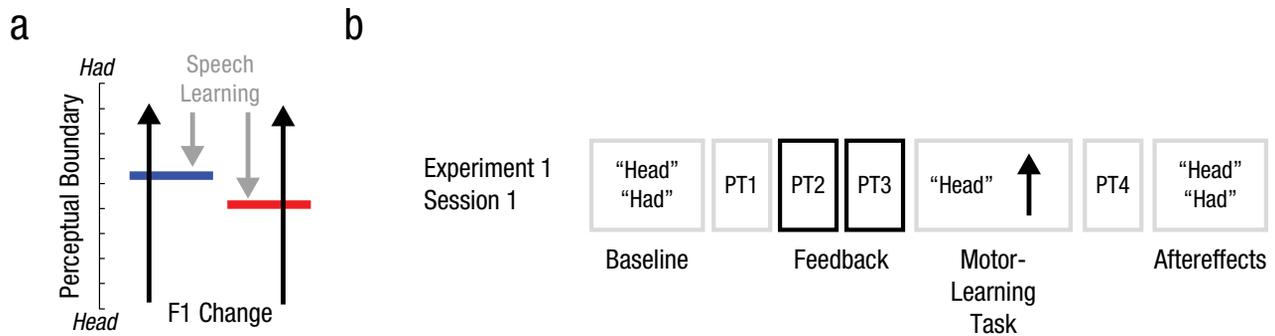


Fig. 1. The hypothesis and design of Experiment 1. When the first formant frequency (F1) of the vowel sound in *head* is increased in real time so that it sounds more like *bad* than like *head* (a; black arrows), subjects should compensate by decreasing the frequency of their produced F1 until their productions fall within the perceptual range of *head* (gray arrows). The amount of compensation should depend on the point of perceptual distinction between *head* and *bad* (horizontal bars). Thus, prior training that alters the perceptual boundary should affect the amount of compensation. The first session of Experiment 1 consisted of five phases (b). First, subjects spoke the words *head* and *bad* (45 times each) without any manipulation (baseline). They then performed three perceptual tests designed to measure (PT1) and then alter (PT2 and PT3) the perceptual boundary between *head* and *bad*. Next, subjects spoke the word *head* 135 times, with the sound of their voices altered in real time (i.e., an increase in F1) so that the word sounded more like *bad*. In a fourth perceptual test (PT4), the perceptual boundary between the two words was measured again. Finally, subjects spoke the two words 45 times each without any manipulation so that aftereffects of the motor learning could be assessed.

speech perception seem to have only a small impact on speech production (Kraljic, Brennan, & Samuel, 2008; Samuel & Kraljic, 2009) even after a considerable amount of speech perceptual training (Rvachew, 1994). Perceptual training can affect speech production in the case of second-language learning but, again, only after days of training (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Callan et al., 2003; Wang, Jongman, & Sereno, 2003).

Recent research on the motor control of speech and limb movements has shown that perceptual change is coupled to motor learning (Cressman & Henriques, 2009; Haith, Jackson, Miall, & Vijayakumar, 2008; Ostry, Darainy, Mattar, Wong, & Gribble, 2010; Shiller, Sato, Gracco, & Baum, 2009; Vahdat, Darainy, Milner, & Ostry, 2010). Studies of speech development show that changes in speech perception precede speech learning (Kuhl, 2004; Tsao, Liu, & Kuhl, 2004). In the study reported here, we examined the impact of perceptual change on adults' capacity for speech learning in their first language. We paired a perceptual-training task with a motor-learning task to test whether altering the perceptual distinction, or boundary, between two vowel sounds significantly influenced the degree to which participants learned to adapt their speech movements in response to perceived errors in producing these sounds. In support of prior research, our results showed that previously learned speech movements were left unchanged by perceptual training. Alterations in the perception of speech did, however, have immediate consequences for adults' speech motor learning.

Figure 1a lays out the experimental hypothesis, and Figure 1b shows the design of Experiment 1 used to test this hypothesis. When the first formant frequency (F1) of

the vowel sound in *head* is increased in real time so that subjects hear themselves saying something closer to the vowel sound in *bad*, they should compensate by decreasing the frequency of their produced F1 until their productions fall within the perceptual range of *head*. If speech perceptual training manipulates their perception of the boundary between *bad* and *head*, the alteration should affect the amount of compensation in a subsequent test of speech motor learning.

Method

Subjects and apparatus

Sixty-four women (age range = 18–30 years) who were native English speakers with normal hearing and speech participated in the study. Forty-four participated in Experiment 1, and 20 participated in Experiment 2. Only women were tested because of the large difference between men's and women's F1 frequencies. The sample sizes were chosen on the basis of our previous speech-motor-learning experiments, which demonstrated significant group differences with 10 to 20 participants in each condition (Lametti, Nasir, & Ostry, 2012; Rochet-Capellan & Ostry, 2011; Rochet-Capellan, Richer, & Ostry, 2012). Results from 2 subjects in Experiment 1 were excluded from the final analysis. In the first case, the subject's perceptual responses differed by more than 2 standard deviations from the group mean; in the second case, the subject's baseline F1 differed by more than 2 standard deviations from the group mean. Testing was performed in a sound-attenuating chamber. Subjects wore headphones (STAX, Miyoshi, Japan), and a directional microphone

(Sennheiser, Wedemark, Germany) was used to record speech. Speech was altered using an acoustical effects processor (VoiceOne; TC Helicon, Victoria, BC, Canada) and a dual-channel analog audio filter. Data analysis was performed in MATLAB (The MathWorks, Natick, MA). The McGill University Faculty of Medicine Institutional Review Board approved the experiments.

Experimental procedure

Experiment 1 began with a measurement of baseline speech production. The words *head* and *bad* appeared on a computer screen 45 times each in random order. Subjects were instructed to say each word in a clear voice. After the subjects produced the displayed word, it was removed from the screen, and the next word was displayed. The first perceptual test was then performed to measure the perceptual boundary between the words *head* and *bad* (see Measuring Speech Perception). Subjects then performed perceptual tests with feedback designed to systematically shift their perceptual boundaries (see Perceptual Training).

One group of 21 subjects received feedback that moved their perceptual boundaries toward *head* (*head*-shift group), and a second group of 21 subjects received training that moved their perceptual boundaries toward *bad* (*bad*-shift group). After this training, subjects performed the motor-learning task: They spoke the word *head* 135 times with the sound of their voices altered in real time. To do this, we used acoustical signal processors and filters to shift F1 of the vowel sound in *head* up in frequency; the remaining formants were left unchanged (see Real-Time Alterations of Speech). A fourth perceptual test without feedback was then performed. Finally, subjects spoke *head* and *bad* 45 times each with unaltered speech so that we could examine aftereffects associated with speech motor learning. In a post hoc addition to Experiment 1, participants were invited back to the lab for a second session of testing; 28 returned (13 in the *head*-shift group and 15 in the *bad*-shift group). The time between the first and second testing sessions averaged 8.85 days ($SD = 2.6$; range = 7–14 days). The second session was the same as the first except that the perceptual tests with feedback (i.e., perceptual training) were omitted.

In Experiment 2, 20 new subjects were divided into two groups (10 in the *head*-shift group and 10 in the *bad*-shift group). During an initial testing session, subjects performed the baseline production task and three perceptual tests, the last two with feedback. Subjects returned to the lab 2 days later for a second session. After a second baseline production task, they performed a fourth perceptual test (without feedback), the motor-learning task, and a fifth perceptual test (also without feedback). Finally, they spoke *head* and *bad* 45 times each with unaltered speech.

Measuring speech perception. Perception was measured using 10 words that spanned the perceptual continuum from *head* (Stimulus 1) to *bad* (Stimulus 10). The words were based on utterances of a Canadian man. To create the stimuli on this 10-step continuum, we took the first two formants (F1 and F2) from the word *head* and shifted them in equal steps toward the formant values in *bad*. F1 and F2 for *head* were 560 and 1745 Hz, respectively. F1 and F2 for *bad* were 768 and 1648 Hz, respectively. During perceptual testing, the entire stimulus-set was played in a random order, one word at a time. This process was repeated 21 times. After subjects heard a stimulus, they were prompted by text on a computer screen to indicate, by pressing a key on a keyboard, whether the stimulus sounded more like *head* or more like *bad*. Pressing the space bar triggered the next stimulus. The proportion of “had” responses for each stimulus was computed on a per-subject basis for each perceptual test. Psychometric functions were fit to these proportions using the binomial-distribution-fitting method (glmfit in MATLAB). The perceptual boundary—the point on the continuum at which *head* was perceived 50% of the time—was calculated from the psychometric function for each subject.

Perceptual training. For the perceptual tests with feedback (i.e., perceptual training), a new perceptual boundary was set for each subject. This boundary was either one stimulus lower (half the subjects) or one stimulus higher (other half of subjects) on the continuum than the subject’s original, rounded-to-the-nearest-integer perceptual boundary. Feedback was then given based on this new boundary. If the new perceptual boundary was Stimulus 6, for instance, “CORRECT” was displayed on the screen if the subject indicated that she had heard *head* for any stimulus from 1 through 5 and *bad* for any stimulus from 6 through 10; otherwise, “INCORRECT” was displayed on the screen.

Incorrect responses added a point to an error counter at the bottom right of the screen. Subjects were instructed to minimize errors. After completion of the first perceptual test with feedback, the number of errors made was displayed on the screen along with an instruction to reduce this number. The error counter was then reset to zero, and subjects made another 210 perceptual choices with feedback, for a total of 420 choices with feedback. Pre-motor learning perceptual testing (PT1) and training (PT2 and PT3) took approximately 18 min.

Analysis of perceptual data. To compute the perceptual boundary on a unitless scale (see Auditory Analysis) used to relate speech motor learning to baseline production, the perceptual stimuli were represented as a ratio of the F1 frequency for each stimulus relative to the F1 frequency of Stimulus 1 (*head*). Thus, the value of

Stimulus 1 was 1.0 (560/560 Hz), the value of Stimulus 2 was 1.04 (582/560 Hz), and so on toward Stimulus 10, which had a value of 1.37 (768/560 Hz). The psychometric function was fit to the proportion of “had” responses at each of these values, and the perceptual boundary for each perceptual test was found from this function. The distance to the perceptual boundary was computed as the difference between “had” on this unitless scale and the value of the perceptual boundary computed as described earlier. Changes in perceptual boundaries were assessed using split-plot analyses of variance (ANOVAs) with Bonferroni-corrected post hoc tests. To examine changes in perception over time, we calculated the proportion of “had” responses for each block of 10 perceptual decisions. Exponential functions of the form $y = a + b(1 - c)^x$ were fit to the mean proportion of “had” responses from the last block of 10 perceptual choices in the baseline test.

Real-time alterations of speech. During the motor-learning task, acoustical signal processors and filters were used to shift F1 of the vowel sound in *head* up in frequency; the remaining formants were unchanged (Rochet-Capellan & Ostry, 2011). The altered signal was mixed with 70-dB, speech-shaped masking noise and played back to subjects through the headphones with a delay of 11 ms. Subjects thus spoke the word *head* but heard a word with an F1 closer to that in *bad*. In Experiment 1, the baseline F1 frequency of subjects averaged 739 Hz, and there was no difference between the two groups in baseline F1 frequency, $t(40) = 1.35, p > .15$. The signal processor increased F1 frequency by approximately 24%, for a total F1 increase of 174 Hz ($SD = 22$ Hz). The F1 shift was calculated separately for each subject, and then an average across the group was calculated. There was no difference in the amount of F1 shift for subjects in the two perceptual-training groups, $t(40) = 0.77, p > .45$. In Experiment 2, the baseline F1 frequency of subjects averaged 729 Hz, and there was again no difference in baseline F1 frequency between the two groups, $t(18) = 0.07, p > .9$. The signal processor increased baseline F1 frequency by approximately 26%, for a total F1 increase of 186 Hz ($SD = 21$ Hz). As in Experiment 1, there was no difference in the amount of shift between the two perceptual-training groups, $t(18) = 0.24, p > .80$.

Auditory analysis

Speech was recorded at a sample rate of 44.1 kHz and a bit depth of 16 bits per sample. The software package Praat was used to detect (Boersma & Weenink, 2014) vowel boundaries and calculated F1 frequencies from a 30-ms window at the center of the vowel (Rochet-Capellan & Ostry, 2011; Shum, Shiller, Baum, & Gracco, 2011). In both experiments, to examine changes in F1 related to

altered auditory feedback, we divided the F1 frequency of each utterance by the mean F1 of the last 30 *head* utterances of baseline production from the first session of testing (pretraining production). We calculated the mean of this normalized measure of F1 frequency for the last 45 utterances of altered auditory feedback and the first 15 utterances of aftereffect trials. For the subjects who returned to the lab after initial testing, we calculated mean normalized F1 frequency for the last 30 utterances of the second session of baseline production, the last 45 utterances of the second session of altered auditory feedback, and the first 15 utterances of the second session of aftereffect trials. These means were compared using split-plot ANOVAs with Bonferroni-corrected post hoc tests. Exponential functions of the form $y = a + b(1 - c)^x$ were fit to the mean normalized F1 values calculated from blocks of five utterances taken from the altered feedback (i.e., motor learning) phase of the experiment.

Results

Figure 2a shows the average of the psychometric functions fit to perceptual responses before and during perceptual training. Perceptual training caused a shift in the psychometric curves either toward *head* or toward *bad* on the continuum. The mean R^2 for the psychometric fits was .98 (range = .88–.99). Figure 2b shows the proportion of “had” responses averaged across subjects computed from blocks of 10 perceptual judgments made with and without feedback. To help visualize the speed of perceptual change, we fit exponential functions to the data (Fig. 2b). The coefficient of determination, R^2 , was .49 for the *head*-shift group and .32 for the *bad*-shift group. As computed from the fit functions, perceptual change reached 90% of asymptote by the 88th trial for the *head*-shift group and by the 44th trial for the *bad*-shift group.

Figure 3a shows the perceptual boundary in units of F1 frequency relative to baseline for each perceptual test in the first session of Experiment 1. Perceptual training moved the boundary of the *head*-shift group toward *head* and the boundary of the *bad*-shift group toward *bad* ($p < .001$ in each case). This change in the perceptual boundary was also observed in the perceptual test that followed speech motor learning (i.e., the fourth perceptual test; $p < .001$). In Figure 3b, F1 frequency of the vowel sound in *head* relative to baseline production of *head* is plotted over the course of the experiment. After perceptual training, subjects produced the word *head* with the signal processor turned on such that the F1 frequency for the vowel was increased to a value closer to that in *bad* (i.e., the motor-learning task). Subjects compensated for this alteration by learning to produce F1 at a lower frequency. Figure 3b shows that the *head*-shift group learned to compensate more for the speech

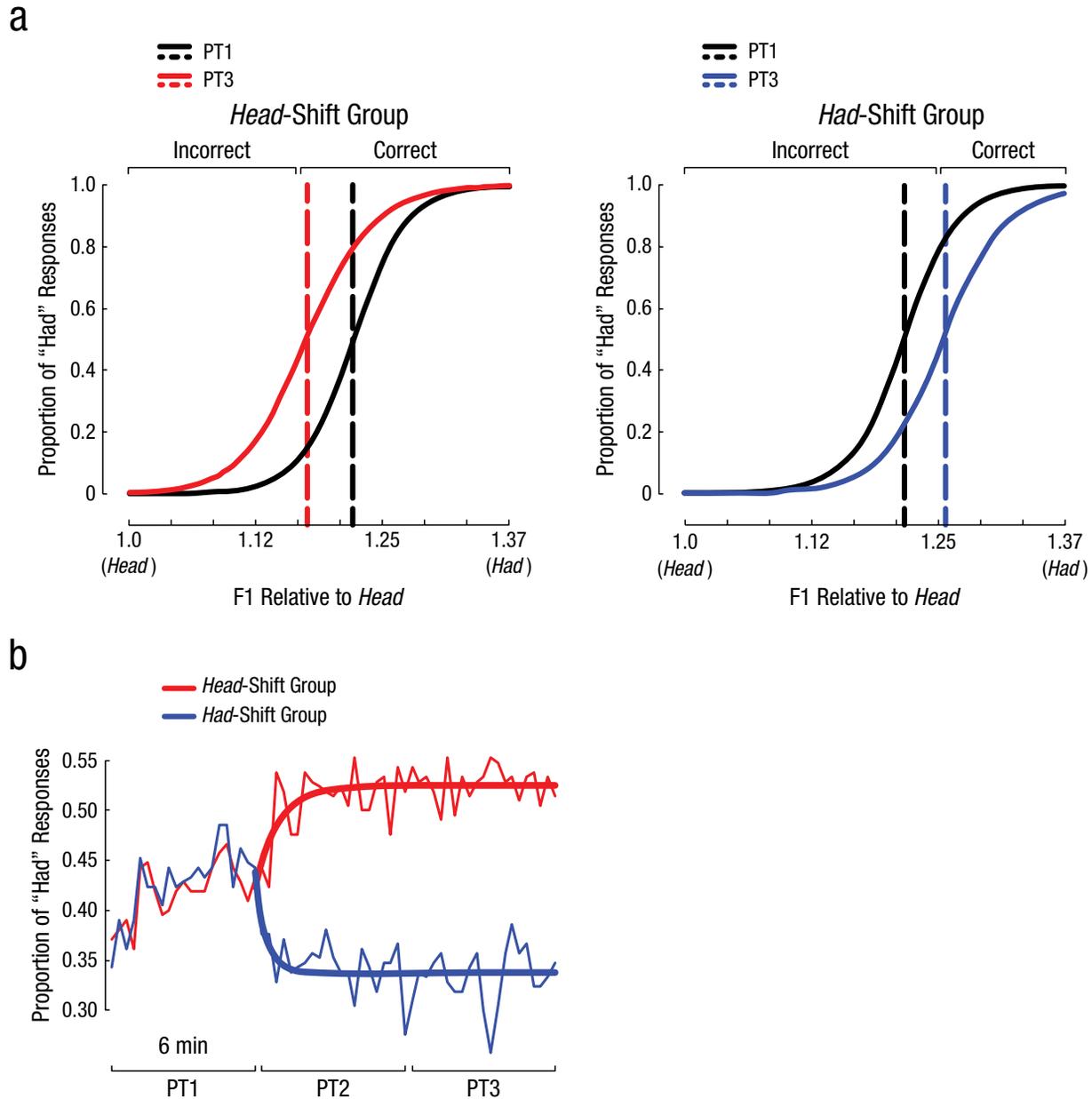


Fig. 2. The impact of perceptual training in Experiment 1. In (a), the proportion of “had” responses is plotted as a function of the frequency of the vowel’s first formant (F1 frequency) relative to Stimulus 1 (“head”). The dashed vertical lines show the perceptual boundaries between *head* and *bad* (calculated in F1 units relative to Stimulus 1) before training (as measured from the first perceptual test, PT1) and after training (as measured from the third perceptual test, PT3). The black curves are based on responses from PT1. The colored curves are based on responses from PT3, which included feedback. “Correct” and “Incorrect” refer to feedback given during PT2 and PT3 for a “had” response. In (b), exponential functions (heavy curves) were fit to the data (thin curves). The proportion of “had” responses is plotted as a function of perceptual test (PT1, PT2, and PT3). The thin colored lines join data points that represent the average of 10 perceptual responses. Exponential functions (heavy colored lines) were fit to the data. Each perceptual test took about 6 min.

alteration than the *bad*-shift group ($p < .04$). The *head*-shift group also showed greater learning-related aftereffects when the voice alteration was removed ($p < .02$).

The amount of speech motor learning in response to the voice alteration depended on the distance from *bad* to the acquired perceptual boundary measured

during the third perceptual test ($r = .52, p < .0005$; see Fig. 3c). Significant correlations between these measures were also found within each group (*head*-shift group: $r = .49, p < .03$; *bad*-shift group: $r = .51, p < .02$). Furthermore, a negative correlation was observed between training-related changes in perception and the amount

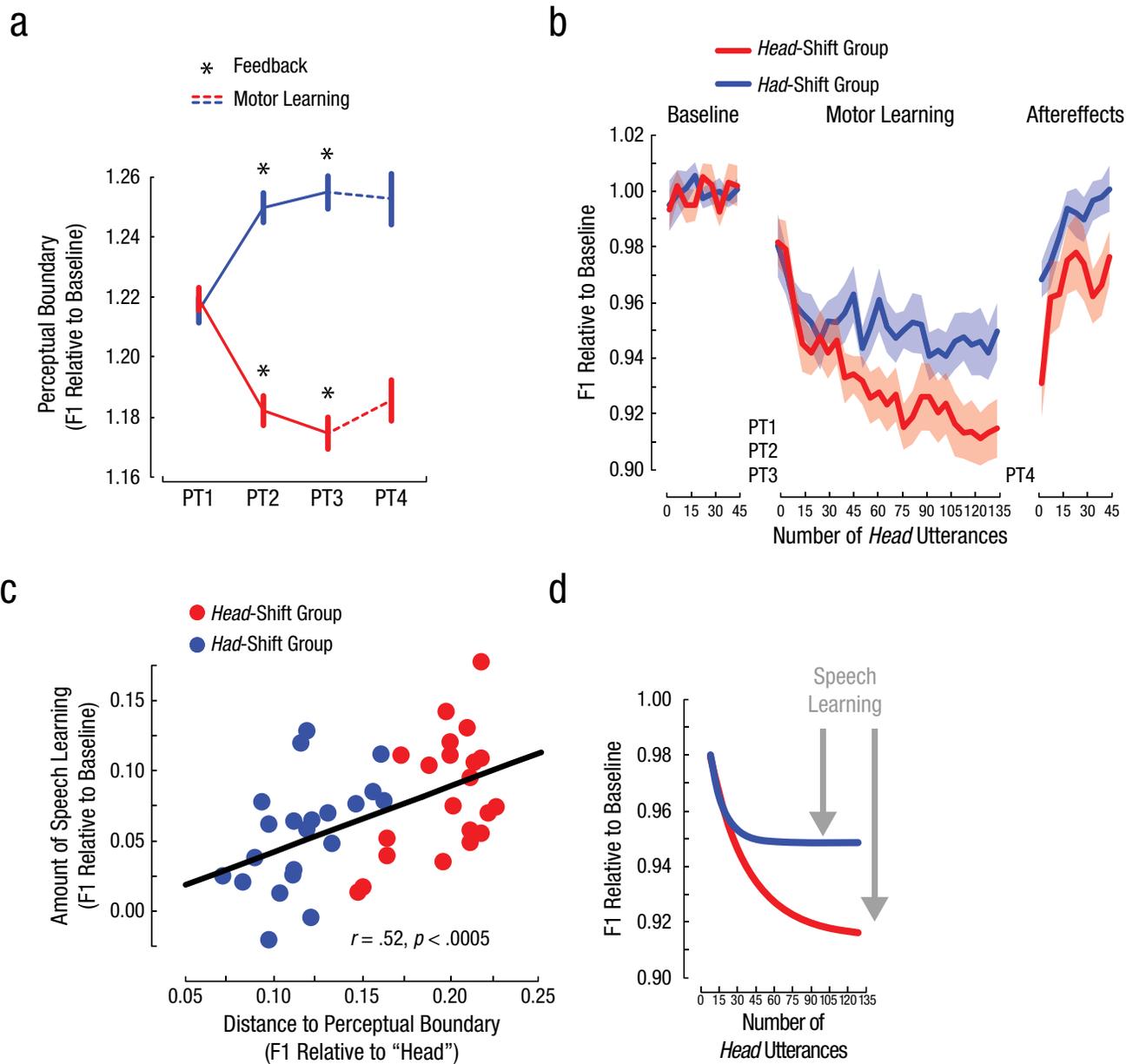


Fig. 3. The impact of perceptual training on motor learning in the first session of Experiment 1. In (a), the perceptual boundary between *head* and *bad* (calculated as the frequency of formant 1, or F1, relative to the frequency of F1 in *head*) is plotted as a function of perceptual test (PT). Error bars represent ± 1 SE. In (b), the produced frequency of F1 relative to the baseline frequency of F1 in *head* is plotted against number of utterances of *head*. In the baseline phase, subjects uttered *head* 45 times. After PT1, PT2, and PT3, subjects said *head* 135 times with altered feedback to induce motor learning; this was followed by PT4. In the final production phase, subjects said *head* 45 times so that aftereffects could be measured. The shaded areas around the curves represent ± 1 SE, and the curves join averages computed from blocks of five utterances. In (c), the amount of speech motor learning (i.e., the change in produced F1) is plotted against the distance between the F1 frequency in *bad* and the trained perceptual boundary. In (d), produced F1 frequency relative to baseline is plotted against number of utterances of *head*. Exponential functions were fit to the data to illustrate the effect of perceptual training on speech motor learning (arrows).

of speech motor learning ($r = -.37, p < .02$). Shifts in the perceptual boundary toward *head* were associated with greater speech motor learning whereas shifts toward *bad* were associated with less speech motor learning. The results suggest that perceptual training predictably altered speech motor learning.

Figure 3d shows exponential functions fit to the patterns of motor learning shown in Figure 3b for each of the two groups. The coefficient of determination, R^2 , was .93 for the *head*-shift group and .66 for the *bad*-shift group. As computed from the functions, the curve for the *head*-shift group reached asymptote at 0.91, 95%

confidence interval (CI) = [0.909, 0.918], in units of F1 frequency relative to baseline, and the curve for the *bad*-shift group reached asymptote at 0.95, 95% CI = [0.946, 0.951], in units of F1 frequency relative to baseline. It is thus unlikely that the two groups would have achieved the same amount of learning with more training. Furthermore, there was no difference in the starting points of the curves. The curve for the *head*-shift group started at 0.99, 95% CI = [0.974, 1.004], and the curve for the *bad*-shift group started at 1.0, 95% CI = [0.970, 1.022]. An empirical examination of the first utterance with altered auditory feedback revealed no difference between the two groups in F1 frequency relative to baseline ($p > .5$). This value was 0.98 ($SD = 0.06$) in the case of the *head*-shift group and 0.99 ($SD = 0.09$) in the case of the *bad*-shift group. This result suggests that perceptual training altered the amount of speech motor learning without significantly altering baseline production.

Twenty-eight of the subjects who participated in Experiment 1 returned to the lab approximately 9 days later. The subjects repeated Experiment 1 minus perceptual training (Fig. 4a). Nine days after perceptual training, there were still differences between the two groups' perceptual boundaries (Fig. 4b, $p < .01$ for both perceptual tests). But only the *head*-shift group maintained a boundary change that differed from baseline ($p < .05$). Even so, the *head*-shift group showed greater learning-related after effects (Fig. 4c) than the *bad*-shift group ($p < .02$). A brief period of perceptual training thus seemed to have a long-lasting impact on at least one measure of speech motor learning. However, the difference in aftereffects observed during the return session may have been driven by a perceptual-training-induced difference in baseline speech production or by a failure, in the case of the *head*-shift group, to completely eliminate motor learning. Indeed, when the patterns of speech motor learning were normalized to baseline production during the return session of testing, the between-group difference in aftereffects was reduced and no longer significant ($p = .076$).

In Experiment 2, the subjects were divided into two groups that underwent speech perceptual training as in Experiment 1 but did not perform the speech motor-learning task until 2 days later, after a period of baseline production (Fig. 4d). This new experiment was designed to examine the durability of the effect of perceptual training on motor learning in an experiment involving a single session of motor learning. It also allowed for the direct examination of the effect of perceptual training on subsequent baseline production.

As in Experiment 1, perceptual training altered subjects' perceptual boundaries (Fig. 4e). Two days later, both groups still showed a boundary that was different from the baseline boundary, as measured by a perceptual test without feedback ($p < .02$ in each group). This perceptual test was followed by speech-motor-learning trials involving

production of *head*. Figure 4f shows that the *head*-shift group (red data) learned to compensate more ($p < .05$) for the voice alteration than the *bad*-shift group (blue data). They also showed greater learning-related aftereffects when the voice alteration was removed ($p < .01$).

After perceptual training, there was no difference between the groups in baseline F1 frequency ($p > .3$), and we observed the same effect of perceptual training on motor learning even after normalizing the data to posttraining baseline production ($p < .05$). For the *head*-shift group, perceptual training caused a +0.2% change in baseline F1 frequency ($p > .5$); for the *bad*-shift group, perceptual training caused a +1.7% change in baseline F1 frequency ($p > .08$). The results of Experiment 2 show that perceptual training altered speech motor learning 2 days later without significantly altering unperturbed speech. A brief period of perceptual training can thus cause long-lasting changes in the perceptual targets that guide speech motor learning

Discussion

We tested the idea that perceptual training could be used to shape adults' speech motor learning. Speech perception is notably malleable in adults (Bertelson, Vroomen, & de Gelder, 2003; Clarke & Garrett, 2004; Dupoux & Green, 1997; Norris, McQueen, & Cutler, 2003); however, previous work suggests that experimentally induced changes in speech perception transfer quite slowly to production if they transfer at all (Bradlow et al., 1997; Kraljic et al., 2008; Rvachew, 1994; Wang et al., 2003). Our results largely support the prior findings, in that we saw little impact of perceptual training on subsequent baseline speech production. However, training-induced changes in the perceptual boundary immediately caused predictable and long-lasting changes in the amount of speech motor learning. Thus, manipulations of speech perception in adults can have an immediate impact on speech motor learning.

We hypothesized that the perceptual boundary between vowels acts as a guide that influences the amount of speech motor learning when perturbations drive production past this boundary point. A recent study supports this hypothesis: Niziolek and Guenther (2013) examined compensation for unpredictable perturbations of vowel sounds and found that compensation was substantially greater for perturbations that pushed productions into a new perceptual category (e.g., *bed* to *bad*) than for perturbations that did not. This finding suggests that alterations in the perceptual boundary between vowels will significantly affect the amount of learned compensation when vowel productions are predictably perturbed, which was exactly the result observed here.

Changes in the perceptual boundary in the current study were driven with only 42 repetitions of the 10-step perceptual continuum (12 min of perceptual training).

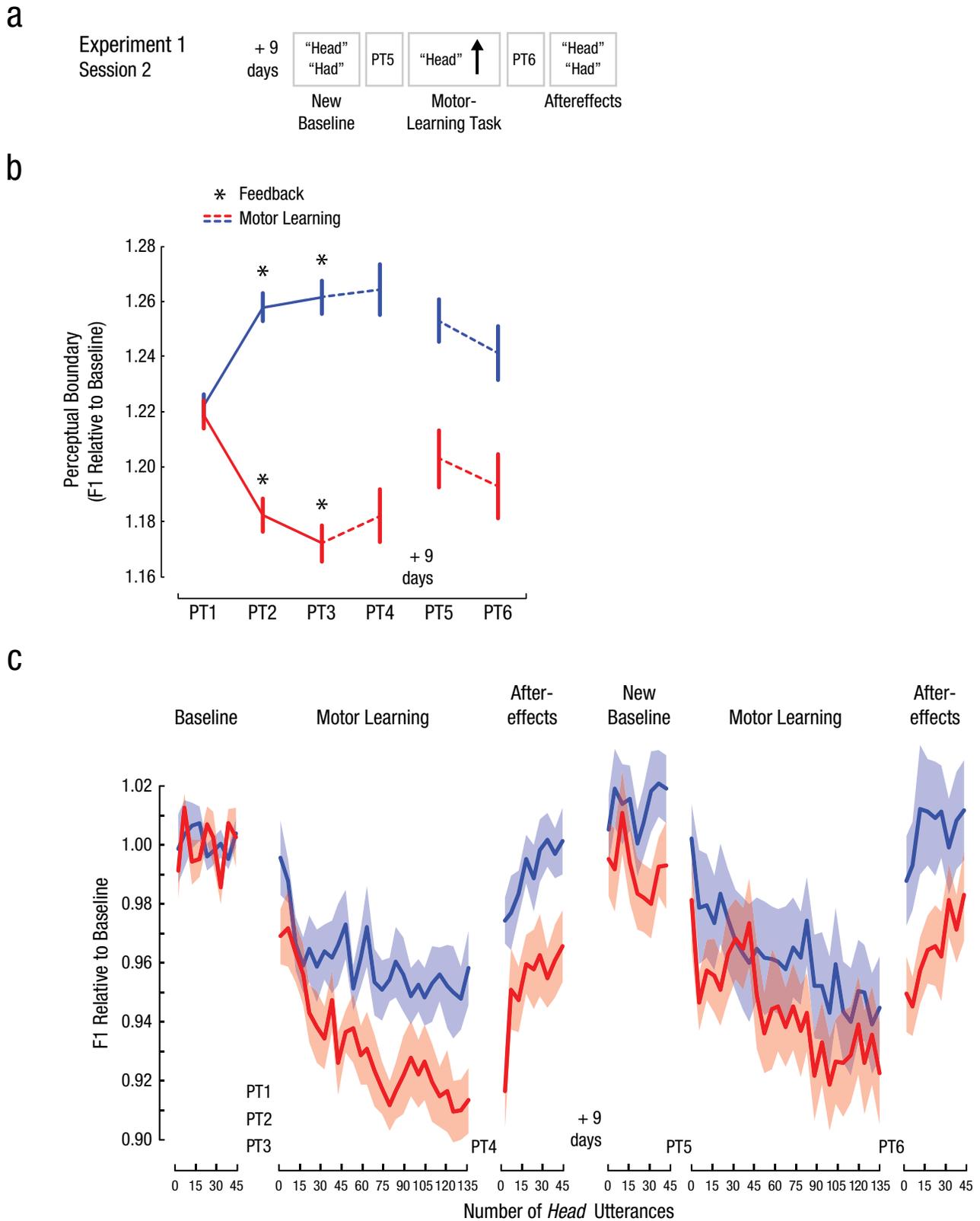


Fig. 4. Procedure and results for the second session of Experiment 1 (a–c) and for Experiment 2 (d–f). A subset of participants from Experiment 1 returned about 9 days later and repeated the experiment, minus perceptual training (a). In (b), the perceptual boundary between *head* and *bad* (calculated as the frequency of formant 1, or F1, relative to the frequency of F1 in *head*) is plotted as a function of perceptual test (PT). Error bars represent ± 1 SE. In (c), for the subjects who returned for the second session of testing, the produced frequency of F1 relative to the baseline frequency of F1 in *head* is plotted against the number of utterances of *head* in both sessions of testing (the procedure for Session 1 is described in Fig. 3b). The shaded areas around the curves represent ± 1 SE, and the curves join

(continued)

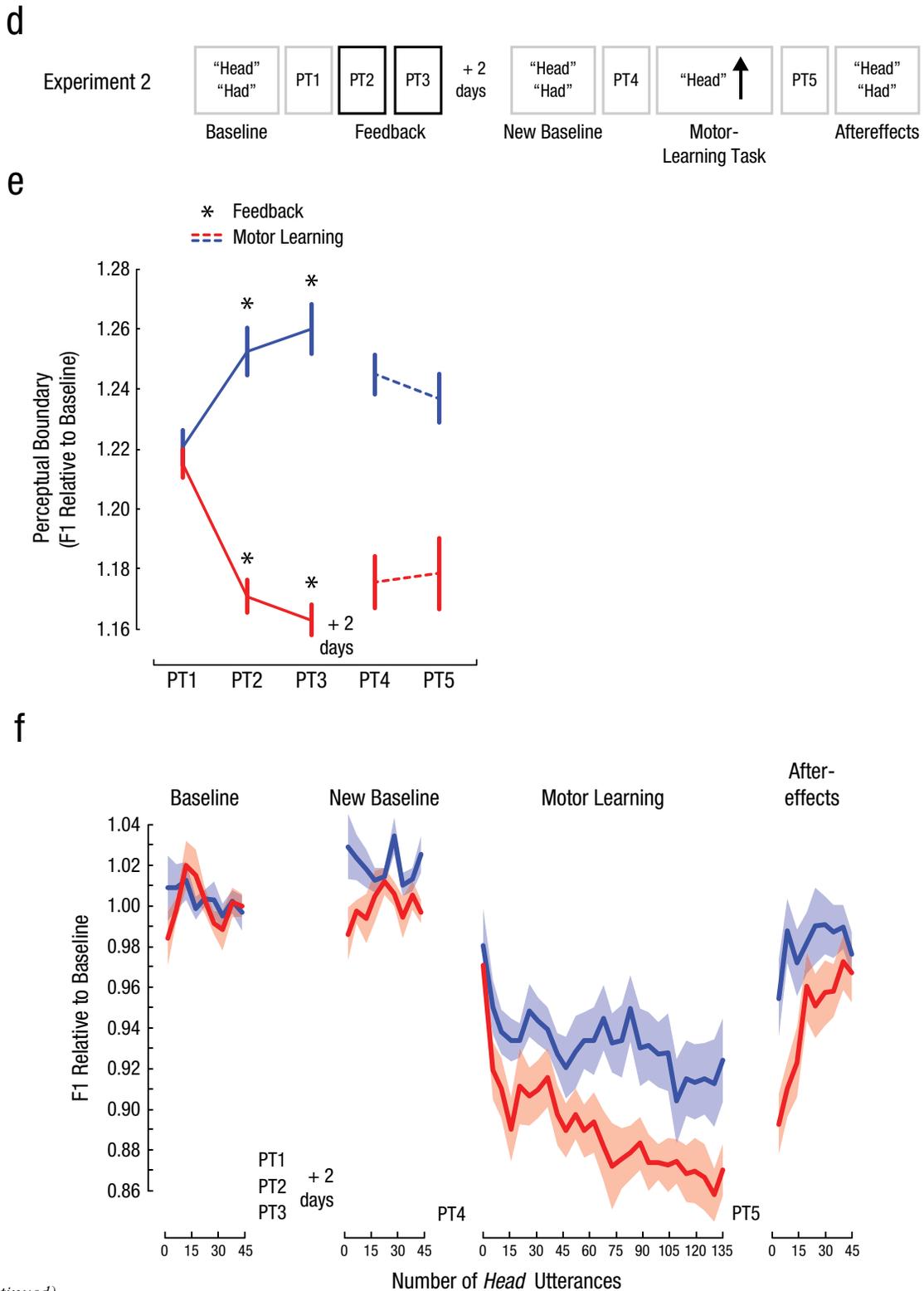


Fig. 4. (continued)

averages computed from blocks of five utterances. The first session of Experiment 2 (d) consisted of two phases. First, subjects spoke the words *head* and *had* (45 times each) without any manipulation (baseline). They then performed three perceptual tests designed to measure (PT1) and then alter (PT2 and PT3) the perceptual boundary between *head* and *had*. Two days later, subjects returned and spoke the word *head* 45 times each to establish a new baseline. This was followed by another perceptual test (PT4). Next, subjects spoke the word *head* 135 times, with the sound of their voices altered in real time (i.e., an increase in F1 so that the word sounded more like *bad*). In a fifth perceptual test (PT5), the perceptual boundary between the two words was measured again. Finally, subjects spoke the two words 45 times each without any manipulation so that aftereffects of the motor learning could be assessed. In (e), the perceptual boundary between *head* and *had* (calculated as the frequency of F1 relative to the frequency of F1 in *head*) is plotted against perceptual tests. Error bars represent ± 1 SE. In (f), the produced frequency of F1 relative to the baseline frequency of F1 in *head* is plotted against the number of utterances of *head* in both sessions of testing associated with Experiment 2.

Given the speed of adaptation, it seems important to question whether the acquired perceptual boundary reflects a true change in perception or simply a response alteration to follow the feedback. Across sensory systems, perceptual learning is typically defined as a long-lasting change in perception that improves an organism's ability to respond to its environment (Goldstone, 1998; Samuel & Kraljic, 2009). The feedback-driven change in the perceptual boundary we observed, and the persistence of this change days after feedback was removed, suggests that our participants' perception of the boundary between *head* and *bad* was altered. But, most importantly, perceptual training also caused differences in speech motor learning. Learned compensation for altered auditory feedback of vowel sounds is known to be unaffected by cognitive strategy. Subjects specifically instructed not to adjust their speech when their production of the word *head* is made to sound like *bad* show as much speech motor learning as those given no instruction (Munhall, MacDonald, Byrne, & Johnsrude, 2009). A response strategy adopted to meet the demands of perceptual training would have had little impact on subsequent speech motor learning.

Whether the perception of others' speech affects the speech motor learning of the listener was not the central question of the study, but the results suggest (with some caveats) that it does. That is, the *head-to-bad* continuum used in perceptual training was based on exemplars taken from a Canadian man, and we saw immediate and stable transfer to the speech motor learning of our 62 female listeners. This result, although in contrast with previous findings suggesting that perceptual learning of speech sounds is speaker-specific and does not cause a global change in the perception of the listener (Eisner & McQueen, 2005), fits nicely with the established idea that speech is learned from a tutor (Doupe & Kuhl, 1999). The perceptual targets that define adults' speech motor learning can be acquired, it seems, through listening. Even so, it remains unclear how much similarity between the speech of the tutor and the listener—in accent, for instance—is required for perceptual training to affect speech motor learning. A different result might have been obtained if the tutor in this study had a foreign accent.

It is also worth testing the extent to which transfer between perceptual training and speech motor learning depends on the perceptual similarity between the trained word and the produced word (Reinisch & Holt, 2013). In our study, perceptual retuning on a *head-to-bad* continuum altered productions of *head*. That is, the trained phonetic contrast included the produced word. The impact of perceptual training on speech motor learning may have been reduced or eliminated if participants had produced a different vowel (e.g., *hid*) during altered feedback. Finally, perceptual retuning in our study was driven using explicit feedback. Previous work has found

that implicit perceptual learning does not seem to affect speech production (Kraljic et al., 2008). Thus, the manner by which speech perception is altered may affect the transfer of perceptual change to speech production.

How tightly is speech production coupled to speech perception? The answer seems to depend on the circumstances. The results of our study suggest that perceptual change immediately drives changes in speech motor learning but has little impact on previously learned speech. Another instance in which speech perception and production appear linked occurs in the phenomenon of phonetic convergence. In this case, a rapid increase in the similarity of different acoustic properties of speech (e.g., voice-onset time, pitch, intensity, formant frequency) is observed when talkers interact (Pardo, 2013). However, the extent of phonetic convergence between acoustic measures across studies is highly variable, and the phenomenon may be driven by idiosyncratic traits of the interacting talkers, such as how attractive they find each other (Babel, 2012). More generally, one's daily acoustic environment can also drive more gradual changes in speech production. Harrington, Palethorpe, and Watson (2000) found that, over a 30-year period, Queen Elizabeth's vowel-sound production came to match that of younger, less socially refined English speakers. Of course, changes in speech perception occur in isolation of production change. As we noted earlier, people adapt their perception of speech to foreign accents without adopting those accents in their own speech. Thus, the relationship between speech perception and production is not fixed.

In the context of motor control, our experiments show that plasticity in adults' perceptual systems can have a marked effect on the outcome of motor learning, even if the perceptual change occurs in the absence of movement. Motor learning is typically studied by examining compensation patterns for perturbations that drive behaviors away from well-defined sensory targets. During the act of reaching, for instance, learning can be observed in both humans and nonhuman primates when the motion path of the limb is predictably perturbed (Krakauer, Ghilardi, & Ghez, 1999; Li, Padoa-Schioppa, & Bizzi, 2001; Shadmehr & Mussa-Ivaldi, 1994). Error-based motor learning of a similar kind is found both in birdsong models of vocal learning (Sober & Brainard, 2009) and in speech production (Houde & Jordan, 1998; Lametti et al., 2012), as demonstrated by the current study. In perturbation-based studies of motor learning, the nervous system detects that a sensory target has not been met, and motor commands are systematically adjusted to compensate for the error (Shadmehr, Smith, & Krakauer, 2010). These experimental models of motor learning thus explain the maintenance of behavior in response to well-defined sensory targets. But how were those sensory goals acquired in the first place?

The literature on limb motor learning has largely handled the question of how sensory targets are established in the context of movement—that is, the perceptual targets that guide movements are acquired by making movements, and then updated by new learning and experience (Körding & Wolpert, 2004; Wolpert, Diedrichsen, & Flanagan, 2011). However, during development, purely perceptual learning plays an integral role in defining the sensory targets that come to guide speech (Kuhl, 2004; Tsao et al., 2004). In this study, we tested whether the same is true for adults by experimentally separating perceptual learning and motor learning. The perceptual systems that support speech are notably plastic, and the results of this study provide further support for this idea. Most notably, however, changes in perception were immediately used by the motor system to shape how a new behavior was learned. Plasticity in sensory function that occurs in the absence of movement can thus play a significant role in motor learning.

Author Contributions

D. R. Lametti and D. J. Ostry designed the experiments. D. R. Lametti and S. A. Krol tested the participants. D. R. Lametti analyzed the data and wrote the first draft of the manuscript with help from S. A. Krol. The manuscript was then edited by D. R. Lametti, D. M. Shiller, and D. J. Ostry.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The research was supported by National Institute on Deafness and Other Communication Disorders Grant DC012502 and by the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT). S. A. Krol was supported by a Natural Sciences and Engineering Research Council of Canada Undergraduate Student Research Award. D. R. Lametti was supported by a postdoctoral fellowship from FQRNT.

References

- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*, 177–189.
- Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*, 592–597.
- Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer (Version 5.3.67) [Computer program]. Retrieved from <http://www.praat.org/>
- Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*, 977–985.
- Bradlow, A., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, *19*, 113–124.
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*, 3647–3658.
- Cressman, E. K., & Henriques, D. Y. P. (2009). Sensory recalibration of hand position following visuomotor adaptation. *Journal of Neurophysiology*, *102*, 3505–3518.
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567–631.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 914–927.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*, 224–238.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.
- Haith, A., Jackson, C., Miall, R., & Vijayakumar, S. (2008). Unifying the sensory and motor components of sensorimotor adaptation. *Advances in Neural Information Processing Systems*, *21*, 593–600.
- Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature*, *408*, 927–928.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*, 1213–1216.
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.
- Krakauer, J. W., Ghilardi, M. F., & Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nature Neuroscience*, *2*, 1026–1031.
- Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*, 54–81.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.
- Ladefoged, P. (1975). *A course in phonetics*. New York, NY: Harcourt Brace Jovanovich.
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, *32*, 9351–9358.
- Li, C. S., Padoa-Schioppa, C., & Bizzi, E. (2001). Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron*, *30*, 593–607.
- Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*, 543–562.
- Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time

- formant perturbation even when instructed not to compensate. *Journal of the Acoustical Society of America*, *125*, 384–390.
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, *33*, 12090–12098.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.
- Ostry, D. J., Darainy, M., Mattar, A. A., Wong, J., & Gribble, P. L. (2010). Somatosensory plasticity and motor learning. *Journal of Neuroscience*, *30*, 5384–5393.
- Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*, Article 559. Retrieved from <http://journal.frontiersin.org/Journal/10.3389/fpsyg.2013.00559/full>
- Reinisch, E., & Holt, L. L. (2013). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi:10.1037/a0034409
- Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory–motor transformations in speech. *Journal of Neuroscience*, *31*, 2657–2652.
- Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, *107*, 1711–1717.
- Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research*, *37*, 347–357.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*, 1207–1218.
- Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, *14*, 3208–3224.
- Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, *33*, 89–108.
- Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America*, *125*, 1103–1113.
- Shum, M., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2011). Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *European Journal of Neuroscience*, *34*, 1817–1822.
- Sober, S. J., & Brainard, M. S. (2009). Adult birdsong is actively maintained by error correction. *Nature Neuroscience*, *12*, 927–931.
- Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*, 1067–1084.
- Vahdat, S., Darainy, M., Milner, T. E., & Ostry, D. J. (2011). Functionally specific changes in resting-state sensorimotor networks after motor learning. *Journal of Neuroscience*, *31*, 16907–16915.
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*, 1033–1043.
- Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*, 739–751.