HHS Public Access

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

Daniel R. Lametti[1,2], **Sonia A. Krol**[1], **Douglas M. Shiller**[3,4,5], and **David J. Ostry**[1,6,*]

[1]Department of Psychology, McGill University, Montreal, Quebec, Canada, H3A 1B1

[2]Institute of Neurology, University College London, London, UK, WC1N 3BG

[3]School of Speech-Language Pathology & Audiology, l'Université de Montréal, Montreal, Quebec, Canada, H3N 1X7

[4]CHU Sainte-Justine Research Centre, Montreal, Quebec, Canada, H3T 1C5

[5]Centre for Research on Brain, Language & Music, Montreal, Canada, H3G 2A8

[6]Haskins Laboratories, New Haven, CT, 06511

Abstract

The perception of speech is notably malleable in adults, yet alterations in perception seem to have little impact on speech production. 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 adult speech perception can have immediate consequences for speech production in the context of speech learning.

Introduction

Our perception of speech is remarkably plastic, yet alterations in speech perception seem to have little immediate impact on speech production. We quickly come to understand foreign English accents, for instance, but this perceptual change does not cause us to suddenly adopt a foreign accent in our own speech. This phenomenon contrasts with other behaviors like reaching, where increased visual acuity from, say, a new pair of glasses would immediately be utilized by the brain to make more accurate movements. Here we provide initial evidence that alterations in speech perception impact adults in the same manner that they impact young children: during speech learning.

*Corresponding Author, David J. Ostry, david.ostry@mcgill.ca, (514) 398-611.

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 constitute 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. Both anecdotally and experimentally it is apparent that we adapt our speech perception to the differing acoustic properties of foreign accents (Clarke and Garrett 2004; Maye et al. 2008). Furthermore, computer-altered speech is quickly understood (Dupoux and Green 1997). Within one's first language, however, changes in speech perception seem to have a negligible impact on speech production (Kraljic et al. 2008; Samuel and Kraljic 2009) unless the perceptual change is driven by a considerable amount of training (Rvachew 1994). Perceptual training can impact speech production in the case of second language learning, but, again, only after days of training (Bradlow et al. 1997; Bradlow et al. 1999; Wang et al. 2003).

Recent work on the motor control of speech and limb movements has shown that perceptual change is coupled to motor learning (Shiller et al. 2009; Haith et al. 2009; Cressman and Henriques 2009; Ostry et al. 2010), while studies of speech development show that changes in speech perception precede speech learning (Kuhl 2004; Tsao et al. 2004). Here we have examined the impact of perceptual change on adults' capacity for speech learning in their first language. By pairing a perceptual training task with a motor learning task, we show that altering the perceptual distinction or boundary between two vowel sounds significantly influences the degree to which participants learn to adapt their speech motor control to perceived production errors when producing these sounds. In support of prior work, previously learned speech movements were left unchanged by perceptual training. The experiments demonstrate that alterations in the perception of speech do in fact have immediate consequences for adult speech production, only not in a way that has been previously considered.

Figure 1A lays out the experimental hypothesis. When the first formant frequency (F1) of the vowel sound in "head" is increased in real-time so that subjects hear something closer to the vowel sound in "had", individuals compensate by decreasing the frequency of produced F1 until their heard production once again falls within the perceptual range of "head". If speech perceptual training manipulates perception of the boundary between had and head, the alteration should thus impact the amount of compensation in a subsequent test of speech motor learning.

Methods

Subjects and Apparatus

Sixty-four native-English-speaking females (18–30 yrs) with normal hearing and speech participated in the study. Forty-four females participated in the first experiment and twenty females participated in the second experiment. The sample sizes were chosen based on our previous speech motor learning experiments that demonstrated significant group differences with ten to twenty participants in each condition (Rochet-Capellan and Ostry 2011; Lametti et al. 2012; Rochet-Capellan et al. 2012). The McGill University Faculty of Medicine

Institutional Review Board approved the experiments. Testing was performed in a sound-attenuating chamber. Subjects wore headphones (Stax) and a directional microphone (Sennheiser) recorded speech. Two 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.

Experimental Procedures

The experiments involved an initial measurement of baseline speech production followed by perceptual testing and training, and then speech motor learning (Figure 1B). Speech production was prompted by the appearance of "head" or "had" on a computer screen. Subjects were instructed to say the word that appeared in a clear voice. Once the word was produced it was removed from the screen and the next word was displayed. During baseline production, subjects said "head" and "had" 45 times each in a random order. The first perceptual test (PT1) was then performed to measure the perceptual boundary between the words "head" and "had" (see *Measuring Speech Perception*). Based on this measurement, each subject experienced two perceptual tests (PTr2 and PTr3) with feedback designed to systematically shift their perceptual boundary (see *Perceptual Training*). Experiment 1 then had all forty-four subjects produce the word "head" 135 times with the sound of their voice altered in real-time (see *Alterations of Speech*). A second, perceptual test (PT4) without feedback was performed after altered auditory feedback. Subjects then produced "head" and "had" 45 times with unaltered speech to examine after-effects associated with speech motor learning.

Subjects tested in Experiment 1 were invited back to the lab for a second session of testing. Twenty-eight of the subjects returned; the amount of time between the first day of testing and the second session of testing averaged 8.85 days (SD = 2.6 days, range = 7 to 14 days). Those who returned repeated Experiment 1 minus perceptual training (Figure 4A). In Experiment 2, twenty new subjects were recruited. After an initial session of testing that consisted of baseline tests of production and perception followed by perceptual training, subjects returned to the lab two days later for a second measurement of baseline production. This was followed by two more measurements of speech perception before and after speech motor learning (Figure 4D).

Measuring Speech Perception

Perception was measured using a continuum of ten words that spanned the perceptual distinction between "head" and "had" based on utterances provided by a Canadian male. To create the ten-step continuum, the first two formants taken from the word "head" were shifted in equal steps towards formant values in "had". F1 and F2 for "head" were 560 Hz and 1745 Hz, respectively. F1 and F2 for "had" were 768 Hz and 1648 Hz. During perceptual testing, each stimulus was played once in a random position within a block of the entire set of stimuli. This was repeated 21 times in each perceptual test. Upon hearing a stimulus, subjects were prompted by text on a computer screen to indicate whether the stimulus sounded more like "head" or more like "had" by pressing keys on a keyboard. The spacebar 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 where "head" is perceived 50% of the time—was calculated from the psychometric functions.

Perceptual Training

After the initial perceptual test, the perceptual boundary was computed based on each subject's responses. For half the subjects, a new perceptual boundary was then set one stimulus lower than their original, rounded-to-the-nearest-integer perceptual boundary (Figure 2A). Feedback during perceptual training was given around this new boundary. If the new perceptual boundary was set to stimulus 6, for instance, "CORRECT" was displayed on the screen if the subject indicated hearing "head" for stimuli 1 through 5 and "had" for stimuli 6 through 10; "INCORRECT" was displayed on the screen if the subject reported hearing "had" for stimuli 1 through 5 and "head" for stimuli 6 through 10. For the remaining subjects, a new perceptual boundary was set one stimulus higher than their original, rounded-to-the-nearest-integer perceptual boundary, and feedback was given in a similar manner. "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 (PTr2), the number of errors made was displayed on the screen with the instruction to reduce this number. The error counter was then reset to zero and subjects made another 210 perceptual choices with feedback. Perceptual testing and training took approximately 18 minutes.

Analysis of Perceptual Data

To compute the perceptual boundary on the same unitless scale used to relate speech motor learning to baseline production, the perceptual stimuli were represented as F1 frequency relative to the F1 frequency of stimulus 1 ("head"). Thus, stimulus 1 was 560 Hz / 560 Hz or 1.0, stimulus 2 was 582 Hz / 560 Hz or 1.04 and so on towards stimulus 10, which was 768 Hz / 560 Hz or 1.37. The psychometric function was fit to the proportion of had responses at each of these values and the perceptual boundary was found for each perceptual test from this function. For Figure 3C, 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 above. Changes in perceptual boundaries were assessed using split-plot ANOVAs with Bonferroni corrected post-hoc tests. To examine changes in perception over time, the proportion of "had" responses was calculated based on blocks of ten perceptual decisions (Figure 2B). 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

Acoustical signal processors and filters were used to shift F1 of the vowel sound in "head" up in frequency; the remaining formants were left unchanged (Rochet-Capellan and 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 produced the word "head" but heard a word with an F1 closer to that in "had". In Experiment 1, the signal

processor increased F1 frequency by approximately 24% resulting in a +174 Hz (SD, 22 Hz) change in F1. There was no difference in the amount of F1 shift for subjects in the two perceptual training directions (t(40) = 0.77, p > 0.45). The baseline F1 frequency of subjects in Experiment 1 averaged 739 Hz, and there was no difference between the two groups in baseline F1 frequency (t(40) = 1.35, p > 0.15). In Experiment 2, the signal processor increased baseline F1 frequency by approximately 26% resulting in a +186 Hz (SD 21 Hz) change in F1. As in Experiment 1, there was no difference in the amount of shift between the two perceptual training directions (t(18) = 0.24, p > 0.80). The baseline F1 frequency of subjects in Experiment 2 averaged 729 Hz, and there was no difference in baseline F1 frequency between the two groups (t(18) = 0.07, p > 0.9).

Auditory Analysis

Speech was recorded at 44.1 kHz (16-bit). As there are large differences in F1 frequency between males and females, only females were tested. The software package Praat detected vowel boundaries and calculated F1 frequencies based on a 30 ms window at the center of the vowel (Rochet-Capellan and Ostry 2011; Shum et al. 2011). In all experiments, to examine changes in F1 related to altered auditory feedback, the F1 frequency of each utterance was divided by the mean F1 of the last 30 "head" utterances of baseline production from the first session of testing (pre-training production). The mean of the last 45 utterances under altered auditory feedback and the first 15 utterances of after-effect trials was found for this measure of normalized F1. For the subjects that returned to the lab after initial testing, mean normalized F1 was found for the last 30 utterances of the second session of baseline production, as well as the last 45 utterances of the second session of altered auditory feedback, and the first 15 utterances of the second session of after-effect 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 of normalized F1 values based on blocks of five utterances taken from the altered feedback phase of the experiment.

Results

In Experiment 1, one group of 21 subjects (red data) received feedback that moved their perceptual boundary towards "head". A second group of 21 subjects (blue data) received feedback that moved their perceptual boundary towards "had". Figure 2A shows the average of psychometric functions fit to perceptual responses before (black curves, from PT1) and during (red and blue curves, from PTr3) perceptual training. Perceptual training caused a shift in the psychometric curves either towards "head" or towards "had" on the continuum. The mean R^2 for the psychometric fits was 0.98 (range = 0.88 to 0.99). Figure 2B shows the proportion of "had" responses averaged across subjects computed from blocks of 10 perceptual judgments made with (PTr2 and PTr3) and without feedback (PT1). To help visualize the speed of perceptual change, exponential functions were fit to the data (Figure 2B). The coefficient of determination, R^2, was equal to 0.49 for the red curve, and 0.32 for the blue curve. As computed from the fit functions, perceptual change reached 90% of asymptote by the 88th trial for the red group and by the 44th trial for the blue group.

Figure 3A shows the perceptual boundary in units of F1 relative to baseline for each perceptual test in Experiment 1. Perceptual training moved the boundary of the red group towards "head" and the blue group towards "had" ($p < 0.001$, in each case). This change in the perceptual boundary was also observed in the perceptual test that followed speech motor learning (PT4, $p < 0.001$). Figure 3B plots F1 frequency of the vowel sound in "head" relative to baseline production over the course of the experiment. Following perceptual training, subjects produced the word "head" with the signal processor turned on such that F1 for the vowel was increased to a value closer to "a". Subjects compensated for this alteration by learning to produce F1 at a lower frequency. Figure 3B shows that the group of subjects who had their perceptual boundary shifted towards "head" learned to compensate more for the speech alteration than the group of subjects who had their perceptual boundary shifted towards "had" ($p < 0.04$). They also showed greater learning-related after effects when the voice alteration was removed ($p < 0.02$). Figure 3C shows that the amount of speech motor learning in response to the voice alteration depended on the distance from "had" to the acquired perceptual boundary measured during the third perceptual test (PTr3) ($r = 0.52$, $p < 0.0005$). Significant correlations between these measures were also found within each group ($r = 0.49$, $p < 0.03$ for the red data, and $r = 0.51$, $p < 0.02$ for the blue data). Furthermore, a negative correlation was observed between training-related changes in perception (PTr3-PT1) and the amount of speech motor learning ($r = −0.37$, $p < 0.02$). Shifts in the perceptual boundary towards "head" were associated with greater speech motor learning, while shifts towards "had" were associated with less. The results suggest that perceptual training predictably altered speech motor learning.

To visualize differences in speech motor learning caused by perceptual training, Figure 3D shows exponential functions fit to the patterns of sensorimotor learning shown in Figure 3B for each of the two groups. The coefficient of determination, R^2, was equal to 0.93 for the red curve, and 0.66 for the blue curve. As computed from the functions, the red curve reaches asymptote at 0.91 (95% CI: 0.909 to 0.918) in units of F1 relative to baseline, and the blue curve reaches asymptote at 0.95 (95% CI: 0.946 to 0.951) in units of F1 relative to baseline. It is thus unlikely that the two groups would have achieved the same amount of learning with more training. There was also no difference in the starting point of the curves. The red curve started at 0.99 (95% CI: 0.974 to 1.004) and the blue curve started at 1.0 (95% CI: 0.970 to 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 ($p > 0.5$). This value was 0.98 (0.06 SD) in the case of the red group, and 0.99 (0.09 SD) in the case of the blue group. This suggests that perceptual training altered the amount of speech motor learning without significantly altering baseline production.

Twenty-eight of the subjects who participated in the first experiment returned to the lab approximately 8 days later. The subjects repeated Experiment 1 minus perceptual training (Figure 4A). Eight days after speech perceptual training there were still differences in the perceptual boundary between the two groups (Figure 4B: red versus blue for PT5 and PT6, $p < 0.01$ in each case). But only the group that had their perceptual boundary shifted towards "had" maintained a boundary change that differed from baseline ($p < 0.05$). Even so, the group of subjects who had their perceptual boundary shifted towards "head" eight days earlier showed greater learning-related after-effects (Figure 4C) than the group of subjects

who had their perceptual boundary shifted towards "had" (p < 0.02). 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 after-effect observed during the return session of testing may have been driven by a perceptual-training-induced difference in baseline speech production, or a failure to completely washout semsorimotor learning in the case of the red group from the first session of testing. Indeed, when the patterns of speech motor learning were normalized to baseline production during the return session of testing the difference in after-effect was reduced and no longer significant (p = 0.076).

20 new female subjects were recruited for Experiment 2. These 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 two days later, after a period of baseline production (Figure 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 sensorimotor learning. It also allowed for the direct examination of the effect of perceptual training on baseline production.

As in Experiment 1, perceptual training altered the perceptual boundary in the new group of subjects (Figure 4E). Two days later both groups still showed a boundary change that was different from baseline as measured by a perceptual test without feedback (PT4, p < 0.02 in each case). This perceptual test was followed by speech motor learning trials involving production of the word "head". Figure 4F shows that the group who had their perceptual boundary shifted towards "head" (red data) learned to compensate more (p < 0.05) for the voice alteration than the group who had their perceptual boundary shifted towards "had" (blue data). They also showed greater learning-related after effects when the voice alteration was removed (p < 0.01). Following perceptual training, there was no difference between the groups in baseline F1 frequency (p > 0.3), and the same impact of perceptual training on motor learning was observed even if the data were normalized to post-training baseline production (p < 0.05). For the red group, perceptual training caused a +0.2% change in baseline F1 (p > 0.5); for the blue group, perceptual training caused a +1.7% change in baseline F1 (p > 0.08). The results of Experiment 2 show that perceptual training altered speech motor learning two 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 adult speech motor learning. Speech perception is notably malleable in adults (Dupoux and Green 1997, Clarke and Garrett 2004, Bertelson et al. 2003; Norris et al. 2003); however, previous work suggests that experimentally induced changes in speech perception transfer quite slowly to production if they transfer at all (Rvachew 1994; Bradlow 1997; Wang et al. 2002; Kraljic et al. 2008). Our results largely support this work in that we see little impact on 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. This hypothesis is supported by a recent study: Niziolek and Guenther (2013) examined compensation for unpredictable perturbations of vowel sounds in relation to the perceptual boundaries between the altered vowels. They found that compensation was substantially greater for perturbations that pushed perceived productions into a new perceptual category (e.g. "bed" to "bad"). This suggests that alterations in the perceptual boundary between vowels will have a significant impact on the amount of learned compensation when vowel productions are predictably perturbed—exactly the result observed here.

Changes in the perceptual boundary in this study were driven with only 42 repetitions of the ten-step perceptual continuum, or 12 minutes of perceptual training. Given the speed of adaptation it seems important to ask 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 (Samuel and Kraljic 2008). The feedback driven change in the perceptual boundary observed here, and the persistence of this change days after feedback was removed, suggests that our participants perception of the boundary between "head" and "had" was altered. But more convincing of a true perceptual change, perceptual training 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 respond when their production of the word "head" is made to sound like "had" show as much speech motor learning as those given no instruction (Munhall et al. 2009). A response strategy adopted to meet the demands of perceptual training would have had little impact on subsequent speech motor learning.

Although not the central question of the study, the results suggest (with some caveats) that the perception of others' speech affects the speech motor learning of the listener. That is, the head-to-had continuum used in perceptual training was based on exemplars taken from a Canadian male, and we saw immediate and stable transfer to the speech motor learning of our listeners, sixty-two females. This result, although in contrast with previous work suggesting that perceptual learning of speech sounds is speaker specific and does not cause a global change in the perception of the listener (Eisner and McQueen 2005), fits nicely with the established idea that speech is learned from a tutor (Doupe and Kuhl 2003). The perceptual targets that define adult speech motor learning can, it seems, be acquired through listening. Even so, it remains unclear how much affiliation between the tutor and the listener —in accent, for instance—is required for perceptual training to impact speech motor learning. A different result may 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 and Holt 2013). Here, perceptual retuning occurred on a head-to-had continuum, altering adaptation to altered feedback during 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 this study was driven using explicit feedback. Previous work has found that implicit perceptual learning does not seem to impact speech production (Kraljic et al. 2008). How speech perception is altered may impact the transfer of perceptual change to speech production.

How tightly is speech production coupled to speech perception? It seems to depend on the circumstances. The results of this 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 (VOT, pitch, intensity, formant frequency) is observed when talkers interact (Pardo et al. 2013). However, there is a high degree of variability in the amount of phonetic convergence between acoustic measures and studies, and the phenomenon may be driven by idiosyncratic traits of interacting talkers, such how attractive they find each other (Babel 2012). More generally, ones' daily acoustic environment can also drive more gradual changes in speech production. Harrington et al. (2000) found that over a thirty-year period Queen Elizabeth's vowel sound production came to match that of younger, socially less refined English speakers. Of course, changes in speech perception occur in isolation of production change. As previously noted, we adapt our perception of speech to foreign accents without adopting the accent in our own speech. Thus, the relationship between speech perception and production is not fixed.

In the context of motor control, the experiments show that plasticity in adult perceptual systems can have a marked affect 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 disturbances that drive behaviors away from well-defined sensory targets. During reaching, for instance, learning can be observed in both humans and nonhuman primates when the motion path of the limb is predictably perturbed (Shadmehr and Mussa-Ivaldi 1994; Krakauer et al. 1999; Li et al. 2001). Error-based motor learning of a similar kind is found in both bird-song models of vocal learning and speech production, as demonstrated here (Houde and Jordan 1999; Sober and Brainard 2009; Lametti et al. 2012). In these paradigms, the nervous system detects that a sensory target has not been met and motor commands are systematically adjusted to compensate for the error (Shadmehr et al. 2010). These experimental models of motor learning thus explain the maintenance of behavior in relation 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 and Wolpert 2004; Wolpert et al. 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). Here we tested whether the same is true for

adults through an experimental separation of perceptual learning and motor learning. The perceptual systems that support speech are notably plastic, and the results of this study support this idea. Most notably, though, changes in perception were immediately utilized 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 adult motor learning.

Acknowledgements

References

Babel M. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. J. Phon. 2012; 40:177–189.

Bertelson P, Vroomen J, de Gelder B. Visual recalibration of auditory speech identification: A McGurk after effect. Psychol. Sci. 2003; 14:592–597. [PubMed: 14629691]

Bradlow AR, Pisoni DB, Akahane-Yamada R, Tohkura Y. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. J Acoust Soc Am. 1997; 101(4):2299–2310. [PubMed: 9104031]

Bradlow A, Akahane-Yamada R, Pisoni D, Tohkura Y. Training Japanese Listeners to identify English/r/and/l: Long-term retention of learning in perception and production. Perception & psychophysics. 1999; 61:977–985. [PubMed: 10499009]

Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. Neuroimage. 2003; 19(1):113–24. [PubMed: 12781731]

Clarke CM, Garrett MF. Rapid adaptation to foreign-accented English. J Acoust Soc Am. 2004; 116(6):3647–3658. [PubMed: 15658715]

Cressman EK, Henriques DYP. Sensory recalibration of hand position following visuomotor adaptation. J Neurophysiol. 2009; 102:3505–3518. 2009. [PubMed: 19828727]

Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci. 1999; 22:567–631. [PubMed: 10202549]

Dupoux E, Green K. Perceptual adjustment to highly compressed speech: effects of talker and rate changes. J Exp Psychol Hum Percept Perform. 1997; 23(3):914–927. [PubMed: 9180050]

Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. Percept Psychophys. 2005; 67(2):224–238. [PubMed: 15971687]

Goldstone RL. Perceptual learning. Ann Rev of Psych. 1998; 49:585–612.

Hazen B, Barret S. The development of phonemic categorization in children aged 6–12. Journal of Phonetics. 2000; 28:377–396.

Haith A, Jackson C, Miall R, Vijayakumar S. Unifying the sensory and motor components of sensorimotor adaptation. Adv Neural Inf Process Syst. 2008; 21:593–600.

Harrington J, Palethorpe S, Watson CI. Does the Queen speak the Queen's English? Nature. 2000; 408:927–928. [PubMed: 11140668]

Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science. 1998; 279(5354):1213–1216. [PubMed: 9469813]

Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. Nature. 2004 Jan 15; 427(6971):244–247. 2004. [PubMed: 14724638]

Kraljic T, Brennan SE, Samuel AG. Accommodating variation: Dialects, idiolects, and speech processing. Cognition. 2008; 107:54–81. [PubMed: 17803986]

Krakauer JW, Ghilardi MF, Ghez C. Independent learning of internal models for kinematic and dynamic control of reaching. Nat Neurosci. 1999; 2(11):1026–1031. [PubMed: 10526344]

Kuhl PK. Early language acquisition: cracking the speech code. Nat Rev Neurosci. 2004; 5(11):831–843. [PubMed: 15496861]

Ladefoged. A course in phonetics. Orlando: Harcourt Brace; 1975. ISBN 0-15-507319-2. 2nd ed 1982, 3rd ed. 1993, 4th ed. 2001, 5th ed. Boston: Thomson/Wadsworth 2006, 6th ed. 2011 (co-author Keith Johnson) Boston: Wadsworth/Cengage Learning

Lametti DR, Nasir SM, Ostry DJ. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. J Neurosci. 2012; 32(27):9351–9358. [PubMed: 22764242]

Li CS, Padoa-Schioppa C, Bizzi E. Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. Neuron. 2001; 30(2):593–607. [PubMed: 11395017]

Maye J, Aslin RN, Tanenhaus MK. The weckud wetch of the wast: lexical adaptation to a novel accent. Cogn Sci. 2008 Apr 5; 32(3):543–562. 2008. [PubMed: 21635345]

McClelland JL, Fiez JA, McCandliss BD. Teaching the /r/-/l/discrimination to Japanese adults: behavioral and neural aspects. Physiol Behav. 2002; 77(4–5):657–662. [PubMed: 12527015]

Munhall KG, MacDonald EN, Byrne SK, Johnsrude I. Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. Acoust Soc Am. 2009; 125(1):384–390.

Niziolek CA, Guenther FH. Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. J Neurosci. 2013; 33(29):12090–12098. [PubMed: 23864694]

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. Cog. Psychol. 2003; 47:204–238.

Ostry DJ, Darainy M, Mattar AA, Wong J, Gribble PL. Somatosensory plasticity and motor learning. J Neurosci. 2010; 30(15):5384–5393. [PubMed: 20392960]

Pardo JS. Measuring phonetic convergence in speech production. Front Psychol. 2013; 4:559. 2013. [PubMed: 23986738]

Reinisch E, Holt LL. Lexically Guided Phonetic Retuning of Foreign-Accented Speech and Its Generalization. J Exp Psychol Hum Percept Perform. 2013 Sep 23. 2013 [Epub ahead of print].

Rochet-Capellan A, Ostry DJ. Simultaneous acquisition of multiple auditory-motor transformations in speech. J Neurosci. 2011; 31(7):2657–2662. [PubMed: 21325534]

Rochet-Capellan A, Richer L, Ostry DJ. Nonhomogeneous transfer reveals specificity in speech motor learning. J Neurophysiol. 2012; 107(6):1711–1717. [PubMed: 22190628]

Rvachew S. Speech perception training can facilitate sound production learning. J Speech Hear Res. 1994; 37(2):347–357. [PubMed: 8028316]

Samuel AG, Kraljic T. Perceptual learning for speech. Atten Percept Psychophys. 2009; 71(6):1207–1218. [PubMed: 19633336]

Shadmehr R, Mussa-Ivaldi FA. Adaptive representation of dynamics during learning of a motor task. J Neurosci. 1994; 14:3208–3224. [PubMed: 8182467]

Shadmehr R, Smith MA, Krakauer JW. Error correction, sensory prediction, and adaptation in motor control. Annu Rev Neurosci. 2010; 33:89–108. [PubMed: 20367317]

Shiller DM, Sato M, Gracco VL, Baum SR. Perceptual recalibration of speech sounds following speech motor learning. J Acoust Soc Am. 2009; 125(2):1103–1113. [PubMed: 19206885]

Shum M, Shiller DM, Baum SR, Gracco VL. Sensorimotor integration for speech motor learning involves the inferior parietal cortex. Eur J Neurosci. 2011; 34(11):1817–1822. [PubMed: 22098364]

Tsao FM, Liu HM, Kuhl PK. Speech perception in infancy predicts language development in the second year of life: a longitudinal study. Child Dev. 2004; 75(4):1067–1084. [PubMed: 15260865]

Vahdat S, Darainy M, Milner TE, Ostry DJ. Functionally specific changes in resting-state sensorimotor networks after motor learning. J Neurosci. 2011; 31(47):16907–16915. [PubMed: 22114261]

Wang Y, Jongman A, Sereno JA. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. J Acoust Soc Am. 2003 Feb; 113(2):1033–1043. 2003. [PubMed: 12597196]

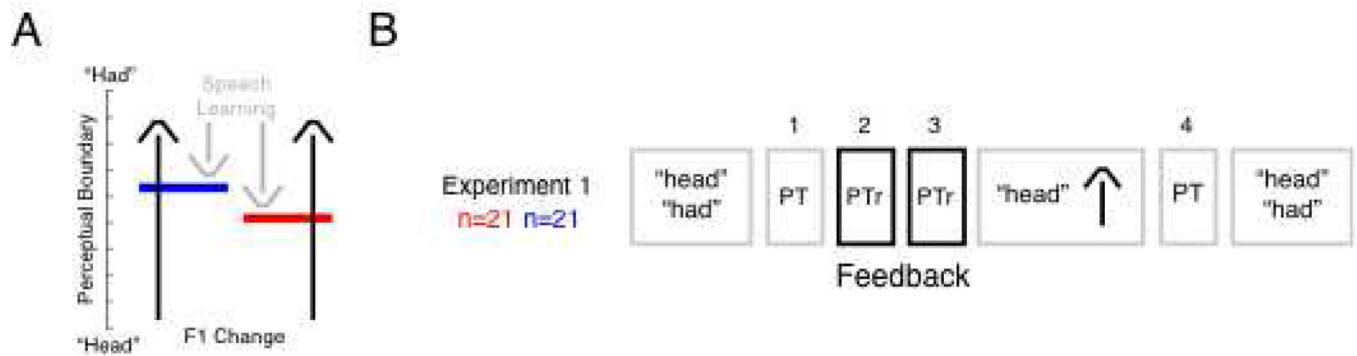Wolpert DM, Diedrichsen J, Flanagan JR. Principles of sensorimotor learning. Nat Rev Neurosci. 2011 Oct 27; 12(12):739–751. 2011. [PubMed: 22033537]

Figure 1.
Hypothesis and Experimental Design (A) An acoustical effects processor was used to alter the first formant frequency (F1) of the vowel sound in "head" so that it sounded more like "had" (black arrows). It was hypothesized that individuals compensate for this alteration by lowering produced F1 (grey arrows) until their heard production once again falls within the perceptual range of "head". Changing the point of perceptual distinction between "head" and "had" should thus alter the amount of compensation (blue versus red bars). (B) Two groups of 21 female subjects were tested. Unaltered production of "head" and "had" (baseline) was followed by three perceptual tests designed to measure (PT1) and then alter (PTr2 and PTr3) the perceptual boundary between "head" and "had". Perceptual training was followed by production of the word "head" with an increase in F1 so that the word sounded more like "had", a final measurement of the perceptual boundary (PT4), and unaltered productions of "head" and "had" to "washout" learning.
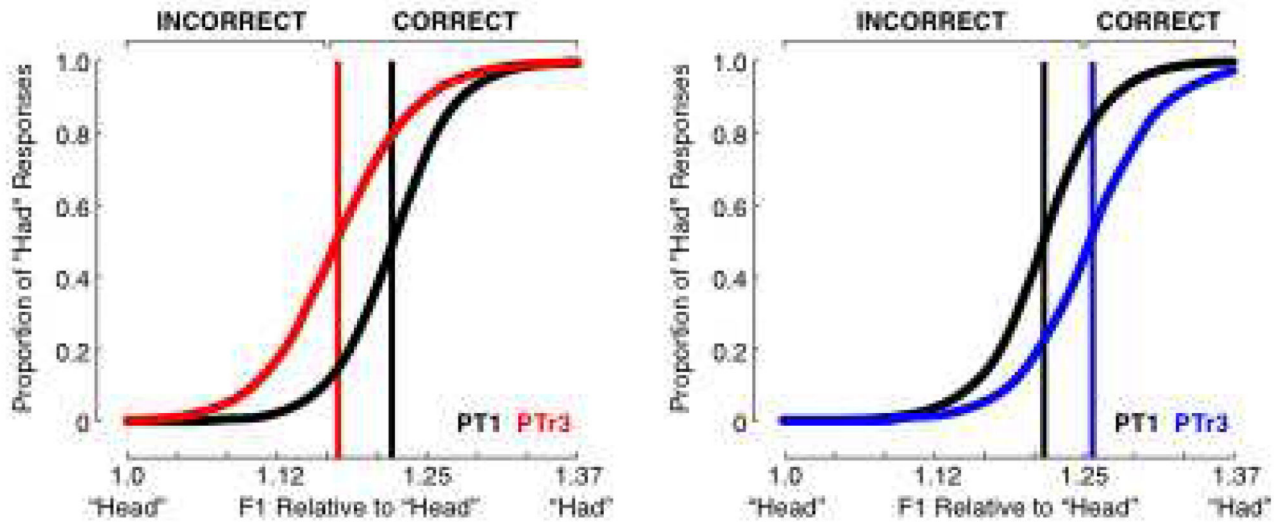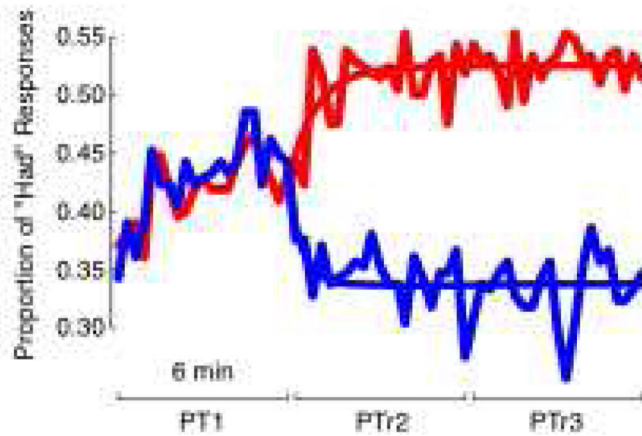
Figure 2.
Perceptual Training Caused a Change in Response (A) Feedback based on the solid colored lines moved the psychometric function from the black curves to the colored curves. The black lines show the initial perceptual boundary; the colored lines show the perceptual boundary following training. Perceptual boundaries were calculated in units of F1 relative to stimulus 1 ("head"). "CORRECT/INCORRECT" feedback refers to a "had" response, in this case. The black psychometric functions are based on responses from the first perceptual test (PT1). The colored psychometric functions are based on responses from the third perceptual test (PTr3), which included feedback. (B) The colored lines join points that represent the average of 10 perceptual responses. Each perceptual test took about 6 minutes. Perceptual training caused a change in the proportion of "had" responses.
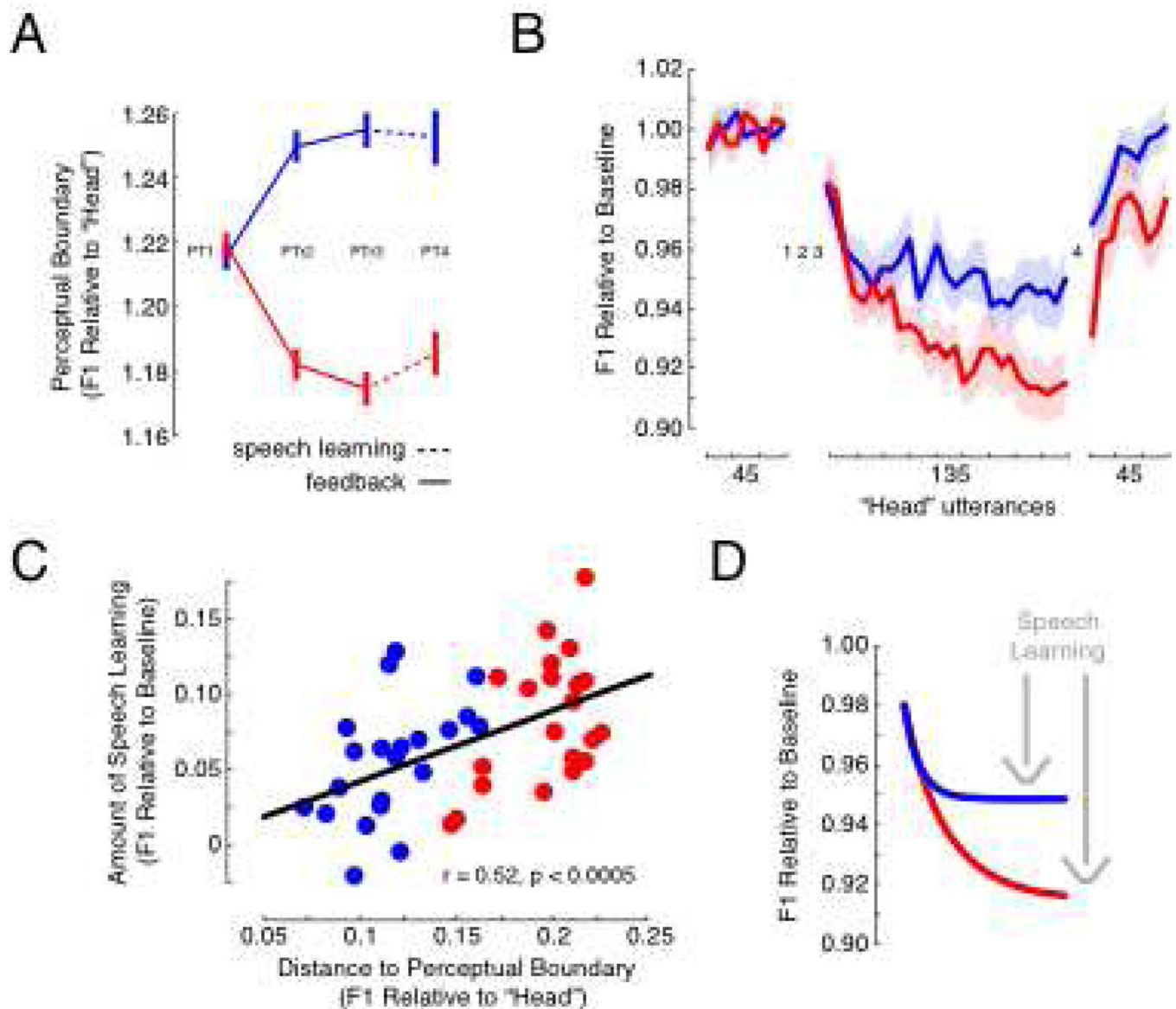
Figure 3.
Perceptual Training Altered Speech Motor Learning (A) During PTr2 and PTr3, feedback moved the perceptual boundary towards "had" (blue data) or "head" (red data) (p < 0.0005). Error bars represent +/− 1 SE. (B) Produced F1 frequency for "head" was normalized to baseline utterances. Subjects compensated for an increase in perceived F1 by decreasing produced F1 (trials 46 to 180). Subjects who had their perceptual boundary shifted towards "head" compensated more for altered auditory feedback than subjects who had their boundary shifted towards "had" (p < 0.04). They also showed greater compensation-related after-effects (p < 0.02). Error clouds represent +/− 1 SE, and the learning curves join averages computed from blocks of five utterances. (C) A significant correlation was observed between the distance from stimulus 10 ("Had") to the trained perceptual boundary and the amount of compensation for altered auditory feedback. (D) Exponential functions

were fit to compensation patterns for altered auditory feedback to visualize the effect of perceptual training on speech motor learning (grey arrows).
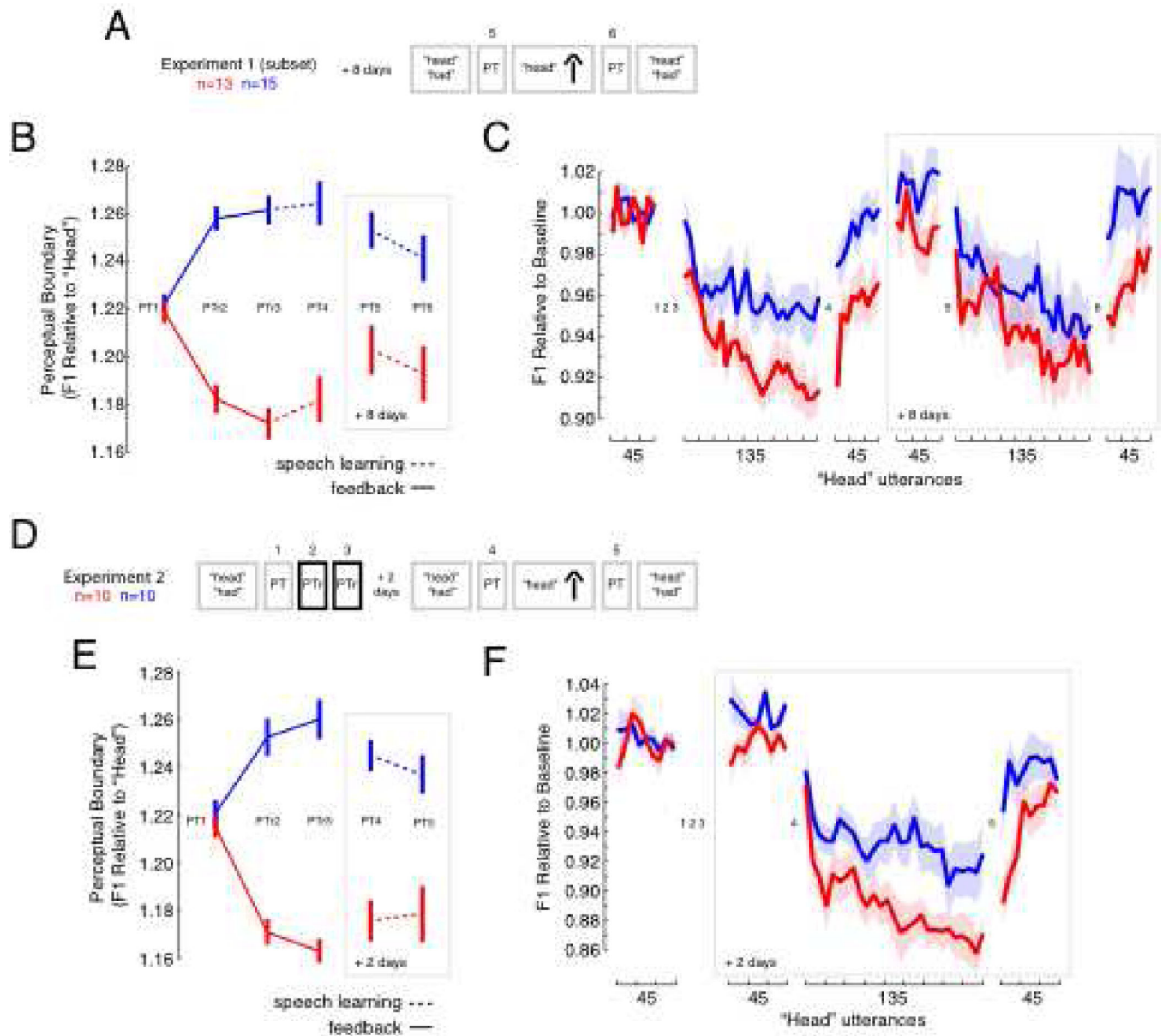
Figure 4.
The Effect of Perceptual Training on Motor Learning Lasts for Days (A) A subset of
participants returned about 8 days later and repeated the experiment, minus perceptual
training. (B) Perceptual boundary differences were present 8 days after training. Error bars
represent +/− 1 SE. (C) Produced F1 frequency for "head" was normalized to baseline
utterances from Day 1. Subjects compensated for an F1 increase by decreasing the
frequency of produced F1 (trials 46 to 180; eight days later, trials 271 to 405). The group
that had their perceptual boundary shifted towards "head" showed greater learning-related
after-effects eight days after training (p < 0.02). Error clouds represent +/− 1 SE, and the
learning curves join blocks of five utterances. (D) 20 new female subjects participated in
Experiment 2. They experienced baseline production, perceptual tests, and altered feedback
2 days after perceptual training. (E) 10 subjects had their perceptual boundary shifted

towards "head" (red data), and 10 subjects had their boundary shifted towards "had". Alterations in the perceptual boundary were present 2 days after training ($p < 0.05$). Error clouds represent +/− 1 SE, and the learning curves join blocks of five utterances. (F) Produced F1 frequency was normalized to Day 1 baseline measures. After a second session of baseline production two days later, subjects compensated for an increase in F1 by decreasing the frequency of produced F1 (trials 91 to 215). Subjects who had their perceptual boundary shifted towards "head" compensated more ($p < 0.05$) for altered auditory feedback. They also showed greater compensation-related after-effects ($p < 0.01$). Error clouds represent +/− 1 SE.