Talking under conditions of altered auditory feedback: Does adaptation of one vowel generalize to other vowels?

E.J.S. Pile[1], H.R. Dajani[1], D.W. Purcell[2] & K.G. Munhall[1]

[1]QUEEN'S UNIVERSITY, KINGSTON, CANADA;
[2]UNIVERSITY OF WESTERN ONTARIO, LONDON, CANADA
elizabeth@elizabethpile.com, h.dajani@utoronto.ca, purcelld@nca.uwo.ca, kevin.munhall@queensu.ca

ABSTRACT

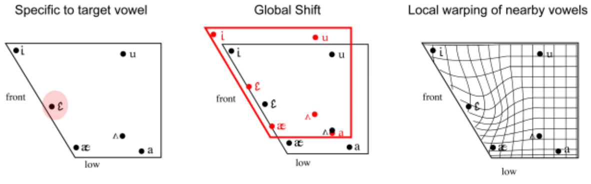
Evidence of perceptual learning has been found in various sensory systems, including the auditory system, but little research has examined the specificity of such learning. In the current study, participants' auditory feedback was altered in real time such that they heard their production of /ɛ/ shifted completely to sound like /æ/. This feedback modification induces a compensatory change in speech production. Following a period of training with this auditory feedback, subjects were tested on the vowels /ɪ/ and /e/ to determine whether learning on one vowel generalized to nearby vowels. All participants produced a reliable compensation to the altered feedback. This compensation did not disappear immediately with the return to normal feedback, indicating that learning had occurred. There was no transfer of this compensation to the other vowels, and production of these nearby vowels also had no effect on the unlearning of the trained vowel. The learning specificity shown here replicates previous findings in the visuomotor and force field learning literatures, and also provides further evidence for the categorical representation of vowels in the vowel space.

Keywords: Sensorimotor Learning, Vowel Production, Auditory Feedback, Perturbation

1. INTRODUCTION

Evidence of sensorimotor learning has been found in several sensory systems, including the visual [1], auditory [2], and kinesthetic systems [3]. Recently, a series of empirical studies have examined learning in speech, which occurs when participants receive altered auditory feedback in real time [4, 5, 6, 7]. These "perturbation" studies change the feedback given to participants such that when the participant says /hɛd/ they hear themselves saying /hæd/. Instead of continuing with the correct production, individuals compensated for the change in feedback by pushing their productions in the direction opposite to that of the perturbation. These compensations persist following the return to normal feedback and thus indicate that short-term learning takes place within the experiment. The current study uses the same perturbation paradigm to investigate how changes in production in response to altered auditory feedback for one vowel generalize to nearby vowels in the vowel space. Participants were trained on the vowel /ɛ/, and sensorimotor learning was induced by perturbing the first (F1) and second (F2) formants of their productions. After this exposure to altered feedback with /ɛ/, participants' productions of two nearby vowels, /ɪ/ and /e/, were tested for any changes from their baselines. Figure 1 depicts three possibilities for the influence of changing /ɛ/ on the rest of the vowel space. As is depicted on the left, the effect could be confined to the trained vowel /ɛ/. The central schematic shows that the entire vowel space could be affected or alternately on the right the region around /ɛ/ could be warped and close vowels could be changed as a function of their proximity to the trained vowel.

Figure 1. Three possible generalization effects of perturbation training on /ɛ/.



In a second test of generalization, we examined how the learning associated with the manipulated vowel /ɛ/ dissipated. We tested whether experience with the manipulated vowel was necessary for recovery of normal formant values to take place. Participants' productions of /ɛ/ were tested following their productions of /ɪ/ or /e/, or after experiencing a period of silence, to determine the influence of other vowels and time on the unlearning of the compensation.

2. METHOD

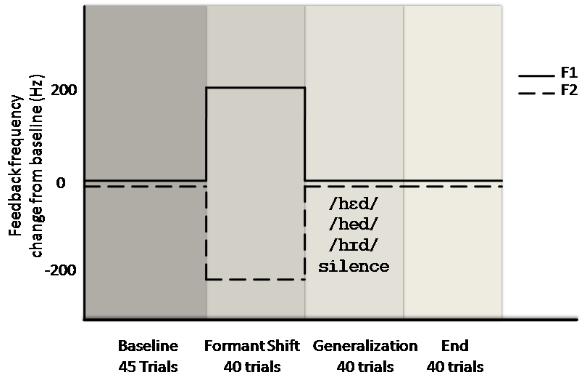
2.1. Participants

Fifty-nine female undergraduate students, 18 to 24 years old (mean 19.8 years) and all native speakers of English, participated in this study. Participants had no known history of speech or hearing impairments, and the majority had hearing thresholds of 20 dBHL or less, for 500, 1000, 2000, and 4000 Hz tones in both ears (two individuals had thresholds that exceeded 20 dBHL at two frequencies in one ear).

2.2. Experimental Conditions

Four between-subjects experimental conditions were employed: Trained, Generalization /e/, Generalization /ɪ/ and Silence (see Fig. 2). All conditions consisted of 165 trials divided into four repeated measures phases: Baseline (45 trials), Formant Shift (40 trials), Generalization (40 trials), and End (40 trials). All four conditions were treated the same for the baseline and formant-shift phases. During Baseline, participants said "head", "hayed", and "hid" 15 times each in random order while receiving unaltered feedback through the headphones. During Formant Shift, participants said "head" while receiving altered feedback, in which both F1 and F2 were shifted such that the feedback approximated the vowel /æ/ for that participant. During the Generalization phase, participants' feedback returned to normal, and differences in the four different conditions were implemented. Participants in the Trained condition continued to say "head", the utterance that was trained during the Formant Shift phase. In the Generalization /e/ condition, another group of participants said "hayed", and in the Generalization /ɪ/ condition, a third group said "hid". In the Silence condition, a final group was prompted to wait in silence for the duration of the Generalization phase. Finally, during the End phase all conditions were once again treated identically, and participants received normal feedback as they said "head".

Figure 2. Schematic representation of experimental design and feedback changes



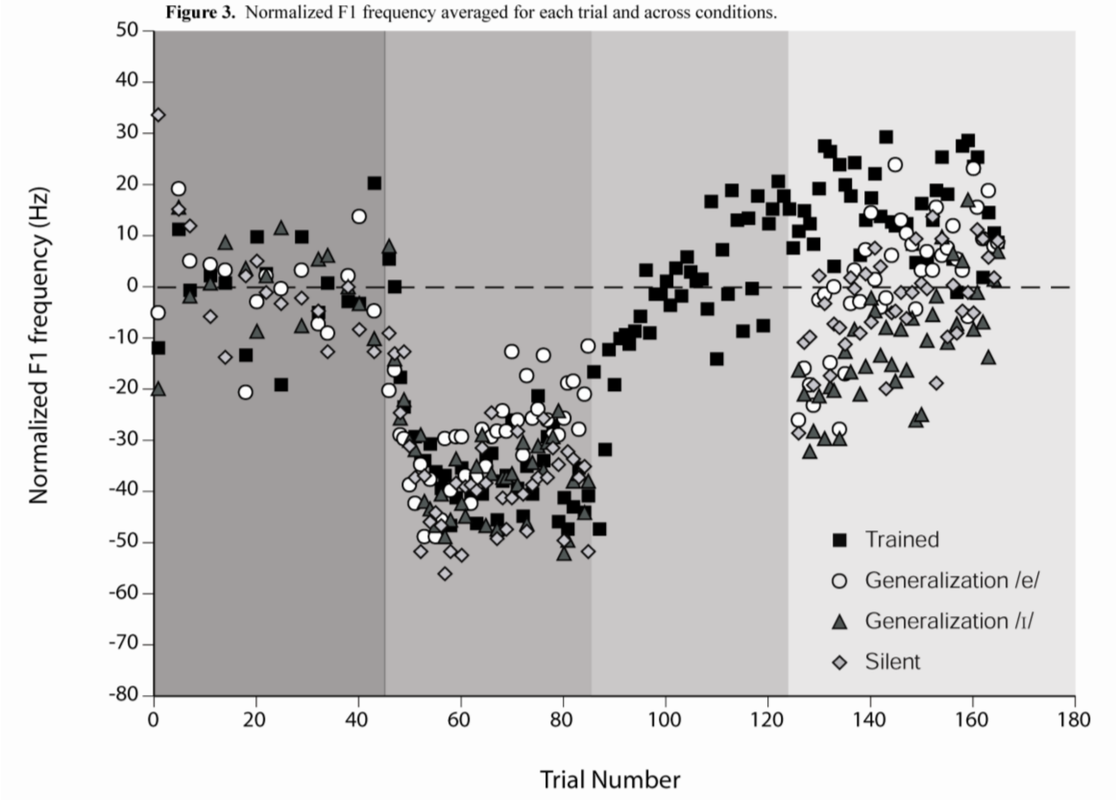
2.3. Experimental Protocol

Participants sat in a sound-proof booth wearing headphones and a microphone, and a warm-up procedure was employed to accustom participants to speaking while hearing feedback of their voice through the headphones, using randomized /hVd/ tokens with the vowel one of /ɪ, ɛ, e, æ, ɑ, u, i/. Then in a vowel-screening test, participants were prompted say a series of seven /hVd/ tokens five times each in random order, extending each token for two seconds. This allowed the computer to determine the best model order for analyzing the formant frequencies of the vowel [8]. From this screening procedure, formant-shift values were calculated for F1 and F2 such that participants' /ɛ/ vowel could be shifted completely to /æ/ during the Formant Shift phase. Upon completion of the screening and warm-up, participants were randomly assigned to one of the four conditions: (all with N = 15 except for Trained N = 14). The interval between word prompts was approximately 1.5 sec and there was no break between the phases.

2.4. Offline Formant Analysis

Each recorded token was segmented manually to determine the beginning and end of the vowel. Formants were then estimated from the segmented tokens 1000 times per second using an iterative Burg algorithm [8], with the model order determined for each participant during screening. Approximately 2% of trials were discarded because formants could not be tracked.

Figure 3. Normalized F1 frequency averaged for each trial and across conditions.



3. RESULTS AND DISCUSSION

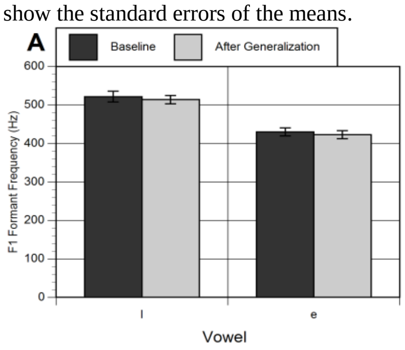
In general, participants were influenced by the modified acoustic feedback and produced compensations that acted so as to reduce the effects of the perturbation. These compensations, however, did not appear to generalize to other vowels, nor did the unlearning of the compensations appear to be influenced by the production of other vowels or the passage of time. Both F1 and F2 data were analyzed but only F1 will be reported here.

3.1. Evidence of Learning and Unlearning

In order to adjust for individual differences in formant values between the groups, the data were normalized with reference to each condition's baseline. A 3 x 4 ANOVA showed that there was a main effect of phase, $F(2, 110) = 32.96$, $p < .001$. Paired samples t-tests using a Bonferroni correction of $\alpha = .008$ confirmed the significance of the compensation effect in response to the altered feedback (see Fig. 3). Across conditions, Baseline F1 productions were significantly different from the asymptote level in the Formant-Shift phase, $t(58) = 5.87$, $p < .001$, but were not significantly different from the asymptote level of the End phase, $t(58) = -1.20$, $p < .24$. Formant-Shift productions were significantly lower than both the productions of /ɛ/ immediately following altered feedback, $t(58) = -4.30$, $p < .001$, and the asymptote level of the End phase, $t(58) = -6.82$, $p < .001$. Finally, productions of /ɛ/ immediately following altered feedback were significantly lower than the asymptote level of the End phase, $t(58) = 5.201$, $p < .001$. The 3 x 4 ANOVA also showed no significant interaction between phase and condition, $F(6, 110) = .63$, $p = .66$, using a Greenhouse-Geisser correction for sphericity.

3.2. Influence of Trained Vowel on Nearby Vowels

Figure 4. Average F1 values for /ɪ/ and /e/ before and after generalization. Error bars show the standard errors of the means.



Learning in response to perturbation on /ɛ/ had no influence on participants' later productions of /ɪ/ and /e/ (see Fig. 4). For each vowel, a repeated-measures ANOVA was conducted with phase (2 levels: Baseline, Generalization) as the within-subjects variable. No main effect of phase was found for either Generalization /e/, $F(1, 14) = 1.96$, $p = .18$, or Generalization /ɪ/, $F(1, 14) = .84$, $p = .37$.

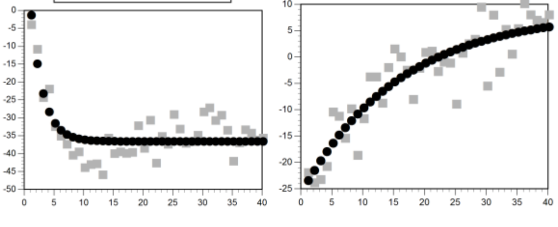
3.3. Influence of Untrained Vowels on Unlearning

For both F1 and F2, neither producing other vowels nor experiencing silence during the Generalization phase had any influence on the unlearning of the trained vowel, /ɛ/. For F1, comparing the first 15 productions of /ɛ/ following altered feedback (productions in the Generalization phase for the trained condition and in the End phase for the other three conditions) revealed no significant differences between the four conditions, $F(3, 55) = .47$, $p = .71$. However, productions of /ɛ/ for the three generalization conditions were significantly different from the productions of /ɛ/ in the trained condition at the beginning of the End phase, $F(1, 57) = 10.43$, $p = .002$, with the formants of productions in the trained condition being higher than the formants of productions in the other three conditions. By the end of the experiment, productions across conditions were no longer different from the Baseline phase, $t(58) = -1.20$, $p = .24$, and did not differ between conditions, $F(1, 57) = 1.382$, $p = .24$.

3.4. Curve Analysis

Figure 5 depicts an exponential curve fitted to the data averaged over all four conditions. The curve shows a steeper slope for the learning (A) phase of the experiment than for the unlearning (B) phase.

Figure 5. (A) Exponential curve for learning in response to altered auditory feedback. (B) Exponential curve for unlearning in response to altered auditory feedback.



When the average data for each condition were fit with exponential curves, the rate parameters of the exponential fits were consistently higher for learning (0.36, 0.86, 0.48 and 0.39 for the Trained, Generalization /e/, Generalization /ɪ/, and Silent conditions respectively) than for unlearning (0.08, 0.09, 0.01 and 0.12 for the Trained, Generalization /e/, Generalization /ɪ/, and Silent conditions respectively).

4. CONCLUSION

Adaptation to the altered auditory feedback was shown in all subject groups. However, this learning had no impact on the neighboring vowels. Similarly, the unlearning of this change was shown to require direct experience with the trained vowel. These findings suggest that sensorimotor learning in speech is specific to the target vowel, as depicted in the first panel in Fig. 1. This is consistent with studies of limb movement that suggest that motor learning under most conditions is quite specific. This fact may have significance for speech motor rehabilitation.

5. REFERENCES

[1] Field, D.F., Shipley, T.F., Cunningham, D.W. 1999. Prism adaptation to dynamic events. *Percept. Psychophys.* 61(1), 161-176.

[2] Lahav, A., Saltzman, E., Schlaug, G. 2007. Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *J. Neuroscience.* 27(2), 308-314.

[3] Tremblay, S., Shiller, D.M., Ostry, D.J. 2003. Somatosensory basis of speech production. *Nature* 423, 866-869.

[4] Houde, J.F., Jordan, M.I. 2002. Sensorimotor adaptation of speech I: Compensation and adaptation. *J. Speech Lang. Hear. R.* 45, 295-310.

[5] Jones, J.A., Munhall, K.G. 2005. Remapping auditory-motor representations in voice production. *Current Biology* 15, 1768-1772.

[6] Perkell, J.S. 2006. Achieving Speech Motor Goals: Feedforward and Feedback Control. *Proceedings of ISSP 2006 7th International Seminar on Speech Production.*

[7] Purcell, D.W., Munhall, K.G. 2006b. Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *J. Acoust. Soc. Am.* 120(2), 966-977.

[8] Orfandidis, S.J. 1988. *Optimum Signal Processing, An introduction*. New York, NY: MacMillan.

ACKNOWLEDGEMENTS

This research was supported by the National Institute of Deafness and Communicative Disorders Grant DC-08092 and the Natural Sciences and Engineering Research Council of Canada.