Increased vowel contrast induced by adaptation to a non-uniform auditory perturbation in speech

Benjamin Parrell & Caroline A. Niziolek**
*University of Wisconsin–Madison, *equal contribution*

Introduction: When auditory feedback is perturbed in a consistent way, speakers learn to adjust their speech to compensate, a process known as sensorimotor adaptation.[1] Typically, feedback perturbation experiments employ a transformation that targets a single vowel, or that affects all vowels in the same way, resulting in a uniform change across the vowel space (e.g., raising the first formant).[2,3] Here, we examine speakers' ability to compensate for a non-uniform perturbation field in which the direction of applied formant shift depended on the vowel being produced. This perturbation field was explicitly designed to affect vowel distinctiveness, causing different vowels to sound more similar to each other. We hypothesized that speakers could adapt to the applied feedback shift by learning to produce greater contrast between vowels, resulting in an expansion of their working vowel space.

Methods: Twenty-five speakers of American English participated in the study. Participants were visually cued to produce 125 repetitions of four monosyllabic English words containing the four corner vowels /i/, /æ/, /ɑ/, and /u/ in a /bVd/ context: bead, bad, bod, and booed. As they spoke, participants heard their own auditory feedback via a modified version of Audapter[4]. After a calibration phase, speakers produced 60 trials with normal feedback ("baseline" phase) from which the centroid of each participant's vowel space was calculated. Speakers were then exposed to a "vowel centralization" feedback perturbation in which the first two formant frequencies (F1 and F2) were shifted towards this centroid, making all vowels sound more like schwa. This vowel centralization was held constant for 320 trials ("hold" phase), after which feedback was returned to normal for 40 trials ("washout" phase). Finally, after a 10-minute delay period, speakers produced an additional 40 trials with normal feedback ("retention" phase) (Figure 1B).

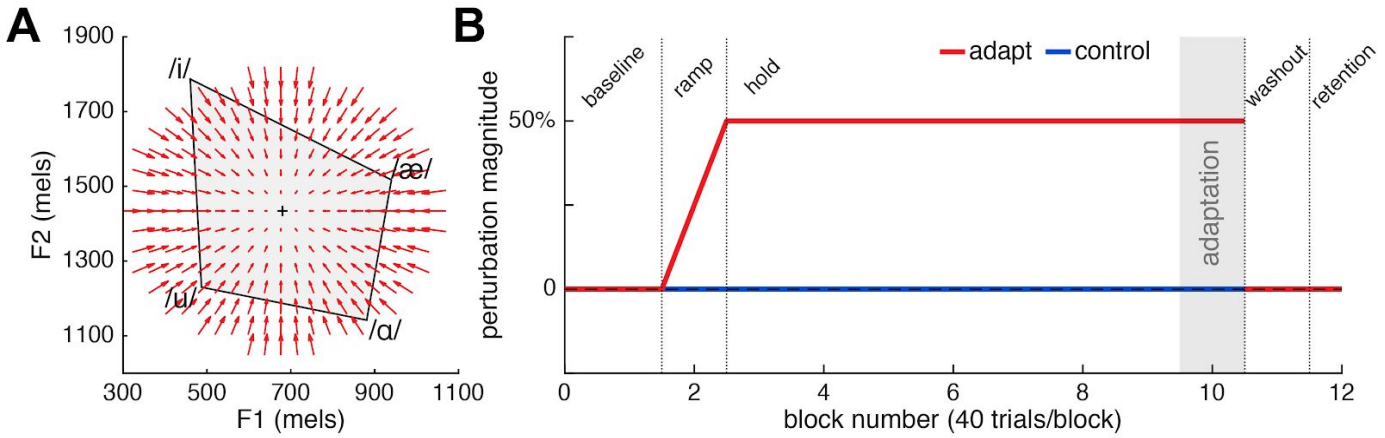

*Figure 1. **A**: Feedback perturbation field for an example participant. **B**: Experiment phases.*

Each participant also completed a control session with an identical procedure but with no alteration to feedback. The sessions took place an average of 1 week apart, and session order was counterbalanced across participants.

F1 and F2 were tracked for each utterance, averaged across the steady-state portion (middle 50%) of the vowel, and converted to the mel scale. To measure adaptation to the vowel centralization perturbation, we calculated how the spacing between produced vowels changed over the course of the experiment. Average

vowel spacing (AVS) was defined as the average of the pairwise distances between the four corner vowels and was compared between the baseline phase, the last 40 trials of the hold phase, and the 40 trials in each of the washout and retention phases.

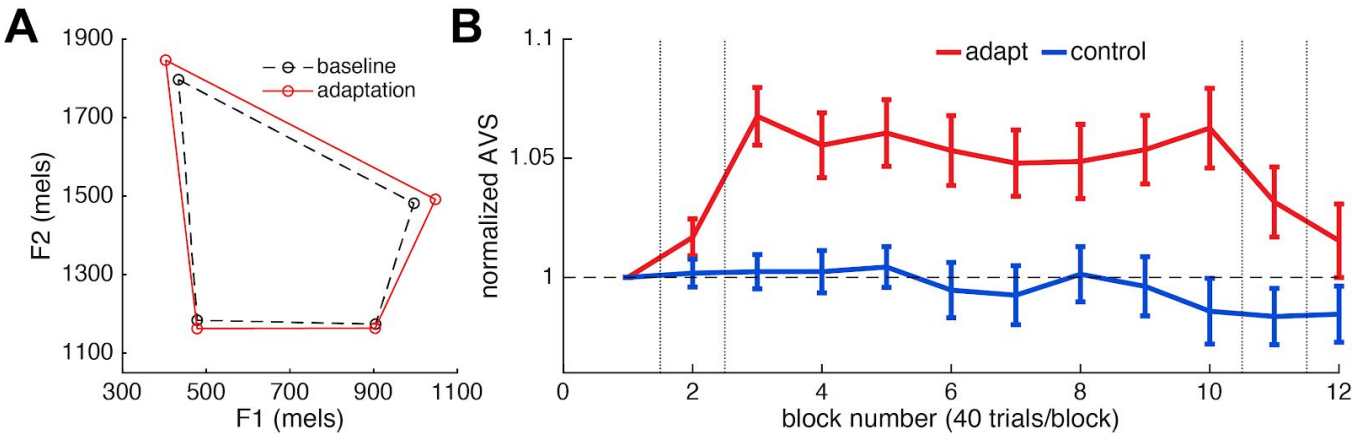

*Figure 2. **A**: Formant values before and after exposure to vowel centralization feedback for the example participant in Fig. 1A. **B**: Group average AVS in adaptation (red) and control (blue) sessions.*

Results: Figure 2A shows the vowel space of an example participant, delineated by the formants of the four corner vowels as produced during baseline (black dashed line) and end of the hold phase ("adaptation", red line). During the centralization perturbation, each of the corner vowels becomes hyperarticulated, moving farther from the center and leading to an increase in the working vowel space area. While not every speaker opposed the perturbation in all four vowels, as a group, speakers increased the global acoustic contrast between vowels as measured by AVS (Figure 2B). This increase in AVS remained significant in the washout phase, after the feedback shift was removed, and persisted even in the retention phase, after a 10-minute silent period. No increase in AVS occurred during the control session.

Discussion: Speakers exposed to a vowel centralization feedback perturbation learned to produce corner vowels with increased contrast to partially overcome the apparent centralization. These findings show that speakers are capable of adapting to non-uniform shifts of vowel formants. Furthermore, they establish the validity of sensorimotor adaptation paradigms to lead to increases in vowel contrast, an outcome that has the potential to enhance intelligibility.

1. Houde, J. F. & Jordan, M. I. Sensorimotor Adaptation in Speech Production. *Science* **279**, 1213–1216 (1998).
2. Mitsuya, T., MacDonald, E. N., Munhall, K. G. & Purcell, D. W. Formant compensation for auditory feedback with English vowels. *J. Acoust. Soc. Am.* **138**, 413–424 (2015).
3. Lametti, D. R., Smith, H. J., Watkins, K. E. & Shiller, D. M. Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Curr. Biol.* **28**, 3106-3113.e2 (2018).
4. Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H. & Perkell, J. S. A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. *Proc. 8th Intl Semin. Speech Prod.* 65–68 (2008).