Toward understanding the limiting factors in speech auditory-motor adaptation: a new look at perceptual targets

Elise LeBovidge[1]
Churan Li[1]
Ludo Max[1,2]
[1]University of Washington
[2]Haskins Laboratories

Rationale and purpose
It is well known that typical speakers monitor their acoustic speech output and gradually adapt, over several trials, to compensate for altered auditory feedback. However, it is still unclear why typical speakers only partially compensate for an experimentally applied auditory perturbation such as a shift of the formant frequencies of the feedback signal. One suggestion has been that adaptation may be limited because exposure to the altered feedback may induce a change in the speaker's intended perceptual targets. Indeed, perceptual adaptation has been reported to occur in parallel with motor adaptation such that the perceptual boundaries between speech sounds shift over the course of an adaptation experiment (e.g., Shiller et al., 2009; Lametti et al., 2014). The goal of this study was to investigate (a) whether the speaker's perceptual target—rather than merely the boundary between targets—shifts during adaptation, (b) whether auditory presentation of the speaker's own baseline production during each trial could prevent this type of perceptual target shift, and (c) whether the amount of auditory-motor adaptation increases with such repeated presentations of the self-generated baseline production.

Methods
The experiment consisted of three main parts: a pre-test, an adaptation task, and a post-test. Participants heard their own auditory feedback and all other stimuli through insert earphones. The goal of the pre-test was to select representative baseline productions for the participant as they repeatedly read three test words (*tech*, *tuck*, *talk*) with unaltered auditory feedback. For each of the test words, we chose the production closest in Euclidean distance in F1/F2 space to the calculated median F1 and F2 values. The subsequent adaptation task involved reading the same individual words out loud, but this time an initial baseline phase with unaltered feedback was followed by a ramp phase during which the formant frequencies of the speaker's auditory feedback were gradually increased up to a maximum of +2.5 semitones. In this task, participants in the *external targets group* (n = 20) heard their own, most typical pre-test production on each trial immediately before producing the word. This played-back stimulus served as a perceptual reminder, or anchor, of the speaker's typical production and, thus, the perceptual target for the test word. Participants in the *no external targets group* (n = 20) did not receive this repeated anchor presentation during any of the trials in the adaptation task.

Each participant also completed a post-adaptation perceptual test. For this perceptual test, the participant manually selected what they believed to be their "best" utterance using a touch-screen monitor that allowed scrolling through—and listening to—various formant-shifted versions of the most typical production from the pre-test. The stimuli were shifted in gradation from -2.5 semitones to +2.5 semitones, including the non-shifted original production. The post-test was conducted immediately after the adaptation task, without interruption (all instructions were

provided prior to the adaptation task to minimize the decay of any potential target shift that may have occurred).

In addition to the two experimental groups described above, three control groups were included for this post-test task. The participants in one post-test control group (n = 20) first completed the same number of active productions of the three test words (i.e., as in the adaptation task) but without any auditory perturbation. Participants in a second control group (n = 20) simply sat in the sound booth for the same amount of time as the adaptation task, but without producing any words before their post-test. Participants in the third control group (n = 20) completed the post-test immediately after the pre-test. This last group was included to determine what individuals would select as their "best" word productions if there had been no auditory perturbation, no intervening speaking, and no additional time between the two tasks.

Results and discussion

Participants typically counteract an upward shift in the formant frequencies of their auditory feedback by gradually lowering the formant frequencies of their subsequent productions. Thus, to quantify adaptation, all utterances were analyzed to compare their F1 and F2 values throughout the duration of the adaptation task. For the production data from the adaptation task, analyses to date indicate that participants who heard their own typical production played back before each trial did not adapt more than participants who completed the same adaptation task without played-back stimuli. For the post-test perceptual task, analyses to date indicate that, on average, participants in both groups tended to select versions of their own productions that had slightly upshifted formant frequencies, but inter-individual variability was very large. Analyses of the perceptual data from the three additional control groups are ongoing, but will allow us to start dissociating the influence on target identification of exposure to manipulated auditory feedback, repeated production of a small set of test words, and habituating to the general sound characteristics of the insert earphones themselves.

Funded by NIH/NIDCD R01DC014510.

References

Lametti, D., Rochet-Capellan, A., Neufeld, E., Shiller, D., and Ostry, D. (2014). Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, 34(31): 10339-10346.

Shiller, D., Sato, M., Gracco, V., and Baum, S. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125(2): 1103-1113.