Published in final edited form as:

Neurosci Lett. 2015 March 30; 591: 25–29. doi:10.1016/j.neulet.2015.02.012.

Feedback delays eliminate auditory-motor learning in speech production

Ludo Max^{1,2} and Derek G. Maffett¹

¹Department of Speech and Hearing Sciences, University of Washington, 1417 N.E. 42nd St., Seattle, WA 98105-6246, USA

²Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511, USA

Abstract

Neurologically healthy individuals use sensory feedback to alter future movements by updating internal models of the effector system and environment. For example, when visual feedback about limb movements or auditory feedback about speech movements is experimentally perturbed, the planning of subsequent movements is adjusted – i.e., sensorimotor adaptation occurs. A separate line of studies has demonstrated that experimentally delaying the sensory consequences of limb movements causes the sensory input to be attributed to external sources rather than to one's own actions. Yet, similar feedback delays have remarkably little effect on visuo-motor adaptation (although the rate of learning varies, the amount of adaptation is only moderately affected with delays of 100-200 ms, and adaptation still occurs even with a delay as long as 5000 ms). Thus, limb motor learning remains largely intact even in conditions where error assignment favors external factors. Here, we show a fundamentally different result for sensorimotor control of speech articulation: auditory-motor adaptation to formant-shifted feedback is completely eliminated with delays of 100 ms or more. Thus, for speech motor learning, real-time auditory feedback is critical. This novel finding informs theoretical models of human motor control in general and speech motor control in particular, and it has direct implications for the application of motor learning principles in the habilitation and rehabilitation of individuals with various sensorimotor speech disorders.

Keywords

motor learning; sensorimotor adaptation; feedback; speech production; motor control

© 2015 Published by Elsevier Ireland Ltd.

Correspondence should be addressed to: Ludo Max, Ph.D., Department of Speech and Hearing Sciences, University of Washington, 1417 N.E. 42nd St., Seattle, WA 98105-6246, USA. LudoMax@uw.edu. Phone: +1 206 543 2674..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Experimentally perturbing visual feedback associated with limb movements [1,2] or auditory feedback associated with speech movements [3-5] causes subjects to gradually adjust the planning of subsequent movements (i.e., sensorimotor adaptation). Conceptually, the involved sensorimotor controller is typically modeled as a hybrid feedforward-feedback system in which (a) the brain predicts sensory consequences of planned movements, and (b) a mismatch between predicted and actual consequences drives the updating of internal models of the effector system and environment [6-12].

A potential problem arises in the presence of feedback delays, because it is also known that perceiving the consequences of one's own actions with a delay results in the sensory information being processed as if it were externally generated rather than self-generated [13]. Hence, feedback delays could be expected to prevent adaptation (due to external credit assignment) and direct sensory evidence of incorrect planning (internal credit assignment) might be a prerequisite for sensorimotor learning. Interestingly, motor learning for the upper limb seems rather robust to such feedback delays. Although the initial *rate* of adaptation (and thus the coefficients of a learning function) may be affected more substantially, the *extent* of visuo-motor adaptation and initial after-effect (relative to a control condition with no additional delay beyond that inherently present in the feedback display system, which by itself can be as large as 60 ms [14]) may be either essentially unchanged [15] or reduced by ~35% [16] with a feedback delay of 100 ms, reduced by ~20% with a delay of 200 ms [14,16], and reduced by ~60% with a delay of 5000 ms [16] (note that in all these studies visual feedback was blocked during the movement, with only feedback about final hand position provided with or without delay after completion of the movement; thus the true delays relative to movement execution were even larger than reported).

Hypothesizing that the robustness of sensory-motor learning to delayed feedback transmission might depend on the involved neural substrates as well as the specific effector and sensory receptor systems, we investigated the effects of delays in auditory feedback on auditory-motor adaptation in speech production. Eight subjects participated in a paradigm in which the formant frequencies in the subject's speech were incrementally ramped up to a 2.5 semitones (ST) shift, and this altered signal was provided as auditory feedback with delay intervals ranging from 0–500 ms. We predicted that speech auditory-motor adaptation to the shifted formants would become increasingly more limited with longer feedback delays.

Material and Methods

Auditory-motor adaptation experiment

Eight adult subjects (4 male, 4 female, mean age 23.8 years) with no speech or hearing problems participated in the study after providing informed consent. All experimental procedures were approved by the University of Washington's Institutional Review Board.

Subjects repeatedly produced the consonant-vowel-consonant (CVC) words “talk,” “tuck,” and “tech” (180 trials per condition, with the order randomized within each epoch of 3 words) while hearing their own speech through insert earphones (ER-3A, Etymotic

Research) after it was routed through a digital vocal processor (VoiceOne, TC Helicon) (Figure 1A). The inherent delay in this auditory feedback set-up was 10 ms. Subjects were provided with visual feedback to aid in maintaining a speech output intensity between 72 and 78 dB SPL measured 15 cm from the mouth. To allow sufficiently loud masking of bone-conducted feedback, the entire microphone-to-earphones audio system was calibrated such that speech with an intensity of 75 dB SPL at a distance of 15 cm from the mouth also resulted in an intensity of 75 dB SPL in the insert earphones. Pink noise was mixed with this feedback signal at a fixed intensity of 68 dB SPL to mask the bone-conducted signal.

Each subject completed four conditions in which all formants (vowel-specific resonance frequencies) in the auditory feedback were incrementally shifted up and then maintained at +2.5 semitones as shown in Figure 1B. The conditions differed with respect to the length of an additional delay (besides that inherent in the overall system) that was implemented by the vocal processor: 0, 100, 250, or 500 ms. Order of the conditions was counterbalanced across subjects. Subjects' speech was digitally recorded for offline analysis with a combination of custom routines in Praat [17] and Matlab (The Mathworks, Natick, MA) that, respectively, selected vowel segments and implemented a linear predictive coding (LPC) technique [18,19]. For each production, the frequencies of the first two formants (F1, F2) were extracted at time points 40% and 50% into the vowel, and each formant's average value across the two measurement points was used as the measured formant frequency. Within each condition, the F1 and F2 measurements in Hertz were then normalized for individual speaker vocal tract differences by converting to semitones units expressed relative to the speaker's average F1 and average F2 formant frequencies across the same word's 10 trials produced in the baseline phase with unaltered auditory feedback.

Auditory-motor adaptation would be reflected in significant changes in subjects' produced formant frequencies relative to baseline. Therefore, one-tailed paired *t*-tests were used to compare the F1 and F2 data in blocks of 5 successive trials (i.e., data averaged within-subjects across the 5 trials) from the ramp, full shift, and after-effects phases with their average values in the second block of 5 successive trials in the baseline phase of the same condition (to account for multiple comparisons within each condition, α was set at $.05/10 = 0.005$).

Control experiment

A control experiment was conducted to examine how much unaltered auditory feedback subjects received from the bone-conducted signal that had been masked with 68 dB SPL pink noise in the main experiment. Six different subjects (3 male, 3 female) completed 120 trials with the same test words and equipment as in the formant-shift experiment. First, subjects read the presented word out loud while 68 dB SPL pink noise was played through the insert earphones and speech feedback was perceived only through bone conduction (i.e., no speech signal was presented through the earphones). This corresponds to the same situation that in the formant-shift experiment occurred briefly (for 100, 250, or 500 ms) before the delayed air-conducted signal was presented (i.e., the time period during which bone conduction may have been helpful for subjects to evaluate the accuracy of their production). Subsequently, subjects heard the exact same production played back as an air-

conducted signal through the earphones, again mixed with 68 dB SPL pink noise. Across trials, the intensity of the played-back speech signal was randomly varied at levels from -5 to -30 dB SPL relative to the level at which the air-conducted signal would have been presented during the formant-shift experiment (according to the input-output calibration described above). At the end of each trial, subjects judged whether the played-back air-conducted signal was louder than the bone-conducted signal heard when they spoke the word (the latter corresponding to the bone conduction level heard during the main experiment). To estimate the level at which bone conduction was perceived, a maximum likelihood procedure was used to fit a logistic psychometric function to the data.

Results

Auditory-motor adaptation experiment

After separating the trials for each of three words (“talk,” “tuck,” “tech”) spoken in each condition (0, 100, 250, 500 ms delay), group data consisting of subjects’ produced first (F1) and second (F2) formant frequencies were organized into 12 blocks of 5 trials each. Figure 2 shows all results. Trial blocks in which F1 or F2 were statistically significantly different from baseline (one-tailed t tests with Bonferroni corrections) are marked with shaded backgrounds that are color-matched with the corresponding condition’s time series data.

Data from the 0 ms delay condition replicate our previous report [3] that (a) non-delayed auditory feedback with a frequency shift applied to all formants leads to auditory-motor adaptation, (b) adaptation to this type of shift is largest for /ɔ/ and /ʌ/ and smallest for /ε/, (c) adaptation is more extensive for F2 than for F1, and (d) after-effects occur after non-manipulated feedback is restored. As an important novel finding, directly addressing the hypothesis of the present study, it is clear from Figure 2 and from the represented statistical results that auditory-motor adaptation is not merely diminished but, instead, entirely abolished in *all* conditions with delayed auditory feedback.

Although a single block of trials (block 4) early in the ramp phase of the 250 ms condition for the target word “talk” reached statistical significance for F2, this isolated result clearly was not part of a pattern of adaptation, and a formant deviation from baseline did not occur in any other trial blocks during either the ramp or the full-shift phases of this condition. Moreover, none of the other delay conditions contained any trial blocks at all in which either F1 or F2 in the subjects’ productions were ever significantly different from baseline.

Control experiment

Figure 3 shows group average results for the subjects who judged the intensity of their bone-conducted speech feedback when masked with 68 dB SPL pink noise (presented through insert earphones) in comparison with the air-conducted signal associated with an immediately following played-back version of the same production that was presented at varying intensities and also masked with 68 dB SPL pink noise. The data indicate that bone-conducted speech feedback was perceived equally loud as an air-conducted signal 17.3 dB SPL lower than that used for AC feedback during the formant-shift experiment. Consequently, for an average production (75 dB SPL) in the formant-shift experiment, the

bone conduction component was perceived at a signal-to-noise ratio (SNR) of -10.3 dB (68 dB SPL noise, 57.7 dB SPL bone-conducted speech) whereas the air conduction component was perceived at an SNR of $+7$ dB (68 dB SPL noise, 75 dB SPL air-conducted speech). Given this poor SNR ratio for the bone conduction signal, it can be ruled out that subjects were able to extract very precise spectral information from this channel to evaluate the accuracy of their productions (by shifting all formants in the same direction, our protocol implemented subtle deviations that do not result in the perception of phonemic errors).

Discussion and Conclusions

Whereas our hypothesis predicted increasingly less adaptation with increasingly longer delays in the auditory feedback signal, the obtained group-based results support an even stronger version of the hypothesis that real-time feedback is critical for speech motor learning: averaged across subjects, auditory-motor adaptation was completely eliminated in *all* delay conditions, even when the delay was as short as 100 ms. Given that studies of reaching have shown only modest reductions in visuo-motor adaptation with similar and much longer delays, our findings suggest that the neural systems underlying speech articulation—which is characterized by faster movements of shorter duration [20]—is more likely to increase the weighting of proprioceptive feedback [19] when exteroceptive feedback is affected by delayed transmission.

The potential alternative explanation that the presence of delays did not result in auditory feedback being ignored in favor of proprioceptive feedback, but that only the delayed air-conducted signal itself was ignored was ruled out in a control experiment (in addition, if this alternative hypothesis was correct, it would be unclear why the very different spectral characteristics of hearing only the bone-conducted signal instead of the typical combination of air conduction *and* bone conduction would not lead to at least some amount of adaptation). Thus, the presence of minimal (masked) but unaltered bone conduction feedback prior to onset of the delayed air conduction signal was not the reason for the lack of auditory-motor adaptation in the various delay conditions.

In conclusion, these new results demonstrate that the neural systems underlying sensorimotor control of speech articulation give little or no weight to delayed auditory feedback for updating internal models of the vocal tract during auditory-motor learning. In our previous work, we have shown that both somatosensory *and* auditory feedback from previous trials are considered in the planning of speech movements: although auditory feedback is “dominant,” somatosensory feedback also drives motor learning if the provided information does not conflict with the auditory modality [19]. The present findings are fully compatible with such a speech motor controller that flexibly adjusts the relative weights assigned to auditory and somatosensory error signals. Here, when auditory information was available (no-delay condition), the controller gradually adjusted movement planning across trials based on errors signaled by this real-time auditory input. When auditory feedback was available only after most, or all, of the target vowel had already been produced, the controller appears to have relied on somatosensory input (which signaled accurate productions with no error), and, consequently, no articulatory adjustments were made across trials.

The high velocities and short durations of articulatory speech movements, together with the fact that even the production of a simple CVC word requires sequential actions within and across articulators, may be the critical reason why even short delays in auditory feedback have such a detrimental effect on speech auditory-motor learning whereas comparable visual feedback delays have limited effects on visuo-motor learning in the case of slower and discrete reaching movements [14,15,16]. In our experimental task, vowel durations across all conditions were on average only ~185 ms in duration (and note that this acoustic vowel duration spans across both the oral opening and closing movements, with kinematic measures of each component separately often indicating durations as short as ~75–180 ms [21]). In addition, all three target words required an upward velum movement to close the nasal cavity, an upward and then downward tongue tip movement to create and release an alveolar obstruction, a tongue body movement to reach the correct amount of oral opening, and an upward and then downward tongue dorsum movement to create and release a velar obstruction. For a neuromuscular system whose relevant sensory consequences are of very short duration and whose actions necessarily involve sequences of coordinated individual movements, it may be highly inefficient to rely on a control policy that would update the planning of future movements also on the basis of delayed auditory information, in particular given the temporal mismatch between auditory input and motor action and, thus, also between auditory and somatosensory input.

Acknowledgment

Supported by grants R01DC007603 and P30DC004661 from the National Institute on Deafness and Other Communication Disorders. The authors thank Ayoub Daliri for assisting with data analysis.

References

1. Ghahramani Z, Wolpert DM. Modular decomposition in visuomotor learning. *Nature*. 1997; 386:392–395. [PubMed: 9121554]
2. Flanagan JR, Nakano E, Imamizu H, Osu R, Yoshioka T, Kawato M. Composition and decomposition of internal models in motor learning under altered kinematic and dynamic environments. *J. Neurosci*. 1999; 19:RC34. [PubMed: 10516336]
3. Max, L.; Wallace, ME.; Vincent, I. Sensorimotor adaptation to auditory perturbations during speech: acoustic and kinematic experiments. In: Solé, MJ.; Recasens, D.; Romero, J., editors. *Proc. Internat. Congr. Phonet. Sci. Futurgraphic*; Barcelona (Spain): 2003. p. 1053-1056.
4. Houde JF, Jordan MI. Sensorimotor adaptation in speech production. *Science*. 1998; 279:1213–1216. [PubMed: 9469813]
5. Purcell DW, Munhall KG. Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J. Acoust. Soc. Am*. 2006; 120:966–977. [PubMed: 16938984]
6. Desmurget M, Grafton S. Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci*. 2000; 4:423–431. [PubMed: 11058820]
7. Wolpert DM, Kawato M. Multiple paired forward and inverse models for motor control. *Neural Netw*. 1998; 11:1317–1329. [PubMed: 12662752]
8. Wolpert DM, Miall RC, Kawato M. Internal models in the cerebellum. *Trends Cogn. Sci*. 1998; 2:338–347. [PubMed: 21227230]
9. Shadmehr R, Mussa-Ivaldi FA. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci*. 1994; 14:3208–3224. [PubMed: 8182467]
10. Krakauer JW, Mazzoni P. Human sensorimotor learning: adaptation, skill, and beyond. *Curr. Opin. Neurobiol*. 2011; 21:636–644. [PubMed: 21764294]

11. Tseng YW, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ. Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *J. Neurophysiol.* 2007; 98:54–62. [PubMed: 17507504]
12. Mazzoni P, Krakauer JW. An implicit plan overrides an explicit strategy during visuomotor adaptation. *J. Neurosci.* 2006; 26:3642–3645. [PubMed: 16597717]
13. Blakemore SJ, Wolpert D, Frith C. Why can't you tickle yourself? *Neuroreport.* 2000; 11:R11–R16. [PubMed: 10943682]
14. Honda T, Hirashima M, Nozaki D. Adaptation to visual feedback delay influences visuomotor learning. *PLoS One.* 2012; 7 doi:10.1371/journal.pone.0037900.
15. Tanaka H, Homma K, Imamizu H. Physical delay but not subjective delay determines learning rate in prism adaptation. *Exp. Brain Res.* 2011; 208:257–268. [PubMed: 21076819]
16. Kitazawa S, Kohno T, Uka T. Effects of delayed visual information on the rate and amount of prism adaptation in the human. *J. Neurosci.* 1995; 15:7644–7652. [PubMed: 7472515]
17. Boersma P. Praat: a system for doing phonetics by computer. *Glott Internat.* 2001; 5:341–345.
18. Vallabha GK, Tuller B. Systematic errors in the formant analysis of steady-state vowels. *Speech Commun.* 2002; 38:141–160.
19. Feng Y, Gracco VL, Max L. Integration of auditory and somatosensory error signals in the neural control of speech movements. *J. Neurophysiol.* 2011; 106:667–679. [PubMed: 21562187]
20. Grimme B, Fuchs S, Perrier P, Schöner G. Limb versus speech motor control: a conceptual review. *Motor Control.* 2011; 15:5–33. [PubMed: 21339512]
21. Max L, Caruso AJ, Gracco VL. Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering individuals. *J. Speech Lang. Hear. Res.* 2003; 46:215–232. (2003). [PubMed: 12647900]

Highlights

- We investigated adaptation to altered auditory feedback during speech articulation.
- Speech audio-motor adaptation is completely eliminated with delays of 100 ms or more.
- Thus, for speech sensorimotor learning, real-time auditory feedback is critical.

Figure 1.

A. Subjects' speech was routed through a vocal processor and played back to the subject through insert earphones. The processor shifted the frequencies of all formants up during the ramp and full shift phases of each condition (0, 100, 250, 500 ms delay). B. Time course for each condition. Each condition included 180 trials: 60 sets of the words "talk," "tuck," and "tech" with the order of these words randomized within each set. Baseline: 10 sets with no formant shift. Ramp shift: 20 sets with formants incrementally shifted from 0 to +2.5 semitones. Full shift: 20 sets with + 2.5 semitones formant shift. After-effects: 10 sets with no formant shift.

Figure 2.

Group average data (with error bars indicating standard errors of the mean) for auditory-motor adaptation in the first (F1, left) and second formant (F2, middle) and in 2D vowel space (right). Rows contain separate results for “talk,” “tuck,” and “tech” trials. Data are shown in semitones (ST) relative to baseline. Shaded areas indicate blocks of 5 trials that were statistically significantly ($p < 0.005$) different from baseline for the color-matched condition (there were no blocks of trials in which more than one condition reached statistical significance). In vowel space, the base of each arrow corresponds to the group average F1 and F2 during baseline; the tip of each arrow corresponds to the group average F1 and F2 at the end of full shift. Auditory-motor adaptation occurred only in the 0 ms delay condition (red lines and arrows).

Figure 3.

Group average data (with error bars indicating standard errors of the mean) from a control experiment in which subjects judged whether the masked (68 dB SPL pink noise) air-conducted play-back of a self-produced word sounded louder than the masked (68 dB SPL pink noise) bone-conducted feedback associated with prior production of that word. The level of 50% “yes” answers was achieved when the air conduction play-back signal for a trial produced at a given intensity was presented 17.3 dB SPL lower than the level at which air conduction feedback in the formant-shift experiment would have been presented for a production of that same intensity (e.g., for a trial produced at 75 dB SPL, air-conducted feedback would have been presented at 75 dB SPL whereas bone-conducted feedback would have been perceived as equivalent to ~57.7 dB SPL). This result demonstrates that, due to its poor SNR, bone-conducted feedback was unlikely to provide subjects’ with relevant spectral information about the accuracy of their productions in the formant-shift experiment.