Neuroscience Letters 591 (2015) 25–29

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research article

Feedback delays eliminate auditory-motor learning in speech production

Ludo Max[a,b,*], Derek G. Maffett[a]

[a] Department of Speech and Hearing Sciences, University of Washington, 1417 N.E. 42nd St. Seattle, WA 98105-6246, USA
[b] Haskins Laboratories, 300 George St. Suite 900, New Haven, CT 06511, USA

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.

ARTICLE INFO

Article history:
Received 2 August 2014
Received in revised form 18 January 2015
Accepted 6 February 2015
Available online 9 February 2015

Keywords:
Motor learning
Sensorimotor adaptation
Feedback
Speech production
Motor control

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.

1. 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

---

* Corresponding author at: Department of Speech and Hearing Sciences, University of Washington, 1417 N.E. 42nd St. Seattle, WA 98105-6246, USA.
Tel.: +1 206 543 2674.
E-mail address: LudoMax@uw.edu (L. Max).

http://dx.doi.org/10.1016/j.neulet.2015.02.012

(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.

2. Material and methods

2.1. 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) (Fig. 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 Fig. 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.

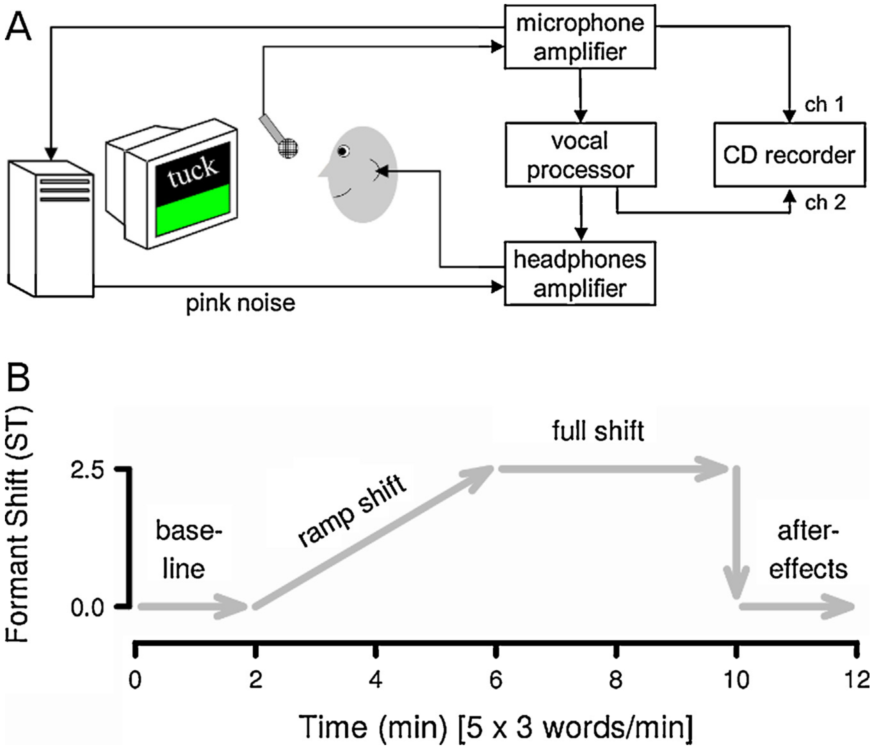

Fig. 1. (A) Subjects' speech was routed through a vocal processor and played back to the subject through insert earphones. The processor shifted the frequencies of all formants up during the ramp and full shift phases of each condition (0, 100, 250, 500 ms delay). (B) Time course for each condition. 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.