2010

Multiple Instances of Vocal Sensorimotor Adaptation to Frequency-Altered Feedback within a Single Experimental Session

Colin S. Hawco
Wilfrid Laurier University

Jeffery A. Jones
Wilfrid Laurier University, jjones@wlu.ca

Multiple instances of vocal sensorimotor adaptation to frequency-altered feedback within a single experimental session

Colin S. Hawco
Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada
hawc2550@wlu.ca

Jeffery A. Jones[a]
Department of Psychology and Centre for Cognitive Neuroscience, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada
jjones@wlu.ca

Abstract: Vocal sensory-motor adaptation is typically studied by introducing a prolonged change in auditory feedback. While it may be preferable to perform multiple blocks of adaptation within a single experiment, it is possible that a carry-over effect from previous blocks of adaptation may affect the results of subsequent blocks. Speakers were asked to vocalize an /a/ sound and match a target note during ten adaptation blocks. Each block represented a unique combination of target note and shift direction. The adaptation response was found to be similar for all blocks, indicating that there were no carry-over effects from previous blocks of adaptation.
© 2010 Acoustical Society of America

1. Introduction

Understanding the contribution of auditory feedback to speech production is important for theories of speech motor control. The role of auditory feedback during speech production has been investigated by altering feedback during ongoing vocalizations. When a random feedback alteration is introduced during a vocalization, there is an online compensation response in the direction opposite the feedback alteration (Larson, 1998; Burnett *et al.*, 1998; Natke and Kalveram, 2001). For example, Natke and Kalveram (2001) had participants vocalize a nonsense word and shifted the fundamental frequency (F0, vocal pitch) of auditory feedback down on 20% of trials. They found an online compensation response in the direction opposite to the F0 shift. This online compensation response uses auditory feedback to correct vocal output and maintain vocal stability, and is delayed by approximately 100–150 ms after a feedback alteration is introduced (Burnett *et al.*, 1998).

When a feedback alteration is introduced and left in place for a prolonged period, sensory-motor adaptation occurs as the system adjusts to compensate for this novel feedback environment. Houde and Jordan (1998, 2002) examined sensory-motor representations for formants by shifting F1 and F2. Participants compensated for the feedback alterations by modifying their formant production in the direction opposite the shift, with these modifications persisting when auditory feedback was removed. Other studies examining vocal control of formants have found similar after-effects when a feedback alteration is removed (Purcell and Munhall, 2006; Villacorta *et al.*, 2007; Tourville *et al.*, 2008). Jones and Munhall (2000, 2002) also noted prominent after-effects after participants heard their F0 shifted over many trials.

---

[a]Author to whom correspondence should be addressed.

These after-effects are evidence of sensory-motor adaptation; because of the sensory-motor remapping during adaptation, the motor system must re-adjust to the original, pre-shift sensory-motor mapping when the feedback alteration is removed.

The results of studies on auditory feedback indicate that feedback is important for online correction during vocalization (Burnett et al., 1998; Natke and Kalveram, 2001) and the maintenance of stored motor commands for vocal production (Guenther, 2006; Jones and Munhall, 2000, 2002; Purcell and Munhall, 2006). Auditory feedback therefore has a dual role in vocal production: to stabilize vocalizations as they are occurring and to modify motor plans for future vocalizations to accommodate for novel feedback environments.

Many previous studies examining sensory-motor adaptation have used a small number of post-shift "test" trials to determine if adaptation had occurred (Houde and Jordan 1998, 2002; Jones and Munhall, 2000, 2002). In studies examining the control of F0, the F0 value at utterance onset has more recently been used as a measure of vocal sensory-motor adaptation (Keough and Jones, 2009; Hawco and Jones, 2009). This measure allows for an online analysis of the time course of adaptation without the need to examine after-effects. Since feedback-based F0 control is typically delayed by 100–150 ms (Burnett et al., 1998), an examination of the F0 data at utterance onset shows changes in the initiation of vocalization, which are not affected by online auditory feedback control. Although it allows for an examination of the course of adaptation, this technique suffers the same lack of power as experiments utilizing post-shift test trials, as there is only a single instance of the critical trials in the experiment (the first shift trials, first post-shift test trials, etc.).

Although several studies have examined sensory-motor adaptation in the vocal system, there are still many questions to be investigated. Some studies have suggested that there may be differences between singers, who are highly trained in vocal control, and non-singers when they adapt to changes in F0 (Jones and Keough, 2008; Keough and Jones, 2009). There are several other special populations who are of interest in studies of sensory-motor adaptation, including Parkinson patients, amusics, schizophrenics (who may have a disruption in the signals sent to auditory cortex to identify speech as self-generated; Ford and Mathalon, 2004), and children at various stages of development. In addition, the neurological mechanisms involved in feedback control of vocalization, and how these mechanisms change during adaptation, are not well understood.

Most studies of sensory-motor adaptation during vocalization involve a single block of trials. With only a single set of critical trials, there is a limit to the statistical power of the analysis, as only a small number of trials for each participant can be used to test for adaptation. Using multiple blocks of adaptation trials would overcome this issue, but it is possible that carry-over effects may occur during repeated blocks of adaptation, resulting in changes to the sensory-motor adaptation observed in subsequent blocks. While many questions regarding sensory-motor adaptation during vocalizations could be addressed using a between-subject design, it may be preferable to be able to perform multiple adaptation blocks within a single participant. This not only reduces the number of participants that would be needed if a between-subject design were used (a particularly important issue when dealing with special populations) but it may also be critical to examining vocal sensory-motor adaptation using neuroimaging techniques [e.g., event-related potentials (ERP), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI)]. However, if there are carry-over effects across multiple blocks of adaptation (for example, if the motor system becomes faster to adapt after repeated exposure to adaptation), it may be difficult to examine experimental differences between blocks as they would be confounded with these carry-over effects.

The present study demonstrates the feasibility of using multiple blocks of adaptation trials in a within-subject frequency-altered feedback design. In order to determine if we could repeat blocks of adaptation without carry-over effects, we attempted to minimize possible carry-over effects in two ways. First, in each block, feedback-shifted trials were presented in the middle of the block, with unshifted trials present at the beginning and end of each block. Having no feedback alterations at the end of each block should allow a "de-adaptation" response, where the participants re-adjust to unaltered feedback prior to the initiation of the next block. We also

used different target notes for vocalization in each block and shifted feedback either up or down in frequency. If the adaptation response does not show a change over blocks, this would suggest that no carry-over effects from previous exposure to adaptation have occurred, and future studies can use blocked adaptation designs without confounds from repeating blocks of adaptation. Data were collected as part of a larger ERP study that will not be presented here, as the ERP findings are not relevant to the current discussion.

2. Methods

2.1 Participants

Data from 14 participants (ages 18–21, 7 males). All participants were undergraduate students at Wilfrid Laurier University participating for course credit and reported no formal training as singers. Informed consent was obtained from all participants.

2.2 Apparatus and procedure

Participants were seated in a comfortable chair in a closed room. They were given a headset with attached boom microphone (Sennheiser HMD 280-13). Vocalizations were sent from the microphone to a digital mixer (828 mkII, MOTU), which passed the voice signal to a digital signal processor (VoiceOne, TC-Helicon). The participant's auditory feedback was mixed with multispeaker babble (20 speakers simultaneously reading different passages; Auditec, St. Louis, MO) played at 90 dB sound pressure level (SPL) and returned to the participant as auditory feedback. The multispeaker babble served to mask air and bone-conducted feedback. The unaltered voice signal was digitally recorded (H4, Zoom) at a sampling rate of 44.1 kHz. Participant's vocalizations were amplified such that a 75 dB SPL production (measured approximately 5 cm from the mouth) was played at approximately 90 dB SPL. Participants vocalized above 75 dB SPL (resulting in feedback of over 90 dB SPL), and participants indicated they could clearly hear their auditory feedback over the multi-speaker babble.

At the beginning of each trial, an auditory cue was played at approximately 92 dB SPL for 1 s. The cue was a male or female voice producing the vowel sound /a/ at a specific target frequency. Target notes for female participants were G4 (392 Hz), E4 (329.63 Hz), D4 (293.66 Hz), B3 (246.94 Hz), and A3 (220 Hz). The target notes for males were G3 (192 Hz), E3 (164.83 Hz), D3 (146.83 Hz), B2 (123.47 Hz), and A2 (110 Hz). Cues were recorded from trained singers who were asked to match a specific pitch. Their productions were processed using the speech modification algorithm speech transformation and representation using the adaptive interpolation of weighted spectrum (Kawahara *et al.*, 1999) to ensure that the mean F0 of the cue was equal to the desired target. Participants were instructed to wait until the cue had ended, and to then produce an /a/ sound, matching the pitch of the cue, until a tone sounded (2 s after utterance onset), resulting in approximately 2 s of vocalization per trial.

The experiment was divided into ten blocks of 42 trials. Each block had a specific target note which was played throughout that block. A trial consisted of an auditory cue (of the target note for that block) and a vocalization. Each block contained 9 or 12 trials with normal, unshifted auditory feedback (pre-shift baseline trials), 18 or 21 trials in which feedback was shifted up or down 100 cents for the entire vocalization (shift trials), followed by 9 or 12 trials of normal, unshifted feedback (post-shift test trials). The different number of trials in each block served to make the onset of shifting in each block less predictable. At the end of each block, there was a break, during which participants could rest and drink water. A schematic diagram of the design is presented in Fig. 1, showing the target note and shift direction of each block.

Voice data were segmented into trials, and F0 was calculated for the first 1500 ms of vocalization using the autocorrelation algorithm in PRAAT (Boersma, 2001), with an F0 value calculated for every 5 ms of voice data. Data were converted into cents using the formula $100^{*}(39.86^{*} \log 10(F0/\text{target}))$, where "F0" was the F0 produced by the participant, and "target" was the target note. In order to assess adaptation, the median F0 (in cents) was calculated for the initial 50 ms of vocalization. We chose to examine the first 50 ms of production because it is not
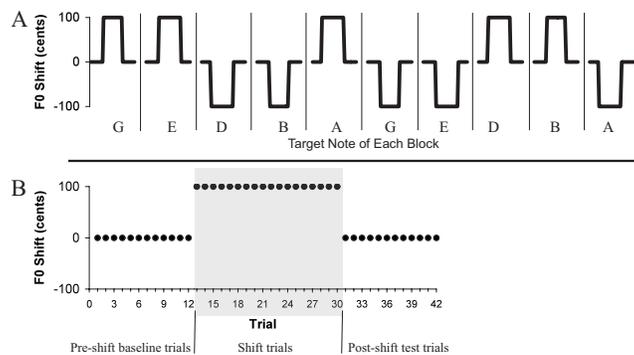
Fig. 1. (A) Schematic diagram of the experimental design, showing the F0 shift in each block (with blocks separated by vertical lines). The target note of each block is shown and was the same for all trials in that block. Target notes for female participants were G4 (392 Hz), E4 (329.63 Hz), D4 (293.66 Hz), B3 (246.94 Hz), and A3 (220 Hz). The target notes for males were G3 (192 Hz), E3 (164.83 Hz), D3 (146.83 Hz), B2 (123.47 Hz), and A2 (110 Hz). Although each target note was presented in two separate blocks, it was shifted in a different direction on each presentation. (B) Detailed view of the first block of trials. Each block consisted of 42 trials, with 9 or 12 pre-shift baseline trials, 18 or 21 shift trials, and 9 or 12 post-shift test trials. Shift trials for the presented block are highlighted in gray. The F0 shift was maintained for the entirety of all shift trials, and there was no F0 shift in any of the pre-shift baseline or post-shift test trials.

affected by volitional feedback control (Burnett *et al.*, 1998). We also calculated the median F0 for the first 1500 ms of vocalization to determine accuracy for hitting the target note.

3. Results

To test for differences in accuracy for hitting the target notes, the median F0 for the first 1500 ms of vocalization for baseline trials 7–9 (the last three baseline trials which were present in all blocks) was analyzed. A 2×5 (repetition by target note) repeated measure analysis of variance (ANOVA) was conducted to test both for any effects on vocal accuracy of repeating the target note in a subsequent block and differences in hitting each target note. No statistically significant differences were found for the target notes or for repetitions of target notes, and there was no interaction between target note and repetition (all $p > 0.1$).

Repeated measure ANOVAs were conducted to test for adaptation. Analyses of upward shifted and downward shifted blocks were first conducted separately. For each shift direction, a 5×5 (block by phase) repeated measure ANOVA was conducted on the median F0 values for the initial 50 ms of the utterances. The factor Phase represented different phases of feedback alteration within the block and included the average of the pre-shift baseline trials 7–9, the shift trials 1–3, shift trials 15–18 (the last three shift trials which were present in all blocks), the post-shift test trials 1–3, and post-shift test trials 7–9. Figure 2 shows the median 50 ms data for each phase for all blocks. For two participants, some trials were missing from the last block (downward shifts). These trials were treated as missing data points. For both upward and downward shifted trials, a main effect of block [$F(4,52)=8.59$, $p<0.001$ and $F(4,44)=6.61$, $p<0.001$, respectively] and phase [$F(4,52)=10.96$, $p<0.001$ and $F(4,44)=6.72$, $p<0.001$, respectively] was found. The main effect of block was caused by different F0 values during the first 50 ms of vocalization for different target notes (see Fig. 2). The main effect of phase indicated adaptation occurred, as utterance onset was changed in response to the shifted feedback. A block by phase interaction was not found for either shift direction, indicating that the pattern of adaptation did not significantly differ across blocks.

To further examine the main effect of phase, planned comparisons were conducted for both upward and downward shifted data. For the 50 ms data, significant differences were found in the upward shifted trials between baseline trials 7–9 and shift trials 15–18 ($p=0.002$) and post-shift test trials 1–3 ($p=0.028$), and in the downward shifted trials between pre-shift baseline trials 7–9 and shift trials 15–18 ($p<0.001$) and post-shift test trials 1–3 ($p<0.001$). For both up-
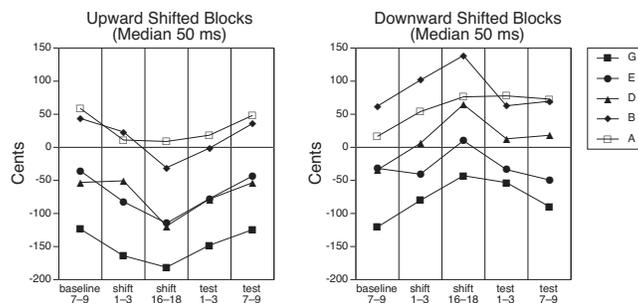
Fig. 2. Median F0 (in cents) for the first 50 ms of production for each target note shifted upward or downward for each phase of shifts used in the statistical analysis. Adaptation responses are in the direction opposite the feedback alteration in order to offset for the altered feedback.

ward and downward shifts, there was no significant difference in the 50 ms data between pre-shift baseline trials 7–9 and post-shift test trials 7–9 ($p > 0.1$), indicating that participants' initial F0 returned to pre-shift baseline values by the end of the block.

To test for the effects of order of the blocks, data were normalized such that the pre-shift baseline phase was equal to zero, and all responses were in the positive direction. First, a $2 \times 5 \times 5$ (shift direction by block by phase) ANOVA was conducted. No main effects of shift direction or block, and no interactions, were observed. There was a main effect of phase [$F(4, 44) = 20.57, p < 0.001$]. A further 10×5 (order by phase) ANOVA was conducted, with no main effect of order and no interaction observed, but again the main effect of phase was significant [$F(4, 44) = 17.88, p < 0.001$].

4. Discussion

The planned comparisons showed a significant difference between the last pre-shift baseline trials and the last shift trials, as well as the last pre-shift baseline trials and the first post-shift test trials, for F0 at utterance onset. This pattern of results shows that adaptation occurred. Participants altered their initial voice F0 to adapt to the feedback alteration, and this change in motor production continued when the feedback alteration was removed. No significant difference was found between the final pre-shift baseline and final post-shift test trials, demonstrating that the F0 at utterance onset returned to the pre-shift baseline level after several trials of normal feedback. This suggests that the sensory-motor mapping at the end of the post-shift test trials was identical to the sensory-motor mapping for the pre-shift baseline trials, indicating that the sensory-motor mapping had returned to the pre-adaptation state before the onset of the next block of the experiment.

We did not observe carry-over effects in adaptation across different blocks. In the current design, each block represented the unique pairing of a target note and shift direction, and therefore a unique context for adaptation to occur. We have shown that when each block is a unique context, adaptation effects do not carry-over across repeated blocks. This is important in that it demonstrates the feasibility of using a within-subject blocked design for experiments on vocal adaptation. Such designs may allow for more efficient experiments to test vocal adaptation to different conditions or different stimuli or to allow for repeated blocks of adaptation for added statistical power. Using repeated blocks can be very important when testing special populations, where a limited number of participants may be available. However, it should be noted that all the participants in this study were non-singers. Given the potential differences in vocal motor control between singers and non-singers (Keough and Jones, 2009), the results of this study should not be applied to trained singers without further testing. Also, it would be interesting to test if the lack of carry-over effects in different vocalization contexts applies to formant control as well as F0 control.

When conducting neuroimaging research (ERP, MEG, fMRI, etc.), it is necessary to have repetitions of the experimental conditions in order to conduct a statistical analysis. If only a single block of adaptation trials was used when conducting imaging studies, it would be difficult to construct a meaningful statistical analysis. Furthermore, any such analysis would be inherently confounded with time, as the pre-shift baseline trials would always precede the shifted trials. This represents a major confound in the contexts of an imaging study on vocal adaptation, where activation patterns may change as the participant adjusts to the novel environment (such as in fMRI).

The lack of a statistical effect of block when the data were normalized suggests that adaptation responses to different target notes are comparable. This, plus the fact that we found no statistical differences across blocks when the data were normalized, suggests that adaptation effects across different notes can be directly compared.

References and links

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot. Int. **5**, 341–345.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Ford, J. M., and Mathalon, D. H. (**2004**). "Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking," J. Psychiatr. Res. **38**, 37–46.

Guenther, F. H. (**2006**). "Cortical interactions underlying the production of speech sounds," J. Commun. Disord. **39**, 350–365.

Hawco, C. S., and Jones, J. A. (**2009**). "Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals," Brain Res. **1276**, 131–139.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jones, J. A., and Keough, D. (**2008**). "Auditory-motor mapping for pitch control in singers and nonsingers," Exp. Brain Res. **190**, 279–287.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–51.

Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.

Keough, D., and Jones, J. A. (**2009**). "The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing," J. Acoust. Soc. Am. **126**, 837–846.

Larson, C. R. (**1998**). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control," J. Commun. Disord. **31**, 489–503.

Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.

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

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (**2008**). "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**, 1429–43.

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**, 2306–2319.