# Compensation following real-time manipulation of formants in isolated vowels

David W. Purcell[a] and Kevin G. Munhall
*Department of Psychology, Queen's University, Kingston, Ontario, K7L 3N6, Canada*

(Received 4 July 2005; revised 23 November 2005; accepted 14 January 2006)

Auditory feedback influences human speech production, as demonstrated by studies using rapid pitch and loudness changes. Feedback has also been investigated using the gradual manipulation of formants in adaptation studies with whispered speech. In the work reported here, the first formant of steady-state isolated vowels was unexpectedly altered within trials for voiced speech. This was achieved using a real-time formant tracking and filtering system developed for this purpose. The first formant of vowel /ɛ/ was manipulated 100% toward either /æ/ or /ɪ/, and participants responded by altering their production with average F1 compensation as large as 16.3% and 10.6% of the applied formant shift, respectively. Compensation was estimated to begin <460 ms after stimulus onset. The rapid formant compensations found here suggest that auditory feedback control is similar for both F0 and formants. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2173514]

PACS number(s): 43.66.Jh, 43.70.Aj [AL] Pages: 2288–2297

## I. INTRODUCTION

Auditory feedback plays an important role in speech production. Post-lingually deafened adults show differences relative to the normal hearing population in the fundamental frequency or F0 of the voice, vocal intensity, and speaking rate (Leder *et al.*, 1987; Waldstein, 1990; Lane and Webster, 1991; Cowie and Douglas-Cowie, 1992; Leder and Spitzer, 1993). There are also differences in the formants of post-lingually deafened speakers after long-term hearing loss (Waldstein, 1990; Schenk *et al.*, 2003), or with shorter-term alteration of auditory feedback through changes in the cochlear implant status (Svirsky and Tobey, 1991; Economou *et al.*, 1992; Perkell *et al.*, 1992; Svirsky *et al.*, 1992). Shorter-term changes in auditory feedback also affect F0 (Binnie *et al.*, 1982; Economou *et al.*, 1992; Perkell *et al.*, 1992; Svirsky *et al.*, 1992). In summary, there is evidence showing both relatively rapid and slower changes in speech parameters like F0 and formants in the absence of auditory feedback. In this paper we investigate whether relatively rapid changes in production can be induced by altering formants in the auditory feedback of normal hearing individuals.

There is some debate, however, about the relative weight of auditory feedback in the control of suprasegmental speech parameters like F0, and segmental aspects of speech like formants (Svirsky *et al.*, 1992; Perkell *et al.*, 1992; Perkell *et al.*, 2000). Perkell *et al.* (2000) suggest that auditory feedback is used differently by a postural system operating on a suprasegmental time scale, and a phonemic settings system at the segmental level. The postural system adjusts average suprasegmental parameters such as SPL, F0, rate, and the degree of prosodic influence on SPL and F0, in order to maintain intelligibility. Phonemic settings control mechanisms responsible for producing the distinctions in different phonemes, and employ auditory feedback intermittently for maintenance purposes. The two systems use feedback on different time scales, and perhaps have different functional organization for different speech attributes. Intelligibility is reduced with post-lingual deafness (Cowie *et al.*, 1982), but the preservation of functionally adequate intelligibility lends evidence to the argument that auditory feedback is not necessary for the moment-to-moment control of segmental speech characteristics. Svirsky and Tobey (1991) note, however, that intelligibility is maintained until relevant acoustic parameters cross phonemic boundaries, making intelligibility a relatively insensitive measure of subtler changes. The present paper is part of a research program that investigates the use of feedback in the control of parameters like F0 and formants.

Two different types of experimental paradigms have been used to investigate the characteristics of the auditory feedback system. In order to study the influence of auditory feedback over longer time spans, a sensorimotor adaptation technique has been used. Drawn directly from earlier literature on visuomotor control (e.g., Held, 1965), the experiments involve the manipulation of sensory feedback. Two aspects of the response are examined: the immediate compensation for the perturbation and, more importantly, the adaptation or persistence of the effect after the perturbation is eliminated.

Using whispered speech, Houde and Jordan (1998; 2002) have shown that gradual manipulation of the feedback of formants F1 through F3 can lead to compensatory responses. For example, a subject may whisper the vowel /ɛ/, but is provided with auditory feedback, where the formants have been adjusted such that the feedback vowel sounds more like the subject's own /i/. Without conscious knowledge of their response, the subject may compensate by producing a vowel closer to /æ/ (Houde and Jordan, 2002). After sufficient training, adaptation was observed when auditory feedback was eliminated using masking (Houde and Jordan, 2002). Similar results have been reported recently for voiced

[a]Author to whom correspondence should be addressed. Electronic-mail: purcelld@nca.uwo.ca

 

speech instead of whispered using improved formant transformation systems (Villacorta *et al.*, 2004; 2005).

Subjects show both compensation and adaptation effects to the perturbation of the fundamental frequency of their voice. Jones and Munhall (2000) in a study of English-speaking subjects showed that gradual changes in pitch feedback were countered by F0 production in the opposite direction, and that an adaptation effect persisted after feedback was returned to normal. This effect has been replicated with native speakers of Mandarin (Jones and Munhall, 2002). Both the formant and F0 adaptation effects have been interpreted as evidence for a central representation or internal model that is used in motor control in a feedforward manner (e.g., Miall and Wolpert, 1996). In this mechanism, feedback's primary role is to tune the "model" that the nervous system uses to predict the effects of movements and thus plan more effectively. This similarity between F0, particularly in English, and the Houde and Jordan vowel formant data suggests that similar systems are involved (or the same system).

The second experimental paradigm that has been used to examine the role of auditory feedback in speech production is the introduction of sudden, unexpected auditory feedback changes. This technique tests auditory feedback as part of a servomechanism that is involved in the moment-to-moment control of articulation. When sudden changes to the pitch of the voice are unexpectedly introduced, most individuals will change their production within about 100 to 225 ms by a shift of F0 in the opposite direction (Burnett *et al.*, 1997). The response magnitude and latency have shown an indifference to stimulus pitch-shift magnitude (Burnett *et al.*, 1998), indicating that the response has generally not been elicited in a linear range. When the duration of the pitch-shift stimulus is extended beyond 500 ms, the existence of a second response is observable, and is more directly under voluntary control (Hain *et al.*, 2002). The rapid response is observed in speakers of Mandarin (Jones and Munhall, 2002) and English (Burnett *et al.*, 1997). In the present paper we extend this work by using sudden perturbations of formant frequency during vowel production.

These two experimental paradigms (employing sudden and gradual manipulations of auditory feedback) can be used to study the potential differences and similarities in the control of segmental and suprasegmental speech characteristics. For F0, the presence of both adaptation effects and compensation to rapid perturbations suggests that the neural mechanisms for its control operate on multiple different time scales. By examining the presence of immediate responses to formant perturbations, this study extends our understanding of the way in which the sound of our voice contributes to fluent speech. If both F0 and formant perturbations show the same sensitivity to immediate feedback, it will suggest that the basic form of the feedback system is similar across the various properties of speech acoustics.

## II. METHODS

### A. Subjects

Subjects were undergraduate students ($n=28$, 21 females), and varied in age from 17 to 28 yr. Each person participated in a single experimental condition completed in one session. Thresholds of hearing were evaluated individually in each ear at octave frequencies from 500 to 4000 Hz. The majority of participants had normal thresholds ($<=20$ dB HL), but four individuals were included who had thresholds as high as 35 dB HL at only one frequency. Other audiometric frequencies were normal in these individuals near the frequency with a high threshold, and they did not have atypical responses to the altered auditory feedback. They were retained to have as much data as possible, and because the accuracy of speech processing is not predictable from the audiogram (Surprenant and Watson, 2001), and the speech feedback was well above threshold across frequency. All subjects learned English as a first language and had no known language or speech impairments. Some subjects beyond the reported 28 were not included in the analysis because they failed one of the above criteria, had difficult to track formants as described below, or their vowels /ɪ/, /ɛ/, and /æ/ were not separated by at least 50 Hz in F1.

### B. Equipment

During the experiment, subjects produced isolated steady-state vowels that were transduced into an electrical signal using a Shure headset microphone type WH20. This microphone signal was amplified using a Tucker-Davis Technologies MA3 microphone amplifier with the +20 dB gain switch active. After amplification, the microphone signal was low-pass filtered with cutoff frequency 4500 Hz using an analog Frequency Devices type 901 filter with a gain of 0 dB. A National Instruments PXI-6052E input/output board mounted in a PXI-1002 chassis digitized this signal at 10 kHz with 16-bit precision. The voice was analyzed and filtered (details explained below) in real time using a National Instruments PXI-8176 embedded controller, which is essentially an Intel Pentium III processor running at 1.26 GHz. The filtered voice signal was converted back to analog by the digitizer at 10 kHz with 16-bit precision. The processed analog signal was again low-pass filtered as above with a second Frequency Devices unit, and subsequently amplified using a Madsen Midimate 622 audiometer where speech noise was added. The microphone MA3 amplifier was adjusted between 30 and 50 dB for each individual during training such that the Madsen input VU meter read approximately 0 dB during vocalization. The Madsen output gain controllers were set such that the vocal auditory feedback at each ear was approximately 80 dBA sound pressure level (SPL), and the speech shaped noise was approximately 50 dBA SPL. Participants heard their vocalization in real time, along with the speech-shaped noise, through Sennheiser "HD 265 linear" headphones with the same signal presented to each ear. These headphones are somewhat acoustically open and strike a reasonable balance between the contradictory requirements of shielding the listener from the airborne sound of their unaltered voice and minimizing the emphasis of bone conducted sound by the occlusion effect (Békésy, 1932; Tonndorf, 1972). The headphones were calibrated using a Brüel & Kjær sound level meter and artificial ear Type 4153. Prompts were shown to the participant on a monitor that also displayed a bar graph representation of their speaking level. This visual feedback allowed the subject to maintain a relatively consistent level throughout the experiment.

### C. Experimental conditions

Two experimental conditions were employed where the first formant (F1) was shifted for the vowel /ɛ/. The altered auditory feedback was always introduced 300 ms after the onset of voicing, and was cross-faded in linearly over 500 ms to replace the normal feedback for the duration of the trial. This slow stimulus onset was chosen to maximize compensation under the assumption that it may be similar to the pitch-shift response. Larson *et al.* (2000) found responses were larger with a 500 ms onset compared to those for step changes in feedback pitch.

The two experimental conditions evaluated whether speakers would compensate for shifts of F1 in either direction. In condition A, 16 participants (13 female) had their F1 shifted upward 100% toward /æ/. In condition B, 12 subjects (8 female) had their F1 shifted downward 100% toward /ɪ/. The actual shift in Hz was determined from the individual's average formant values. The upward shift was the individual difference between F1 of /æ/ and F1 of /ɛ/. Similarly, the downward shift was the difference between /ɛ/ and /ɪ/. For the subjects in this experiment, the average upward shift was +136 Hz (standard deviation SD=46.2 Hz), and the average downward change was −135 Hz (SD=42.7 Hz).

### D. Experimental protocol

Participants were seated in a comfortable chair located in an Industrial Acoustics Company (IAC) sound insulated room for the experiment, which lasted one hour. Participants were instructed to try to produce isolated steady-state vowels using their normal speaking voice. A normal speaking level was used to set microphone gain, and a headphone level was set sufficiently high to maintain a favorable relative level between the contributions of bone-conducted and headphone delivered sounds to the net voice signal at the cochlea. If subjects had increased their speaking level significantly during the experiment, the headphone signal could have clipped and thus not maintained the desired relative level at the cochlea (or fidelity; level feedback was therefore provided to the participants and was monitored by the experimenter to avoid problems of this type). Subjects were prompted during roughly 25 practice trials to produce the isolated vowels /ɛ/, /æ/, and /ɪ/ for the duration of a 2.5 s prompt, as they would sound in the consonant-vowel-consonant (CVC) contexts "head," "had," and "hid," respectively. For formant tracking (via the same method as described in Sec. II E), the spectral model order was always in the range 8 through 12, and was chosen manually based on the practice trials of /ɛ/ such that F1 was most continuous (Vallabha and Tuller, 2002). When the practice period was complete, 10 utterances of each of /ɛ/, /æ/, and /ɪ/ were recorded and the average steady-state F1 of each was determined (for the purpose of setting the individual F1 shift size).

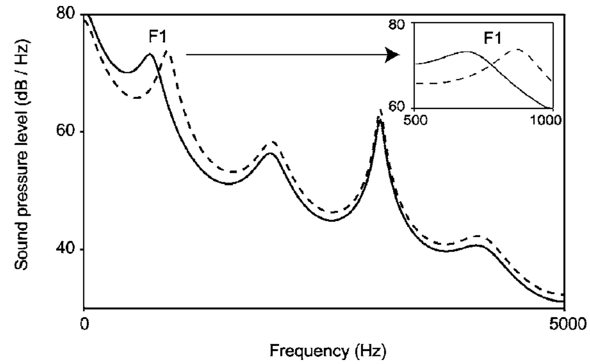


FIG. 1. Example formant shift used in a single test trial for one participant's steady-state utterance of vowel /ɛ/. The solid line shows the spectral envelope of the microphone signal, where the first formant F1 was estimated online at 657 Hz. Formant F1 was shifted upward +142 to 799 Hz for the headphone feedback signal shown with the dashed line. This shift is the difference between the participant's average F1 for vowels /æ/ and /ɛ/. Although she spoke /ɛ/, the perturbation placed the feedback F1 near the F1 frequency of her average production of /æ/. The inset shows the F1 formant perturbation with a smaller frequency range.

The experiment itself consisted of four blocks of 100 trials with an interprompt interval of approximately 1.5 s. Each trial consisted of the presentation of 1 of 11 different vowels in a CVC context (/hVd/) that were chosen to use the entire English vowel space (Ladefoged, 1982). Auditory feedback was normal for all trials except rare test trials, where the target vowel /ɛ/ was perturbed. This approach was intended to minimize learning effects such that perturbed trials could be averaged together. In each block of 100 trials, there were generally five test trials where the target vowel /ɛ/ was manipulated; however, due to random spacing and order constraints, as few as four or as many as six test trials could occur in a block. The test trials were intermingled randomly with the other 10 distractor vowels. The test trials were always separated by at least three instances of the target vowel /ɛ/, where auditory feedback was normal. These target instances with normal auditory feedback were themselves also separated by at least one distractor vowel. Thus, it was presumed that production of the target vowel had returned to normal before auditory feedback was again manipulated.

### E. Online formant shifting and detection of voicing

With the feedback modification system, F1 was shifted by filtering out the signal at the frequency where the formant was estimated, and emphasizing the signal at the new desired formant frequency. This was accomplished using a filter transfer function with a pair of spectral zeros in the numerator to attenuate the energy (harmonics of the glottal fundamental frequency F0 for vowels) near the existing formant, and a pair of spectral poles in the denominator to amplify energy near the new formant. Figure 1 shows the spectral envelope of the microphone and feedback signals for a single trial where F1 of /ɛ/ was pushed upward 142 Hz.

To shift formants, it is necessary to estimate them from the speech signal, and then to filter the speech signal and deliver it to the subject as altered auditory feedback. Online real-time formant tracking and filtering of the speech signal was implemented on the National Instruments PXI computer using the LabView Real-Time language. Formants were tracked online using an iterative Burg algorithm for estimating spectral parameters (Orfandidis, 1988). This method uses a sliding analysis window where the weights applied to older samples decay exponentially. The decay parameter was chosen such that 50% of the area under the weighting curve applied to samples less than 8.6 ms old. While tracking two formants, and shifting the first formant, the system was able to estimate the formants and update the filter coefficients approximately every 900 $\mu$s or every nine speech data points. Although the analysis window was infinitely long, using exponentially decaying weights the effective delay of the system was less than about 20 ms, or two pitch periods for a deep male voice of F0=100 Hz. This approach is slightly different from filtering methods that employ double buffers of fixed length, since here the formants are updated as often as possible with whatever new microphone data points are available after the completion of each estimate. In contrast, Houde and Jordan (1998) synthesized whispered speech by filtering a noise source with formant estimates from windows of whispered speech.

The onset of voicing was detected using a simple statistical amplitude threshold technique. On visual inspection of the microphone record, this triggering technique typically detected voicing within 10 ms of its true start.

### F. Offline formant analysis

The subject's response to altered auditory feedback was evaluated offline using the same formant-tracking algorithm that was employed online. Formants were estimated offline from the recorded speech 1000 times per s (every ten speech samples). Although rare, trials were removed from the analysis if the participant stumbled or produced the wrong vowel, or if the voicing trigger was more than 150 ms early. Early triggers occurred rarely due to lip smacks or other noises that preceded vocalization. Since the manipulation of feedback commenced 300 ms post-trigger, trials with early triggers still provided at least 150 ms of unperturbed baseline data. Continuous and stable formant tracking in the target vowel was a problem for some subjects, and manifested itself as the misinterpretation of F2 as F1. If this occurred excessively during a test trial, then the filter coefficients were not stable and the auditory feedback was garbled. These trials were removed from the analysis, and in some cases a subject (six beyond the reported 28) had to be removed from the analysis if too few usable test trials remained (<10). On average, there were 20 test trials available for further analysis, but 25% of subjects had only 10 to 15 usable test trials.

The formant tracks in the test /ɛ/ trials were averaged together, as were those in the unperturbed /ɛ/ trials immediately preceding the test trials (termed "pretest" trials). Prior to averaging however, the formant tracks were conditioned to remove any individual formant estimates in each trial that were in error by evaluating the histogram of each formant track, and eliminating those estimates that were well outside the norm (estimate bins that had a count <1% of the average count of the mode bin across trials). The average formant tracks and their SDs within a given group (test or pretest) were then determined by averaging the valid formant estimates at each point in time across trials. The average trials were truncated at the end of the vowel to the duration of the shortest individual trial used in the analysis. While the subjects were prompted for 2.5 s, the average trials were typically of 1.5 s duration due to a slow reaction time or shorter vocalization in one or more trials over the course of the experiment.

### G. Response evaluation

To determine if the subjects responded to the altered auditory feedback, the formants in the test trials were compared to those in the pretest trials. Average formant values were calculated for eight consecutive 150 ms blocks of data for each subject beginning at −100 ms relative to the manipulation onset. The manipulation onset will henceforth be defined as time zero. Block averages were used to smooth and reduce the data to a manageable set size for analyses of variance (ANOVA). The block size was chosen to be smaller than average pitch-shift response delays (Burnett *et al.*, 1998; Larson *et al.*, 2000), and smaller than the minimum time for a speaker to interrupt their speech found by Ladefoged *et al.* (1973). The block differences between test and pretest trials were averaged across subjects, and two-way ANOVAs with Greenhouse-Geisser corrections (due to sphericity concerns) were calculated for factors of the formant shift direction (up and down) and block number (one through eight). One ANOVA was determined separately for each formant F1 and F2.

The slow stimulus onset used in this experiment was not designed to measure the minimum response time. However, an upper limit of when the compensatory response began can be estimated using the change point test (Donath *et al.*, 2002; Siegel and Castellan, 1988). The F1 difference between the (unblocked) grand average test and pretest trials was evaluated to determine the time at which the formants began to change.

## III. RESULTS

### A. Formant changes in response to altered feedback

Subjects tended to change F1 in their production in the direction opposite to the manipulation of F1 in the auditory feedback. This was true for both experimental conditions. When F1 of /ɛ/ was made to sound more like /æ/ in the auditory feedback, participants tended to produce vowels with F1 in the direction of /ɪ/. Correspondingly, when F1 of /ɛ/ was pushed toward /ɪ/, subjects tended to produce vowels with F1 in the direction of /æ/. These trends can be observed in Fig. 2 for both manipulation directions. In this figure, the dark lines represent the grand average of F1 for the test trials in each experimental condition after the grand average of F1 for unperturbed pretest trials was subtracted. This normalized the data for trends within vowels and across the experiment. In addition, the one SD band indicates the between-subject variability. On average, the compensation in F1 production was small compared to the change of F1 in the auditory feedback. The variability in the response was also large rela-

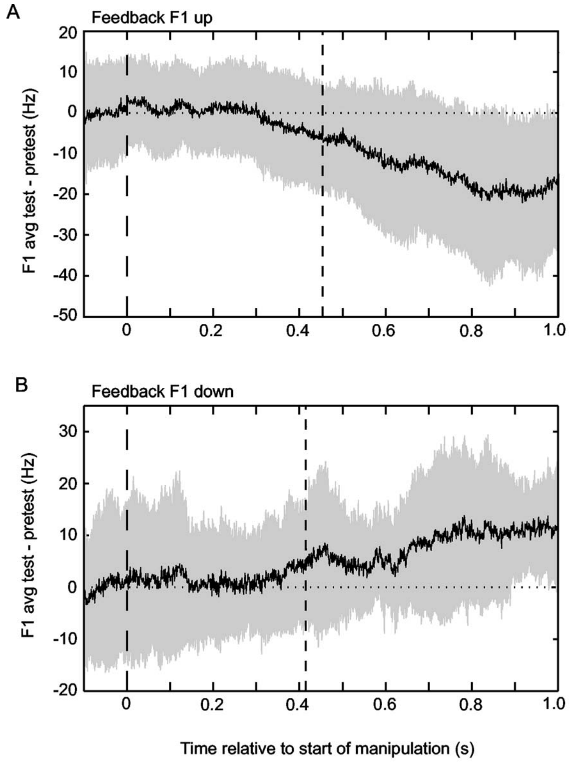


FIG. 2. The F1 formant difference responses are shown for grand averages across subjects (thick black lines) with 1 SD bounds (grey area). These curves are the F1 differences between the test and pretest trials of /ε/ plotted against time with respect to the perturbation onset at 0 s. Panel (a) shows the responses of 16 subjects where F1 was shifted upward toward /æ/. Panel (b) shows the responses of 12 subjects, where F1 was shifted downward towards /ɪ/. The horizontal dotted lines at zero mark no difference between test and pretest trials. The widely spaced dashed vertical lines at 0 s mark the time when altered auditory feedback began linearly cross-fading into the headphones over 0.5 s. The narrowly spaced dash vertical lines mark the change points where a significant change was first detected in the grand average F1 difference curve. For panels (a) and (b), the change points occurred at 0.454 and 0.415 s, respectively.

tive to the size of this compensation. The second formant (F2) showed small concomitant changes toward the opposite vowel category as with F1, but these were not statistically significant and again demonstrated relatively large variability. For the condition where F1 feedback was shifted downward in frequency, one subject correctly identified that /ε/ was being altered occasionally in an exit interview. His data were not atypical and he was not excluded from the analysis. No other participants recognized the manipulation.

The difference between average test and pretest trials was determined for each block number for formants F1 and F2, as well as the percent compensation and percent change relative to the average perturbation. These are given in Table I for the condition where feedback F1 was pushed upward on average +136 Hz, and Table II, where F1 was pushed downward on average −135 Hz. In addition, the tables report the SDs across subjects for the differences between average test and pretest trials. Variability in the absolute estimates of F1 and F2 for /ε/ was also large across trials within each subject. For each subject, SDs of the formant estimates were calculated separately for test and pretest trials. These SDs were then averaged across subjects for both manipulation directions, resulting in an F1 SD=38 Hz, and an F2 SD =53 Hz.

For F1, a two-way ANOVA [time block (sequential blocks of 150 ms, numbered one to eight), and direction of F1 perturbation (upward versus downward)] showed significant main effects of the block number [$F(2.8, 71.7)=3.1$, $p < 0.04$], and manipulation direction [$F(1, 26)=22.2$, $p < 0.001$]. There was also a significant interaction between these two factors [$F(2.8, 71.7)=27.5$, $p < 0.001$]. For F2, the same ANOVA did not quite show a significant main effect of the manipulation direction [$F(1, 24)=3.8$, $p < 0.062$], but had a significant interaction between the direction and block number [$F(2.8, 68)=4.3$, $p < 0.01$]. The interactions between the block number and manipulation direction are because the changes with block number are either negative or positive depending on the manipulation direction. Both manipulation directions had one subject without valid F2 estimates, so the degrees of freedom were lower than in the F1 ANOVA.

For each stimulus condition and formant number, Scheffé's method was used to evaluate whether the test-pretest differences of block one were different from the other blocks two through eight (given in Tables I and II). For the experi-

TABLE I. Grand average block means and standard deviations (SDs) across subjects of the difference between test and pretest trials for stimulus condition A, where F1 was manipulated upward 100% toward /æ/. Blocks are sequential windows of 150 ms. Time zero was chosen as the time when the manipulation commenced, so Block 1 shows the mean formants prior to the response. Percent compensation and percent change were with respect to the group average F1 manipulation of +136 Hz. One subject did not have valid F2 estimates and is not included in the F2 summary.

| | Block 1 | Block 2 | Block 3 | Block 4 | Block 5 | Block 6 | Block 7 | Block 8 |
|---|---|---|---|---|---|---|---|---|
| Block center re manipulation start (s) | −0.025 | 0.125 | 0.275 | 0.425 | 0.575 | 0.725 | 0.875 | 1.025 |
| | | | Formant F1 | | | | | |
| Mean $F1_{test}-F1_{pretest}$ (Hz) | −1.4 | −1.8 | −2.5 | −7.6 | −12.7 | −16.9 | −22.2 | −21.5 |
| SD $F1_{test}-F1_{pretest}$ (Hz) | 10.2 | 8.9 | 10.1 | 11.7 | 15.0 | 16.1 | 17.3 | 14.8 |
| % F1 compensation | 1.0 | 1.3 | 1.8 | 5.6 | 9.4 | 12.4 | 16.3 | 15.8 |
| | | | Formant F2 | | | | | |
| Mean $F2_{test}-F2_{pretest}$ (Hz) | 4.2 | 5.7 | 2.7 | 5.0 | 4.5 | 8.8 | 10.7 | 11.7 |
| SD $F2_{test}-F2_{pretest}$ (Hz) | 11.9 | 15.4 | 12.5 | 14.6 | 16.1 | 16.1 | 20.7 | 22.6 |
| % F2 change | 3.1 | 4.2 | 2.0 | 3.7 | 3.3 | 6.5 | 7.9 | 8.6 |

TABLE II. Grand average block means and standard deviations (SDs) across subjects of the difference between test and pretest trials for stimulus condition B, where F1 was manipulated downward 100% toward /ɪ/. Blocks are sequential windows of 150 ms. Time zero was chosen as the time when the manipulation commenced, so Block 1 shows the mean formants prior to the response. Percent compensation and percent change were with respect to the group average manipulation of −135 Hz. One subject did not have valid F2 estimates and is not included in the F2 summary.

| | Block 1 | Block 2 | Block 3 | Block 4 | Block 5 | Block 6 | Block 7 | Block 8 |
|---|---|---|---|---|---|---|---|---|
| Block center re manipulation start (s) | −0.025 | 0.125 | 0.275 | 0.425 | 0.575 | 0.725 | 0.875 | 1.025 |
| | | | | Formant F1 | | | | |
| Mean $F1_{test}-F1_{pretest}$ (Hz) | 3.3 | 4.2 | 3.6 | 7.5 | 7.9 | 12.2 | 13.3 | 14.3 |
| SD $F1_{test}-F1_{pretest}$ (Hz) | 13.4 | 12.2 | 8.9 | 11.4 | 8.0 | 13.1 | 9.3 | 6.4 |
| % F1 compensation | 2.4 | 3.1 | 2.7 | 5.6 | 5.8 | 9.1 | 9.8 | 10.6 |
| | | | | Formant F2 | | | | |
| Mean $F2_{test}-F2_{pretest}$ (Hz) | −1.7 | 1.0 | 0.4 | −3.3 | −4.9 | −6.9 | −8.0 | −11.0 |
| SD $F2_{test}-F2_{pretest}$ (Hz) | 13.5 | 11.3 | 14.3 | 14.3 | 15.5 | 18.6 | 18.1 | 20.3 |
| % F2 change | 1.2 | −0.8 | −0.3 | 2.4 | 3.6 | 5.1 | 5.9 | 8.2 |

mental condition where feedback F1 was moved upward, blocks five through eight had significantly larger absolute test-pretest formant changes in F1 than block one ($p < 0.05$). There were no significant differences for F2, or for F1 of the downward condition. Given the small response and large variability, the lack of significance in the downward post-hoc tests was probably because there were fewer participants in this condition.

### B. Formant F1 change points and comparison of manipulation direction

Figure 2 also shows the change points calculated for F1 of each experimental condition. In Fig. 2(a), where feedback F1 was pushed upward, a significant change was found in F1 at a delay of 0.454 s with respect to stimulus onset. Figure 2(b), where feedback F1 was pushed downward, shows a significant change point at a delay of 0.415 s with respect to the stimulus onset.

Separate variance t tests showed that the absolute magnitude of the stimulus manipulation was not different between conditions (+136 Hz vs −135 Hz). The absolute response magnitude and percent compensation/change for F1 and F2 were also not different between conditions for the mean formants in block 8 (see Tables I and II). The F1 responses for the up/down directions can be compared qualitatively in Fig. 2 between panels (a) (F1 feedback up) and (b) (F1 feedback down).

Whereas Fig. 2 shows the average trial data, histograms of the compensation in each test trial from all participants are given in Fig. 3. The percent compensation in each test trial was estimated as 100 times the difference between blocks eight and one divided by the size of the manipulation. Note that for both panels, a positive percent compensation indicates that production changed to counter the feedback manipulation. A negative percent compensation on the horizontal axis indicates that F1 production followed the shift in F1 feedback. In panel (a), where feedback F1 was increased, trials with positive percent compensation had a decrease in F1 production. For panel (b), feedback F1 was decreased, so trials with positive percent compensation had an increase in produced F1. A correction was derived for participants from their pretest trials, in order to reduce bias in the percent compensation estimate due to the normal trajectory of F1 over the course of an utterance (the average percent change between blocks eight and one in the pretest trials, relative to the size of the test trial manipulation, was added to the values calculated for the test trials). These histograms show that the size of the response to feedback perturbation varies from trial to trial, as discussed further below. Both histograms are shifted slightly right of zero, indicating that both feedback manipulations were compensated in many trials, but very few trials approach complete compensation (positive values near 100%). Both distributions include zero and negative percent compensation because there were trials with no com-

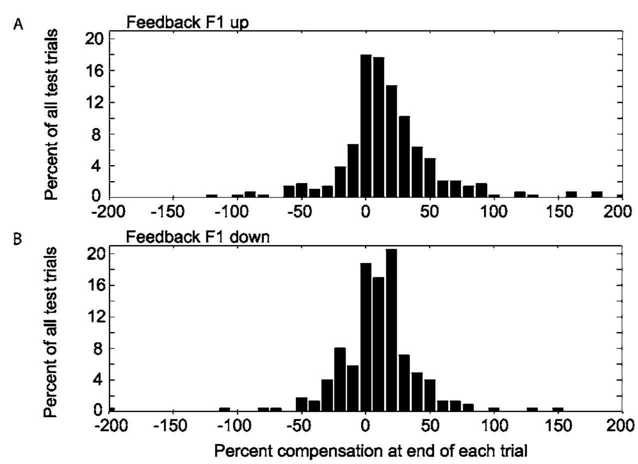


FIG. 3. Histograms of the percent compensation in all test trials from all subjects. The vertical axis shows the bin counts as a percentage of all test trials, rather than the absolute number of trials. Percent compensation on the horizontal axis was calculated as 100 times the F1 difference between blocks eight and one, divided by the (participant dependent) manipulation size. Each bin is 10% wide, where the zero bin spans from −5% to +5%. Positive values of percent compensation indicate that the change in production countered the feedback manipulation. Negative percent compensation indicates that F1 production changed in the same direction as the manipulation (a following response). Panels (a) and (b) show responses for the experimental conditions where feedback F1 was increased and decreased, respectively. Both distributions are slightly right of zero because both conditions elicited a majority of compensatory responses.

pensation, and those where F1 followed the manipulation. Figure 3 does not explicitly separate between- and within-subject variability, however, all participants had mean percent compensations $<30\%$. The spread of the histogram toward higher compensations is due to trial-to-trial variability within subjects, not because of a few participants with consistently high percent compensation.

## IV. DISCUSSION

The data reported here reveal a modest sensitivity to unexpected formant perturbations in auditory feedback. Subjects, on average, produced compensatory changes in the frequency of their first formant in the opposite direction to unexpected changes in the auditory feedback of their own voice. When the manipulation of the feedback raised F1, the compensations were slightly larger on average (although not significantly so) than those in response to perturbations that lowered F1. However, considerable variability in the compensatory responses for both directions was evident within and between participants. An examination of the total distribution of responses (Fig. 3) found that the perturbation was completely compensated in a few of the test trials. In combination with reported rapid F0 compensations (e.g., Burnett *et al.*, 1998), these results suggest a general similarity between F0 and formant behavior when feedback is suddenly modified.

Incomplete compensation has been found previously in pitch-shift studies. For example, Fig. 3(b) of Burnett *et al.* (1998) shows that the response magnitude is largely independent of stimulus magnitude from 25 to 300 cents, leading to a lower percent compensation for the larger stimulus pitch shifts. Recent data (Leydon *et al.*, 2003) suggests that the pitch-shift response may have a linear operating range where a single gain relates output to input, whereas most studies have employed stimulus shifts that saturate the response at the limit of compensation. Since only a single formant shift value was used in this study, a more comprehensive comparison would be required to make any conclusions about the relative responsiveness of the speech motor system to formant modification.

Incomplete compensations have also been observed in the study of formant shifts using whispered steady-state vowels in an adaptation paradigm (Houde and Jordan, 1998; 2002). In the data of the eight reported subjects, the observed compensation varied between about 10% and 110% for a large feedback shift of two vowel categories (the cross-subject mean of their 1998 Fig. 3 is roughly 50%). Houde and Jordan carefully produced formants near hypothetical lines in formant space that connect adjacent vowels. According to Houde and Jordan, this path defines "perceptually salient directions" between different vowels, and identifies what formant frequencies can be realized by the vocal system. Manipulations should produce feedback vowels that are on the linear segments connecting vowel categories, if the manipulations are to be perceived and if compensation is to be possible. In the experiment reported here, only F1 was manipulated, and for most individuals this would push /ε/ off the path advocated by Houde and Jordon. They suggest that feedback vowels with off-path manipulations are projected onto the nearest intervowel category line. If this were the case, then manipulations of F1 alone would be effectively smaller as interpreted by the auditory vocal feedback system because the projected manipulation must be smaller in magnitude than the original manipulation vector.

Figure 3 shows that there were a small percentage of complete compensations and many responses near the zero percent value. One possible explanation is that the perturbations are not always detected by the speech motor system. Kewley-Port and colleagues (e.g., Kewley-Port and Watson, 1994; Kewley-Port and Zheng, 1999; Kewley-Port, 2001) have studied the psychoacoustic thresholds for the detection of changes in single formants. These studies have shown that well-trained individuals can detect changes in F1 as small as 14 Hz under ideal conditions (Kewley-Port and Watson, 1994). While this is significantly larger than the threshold for detecting changes in the frequency of tones (Moore, 1973), it is smaller than the manipulations that were employed here. Lack of training can raise the threshold for detecting F1 formant changes to about 45 Hz (Kewley-Port, 2001), and the presence of noise may increase this 20% to over 300% depending on SNR and noise characteristics (Liu and Kewley-Port, 2004). These thresholds are somewhat smaller than the average manipulations used here of approximately ±135 Hz with naiive listeners and some background noise. It is unknown whether thresholds for the detection of changes in one's own voice during production are similar to those for listening to the voices of others. It is possible that for some individuals the manipulation was subthreshold, but this seems tenuous given that F1 was moved an entire vowel category. However, only one subject identified that /ε/ was being manipulated, and only a few recognized (in retrospect) that there was some change in their vowel sounds after the experiment was explained in detail during an exit interview. It is also unknown whether compensation occurs for manipulations smaller than the psychoacoustic threshold for the detection of changes in one's own voice during production.

The variability in response magnitude observed here does not seem to be unique to the formant shift. While unaveraged individual trial data is unpublished, pitch-shift studies have previously reported that from 5% to 20% of subjects' averaged responses do not meet criteria to be "valid" for all stimulus conditions (e.g., Burnett *et al.*, 1998; Bauer and Larson, 2003; Sivasankar *et al.*, 2005). These invalid (non)responses from individuals are typically dropped from further analysis. In most pitch-shift reports, the between-subject response SDs are a large fraction of the average response magnitudes. Response magnitude varies with experimental design but some example values are approximately 50 (SD=20) cents (Donath *et al.*, 2002), 60 (SD =32) cents (Natke *et al.*, 2003), 38 (SD=33) cents (Larson *et al.*, 2001), and 28 (SD=15) cents (Burnett *et al.*, 1998). These example intersubject response magnitude SD values are relatively large, which is noted by Burnett *et al.* (1998) as "considerable between- and within-subject variability." The histograms in Fig. 3 show that there is a wide range in the percent compensation in individual test trials. A large proportion of test trials exhibit no compensation, whereas some perturbations are robustly compensated. Overall, the distributions are shifted toward compensation (i.e., right of zero in Fig. 3), but in a sizable proportion of trials participants actually followed the manipulation, indicated by negative compensation bins.

A possible explanation why formant and F0 perturbations appear to be similar (other than actually being similar or the same) is that the task used in the sudden perturbation studies may induce a uniformity of response that does not exist in natural speech; the production of prolonged vowels may be influencing the results in these studies. This relatively unnatural speech task might alter the way in which auditory feedback is processed or change the importance of the information. In pitch-shift studies it has been found that task relevance has an influence on the characteristics of the compensatory response. Whereas many studies have applied pitch shifts to single syllable vocalizations that were produced for several seconds, another approach has been the use of multisyllabic nonsense words with different stress patterns (Natke and Kalveram, 2001). The latency of the compensatory response is long enough that it does not affect short, unstressed initial syllables, but it does, however, influence long, stressed initial syllables and subsequent syllables. Interestingly, the response persists after the pitch-shift stimulus is terminated, and influences the following word that may be uttered up to six seconds later (Donath *et al.*, 2002). Responses were found to be larger and the aftereffect longer, when subjects sang nonsense words in tune with a piano compared to speaking them (Natke *et al.*, 2003). This is presumably in part because F0 control is more important in singing than speaking for nontonal languages. Although Jones and Munhall (2002) found similar compensation latencies for speakers of the tonal language Mandarin relative to English speakers, there were no Mandarin speakers with the pitch-following response that has been reported in English participants (Burnett *et al.*, 1997). Comparisons within Mandarin have shown that response latency is shorter and magnitude is greater when pitch shifts are applied during dynamic tones, relative to static tones (Xu *et al.*, 2004). This supports the hypothesis that the response is modulated by the importance of pitch to the task. The relationship is, however, complex. In trained singers who spoke a nontonal language, the response latency was longer and the compensation magnitude smaller in singing when the target pitch was a dynamic glissando instead of a steady value (Burnett and Larson, 2002).

These results raise the possibility that the importance or salience of the auditory feedback can be modulated over time. Such modulation may also account for substantial within-subject variability. Average formant SDs were 38 Hz for F1 and 53 Hz for F2. These are large compared to the average changes between test and pretest trials (i.e., compensation) shown in Tables I and II. The within-subject formant SDs can be compared to values from other investigations of vowel production. The SDs here are in the same range as those reported in the vowel imitation literature (e.g., Kent and Forner, 1979; Repp and Williams, 1987; Vallabha and Tuller, 2004), and vowel production studies (e.g., Pisoni, 1980; Perkell and Nelson, 1985; Beckman *et al.*, 1995).

The pitch-shift literature has demonstrated a relatively rapid response to perturbations. Unfortunately the present study was not designed to measure minimum response time. Rather, the relatively slow onset of altered auditory feedback over 500 ms was chosen to try and maximize any compensatory response, under the assumption that formant feedback monitoring may be similar to that in pitch-shift paradigms (Larson *et al.*, 2000). This slow onset made it difficult to determine response latency since it was not known when the stimulus amplitude crossed the hypothetical detection or response initiation threshold. Response delay was estimated using the change point test, but the true physiological delay should be smaller since the early part of the altered feedback would presumably be subthreshold. Indeed, the mean change point times for F1 are longer than previously reported latencies for the pitch-shift response, such as 100 to 225 ms (Burnett *et al.*, 1997), or means of 192 and 266 ms (Burnett *et al.*, 1998). Larson *et al.* (2000) found mean upward responses were 217 vs 273 ms for downward responses. With the stimulus used here, it was not possible to determine whether formant compensation can occur as quickly as for the pitch-shift response. A more rapid transition to altered auditory feedback would allow a better estimate of response latency. In pilot measurements, a cross-fade of 20 ms was employed, and on average individuals showed both compensation and in some cases following responses.

Although only F1 was manipulated, small concomitant changes were also observed in the second formant, as suggested by trends in the F2 data of Tables I and II. This is not unexpected since formants are created by constrictions in the vocal tract (Fant, 1970), and it is unlikely that the map between articulation and acoustic goals learned as an infant would require or employ independent control of F1 and F2. Rather, a movement of vocal tract constriction would likely change both F1 and F2 because they are coupled in the resonant system (Stevens, 1998). The average data in Tables I and II suggest that the formant changes are appropriate for compensation along trajectories between vowel groups. For example, in Table II, F1 of /ɛ/ was perturbed downward toward /ɪ/. The average responses in the later blocks (e.g., six through eight) show the production of something in the direction of the /æ/ category, with higher F1 and lower F2. This would partially counter the lower F1 and higher F2 of the /ɪ/ category, despite the fact only F1 was manipulated. Similar but reversed trends are seen in Table I where F1 was shifted toward /æ/.

The influence of altered auditory feedback could have been attenuated by direct mouth to ear airborne sound, and through bone-conducted sound. A natural unprocessed voice signal that reaches the cochlea competes with the altered auditory feedback provided by the headphones. In this experiment the headphones themselves provided some isolation from airborne sound of the voice emitted at the mouth. It would be inappropriate, however, to use standard sound attenuation headphones with an acoustically closed back. Such headphones emphasize the bone-conducted signal radiated into the ear canal through the occlusion effect (Tonndorf, 1972). Headphones can be carefully designed to reduce airborne signals in the ear canal and simultaneously avoid the occlusion effect through the use of larger cavities (Békésy, 1932; Khanna *et al.*, 1976), but they are unwieldy in practice. Bone-conducted sound radiated into the ear canal is not the only way for speech vibrations to reach the cochlea; there are other mechanisms of bone conduction (inertial and compressional) that cannot be mitigated with human participants (Tonndorf, 1972). Bone-conducted sound can have a significant influence on the net signal at the cochlea for frequencies in the range of F1 (Porschmann, 2000). To try and control the relative influence of the unprocessed bone-conducted voice, participants were discouraged from speaking above the normal conversational level used to set microphone gain, and headphone feedback was designed to dominate the response by presentation at a relatively high level (80 dBA SPL at each ear). Bone conduction transfer functions are highly individual (Purcell *et al.*, 2003), and it is possible that in some individuals the bone-conducted voice was sufficient to attenuate the desired illusion of production errors.

The present data are generally consistent with the idea that a single type of feedback system governs both F0 and formant production. Under similar testing conditions, both fundamental frequency and formants show rapid responses to feedback changes and show evidence for adaptation (Jones and Munhall, 2000; Houde and Jordan, 1998). The similarity of operational principles that the data implies does not mean that a single system is necessarily controlling both aspects of speech. Indeed, the intermediate vowel /ɛ/ may be more labile than point vowels since it does not benefit from a saturation effect (Svirsky and Tobey, 1991; Perkell *et al.*, 2000). It also can be shown that the nervous system is capable of learning and simultaneously maintaining more than one independent motor controller (Ghahramani and Wolpert, 1997). In order to test this possibility for speech motor control, studies will have to be specifically designed to test learning and interactions between multiple models of auditory feedback (Wolpert and Kawato, 1998). In addition, the conditions that modulate the actions of these feedback systems over time must be addressed.

## V. CONCLUDING REMARKS

The study reported here employed paradigms from the vocal pitch-shift literature to evaluate whether the auditory vocal system compensates for sudden changes in the first formant of vowels. Partial compensation took place similar to that reported for manipulations of F0. The response was quite variable, and was smaller than previously measured in formant adaptation studies with whispered speech. The presence of immediate compensations in F0 and formant frequency, as well as sensorimotor adaptation in both parameters, suggests that the auditory feedback system for both aspects of speech is similar.

## ACKNOWLEDGMENTS

The authors would like to thank Reiner Wilhelms-Tricarico and Richard McGowan for their assistance in the development of the formant tracking and filtering algorithms. Research supported by the National Institute on Deafness and Other Communication Disorders, and the (Canadian) Natural Sciences and Engineering Research Council.

Bauer, J. J., and Larson, C. R. (**2003**). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," J. Acoust. Soc. Am. **114**, 1048–1054.

Beckman, M. E., Jung, T. P., Lee, S. H., Dejong, K., Krishnamurthy, A. K., Ahalt, S. C., Cohen, K. B., and Collins, M. J. (**1995**). "Variability in the production of quantal vowels revisited," J. Acoust. Soc. Am. **97**, 471–490.

Békésy, G. v. (**1932**). "Zur theorie des hörens bei der schallaufnahme durch knochenleitung," Ann. Phys. **13**, 111–136.

Binnie, C. A., Daniloff, R. G., and Buckingham, H. W., Jr. (**1982**). "Phonetic disintegration in a five-year-old following sudden hearing loss," J. Speech Hear. Disord. **47**, 181–189.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch-shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). "Voice F0 responses to pitch-shifted auditory feedback: A preliminary study," J. Voice **11**, 202–211.

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.

Cowie, R., and Douglas-Cowie, E. (**1992**). *Postlingually Acquired Deafness* (Mouton de Gruyter, New York), p. 304.

Cowie, R., Douglas-Cowie, E., and Kerr, A. G. (**1982**). "A study of speech deterioration in post-lingually deafened adults," J. Laryngol. Otol. **96**, 101–112.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice F0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Economou, A., Tartter, V. C., Chute, P. M., and Hellman, S. A. (**1992**). "Speech changes following reimplantation from a single-channel to a multichannel cochlear implant," J. Acoust. Soc. Am. **92**, 1310–1323.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Ghahramani, Z., and Wolpert, D. M. (**1997**). "Modular decomposition in visuomotor learning," Nature **386**, 392–395.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Held, R. (**1965**). "Plasticity in sensory-motor systems," Sci. Am. **213**, 84–94.

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 Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

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.

Kent, R. D., and Forner, L. L. (**1979**). "Developmental study of vowel formant frequencies in an imitation task," J. Acoust. Soc. Am. **65**, 208–217.

Kewley-Port, D. (**2001**). "Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training," J. Acoust. Soc. Am. **110**, 2141–2155.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Kewley-Port, D., and Zheng, Y. (**1999**). "Vowel formant discrimination: Towards more ordinary listening conditions," J. Acoust. Soc. Am. **106**, 2945–2958.

Khanna, S. M., Tonndorf, J., and Queller, J. E. (**1976**). "Mechanical parameters of hearing by bone conduction," J. Acoust. Soc. Am. **60**, 139–154.

Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich, Inc., New York).

Ladefoged, P., Silverstein, R., and Papcun, G. (**1973**). "Letter: Interruptibility of speech," J. Acoust. Soc. Am. **54**, 1105–1108.

Lane, H., and Webster, J. W. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**, 859–866.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparison of voice F0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**, 2845–2848.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). "Effects of pitch-shift velocity on voice F0 responses," J. Acoust. Soc. Am. **107**, 559–564.

Leder, S. B., and Spitzer, J. B. (**1993**). "Speaking fundamental frequency, intensity, and rate of adventitiously profoundly hearing-impaired adult women," J. Acoust. Soc. Am. **93**, 2146–2151.

Leder, S. B., Spitzer, J. B., and Kirchner, J. C. (**1987**). "Speaking fundamental frequency of postlingually profoundly deaf adult men," Ann. Otol. Rhinol. Laryngol. **96**, 322–324.

Leydon, C., Bauer, J. J., and Larson, C. R. (**2003**). "The role of auditory feedback in sustaining vocal vibrato," J. Acoust. Soc. Am. **114**, 1575–1581.

Liu, C., and Kewley-Port, D. (**2004**). "Formant discrimination in noise for isolated vowels," J. Acoust. Soc. Am. **116**, 3119–3129.

Miall, R. C., and Wolpert, D. M. (**1996**). "Forward models for physiological motor control," Neural Networks **9**, 1265–1279.

Moore, B. C. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.

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.

Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.

Orfandidis, S. J. (**1988**). *Optimum Signal Processing, An Introduction* (MacMillan, New York).

Perkell, J., Lane, H., Svirsky, M., and Webster, J. (**1992**). "Speech of cochlear implant patients: a longitudinal study of vowel production," J. Acoust. Soc. Am. **91**, 2961–2978.

Perkell, J. S., and Nelson, W. L. (**1985**). "Variability in production of the vowels /i/ and /a/," J. Acoust. Soc. Am. **77**, 1889–1895.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonetics **28**, 233–272.

Pisoni, D. B. (**1980**). "Variability of vowel formant frequencies and the quantal theory of speech: A first report," Phonetica **37**, 285–305.

Porschmann, C. (**2000**). "Influences of bone conduction and air conduction on the sound of one's own voice," Acta Acust. **86**, 1038–1045.

Purcell, D. W., Kunov, H., and Cleghorn, W. (**2003**). "Estimating bone conduction transfer functions using otoacoustic emissions," J. Acoust. Soc. Am. **114**, 907–918.

Repp, B. H., and Williams, D. R. (**1987**). "Categorical tendencies in imitating self-produced isolated vowels," Speech Commun. **6**, 1–14.

Schenk, B. S., Baumgartner, W. D., and Hamzavi, J. S. (**2003**). "Effect of the loss of auditory feedback on segmental parameters of vowels of postlingually deafened speakers," Auris Nasus Larynx **30**, 333–339.

Siegel, S., and Castellan, N. J. (**1988**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, Boston).

Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (**2005**). "Voice responses to changes in pitch of voice or tone auditory feedback," J. Acoust. Soc. Am. **117**, 850–857.

Stevens, K. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.

Svirsky, M. A., and Tobey, E. A. (**1991**). "Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users," J. Acoust. Soc. Am. **89**, 2895–2904.

Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.

Tonndorf, J. (**1972**). "Bone Conduction," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 195–237.

Vallabha, G. K., and Tuller, B. (**2002**). "Systematic errors in the formant analysis of steady-state vowels," Speech Commun. **38**, 141–160.

Vallabha, G. K., and Tuller, B. (**2004**). "Perceptuomotor bias in the imitation of steady-state vowels," J. Acoust. Soc. Am. **116**, 1184–1197.

Villacorta, V., Perkell, J. S., and Guenther, F. H. (**2004**). "Sensorimotor adaptation to acoustic perturbations in vowel formants," J. Acoust. Soc. Am. **115**, 2430.

Villacorta, V., Perkell, J. S., and Guenther, F. H. (**2005**). "Relations between speech sensorimotor adaptation and perceptual acuity," J. Acoust. Soc. Am. **117**, 2618.

Waldstein, R. S. (**1990**). "Effects of postlingual deafness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**, 2099–2114.

Wolpert, D. M., and Kawato, M. (**1998**). "Multiple paired forward and inverse models for motor control," Neural Networks **11**, 1317–1329.

Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.