Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation

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

Kevin G. Munhall
Department of Psychology and Department of Otolaryngology, Queen's University, Kingston, Ontario K7L 3N6, Canada

(Received 10 February 2006; revised 26 May 2006; accepted 31 May 2006)

Auditory feedback during speech production is known to play a role in speech sound acquisition and is also important for the maintenance of accurate articulation. In two studies the first formant (F1) of monosyllabic consonant-vowel-consonant words (CVCs) was shifted electronically and fed back to the participant very quickly so that participants perceived the modified speech as their own productions. When feedback was shifted up (experiment 1 and 2) or down (experiment 1) participants compensated by producing F1 in the opposite frequency direction from baseline. The threshold size of manipulation that initiated a compensation in F1 was usually greater than 60 Hz. When normal feedback was returned, F1 did not return immediately to baseline but showed an exponential deadaptation pattern. Experiment 1 showed that this effect was not influenced by the direction of the F1 shift, with both raising and lowering of F1 exhibiting the same effects. Experiment 2 showed that manipulating the number of trials that F1 was held at the maximum shift in frequency (0, 15, 45 trials) did not influence the recovery from adaptation. There was a correlation between the lag-one autocorrelation of trial-to-trial changes in F1 in the baseline recordings and the magnitude of compensation. Some participants therefore appeared to more actively stabilize their productions from trial-to-trial. The results provide insight into the perceptual control of speech and the representations that govern sensorimotor coordination. *© 2006 Acoustical Society of America.* [DOI: 10.1121/1.2217714]

PACS number(s): 43.66.Jh, 43.70.Aj [AL] Pages: 966–977

I. INTRODUCTION

Human movement shows a striking adaptability to a variety of conditions. We walk on surfaces that vary in regularity, traction, and compliance. We can control our arms and hands under a variety of lighting conditions and loads. We can speak at different rates and volumes, and even with pens clasped between our teeth. This adaptability could not exist without accurate perception of the movement environment, but it has long been known that strictly sensory-controlled movement could not account for coordination in rapid skills such as speech (e.g., Lashley, 1951). Closed-loop control schemes are too slow and lack stability. In recent years, motor research has focused on the role of "internal models" in accounting for such skilled movement. Internal models are neural representations of the kinematics, dynamics, and sensory consequences of movement that are thought to play a role in motor planning and control (Tin and Poon, 2005). According to this view, sophisticated internal models are learned through practice and are used to facilitate motor control when sensory feedback is insufficient. These representations permit the motor system to predict some of the internal and external conditions that could contribute to variability, and plan accordingly. The adaptability of motor skill over time is attributed to the plasticity of these internal models and their role in sensorimotor learning. In this paper we address how an internal representation of speech acoustics adapts when acoustic feedback is modified during speech production.

The first proposal for an internal model of this kind for speech production was made by Kawato (1989), though similar ideas of a motor program, or motor image or schema, have long histories in speech research. In the years since Kawato's paper, the concept of an internal model has been drawn on numerous times (Guenther, 1995; Perkell *et al.*, 1997; Munhall *et al.*, 2000; Jones and Munhall, 2000; Tremblay *et al.*, 2003). In spite of this popular support for the role of internal models in speech, there are few confirmed specifics about how these hypothetical mechanisms might work (cf. Guenther, 2003).

One of the primary candidates for an internal model in speech is an auditory representation. Auditory feedback is too slow to be used in moment-to-moment control, but hearing your own speech is an essential part of talking. The evidence that hearing has a strong impact on speech motor control comes from both clinical and laboratory studies. Babbling (Oller and Eilers, 1988) and learning to talk (Smith, 1975) are impaired in individuals with congenital hearing impairments. Postlingually deafened adults also show changes in many aspects of their speech (e.g., Wald-

[a]Author to whom correspondence should be addressed. Currently affiliated with the National Centre for Audiology at the University of Western Ontario. Electronic-mail: purcelld@nca.uwo.ca

stein, 1990; Lane and Webster, 1991; Leder and Spitzer, 1993; Schenk *et al.*, 2003). Laboratory-induced manipulations to the loudness (Lane and Tranel, 1971), timing (Smith and Smith, 1962), fundamental frequency (Kawahara, 1995; Burnett *et al.*, 1998), or formant frequencies (Houde and Jordan, 1998, 2002) of auditory speech feedback all produce rapid changes in speech production.

In this paper we extend Houde and Jordan's (2002) approach by examining how individuals respond to manipulations of the first formant frequency of the vowels in monosyllables. Houde and Jordan's work was carried out with whispered speech, but the studies here and elsewhere (Purcell and Munhall, 2006; Villacorta *et al.*, 2004, 2005) have demonstrated the viability of manipulating voiced stimuli successfully. The present studies are modeled after sensorimotor manipulation studies outside of the speech literature, such as the use of prisms to displace the visual field and interfere with the visuomotor control of reaching (e.g., Stratten, 1896; Held, 1965), or the use of novel force fields to study the dynamics of motor control (Shadmehr, 2004). In the presence of a perceptual manipulation in these studies, the motor system compensates to achieve a goal. This compensation, of course, could occur if the movement was only guided in real time by sensory information. However, when the perceptual manipulation is removed an aftereffect or adaptation to the perceptual manipulation is observed. This persistence or learning in the motor system does not disappear immediately, indicating that a representation is involved in the control scheme.

Houde's approach, and the one we follow here, involves producing small incremental shifts in the frequency of the formants of vowels that participants hear themselves producing. Over successive trials the frequency of the produced and heard speech become farther and farther apart. Both compensation and adaptation have been demonstrated for formants (Houde and Jordan, 1998, 2002) and fundamental frequency (Jones and Munhall, 2000, 2002, 2005) with this paradigm. In the two experiments presented in this paper, we test fundamental aspects of the auditory-motor system: the sensitivity of the system to manipulations in different directions in the vowel space (experiment 1), the threshold for response to auditory manipulation (experiments 1 and 2), and the persistence of short-term alterations in the auditory-motor mapping (experiment 2).

II. EXPERIMENT 1

When an individual produces a series of the same vowel, the formant frequencies vary from trial to trial, but cluster around an average frequency, or possible vowel target, for that talker. Evidence shows that auditory feedback influences each of these vowel targets in the talker's vowel space. For example, individuals with cochlear implants may experience expanded vowel spaces following activation of their implants (e.g., Svirsky and Tobey, 1991; Perkell *et al.*, 1992; Economou *et al.*, 1992; Svirsky *et al.*, 1992), and talkers who discriminate auditory vowel contrasts more accurately also produce vowels with less variability in the formant frequencies (Perkell *et al.*, 2004). Thus, some type of auditory-motor mapping must be at play during vowel production.

In this paper we address two aspects of this mapping. First, we tested the threshold formant frequency at which participants began to compensate for a discrepancy between produced and heard vowel feedback. When feedback for F0 is shifted by very small amounts, there is at first no obvious response by the speech production system (Jones and Munhall, 2000). In this pitch-shift experiment, the acoustic feedback was systematically shifted in frequency from the spoken acoustics in small steps across utterances. Initially, participants did not change their production even though there was a widening frequency gap between uttered and heard speech. At some point the discrepancy between spoken acoustics and the feedback appears to cross some threshold and there is evidence of compensation in the participant's speech. We will call this point the **compensation threshold** and it presumably reflects a tolerance bound for sensorimotor control. While a great deal is known about psychoacoustic thresholds for formant frequencies (Kewley-Port and Watson, 1994; Kewley-Port *et al.*, 1996; Kewley-Port and Zheng, 1999; Kewley-Port, 2001; Liu and Kewley-Port, 2004), little is known about compensation thresholds, which must reflect a characteristic of the internal model for the auditory-motor control of speech.

The second goal of this study was to test whether the sensitivity to feedback manipulation is symmetrical around the talker's central formant tendency. Vowel spaces are not homogeneous in the distribution of vowel spacing (Lindblom, 1986; Maddieson, 1984), and factors such as proximity of adjacent vowels many influence the tolerance or compensation thresholds for individual vowels. In addition, the spatial variance of tongue movements during vowel production differs for vowel qualities and differs along the length of the tongue, presumably due to the properties of the tongue and its motor control (Perkell, 1996). All of these factors combined may introduce nonlinearities in the auditory feedback system.

A. Participants

Ten female participants varying in age from 18 to 24 years were tested in a single session. For each ear, hearing thresholds were measured at octave frequencies from 500 to 4000 Hz. All individuals had normal thresholds ($\leqslant$20 dB HL). No participants had known language or speech impairments, and all had learned English as their first language.

B. Equipment

Equipment was similar to that previously reported in Purcell and Munhall (2006). Participants' speech was transduced into an electrical signal with a type WH20 Shure headset microphone. A Tucker-Davis Technologies MA3 microphone amplifier with the +20-dB gain switch active was used to amplify the microphone signal. This signal was low-pass filtered with an analogue Frequency Devices type 901 filter using a cutoff frequency of 4500 Hz and gain of 0 dB. The filtered signal was digitized at 10 kHz with 16-bit precision using a National Instruments PXI-6052E input/output board mounted in a PXI-1002 chassis. Real-time analysis and filtering of the voice was achieved using a National Instruments PXI-8176 embedded controller, and the altered voice was converted back to analogue by the 6052E at 10 kHz with 16-bit precision. A second Frequency Devices unit was employed to low-pass filter the altered feedback voice as above. A Madsen Midimate 622 audiometer added speech noise and amplified the signal for bilateral presentation of the same stimulus through Sennheiser "HD 265 linear" headphones. During practice trials, the microphone MA3 amplifier was adjusted between 30 and 50 dB for each individual so that vocalizations caused the Madsen input VU meter to read approximately 0 dB. Audiometer output gain was set so the headphone voice signal at each ear was approximately 80 dBA sound pressure level (SPL) with background speech shaped noise of approximately 50 dBA SPL. Calibration was performed using a Brüel &Kjær sound level meter and artificial ear Type 4153.

C. Experimental conditions

When normal feedback was not provided, the first formant (F1) of the vowel /ε/ was altered. Participants produced this vowel at normal speed in the CVC context of the word "head," and F1 was altered for the entire vowel (and incidentally the vocalic portion of the consonant "d"). Utterances were collected in sets of trials that will be referred to as blocks. Prior to any alteration of auditory feedback, baseline or "start" trials were collected with normal feedback at the beginning of each experimental block. These were followed by "ramp" trials where the magnitude of the shift was slowly increased in 4-Hz steps over 50 utterances. Shift magnitude was constant throughout each trial and was changed while the subject paused between utterances. Thus the manipulation of F1 was gradually changed from 0 Hz to a maximum of either ±200 Hz over the course of the 50 ramp trials. A positive shift was towards the vowel /æ/, and a negative shift was towards /ɪ/. The ramp trials were followed by "hold" trials where the maximum feedback change of ±200 Hz was employed. Auditory feedback was returned to normal with a step change following the hold trials. These last trials collected in each block with normal feedback were referred to as "end" trials. Both experiments employed naive subjects, and in an exit interview about 40% of participants were aware of some kind of change in the auditory feedback over the course of the experiment. The step change between the maximum formant shift and normal feedback (between the hold and end trials) was the point in the experiment where the manipulation would have been most apparent. Only 8% of individuals correctly identified that their vowels had been changed in the auditory feedback. Noticing the presence of feedback changes was not related to the pattern of compensatory behavior.

D. Experimental protocol

Experiments were performed in an Industrial Acoustics Company (IAC) sound-insulated room with participants seated in a comfortable chair. Individuals were asked to produce the words prompted on a video display using their normal speaking rate and level (vowels were not extended). Speaking level was monitored on the audiometer VU meter and was generally consistent over the course of an experiment. Each prompt lasted 1.5 s, and the interprompt interval was approximately 2.5 s.

At the start of each experiment, an automated screening program was used to prompt the participant to speak five extended tokens of seven English vowels spanning the vowel space (Ladefoged, 1982) in a CVC context (all /hVd/). The screening protocol was used to establish the best parameters for tracking the formants of each participant. The stability of the F1 estimate for the vowel /ε/ was evaluated, and the best linear predictive coding (LPC) model order in the range 8 through 12 (inclusive) was determined (Vallabha and Tuller, 2002). The best model order was selected as the order that produced F1 estimates with the smallest standard deviation (SD). This model order was then used for the rest of the experiment.

E. Online formant shifting and detection of voicing

Detection of voicing and formant shifting were performed as previously described in Purcell and Munhall (2006). Briefly, the manipulation of auditory feedback was achieved by filtering the voice in real-time. A simple statistical amplitude threshold technique was used to detect the onset of voicing in each trial. The first step in manipulating auditory feedback was to determine formants in the speech using an iterative Burg algorithm for estimating spectral parameters (Orfandidis, 1988). The sliding analysis window used in this procedure weighted older samples with an exponential decay chosen such that 50% of the area under the weighting curve applied to samples less than 8.6 ms old. The National Instruments system was capable of performing a new formant estimate every nine speech samples. This estimate of F1 was used to calculate filter coefficients such that a pair of spectral zeros deemphasized the existing formant, and a pair of spectral poles emphasized existing energy in the voice in the frequency region of the new desired formant. These filter coefficients were updated with each new formant estimate about every 900 μs.

F. Offline formant analysis

Prior to estimating formants offline, the record of each utterance was trimmed both before and after the vowel using a supervised semi-automated process. Rare vowels that were shorter than the mean duration minus two SDs were removed from further analysis, as were any trials where the participant clearly stumbled or failed to utter the correct word. The final preprocessing step was to trim all vowels to the duration of the shortest vowel in the set by truncating the tail. Vowel formants were calculated offline by sliding the analysis window ten speech samples (1 ms) per estimate, and used the same algorithm as the on-line experiment. Using the best model order for F1 of /ε/ helped reduce gross errors in tracking, however for some participants F2 was occasionally misinterpreted as F1. These bad estimates were removed using a histogram method where bins were declared unusable if they

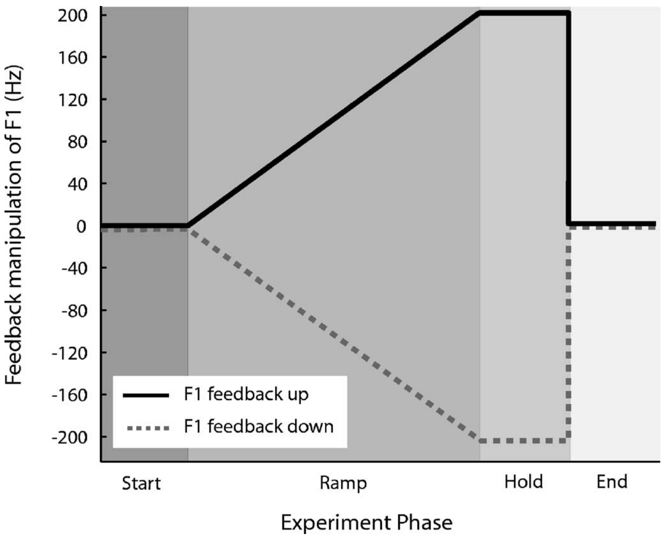

FIG. 1. Diagram of the time course of the manipulations made to the feedback of formant F1 from utterance to utterance. The four phases of the experiment (start, ramp, hold, and end) are indicated with shading of increasing lightness. The solid line is for the experimental condition where F1 of the test vowel /ɛ/ was pushed upwards towards the vowel /æ/, and the dash line indicates where F1 was shifted downwards towards /ɪ/.

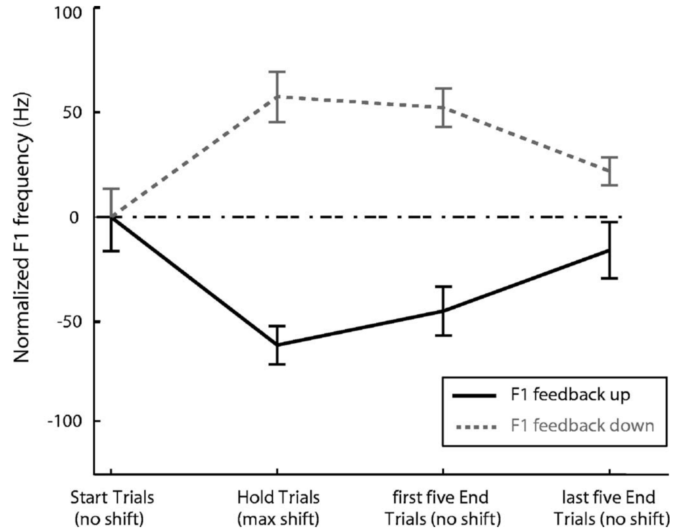

FIG. 2. Average changes in production measured with the microphone in response to the feedback manipulations in Fig. 1. The response is normalized to the average baseline F1 measurement in the start phase of the experiment. Average values are reported for the start, hold, and end phases of the experiment. The end phase has been subdivided into the means of the first and last five trials of that phase. Error bars are ±1 standard error. Note that the change in F1 production was in the opposite direction of the feedback manipulation in Fig. 1.

had counts of less than 5% of the average mode and were at least 150 Hz distant from the average mode. A single "steady-state" F1 value was determined for each trial by averaging F1 estimates in a 20-ms window (i.e., 20 estimates) beginning 50% of the way through the vowel. Mean values reported in the tables and figures were calculated by averaging these steady-state values across the relevant trials and participants.

G. Response evaluation

Changes in the production of F1 were evaluated against the average of the start trials that had normal auditory feedback. Compensation was calculated as the change in the production of F1 between the hold trials with the maximum feedback shift, and the start trials. Adaptation was calculated as the difference between the end trials with normal feedback, and the start trials. In both experiments, the onset of a response was estimated using the change point test (Donath *et al.*, 2002; Siegel and Castellan, 1988) applied to the steady-state F1 estimates.

H. Procedure

The ten participants were prompted to say the word "head" a total of 95 times for each of two shift directions. Half of the participants served first in a condition in which auditory feedback was shifted upward and then they experienced the same design with auditory feedback shifted downward. The order of shift direction was counterbalanced across participants. For each shift direction, auditory feedback was normal for 15 start trials. This was followed by 50 ramp trials where feedback was shifted to a maximum of either ±200 Hz. A further 15 hold trials were collected at the maximum shift, followed by 15 end trials with normal feedback. This is shown schematically in Fig. 1. After a short break, each participant then repeated the experiment, but with the manipulation in the opposite direction.

I. Results and discussion

On average, production of F1 changed in a direction opposite to the manipulation, as though the vocal control system were attempting to compensate for the manipulation. For both directions of F1 frequency manipulation the largest compensation occurred during the hold trials while altered feedback was at its maximum. However, even during the hold trials only partial compensation was observed. When feedback was returned to normal, F1 production did not immediately return to start values. In response to the auditory feedback manipulations in Fig. 1 (the stimulus), the mean F1 compensation and adaptation values are shown in Fig. 2 and Table I.

Repeated measures analysis of variance (ANOVA) using Greenhouse-Geisser corrections with factors of manipulation direction (up or down) and experiment phase (start, hold, and first or last five end trials) showed that there was a significant effect of direction $[F(1,7)=57.2, p<0.001]$ and a significant interaction between direction and experiment phase $[F(1.6,10.9)=14.5, p<0.001]$. There was no main effect of experiment phase because the responses were in opposite directions depending on whether feedback was altered up or down. Scheffé's method was used to detect significant differences between experiment phases for each manipulation direction. Hold trials were significantly different from start trials for both shift directions [up $F(3,27)=9.7$, $p<0.001$; down $F(3,21)=7.0, p<0.01$]. There was also a significant difference between the start and the first five end trials [up $F(3,27)=5.2$, $p<0.01$; down $F(3,21)=5.8$, $p<0.01$], but not the last five. Using a separate variance t test, there was also no significant difference between start trials for the two

TABLE I. Average data from experiment 1. There are ten individuals in the mean for the +200-Hz condition. The formant tracker was unstable for two participants when F1 was pushed −200 Hz, and this average therefore includes data from only eight individuals. The mean change points used one fewer participant in each condition because one person in each had an irrational change point (before the manipulation). All values are in Hz with between-subject standard error in parentheses. Compensation is the difference between mean F1 of the start and hold trials times the sign of the manipulation. Adaptation is the difference between the mean F1 of the start and end trials times the sign of the manipulation.

Condition	F1+200 Hz	F1−200 Hz
Measurement		
Start F1	775 (16.3)	769 (13.6)
F1 change point	64 (12.1)	89 (10.3)
Hold F1	713 (9.2)	827 (12.0)
First five end F1	730 (11.7)	822 (9.2)
Last five end F1	759 (13.5)	791 (6.7)
Compensation	61 (15.4)	58 (16.7)
Adaptation first five end	45 (15.0)	52 (16.6)
Adaptation last five end	16 (14.6)	22 (14.2)

manipulation directions. Although there was no statistical difference between the start trials for the two measurement directions, and no difference between the start trials and the last five end trials for a given direction, the means of the last five end trials in Fig. 2 suggest that production had not yet returned to baseline values. The order in which the manipulation direction was applied might therefore have had a subtle effect on an individual's production in the second part of the test. That is, the baseline F1 in the start trials of the second measurement block may have been slightly shifted in the direction of that block's manipulation as a remnant of the first measurement. This recovery from adaptation or deadaptation is explored in more detail in experiment 2.

The change point, or point during the ramp trials that production first changed significantly from the baseline value of the start trials, was identified using the change point test with the start and ramp trials. The mean of the individual participant change points is given for each manipulation direction in Table I. This measure will be considered a compensation threshold. There was no statistical difference between the mean change points for the different shift directions, as evaluated with a separate variance t test ($p < 0.14$). This compensation threshold (mean of both directions was 76 Hz) is above psychoacoustic formant thresholds observed for untrained individuals with background noise (Kewley-Port and Zheng, 1999; Kewley-Port, 2001) (recall noise level here was 50 dBA). However, this difference should be viewed as preliminary. A within-subject design, in which participants perform both tasks, is required because the responses to altered feedback and psychoacoustic thresholds vary markedly between participants. In the present data, compensation to the 200-Hz manipulation of F1 ranged from 9 to 126 Hz. That is to say, whereas one individual modified their F1 production by 9 Hz in response to this manipulation (essentially no significant change in production given the trial to trial variability), another modified theirs 14 times more. A large range of response magnitude is consistent with that observed in pitch-shift studies (e.g., Burnett *et al.*, 1998; Liu and Kewley-Port, 2004; Sivasankar *et al.*, 2005), but the reasons for such between-participant differences in susceptibility to altered feedback cannot be accounted for at present. In Kewley-Port and Watson's (1994) psychoacoustic difference thresholds for well-trained individuals, there were some large individual differences in performance for F2, but the range in F1 was generally smaller.

For both the size of the compensation and the compensation threshold, there were no statistically significant differences in the response for raising or lowering the frequency of F1 feedback. As shown in Table I, the compensation and adaptation values were similar (in Hz) for both directions. The directional sensitivity of the speech motor system therefore appears to be similar for F1 of the vowel /ɛ/. However, whether this symmetry generalizes to other vowels is unknown. Point vowels such as /æ/ that have no competing vowels in one F1 frequency direction may behave differently. The issue of the size of the "effective" stimulus also must be considered.

Among the front vowels, the difference in F1 between /ɪ/ and /ɛ/ is not the same as it is from /ɛ/ to /æ/. The values reported by Baken and Orlikoff (2002, Table 7-1, p. 260) between /ɛ/ and /ɪ/ for men and women are 140 and 180 Hz, respectively. The F1 differences between /ɛ/ and /æ/ for men and women are −130 and −250 Hz. If the vowel formant space is defined functionally by vowel categories, then linear manipulations like the ±200 Hz used here may not be treated uniformly by the auditory-vocal feedback system. Given the unsymmetrical F1 positioning of /ɪ/ and /æ/ in the vowel space with respect to /ɛ/ for women, it might be expected that the functional size of the manipulation employed here was larger for the shift from /ɛ/ towards /ɪ/. In the start trials, the measured average F1 difference between /ɛ/ and /ɪ/ was 174 Hz. Between /ɛ/ and /æ/ the mean difference was −217 Hz. Therefore on average, the manipulation may have been slightly functionally larger, with respect to this group's vowel categories, for the downward shift of F1 towards /ɪ/. Of course individual F1 spacing of the front vowels varies, so the relative functional manipulation size would change with the speaker. This issue is reminiscent of discussions of the metric for expressing the perceptual distance between vowels (e.g., Lindblom, 1986).

III. EXPERIMENT 2

Experiment 1 showed symmetrical compensation thresholds for raising and lowering of the first formant, and also demonstrated adaptations in both frequency directions that did not immediately return to the baseline level. In this experiment, we explored the dynamics of the deadaptation phase by recording a longer "end" phase in the experiment. The 15 trials recorded in experiment 1 were not enough to track the recovery from manipulation. Here, we also manipulated the length of the hold phase to begin to explore what conditions are influencing the adaptation persistence.

The effects of sensorimotor adaptation can persist for hours, days, and even longer in some cases, depending on the length of exposure to the manipulated sensory conditions, type of transformation, and activity following return to normal sensory conditions (Darainy *et al.*, 2006; Shadmehr and Brashers-Krug, 1997; Caithness *et al.*, 2004). For example, the vestibular-motor system of astronauts adapts during space flight and does not fully return to normal for more than a day after the return to earth's gravitational field (Paloski *et al.*, 2004). On the other hand, under some conditions the recovery from adaptation in saccadic eye movements that have been adapted using a target-jump paradigm can be quite rapid (Gaveau *et al.*, 2005).

There is little evidence from speech adaptation studies on the persistence of the effects. Jones and Munhall (2005) showed that the effect of adapted fundamental frequency during the production of Mandarin tones did not disappear in a short block of trials following return to normal feedback. Houde and Jordan (2002) report the puzzling observation that participants who had participated in an adaptation experiment showed baseline shifts in the direction of adaptation when they returned to the laboratory for recordings one month later. The baselines of the second phase of experiment 1 showed the same trend. This is not evident in the baseline start trials of Fig. 2 due to normalization, but it can be seen that the mean of the last five end trials plotted in Fig. 2 did not quite reattain the baseline, despite normal auditory feedback. In both Houde and Jordan's study and experiment 1, the gap between testing was filled with uncontrolled, natural speaking. Experiment 2 examined deadaptation under controlled conditions in which the number and content of participant productions were part of the design.

A. Participants

Forty-one participants varied in age from 18 to 23 year (nine females and five males in each of the hold 0 and hold 45 conditions, eight females and five males in the hold 15 condition). For each ear, hearing thresholds were measured at octave frequencies from 500 to 4000 Hz. Most individuals had normal thresholds ($\leqslant$20 dB HL), however there were two participants (one male and one female from the hold 45 condition) who had thresholds of 25 dB HL in one ear at one frequency. These individuals had typical responses and normal thresholds otherwise and were therefore included in the analysis. No participants had known language or speech impairments, and all had learned English as their first language.

B. Procedure

In this experiment F1 was shifted only in the upward direction to a maximum of +200 Hz (towards the vowel /æ/), and there were 20 start, 50 ramp, and 115 end trials. The first five start trials were not included in averages so that 15 trials were used as in experiment 1. Three different hold conditions were employed in a between-participants design with 0, 15, or 45 hold trials at the maximum feedback shift (the three conditions employed 14, 13, and 14 individuals, respectively). In the case of the hold 0 condition, the last five ramp trials were used to estimate compensation. All other features of the experiment were the same as in experiment 1.

TABLE II. Average data from experiment 2. There are 14 individuals in the means for the hold 0 and hold 45 conditions, and 13 for the hold 15 condition. The mean change point for each hold condition was the average of one fewer participant because either an individual did not have a statistically significant change point, or the change point was irrational (before the manipulation). All values are in Hz with between-subject standard error in parentheses. Compensation is the difference between the mean F1 of the start and hold trials. Residual adaptation is the difference between the mean F1 of the start trials and the last half of the end trials.

Condition	Hold 0	Hold 15	Hold 45
Measurement			
Start F1	670 (25.6)	693 (29.0)	692 (30.6)
F1 change point	73.5 (8.2)	79.0 (6.9)	93.5 (11.0)
Hold F1	616 (24.9)	624 (27.4)	645 (26.1)
Last half of end trials F1	647 (28.3)	681 (30.7)	672 (27.4)
Compensation	54 (9.5)	69 (8.1)	47 (11.0)
Adaptation last half of end trials	23 (10.2)	12 (7.9)	20 (10.5)

C. Results and discussion

Mean responses for the three different hold conditions are given in Table II. As in experiment 1, the average F1 production changed in the opposite direction of the manipulation of F1 in the auditory feedback. In experiment 2, F1 was raised in frequency (+200 Hz), so production of F1 in the hold trials tended to be lower than in the start trials. This effect extended well into the end trials that had normal feedback. The average response across participants for every trial is plotted in Fig. 3 for the hold 45 condition. Data for the hold 0 and 15 conditions were similar, but necessarily had fewer hold trials with the maximum feedback shift.

In the end trials shown in Fig. 3, F1 returns towards its original value in the start trials but does not reach the baseline level. The distribution of end trials for each of the three hold conditions is plotted in Fig. 4. Data from the different

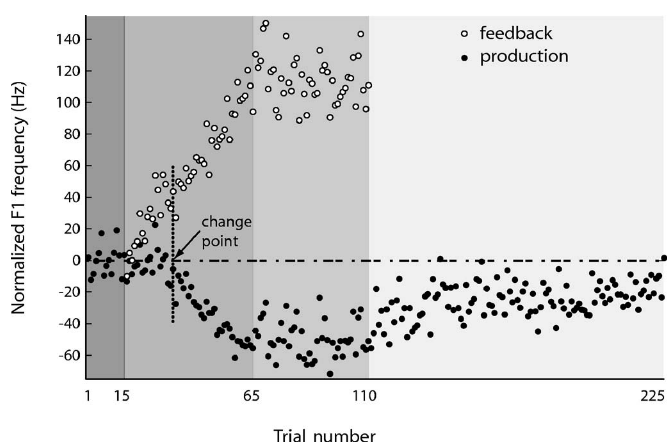

FIG. 3. Average F1 values across participants for each trial from the hold 45 condition of Experiment 2. Different phases of the experiment are again indicated with shading of increasing lightness as in Fig. 1 (from left to right these are start, ramp, hold, and end phases). Formant values are normalized with respect to the average baseline F1 in the start phase. Filled circles indicate production values of F1, whereas the open circles demarcate feedback F1 in the headphones (only shown where feedback differed from production). An arrow indicates the change point where production changed significantly from the baseline value in the start phase. This occurred when the feedback was altered 76 Hz on the 9th trial of the ramp phase.

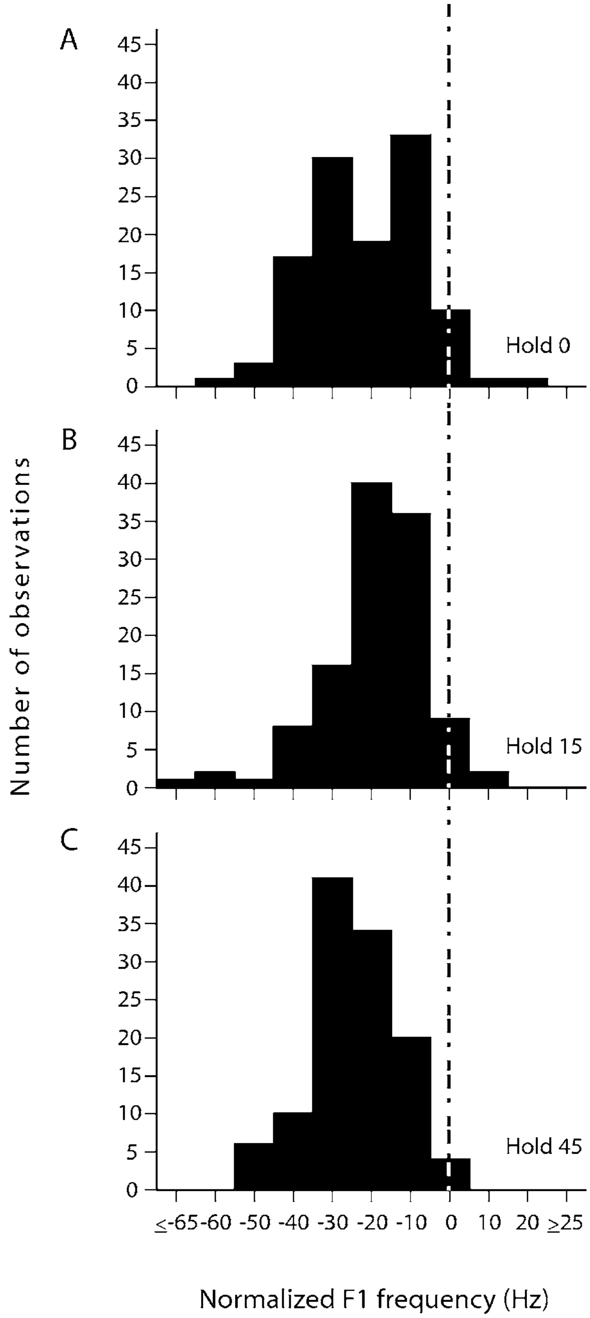

FIG. 4. Histograms of produced F1 in the end trials of all three hold conditions [panels (a), (b), and (c) are for hold conditions 0, 15, and 45, respectively]. Each bin was 10 Hz wide, and bin centers are labeled. Formant values are normalized with respect to the average baseline F1 in the start phase. The distributions of end trials overlap substantially across hold conditions, and all are shifted left (negative, or in the direction of compensation) relative to the average baseline F1 in the start phase.

hold conditions overlap substantially and did not differ in mean value ($p > 0.80$, Scheffé's method applied to within-subject means). In order to analyze the dynamics of recovery from adaptation, the end trials from the three hold conditions were averaged. As can be seen in Fig. 5, adaptation decayed relatively quickly from its maximum over the first 15 end trials with normal feedback. Subsequently, the production of F1 appeared to asymptote at about −20 Hz relative to F1 in the start trials. In this experiment, it was not possible to separate the effects of time and utterance number on the shape of the decay function because timing was rigidly paced. The average end trials were well modeled by a decaying exponential curve fit using a least squares cost function. Deadaptation in other sensorimotor systems shows similar exponential patterns (Davidson and Wolpert, 2004; Hopp and Fuchs, 2004).

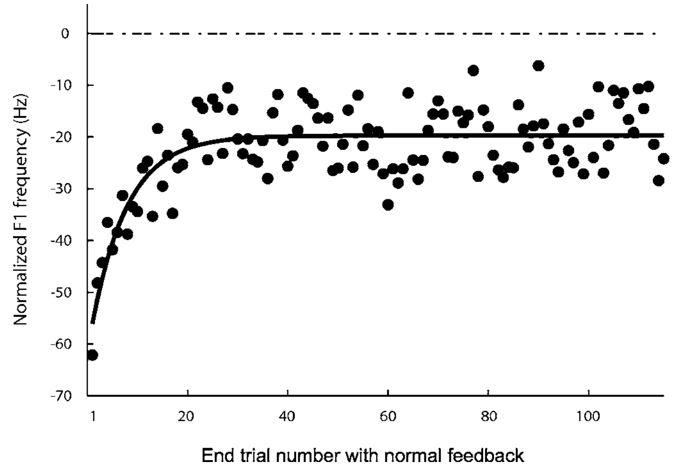

FIG. 5. Average F1 values across participants and hold conditions for each trial from the end phase of experiment 2. Each dot therefore represents mean data from 41 individuals for that trial. End phase trial numbers have been reindexed from 1 to 115 in this figure, but these correspond with trials 111 to 225 shown in Fig. 3. Formant F1 values are normalized with respect to the average baseline F1 in the start phase. An exponential has been fit to the values from each trial using a least squares cost function and is shown with a solid line (asymptote −19.8 Hz, intercept −61.7 Hz, and tau 0.141).

An ANOVA with between-participant factor of hold condition (0, 15, and 45 hold trials), and within-participant factor of experiment phase (start, hold, and last half of end trials) was calculated. The mean of the last half of the end trials was chosen as an adaptation metric to avoid the rapid decay of adaptation during the first 15 end trials as shown in Fig. 5. There was a significant effect of experiment phase [$F(2,76)=55.5, p<0.001$], but not of hold condition. *Posthoc* comparisons using Scheffé's method showed that the hold trials had significantly lower F1 values than the start trials, for all three hold conditions [hold 0, $F(2,26)=14.8$, $p<0.001$; hold 15, $F(2,24)=33.9$, $p<0.001$; hold 45, $F(2,26)=11.2$, $p<0.001$]. Combining the three hold conditions, the last half of the end trials were significantly different from the start trials [$F(2,80)=5.6, p<0.01$]. In summary, the number of hold trials had no apparent effect on the compensation achieved, or the decay of adaptation with normal feedback.

During the ramp trials F1 did not change from baseline as soon as the feedback shift began (e.g., Fig. 3). Rather, the onset of compensation (change point) was delayed until a large enough discrepancy between produced and heard speech occurred. The change point shown in Fig. 3, calculated with the grand average data for this condition, was at a feedback manipulation of 76 Hz (in Table II, the arithmetic average of individual change points is 93.5 Hz for this hold condition). An ANOVA with the factor of hold condition (0, 15, and 45 hold trials) showed no difference in change points

TABLE III. Within-participant F1 production variability estimated as the standard deviation (SD) from token to token in the present analysis and those previously reported. Only select front vowels are given; data for other vowels may be available in the studies listed. Where different age groups were available, those closest in age to the young adults of the present study were chosen.

Study	Vowels	Speakers	Repetitions	SD (Hz)
Present	/ɛ/	41	15	29
Peterson and Barney (1952)	Entire vowel space	61	2	23
Kent (1973)	/ɪ/,/æ/	4	10	On order of 45
Kent and Forner (1979)	/ɪ/,/æ/	5	5	<60
Repp and Williams (1987)	/ɪ/,/æ/	2	4	20-30
Vallabha and Tuller (2004)	Entire vowel space	3	10	27
Pisoni (1980)	/ɪ/,/ɪ/,/ɛ/,/æ/	2	10	17, 23, 24, 38
Perkell and Nelson (1985)	/æ/	1	24	31
Beckman *et al.* (1995)	/ɪ/,/æ/	1	Not explicitly given	21, 45
Hawkins and Midgley (2005)	/ɪ/,/ɛ/,/æ/	5	4	22, 47, 56
Waldstein (1990)	/ɪ/,/ɪ/,/ɛ/,/æ/	7	8	22, 28, 46, 50

between the three hold conditions. This was expected since the experiment was identical until the hold trials themselves were reached.

One obvious feature of the data reported here is the variability in the talkers' formant values. Within-participant trial-to-trial variability in F1 was estimated by calculating SD in the start phase. Across participants the average SD was 29 Hz. In the literature, between-participant variability is generally what is reported, but within-participant values are available and are presented in Table III. Peterson and Barney (1952) recorded two tokens per speaker, and while this is a small set, their raw F1 data from across the English vowel space can be used to underestimate within-participant SD values. Select vowels are also available from the literature on the imitation of speech (e.g., Kent, 1973; Kent and Forner, 1979; Repp and Williams, 1987; Vallabha and Tuller, 2004). Interestingly, Pisoni (1980) found that mean F1 values were correlated between sessions, whereas F1 SD was not. Variability is also available from x-ray microbeam studies with single speakers (e.g., Perkell and Nelson, 1985; Beckman *et al.*, 1995). Individual token data in the appendix of Hawkins and Midgley (2005) can also be used to estimate variability in young male speakers. Additionally, Waldstein (1990) reports SD values for seven normal-hearing controls. The variability observed in the present study was quite similar to these previously reported values, as evident in Table III.

One important aspect of this variability is its sequential dependence. Due to delays in processing auditory feedback, the influence of one trial's perceived formant frequency should not be seen until the next trial. To evaluate this process an autocorrelation analysis was carried out across a range of lags for the start phase of the experiment (the non-stationary frequency changes in the other phases make this analysis complex for any other part of the data). The change in F1 between trials i and $i+1$ was calculated for every consecutive pair of start trials (using all 20 trials; there were 27 participants not missing any start trials included in this autocorrelation analysis), and the mean F1 value was subtracted from these difference scores before the normalized autocorrelation function was calculated at 11 lags, including a lag of zero. The autocorrelation values were then arithmetically averaged across individuals and are shown in Fig. 6(a). The negative lag-one correlation approached significance ($r=-0.37$, $p=0.10$), whereas greater lags had correlations closer to zero. This trend is reminiscent of the finding in learning force fields in which error on a given trial was predicted by the previous trial's error (Thoroughman and Shadmehr, 2000). The strength of this relationship was not uniform across participants with the range of lag-one correlations being from 0.04 to −0.73. The variability can be used to predict the size of F1 compensation during the hold phase of the experimental manipulation [Fig. 6(b)]. Participants who had strongly negative lag-one correlations during the start trials were more likely to produce large compensations during the hold trials, when feedback was maximally changed ($r=-0.44$, $p=0.02$).

IV. GENERAL DISCUSSION

Both studies demonstrated compensation and adaptation to modifications of acoustic feedback of the first formant frequency. When the frequency of F1 was shifted down (experiment 1) or up (experiments 1 and 2), participants compensated by varying F1 in the opposite frequency direction. In all conditions the average compensations were incomplete and only accounted for a portion of the frequency manipulation. Across all conditions in both experiments, the mean compensation was about 29% of the applied absolute formant manipulation of 200 Hz. The size of the compensation did not differ as a function of direction of shift, nor did the threshold for compensation. Adaptation of F1 frequency was observed in both experiments, and the pattern of adaptation was independent of direction of frequency manipulation. Finally, deadaptation showed an exponential pattern that was not influenced by amount of exposure to altered feedback. Within the limited range of conditions tested (0 to 45 hold trials), no differences were observed in F1 during the end trials following return to normal auditory feedback.

The observed compensation and adaptation are consistent with the findings of Houde and Jordan (Houde and Jordan, 1998, 2002) for whispered speech. Both Houde and Jordan and the present studies found similar ranges of

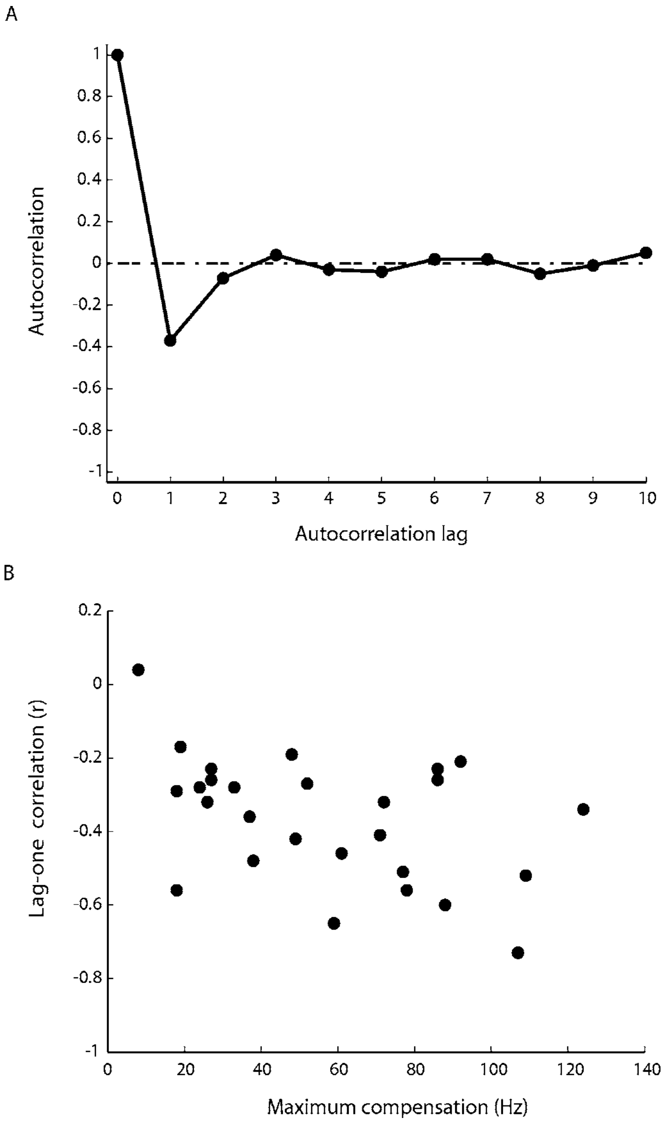

FIG. 6. Panel (a) shows the average autocorrelation of the normalized F1 difference between consecutive trials in the start phase, across subjects and at various lags. The average negative lag-one correlation was $r=-0.37$, whereas higher lags were close to zero. Panel (b) plots the lag-one autocorrelation values against compensation in the hold phase of the experiment (where maximum compensation was observed) for each individual. The cross correlation between these parameters was $r=-0.44$ and was statistically significant.

compensatory behavior and enduring aftereffects. In particular, Houde and Jordan and the studies reported here all show incomplete compensation in response to the formant manipulation. The relatively weak compensation (29%) could be due to a number of factors including nonlinear regions of the feedback response system, the single formant perturbations used in the studies reported here, and the possibility that the importance of auditory feedback may be modulated over time (see Purcell and Munhall, 2006).

Compensation and adaptation have also been demonstrated with manipulations of fundamental frequency feedback (Jones and Munhall, 2000, 2002, 2005), but F0 shows intriguing differences from the formant adaptations shown here and by Houde and Jordan. The F0 studies show negative aftereffects not observed with formants manipulations. For example, when feedback is raised in pitch, participants compensate by lowering their vocal pitch. When normal feedback is returned, the participants overshoot the baseline frequency and produce a higher pitch than baseline. Negative aftereffects such as this are the most common form of adaptation found in perception (e.g., Goldstein, 2006). This is also the form of aftereffect found in other sensorimotor adaptations such as prism adaptation (Held, 1965), and force field adaptations in arm (Lackner and DiZio, 2005) and speech (Tremblay *et al.*, 2003) movement. While the F1 compensation behavior is similar to that seen for F0, the behavior of F1 adaptation is quite different. Formant adaptation is positive, meaning that the produced formant frequencies simply remain near their compensation values. In this sense the phenomena is more like a recalibration of the F1 target or like Helson's (1951, 1964) idea of a change in adaptation level.

Adaptation-level theory as originally proposed (Helson, 1951, 1964) aimed to account for the maintenance of a functional neutral point in perception through computation of a weighted mean of perceived stimuli. While the idea of a shift in the average is consistent with the present data, the statistical pooling of stimulus values to determine this shift is at odds with the findings here and elsewhere. Experiment 2 revealed no relationship between the length of the hold phase and the rate of recovery. Other studies also show results that can not obviously be accounted for by a running average. Houde and Jordan's study involved a substantially more extended manipulation phase (over 2000 trials in the hold condition), but showed similar patterns of data to the present study. At the other extreme, Donath *et al.* (2002) found that production changes were evident in a subsequent utterance about 6 s after manipulation of feedback F0 in a single trial. Finally, Houde and Jordan raise the very surprising possibility that the adaptation persisted for a very long time. They reported that participants showed vestiges of their feedback manipulation when they returned to the laboratory a month following the study. The month of normal talking would surely have washed out any perturbation if a simple running average were at work.

Although it is unclear how long the adaptation lasts, there is consistent evidence about the shape of the recovery function. Both experiments show evidence for an initial period of rapid change. In a more detailed study of end trials, experiment 2 showed an exponential deadaptation towards baseline with an initial fast recovery and a mean F1 that was still below the baseline frequency even 115 trials following the end of the manipulation. This exponential function is consistent with studies of deadaptation following learning to move the arm in novel force fields (Davidson and Wolpert, 2004), and studies of recovery from adaptation of saccadic amplitude (Hopp and Fuchs, 2004). Retention curves in memory research of a very similar shape have been reported for more than 100 years, raising the possibility that the phenomenon being studied here is functionally a memory phenomenon. Rubin and Wenzel (1996), in a meta-analysis of 210 studies of retention curves in memory studies, showed that the loss of information from memory could be fit very well by a range of different kinds of curve-fitting functions similar to the exponential.

The exponential in our analysis was chosen for convenience and was not derived for any theoretical reason. As Rubin and Wenzel (1996) found, it is difficult to distinguish subtly different mathematical functions in data that contain moderate variance. A more promising approach is to constrain the choice of curve-fitting procedures to those that have more appropriate theoretical underpinnings. For example, Wickelgren (1972) proposed an exponential power function for memory decay in which decay was proportional to the similarity of the stimuli being studied. Recent work by Jones and Munhall (2005) suggests that the decay of adaptation for F0 is related to similarity between the adaptation stimuli and the stimuli spoken after the return to normal feedback. When one Mandarin tone was adapted, the fundamental frequency of that tone remained altered longer than a tone that had not been adapted during the manipulation phase. The conditions that influence the decay of adaptation is a promising area for further studies.

Two kinds of variance are present in the current studies. Formant values varied between talkers and also within talkers' data. Second, compensation/adaptation varied significantly between individuals. Part of the variance in formant values can be attributed to formant estimation errors (Vallabha and Tuller, 2002), but it is apparent that there is real variability from token to token of a vowel as it is uttered by an individual. It has been pointed out previously (e.g., Broad, 1976; Kent and Forner, 1979) that this variability is similar to the difference limens found by Flanagan (1955). As mentioned above, more recently Kewley-Port and Watson (1994) have shown that well-trained individuals performing under ideal listening conditions can achieve better performance. However, more ordinary listening conditions, lack of training, and noise can raise thresholds above the estimates of SD reviewed in Table III (Kewley-Port and Zheng, 1999; Kewley-Port, 2001; Liu and Kewley-Port, 2004). Using the change point data here, it is interesting that the response began when manipulations were larger than two SDs beyond the baseline average value of F1. This suggests that some production variability is expected and tolerated by the feedback controller, but when production strays too far (near three SDs) a corrective mechanism is invoked. The fact that compensation was not reliably observed in the present study until the F1 manipulation was >60 Hz is congruent with the idea of managing a speech production system whose output is quite variable. There may be stability costs if a feedback system were to micro-manage off-target productions with deviations smaller than those reasonably expected from that production system (about ± 60 Hz). As noted above, the compensation threshold is higher than the best psychoacoustic performance for detecting changes in F1 (Kewley-Port and Watson, 1994), but under normal listening conditions a somewhat higher threshold may be expected. There may of course also be differences in psychoacoustic and feedback control system thresholds, and between pure listening and listening while speaking.

Why some participants do not compensate for the manipulations is unknown. However, the correlation between the lag-one autocorrelation and the magnitude of compensation observed in experiment 2 suggests that some participants are more actively stabilizing their productions from trial-to-trial even under normal feedback conditions. One possibility is that talkers weight the different types of feedback differently (e.g., Borden, 1979; Perkell, 1980). Some may rely more on kinesthetic feedback and thus are not influenced as much by acoustic feedback. The individual differences in response to manipulations such as delayed auditory feedback are consistent with this possibility (Howell and Archer, 1984). A second possibility is that talkers differ in what the control parameter for the vowel is. In carrying out this work, we have made a number of implicit assumptions about the representation of vowels and their perception: that formants are being tracked by talkers, that the acoustic space is a linear frequency space, and that a static formant estimate is sufficient for characterizing dynamic vowels in natural utterances. Strong arguments have been made against each of these assumptions (e.g., Lindblom, 1986; Strange, 1989), and it may be that our experimental approach does not capture perceptually pertinent patterns for some participants.

In closing, the present experiments have demonstrated that formant production is sensitive to auditory feedback and that some form of representation mediates vowel production. The studies have provided preliminary evidence about the sensitivity of the feedback system and its time constants. However, we still know little about this phonetic representation system, the conditions under which learning and adaptation occur, and its role in the broad context of communication (Pardo and Remez, 2006). Although it has been recognized for a long time that auditory feedback drives learning in speech production, progress in understanding this perception/production relationship has been slow. The approach taken here permits us to address this interaction directly. Beyond the theoretical importance of this work, there are a range of applied problems in which auditory-motor learning is primary. Second-language learning, articulation training of individuals with hearing impairments and cochlear implants, and many types of speech therapy involve perceptual learning and imitation. Thus, progress in this laboratory-based adaptation phenomenon has the potential for significant contributions to enduring problems in rehabilitation science.

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. The authors would also like to thank Bryan Burt for his assistance in data collection, and Courtney Belfour and Sheena Kirkland for their work segmenting vowels from the recorded tokens. Research supported by the National Institute on Deafness and Other Communication Disorders, and the (Canadian) Natural Sciences and Engineering Research Council.

Baken, R. J. and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice* (Singular, San Diego).
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.
Borden, G. J. (**1979**). "An interpretation of research of feedback interruption in speech," Brain Lang **7**, 307–319.
Broad, D. J. (**1976**). "Toward defining acoustic phonetic equivalence for vowels," Phonetica **33**, 401–424.
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.
Caithness, G., Osu, R., Bays, P., Chase, H., Klassen, J., Kawato, M., Wolpert, D. M., and Flanagan, J. R. (**2004**). "Failure to consolidate the consolidation theory of learning for sensorimotor adaptation tasks," J. Neurosci. **24**, 8662–8671.
Darainy, M., Malfait, N., Towhidkhah, F., and Ostry, D. J. (**2006**). "Transfer and durability of acquired patterns of human arm stiffness," Exp. Brain Res. **170**, 227–237.
Davidson, P. R. and Wolpert, D. M. (**2004**). "Scaling down motor memories: de-adaptation after motor learning," Neurosci. Lett. **370**, 102–107.
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.
Flanagan, J. L. (**1955**). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**, 613–617.
Gaveau, V., Alayane, N., Salemme, R., and Desmurget, M. (**2005**). "Self-generated saccades do not modify the gain of adapted reactive saccades," Exp. Brain Res. **162**, 526–531.
Goldstein, E. B. (**2006**). *Sensation and Perception* (Wadsworth, Belmont, CA).
Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.
Guenther, F. H. (**2003**). "Neural control of speech movements," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin), pp. 209–239.
Hawkins, S. and Midgley, J. (**2005**). "Formant frequencies of RP monophthongs in four age groups of speakers," J. Int. Phonetic Assoc. **35**, 183–199.
Held, R. (**1965**). "Plasticity in sensory-motor systems," Sci. Am. **213**, 84–94.
Helson, H. (**1951**). "Perception," in *Theoretical Foundations of Psychology*, edited by H. Helson (Van Nostrand, Toronto), pp. 348–389.
Helson, H. (**1964**). *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior* (Harper & Row, New York).
Hopp, J. J. and Fuchs, A. F. (**2004**). "The characteristics and neuronal substrate of saccadic eye movement plasticity," Prog. Neurobiol. **72**, 27–53.
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.
Howell, P. and Archer, A. (**1984**). "Susceptibility to the effects of delayed auditory feedback," Percept. Psychophys. **36**, 296–302.
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.
Jones, J. A. and Munhall, K. G. (**2005**). "Remapping auditory-motor representations in voice production," Curr. Biol. **15**, 1768–1772.
Kawahara, H. (**1995**). "Transformed auditory feedback: The collection of Data from 1993.1 to 1994.12 by a new set of analysis procedures," TR-H-120 (ATR Human Information Processing Research Laboratories, Kyoto), pp. 1–52.
Kawato, M. (**1989**). "Motor theory of speech perception revisted from the minimum torque-change neural network model," in *8th Symposium on Future Electron Devices*, Tokyo, Japan, pp. 141–150.
Kent, R. D. (**1973**). "The imitation of synthetic vowels and some implications for speech memory," Phonetica **28**, 1–25.
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.
Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. T. (**1996**). "Fundamental frequency effects on thresholds for vowel formant discrimination," J. Acoust. Soc. Am. **100**, 2462–2470.
Lackner, J. R. and DiZio, P. (**2005**). "Motor control and learning in altered dynamic environments," Curr. Opin. Neurobiol. **15**, 653–659.
Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York).
Lane, H. and Tranel, B. (**1971**). "Lombard Sign and Role of Hearing in Speech," J. Speech Hear. Res. **14**, 677–709.
Lane, H. and Webster, J. W. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**, 859–866.
Lashley, K. S. (**1951**). "The problem of serial order in behavior," in *Cerebral Mechanisms in Behavior*, edited by L. A. Jeffress (Wiley, New York).
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.
Lindblom, B. (**1986**). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic P, Toronto), pp. 13–44.
Liu, C. and Kewley-Port, D. (**2004**). "Formant discrimination in noise for isolated vowels," J. Acoust. Soc. Am. **116**, 3119–3129.
Maddieson, I. (**1984**). *Patterns of Sounds* (Cambridge U.P., Cambridge).
Munhall, K. G., Kawato, M., and Vatikiotis-Bateson, E. (**2000**). "Coarticulation and physical models of speech production," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, edited by M. Broe and J. Pierrehumbert (Cambridge U.P., Cambridge), pp. 9–28.
Oller, D. K. and Eilers, R. E. (**1988**). "The role of audition in infant babbling," Child Dev. **59**, 441–449.
Orfandidis, S. J. (**1988**). *Optimum Signal Processing, An introduction* (MacMillan, New York).
Paloski, W. H., Black, F. O., and Metter, E. J. (**2004**). "Postflight balance control recovery in an elderly astronaut: a case report," Otol. Neurotol. **25**, 53–56.
Pardo, J. S. and Remez, R. E. (**2006**). "The perception of speech," in *The Handbook of Psycholinguistics*, edited by M. Traxler and M. A. Gernsbacher (Elsevier, New York), to be published.
Perkell, J. (**1980**). "Phonetic features and the physiology of speech production," in *Language Production*, edited by B. Butterworth (Academic, London), pp. 337–372.
Perkell, J. (**1996**). "Properties of the tongue help to define vowel categories: Hypotheses based on physiologically-oriented modeling," J. Phonetics **24**, 3–22.
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 /ɪ/ and /a/," J. Acoust. Soc. Am. **77**, 1889–1895.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338–2344.
Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J., and Guiod, P. (**1997**). "Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models," Speech Commun. **22**, 227–250.
Peterson, G. E. and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Pisoni, D. B. (**1980**). "Variability of vowel formant frequencies and the quantal theory of speech: a first report," Phonetica **37**, 285–305.
Purcell, D. W. and Munhall, K. (**2006**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.
Repp, B. H. and Williams, D. R. (**1987**). "Categorical Tendencies in Imitat-

ing Self-Produced Isolated Vowels," Speech Commun. **6**, 1–14.
Rubin, D. C. and Wenzel, A. E. (**1996**). "One hundred years of forgetting: A quantitative description of retention," Psychol. Rev. **103**, 734–760.
Schenk, B. S., Baumgartner, W. D., and Hamzavi, J. S. (**2003**). "Effect of the loss of auditory feedback on segmental parameters of vowels of post-lingually deafened speakers," Auris Nasus Larynx **30**, 333–339.
Shadmehr, R. (**2004**). "Generalization as a behavioral window to the neural mechanisms of learning internal models," Hum. Mov. Sci. **23**, 543–568.
Shadmehr, R. and Brashers-Krug, T. (**1997**). "Functional stages in the formation of human long-term motor memory," J. Neurosci. **17**, 409–419.
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.
Smith, C. R. (**1975**). "Residual hearing and speech production in deaf children," J. Speech Hear. Res. **18**, 795–811.
Smith, K. U. and Smith, W. M. (**1962**). *Delayed Sensory Feedback and Behavior* (Saunders, Philadelphia).
Strange, W. (**1989**). "Evolving theories of vowel perception," J. Acoust. Soc. Am. **85**, 2081–2087.
Stratten, G. (**1896**). "Some preliminary experiments on vision without inversion of the retinal image," Psychol. Rev. **3**, 611–617.
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.
Thoroughman, K. A. and Shadmehr, R. (**2000**). "Learning of action through adaptive combination of motor primitives," Nature (London) **407**, 742–747.
Tin, C. and Poon, C. S. (**2005**). "Internal models in sensorimotor integration: Perspectives from adaptive control theory," J. Neural Eng. **2**, S147–S163.
Tremblay, S., Shiller, D. M., and Ostry, D. J. (**2003**). "Somatosensory basis of speech production," Nature (London) **423**, 866–869.
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.
Wickelgren, W. (**1972**). "Trace Resistance and Decay of Long-Term Memory," J. Math. Psychol. **9**, 418–455.