

Auditory-motor mapping for pitch control in singers and nonsingers

Jeffery A. Jones · Dwayne Keough

Received: 20 September 2007 / Accepted: 11 June 2008 / Published online: 1 July 2008
© Springer-Verlag 2008

Abstract Little is known about the basic processes underlying the behavior of singing. This experiment was designed to examine differences in the representation of the mapping between fundamental frequency (F0) feedback and the vocal production system in singers and nonsingers. Auditory feedback regarding F0 was shifted down in frequency while participants sang the consonant-vowel /ta/. During the initial frequency-altered trials, singers compensated to a lesser degree than nonsingers, but this difference was reduced with continued exposure to frequency-altered feedback. After brief exposure to frequency altered auditory feedback, both singers and nonsingers suddenly heard their F0 unaltered. When participants received this unaltered feedback, only singers' F0 values were found to be significantly higher than their F0 values produced during baseline and control trials. These after-effects in singers were replicated when participants sang a different note than the note they produced while hearing altered feedback. Together, these results suggest that singers rely more on internal models than nonsingers to regulate vocal productions rather than real time auditory feedback.

Keywords Auditory feedback · Internal model · Sensorimotor · Fundamental frequency · F0 · Pitch · Singing · Singer

Introduction

It is commonly assumed that the human voice was the first musical instrument. Singing involves producing a succession of musical sounds with the voice and appears to be ubiquitous across all known cultures. Given the uniqueness, universality and importance of singing, it is surprising how little is known about the basic processes underlying the behavior. The core skill required of a singer is accurate control of the fundamental frequency (F0) of the voice. Singers must match their F0 to the frequency of a particular musical note: often in the absence of an external reference, when singing a cappella, or in the presence of other voices or instruments that may or may not be producing the same note.

To produce a particular musical note, a singer must have precise control over intrinsic and extrinsic laryngeal muscles as well as respiratory muscles. This control is achieved by a complex network of cortical and brainstem centers that rely on proprioceptive (Kirchner and Wyke 1965; Wyke 1974; Yoshida et al. 1989) and auditory (Sapir et al. 1983) reflex mechanisms. However, there are nonreflexive mechanisms that respond to auditory input which also play an extremely important role. In fact, numerous clinical and experimental studies have shown auditory feedback to be essential for developing and maintaining normal vocal control in general. For example, the quality of a child's articulations is affected when hearing impairments occur early in life (Oller and Eilers 1988). Moreover, auditory feedback remains important for continued accurate vocal productions throughout life; adults who acquire severe hearing loss often have difficulties controlling their F0, vocal intensity and speaking rate (Cowie and Douglas-Cowie 1992). Altering the auditory feedback that speakers hear in controlled laboratory studies often causes reciprocal

J. A. Jones
Center for Cognitive Neuroscience, Wilfrid Laurier University,
Waterloo, ON, Canada N2L 3C5

J. A. Jones (✉) · D. Keough
Department of Psychology, Wilfrid Laurier University,
Waterloo, ON, Canada N2L 3C5
e-mail: jjones@wlu.ca

changes in speakers' ongoing speech productions. For instance, speakers exposed to increases in masking noises or decreases in side-tone amplitude compensate by increasing their speaking volume and the duration of their utterances (Bauer et al. 2006; Lane and Tranel 1971). Selectively filtering speech frequencies (Garber and Moller 1979), or shifting F0 (Burnett et al. 1997, 1998; Elman, 1981; Jones and Munhall 2000, 2002, 2005; Kawahara 1998) or formant frequencies (Houde and Jordan 1998; Purcell and Munhall 2006b) also elicits compensatory modifications in vocal output.

Although relatively little research has directly investigated the role of feedback during singing, studies have unsurprisingly shown results comparable to those found during speech production. For example, when auditory feedback is masked, intonation accuracy is reduced (Mürbe et al. 2002). Likewise, singers show compensatory responses to frequency-altered feedback (FAF) that are similar to the responses made by participants who are merely speaking (Burnett et al. 1997, 1998; Natke et al. 2003). That is, when singers hear their F0 shifted either up or down in frequency, they shift the frequency of their F0 in the direction opposite of the perturbation. The magnitude of the compensation in both speech and singing is not complete and appears to be limited to corrections of up to half a semitone (Larson et al. 2000).

The compensatory effects that result from altered feedback conditions led to speculation that vocal production is monitored in a closed-loop manner (Fairbanks 1954; Larson et al. 2000; Lee 1950). Such servomechanistic accounts posit that a comparator looks for discrepancies between the intended output of a vocal production and sensory feedback. The observed compensations are initiated to overcome the perceived mismatch. However, most researchers agree that vocal production cannot be exclusively guided in a closed-loop manner. Typically speech rates are too fast for auditory feedback to be processed and the corrections implemented before the next segment is produced (Borden 1979). For example, when a vocalist sings an intended F0 target, sensory feedback is not available until a few milliseconds after vocalization. Thus, prior to vocalization and immediately after phonation begins, vocal fold stiffness and positioning of laryngeal structures (prephonatory tuning; Watts et al. 2003) must be entirely the result of open-loop motor planning.

The consensus therefore is that vocal production is the result of interplay between closed and open-loop control (Guenther and Perkell 2004; Jones and Munhall 2000; Perkell et al. 1997). Indeed, recent efforts to understand the complex control systems underlying vocal production have borrowed from a recent body of work in general motor control that suggests the brain relies on "internal models" during rapid skilled movement. These internal models are

hypothesized as neural maps of the relationships among the motor commands, musculature, environment and sensory feedback (Desmurget and Grafton 2000; Flanagan and Wing 1993; Shadmehr and Mussa-Ivaldi 1994). After formation, these models are used to predict the outcome of a movement and provide internal feedback to the planning and control systems, which guide future movements. The provision of internal feedback effectively avoids the delays inherent to the reliance on sensory feedback (Desmurget and Grafton 2000; Tin and Poon 2005; Wolpert et al. 1995).

Both empirical (Houde and Jordan 1998; Jones and Munhall 2000, 2002, 2003, 2005; Perkell et al. 1997; Purcell and Munhall 2006a, 2006b) and modeling (Guenther 1994; Guenther and Perkell 2004) research addressing acoustic-articulatory mappings has suggested at least three possible roles for auditory feedback. (1) Auditory feedback provides the most important and reliable information regarding target achievement. Indeed this is the case for children learning the sounds of their native language, and for adults learning the sounds of a second language or adapting to a new vocal tract arrangement (e.g., orthodontic braces, dentures, piercings). (2) Feedback becomes important when environmental conditions (e.g., masking noise) reduce the quality of the sound reaching the listener. As a result, speakers will modify future productions to improve intelligibility by enunciating more clearly, increasing amplitude or reducing the speaking rate. (3) The motor planning and control systems use auditory feedback for online calibration of internal models of the speech motor system.

In this paper, we address the importance of this third role for auditory feedback during singing. Specifically, we were interested in whether trained singers, by virtue of their extended experience reproducing pitch targets, would rely more on a well-established internal model than non-singers. Much of the recent work addressing the role of auditory feedback has done so by altering feedback regarding pitch and evaluating changes in production. As previously mentioned, subjects compensate when auditory feedback regarding their own pitch is suddenly raised or lowered artificially (Burnett et al. 1997, 1998; Elman 1981; Jones and Munhall 2000, 2002, 2005; Kawahara 1998). These compensatory responses have typically lent support to the idea that F0 control is reliant on sensory feedback (Larson et al. 2000). However, in a series of studies, Jones and Munhall (2000, 2002, 2005) slowly shifted (by 1 cent increments to 1 semitone) vocal pitch feedback up or down in frequency while speakers produced vowels. Although they were unaware of the feedback manipulation (after testing subjects were asked if they noticed a change in pitch and no subjects indicated that they were aware of any perturbations), speakers modified their produced F0 in the

opposite direction of the shifted feedback. When F0 feedback was returned to normal after this brief exposure to the altered feedback conditions, aftereffects were observed: if speakers heard their F0 feedback shifted higher than normal, their F0 increased, relative to unaltered F0 productions prior to any perturbations, when they were unexpectedly given normal feedback. Conversely, if speakers heard their F0 feedback shifted downward they decreased their pitch when they were given normal feedback. The adaptation indicates that F0 control is not only dependent on sensory feedback but is also reliant on an internal representation of the mapping between pitch output and the motor systems that control it.

The aforementioned research looked specifically at speech. Although both singing and speech involve the same articulators, the principal roles of pitch in singing and language production are clearly different. In speech, F0 can play a role in conveying linguistic, para-linguistic as well as nonlinguistic information. In English for example, F0 varies depending on prosodic pattern as well as emotional context (Zemlin 1981). In tone languages such as Mandarin and Cantonese, F0 patterns are used to differentiate between words and grammatical categories (Yip 1995). Regardless of the role, in speech pitch targets are relative, not absolute, to the speaker's own productions. In singing on the other hand, the usual goal is for F0 to match absolute fundamental frequencies that correspond to musical notes. The notes provide singers with an external reference to which they compare their own productions. Studies have shown that in general, auditory feedback is important for accurate control of F0 during singing; when feedback is masked, pitch-matching accuracy decreases (Elliot and Niemoeller 1970; Ward and Burns 1978). However, trained singers appear to have superior pitch-matching abilities than untrained singers (Murry 1990) and are more resistant to the effects of masking (Watts et al. 2003). Singers' increased reliance on internal models may contribute to this resistance. On the other hand, the effects of training and natural talent may be confounded as talented nonsingers have been shown to perform as well as trained singers on pitch-matching tasks and have performed better than trained singers when auditory feedback was unavailable (Watts et al. 2003).

In the present investigation, we used a FAF paradigm to examine the auditory-motor representation of the mapping between F0 feedback and the vocal production system in singers and nonsingers. In one condition, participants were asked to emulate the note, G4 (ISO 16, concert pitch) while hearing their F0 shifted down 1 semitone. Based on previous research we predicted that overall both singers and nonsingers would compensate for the F0 perturbations by increasing their F0. However, we hypothesized that the magnitude of compensations would be initially smaller in

singers than nonsingers due to their stronger reliance on an internal model for F0 production. During this exposure to FAF, participants' internal model would be recalibrated based on the error detected between production and feedback. If this recalibration did indeed occur, we predicted that after participants heard their feedback returned to normal, aftereffects would be apparent and more pronounced in singers compared to nonsingers. That is, we hypothesized that singers would be less accurate on subsequent vocal productions after having received altered feedback, such that singers' F0 productions would be sharper than during unaltered feedback trials. Moreover, any adaptation observed should generalize to a greater degree in singers as opposed to nonsingers when they were asked to produce another note (F4).

Methods

Subjects

Forty participants (all women) whose first language was North American English participated in the FAF experiment. We chose to exclude men so that all participants could comfortably sing the same pitch. No previous work has demonstrated that men and women differ in their response to FAF and there is no theoretical basis to assume that a gender difference should exist. The participants were between 18 and 27 years of age (mean of 20 years), grew up in English-speaking communities and received their primary education in English. Of the 40 participants, 20 were trained singers (with a mean vocal training of 12 years) recruited from the faculty of music at Wilfrid Laurier University. The remaining 20 participants were nonsingers recruited from Wilfrid Laurier University who reported never receiving formal vocal training and no participation in any form of formal singing (e.g., school or church choirs). All participants received either course credit or financial compensation for their involvement in this study. Participants gave written informed consent and the Wilfrid Laurier University Research Ethics Committee approved the procedures.

Apparatus

Participant recording sessions

The participant recording sessions took place in a double-walled sound attenuated booth. Participants wore headphones (Sennheiser HMD 280-13) and a condenser microphone (AKG C 420^{III} PP) that was maintained at a fixed distance of approximately 3 cm from their mouth. To reduce the amount of natural acoustic feedback, participants

heard multitalker (20 talkers) babble (Auditec, St Louis, MO, USA) at a level of 75 dB SPL through the headphones for the duration of the experiment. In addition to the multi-speaker babble noise, participants also heard the target pitch, which was a female voice singing /ta/, at either 392 Hz or 349 Hz, G4 or F4, respectively. The microphone signals were amplified (MA3 stereo microphone amplifier, Tucker-Davis Technologies) and then sent to a signal processor (VoiceOne 2.0, TC Helicon) that shifted the vocal pitch. Participants maintained similar vocal amplitude throughout a session by monitoring a loudness monitor (PPM, Paul Marshall, Lichfield, England) presented on a computer screen. The frequency-altered speech signal was then mixed (Onyx 1640, Mackie) with the multitalker babble and fed back to the participant. Participants' productions were digitized (44.1 kHz; 828 mkII, MOTU) for later analysis.

Target stimuli recording

The target stimuli were created by recording a trained singer producing G4 and F4. These recordings were then processed using the speech modification algorithm STRAIGHT (Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum; Kawahara et al. 1999) so that F0 for each target was precisely 392 or 349 Hz.

Procedure

Participants produced the consonant-vowel /ta/ (2 s duration) in two blocks of 60 trials. Prior to beginning the experiment participants were acclimated to the task with five practice trials. During these practice trials they heard the multitalker babble and the target, G4, which was presented each time before participants produced the target (this was also the case during testing), while receiving unaltered feedback. They were instructed to sing the target as accurately as possible in pitch (392 or 349 Hz) and duration (approximately 2 s). In addition, participants were asked to produce their utterances with a self-determined and comfortable amplitude, and to monitor their loudness using a loudness monitor visible on a computer monitor. Following practice, participants produced the target during a control condition and an experimental condition. The order of these two conditions was counterbalanced across the participants. During the control condition participants produced the target for 60 trials while their auditory feedback was unaltered. During the experimental condition participants first received unaltered auditory feedback for ten trials, and then produced the target for 30 trials while they heard their F0 shifted down 100 cents (1 semitone). On the final 20 trials, participants again heard their

feedback unaltered. Figure 1 depicts the FAF protocol used during the experimental condition. Note that participants heard their F0 shifted from the beginning until the end of their vocal productions.

Additionally, there were two singing conditions in this study (G–G and G–F). In the G–G condition participants were required to sing G4 (392 Hz) on all utterances, whereas in the G–F condition participants were required to sing G4 on the first 40 trials and F4 (349 Hz) on the remaining 20 trials. Participants were randomly assigned to either the G–G ($N = 20$) or G–F ($N = 20$) condition; they did not participate in both. Trial initiation and the pitch processing were computer controlled. During offline analyses, F0 values for the utterances during each trial were determined using an autocorrelation algorithm included in the Praat program (Boersma 2001). F0 values in Hertz were normalized to the target pitch (G4 or F4) by converting values to cents using the following formula:

$$\text{Cents} = 100(12 \log_2 F/B)$$

In the formula, F is the F0 value in Hertz and B is frequency of the target pitch participants were to sing (392 or 349 Hz).

Results

The mean F0 values for each utterance produced during the control and experimental trials were calculated. Previous research has demonstrated that compensation to perturbations occurs within 130–500 ms after perturbation onset (Burnett et al. 1997, 1998; Jones and Munhall 2002). In this

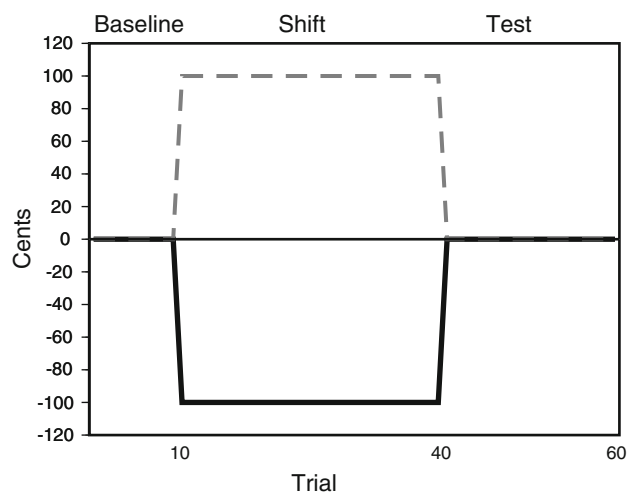


Fig. 1 A schematic depicting the frequency-altered feedback paradigm used in the experiment. The *black line* indicates the magnitude of frequency shift throughout the experimental session. The *gray line* represents the fundamental frequency a participant would need to produce to perfectly compensate for the altered feedback

study, participants experienced FAF from the start of their productions. Because earlier portions of an utterance were more likely to be the result of open-loop control, F0 for the first 1,000 ms of each utterance was analyzed. The G–G and G–F conditions were analyzed separately. These two analyses focused on six different blocks of trials within each condition: baseline trials (1–10), initial-shift trials (11–20), middle-shift trials (21–30), late-shift trials (31–40), initial-test trials (41–50), and late-test trials (51–60). The F0 values for the first five trials and last five trials of each block were averaged and categorized as early and late phases, respectively. Thus, two separate MANOVAs for the G–G and G–F conditions were performed with two (experience: singer and nonsinger) \times two (session: control and experimental) \times six (the six blocks) \times two (phase: early and late) as factors. The Fisher LSD procedure was used to conduct post hoc tests. An alpha level of 0.05 was used for each statistical analysis.

The mean F0 values for each trial in the control and experimental sessions, for the singers and nonsingers during the G–G condition, are presented in Fig. 2a. During the G–G condition, a main effect of session revealed that the mean F0 observed in the control session was much lower than the mean F0 observed in the experimental session, $F(1,18) = 297.82$, $P < 0.05$. A main effect of block and phase was also observed, $F(5,90) = 185.34$, $P < 0.05$ and $F(1,18) = 10.54$, $P < 0.05$, respectively. These main effects were expected and the natural result of compensatory responses made during the shift blocks in the experimental session. Overall, the F0 values during the shift blocks (trials 11–40) were higher than F0 values during the baseline and test blocks. The F0 values observed in the early phases were lower than F0 values in the late phases.

A two-way interaction between experience and phase was also significant, $F(1,18) = 5.23$, $P < 0.05$. Singers increased their F0 on average from the early phase to the late phase ($P < 0.05$). Moreover, two- to three-way interactions existed. Of particular interest, an interaction between experience, session and block was found, $F(5,90) = 12.55$, $P < 0.05$. As well, an interaction between session, block and phase was observed, $F(5,90) = 3.65$, $P < 0.05$. During the baseline trials within the experimental and control sessions, singers' F0 productions were slightly higher than the F0 productions of nonsingers ($P < 0.05$). This pattern was observed in each block of the control session ($P < 0.05$). Within each group of singers and nonsingers, there were no significant differences in the F0 produced across each block in the control session ($P < 0.05$). Similarly, differences observed between baseline F0 values in the experimental and control sessions for both the singers and the nonsingers failed to reach significance ($P > 0.05$). However, F0 values for the first phase of trials of the baseline for both the control and experimental conditions

were lower than F0 values during the final phase of trials across singers and nonsingers ($P < 0.05$). This difference may reflect a period of acclimation to the task.

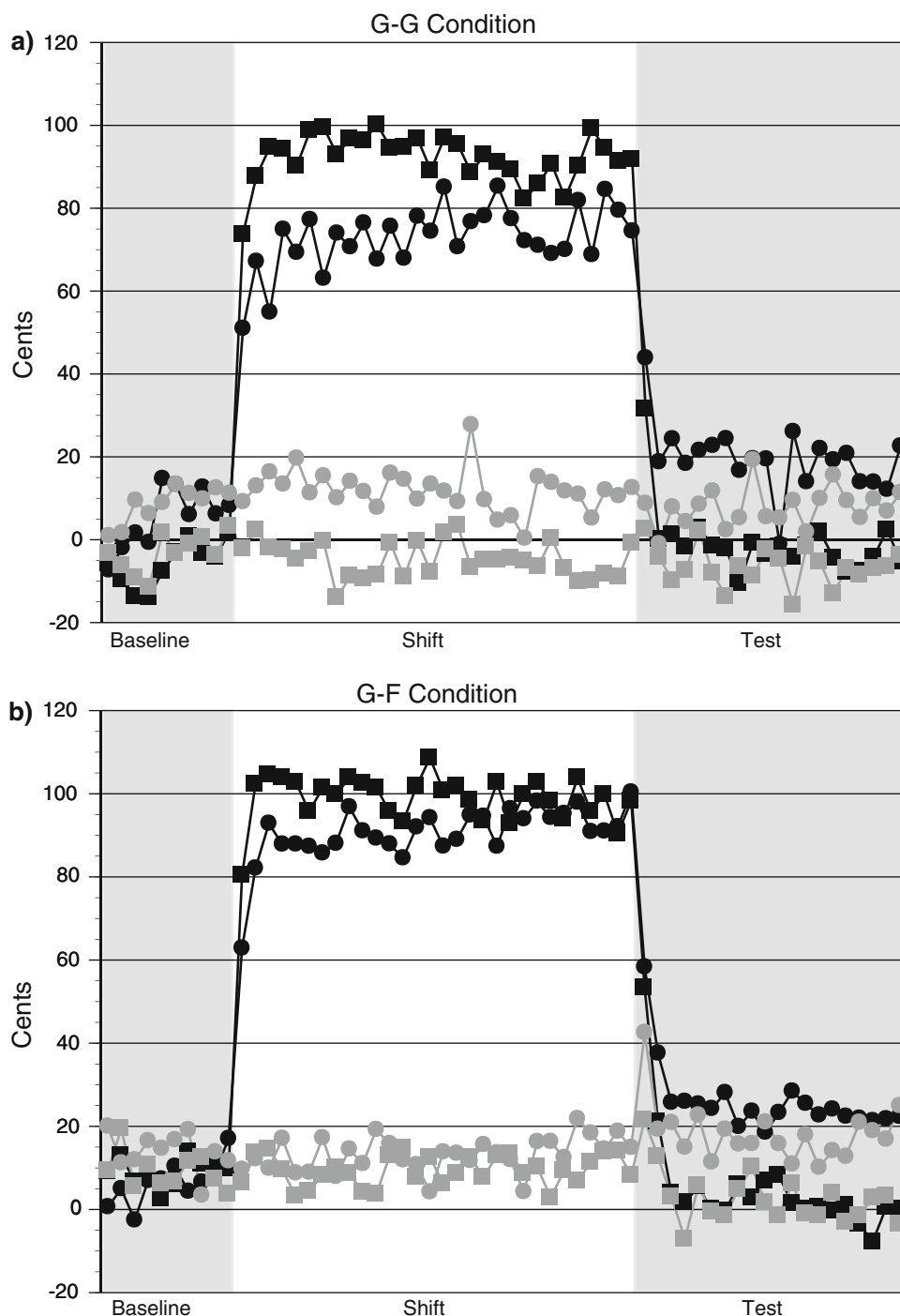
Post hoc analysis of the initial-shift (11–20), middle-shift (21–30) and late-shift (31–40) trials during the G–G condition revealed that both singers and nonsingers compensated for the FAF—mean F0 values during the experimental session were significantly higher than F0 values during the control session and the baseline blocks of the experimental session ($P < 0.05$). However, nonsingers compensated more during these three shift blocks of the experimental session than singers ($P < 0.05$).

For the G–G condition, during the early test trials (41–50) an aftereffect was observed for the singers such that mean F0 values for these early test trials in the session were significantly higher than mean F0 values during the control session ($P < 0.05$). Likewise, these experimental session test trials were significantly higher than F0 values observed in the baseline block of the same session ($P < 0.05$). These aftereffects carried over to the following late-test trials (51–60) relative to the baseline trials in the experimental session ($P < 0.05$). However, when F0 values in these late-test trials were compared to the same trials in the control session, this difference did not reach significance ($P > 0.05$). No other significant main effects or interactions were observed for the G–G condition.

The data for the G–F condition are presented in Fig. 2b. During the G–F condition, main effects of session, $F(1,18) = 118.35$, $P < 0.05$, and block, $F(5, 90) = 98.3$, $P < 0.05$, were found to be significant. As with the G–G condition, these main effects were the result of compensatory responses observed during the shift blocks and the experimental session. F0 values observed in the control session were much lower than those observed in the experimental session; during the shift blocks in the experimental session, F0 values were higher than the values observed during the baseline and test blocks.

Similar to the GG condition, a three-way interaction between session, block and phase were found to be significant in the G–F condition, $F(5, 90) = 3.96$, $P < 0.05$. Across both singers and nonsingers, the early, middle and late-shift trials during the experimental session yielded higher F0 values than during the control session ($P < 0.05$). During the initial-test trials of the control session, F0 values were slightly higher than during the baseline and late-test trials ($P < 0.05$). However, the difference between the F0 values observed during the initial-test trials in the experimental session, and the F0 values of the control session and baseline of the experimental session was even greater ($P < 0.05$). These differences occurred primarily in the early phase of the early test trials. It is during this block and phase that participants were asked to produce F4 after a series of G4 productions.

Fig. 2 The mean fundamental frequency for **a** G–G and **b** G–F conditions during the baseline, shift and test trials of each session. The control session (unaltered feedback) is in gray. The experimental session (FAF during Shift) is in black. Circles represent data for the singers, and data for the nonsingers is represented by squares



A three-way interaction between experience, session and block was also observed, $F(5, 90) = 2.87$, $P < 0.05$. During the control session, both singers and nonsingers consistently produced the same F0 values up to and including the late-shift trials. During early test trials, participants were asked to produce F4 after producing G4. During the early test trials of the control session, singers produced slightly higher F0 values than they did in the initial-shift trials ($P < 0.05$). No other differences were observed for singers in this

respect. Nonsingers on the other hand produced lower F0 values in these late-test trials. These F0 values were significantly lower than all other blocks except the F0 values observed in the early test trials ($P < 0.05$).

Again for the G–F condition, post hoc comparisons revealed that both singers and nonsingers compensated for the FAF. The F0 values observed during the experimental session were significantly higher than values observed during the control session and the baseline blocks of the

experimental session ($P < 0.05$). Nonsingers' productions were higher than the singers F0 values during the initial-shift (11–20) and middle-shift (21–30) trials ($P < 0.05$), but were equivalent to singers F0 values during the late-shift (31–40) trials ($P > 0.05$). Singers' increase their F0 values from the initial-shift to the late-shift trials ($P < 0.05$), but nonsingers' productions were consistent across the shift trials ($P > 0.05$).

Similar to the pattern observed for the G–G condition, singers' mean F0 values for the early test trials of the experimental session in the G–F condition were significantly higher than F0 values produced in the baseline block of the same session ($P < 0.05$). These experimental session test trials were likewise significantly higher than F0 values observed during the control session ($P < 0.05$). According to post hoc tests the F0 productions for singers during the late-test trials did not significantly differ from early test trials ($P > 0.05$). Moreover, these late-test trials were significantly higher than those observed during the baseline trials of the experimental session ($P < 0.05$). However, the difference between the F0 values for the late-test trials in the experimental session and the control session failed to reach significance ($P > 0.05$). By contrast, no aftereffects were observed for the nonsingers. F0 values for the early and late-test trials were not significantly different than the baseline trials in the experimental session or the test trials in the control session. No other significant main effects or interactions were observed for the G–F condition.

Discussion

The purpose of the present study was to explore whether trained singers rely more on a well-established internal model for F0 control than nonsingers. Based on that assumption, we predicted that nonsingers would utilize auditory feedback more readily than singers and would initially produce larger compensations to the FAF. However, we also predicted that the brief exposure to FAF would induce stronger adaptation effects in singers than nonsingers. In accordance with our expectations, we found that during the initial FAF trials, singers compensated to a lesser degree than nonsingers. Moreover, F0 values were generally lower for singers than nonsingers during the shift trials of the experimental sessions. However, although singers overall compensated to a lesser degree than nonsingers, aftereffects existed for singers but not for nonsingers. That is, the F0 values observed in the initial-test trials were higher than baseline and control values for singers but not for nonsingers. This pattern of aftereffects was replicated when singers were asked to produce a different note than the note they produced during FAF.

Combined, the observations that singers compensated to a lesser degree than nonsingers, yet exhibited significant aftereffects suggests that they rely more on an internal model for F0 production during singing. Although the majority of previous studies were not designed to investigate aftereffects, they have shown that individuals performing speech tasks, like singing tasks, often produce opposing responses when exposed to FAF (Burnett et al. 1997, 1998; Elman 1981; Kawahara 1998). These observations led to suggestions that F0 production relies on closed-loop control (Larson et al. 2000; Natke et al. 2003). However, Jones and Munhall (2000, 2002, 2005) demonstrated that short-term exposure to FAF modified an internal representation of the mapping between pitch output and the motor systems that control it. The present study shows that this online recalibration occurs even when the participants are aware of the altered feedback conditions. More importantly, the study reveals that these internal representations become more entrenched for trained singers as a result of experience achieving an absolute pitch-target in the form of singing. This increased reliance is presumably due to singers' extensive practice in the form of vocal exercises and performance. Nevertheless, despite this relatively strong reliance on an internal model for F0 control, an extremely brief exposure (30 trials) to altered feedback conditions can cause partial remapping of the representation of the relationship between motor commands and their expected feedback consequences.

The differences we found between singers' and nonsingers' responses to FAF are consistent with the observations made in a recent functional magnetic resonance imaging (fMRI) study (Zarate and Zatorre 2005; see also Zarate and Zatorre 2008). Zarate and Zatorre had singers (individuals with more than 3 years musical experience) and nonsingers (individuals with less than 3 years musical experience) sing target notes while receiving normal auditory feedback and with feedback shifted up or down 2 semitones. Singing with normal unaltered feedback resulted in enhanced activation in bilateral auditory and motor cortices, supplementary motor area, anterior cingulate cortex (ACC), thalamus, insula and the cerebellum for both singers and nonsingers. This pattern of activation was similar to the pattern of activations observed in previous studies (Jeffries et al. 2003; Perry et al. 1999; Riecker et al. 2000).

When Zarate and Zatorre (2005) exposed participants to FAF and asked them to *ignore* the feedback both singers and nonsingers showed enhanced activity in the inferior parietal lobule (IPL) compared to the normal feedback condition. Zarate and Zatorre posited that the IPL activation represented error processing. However, singers showed relatively more activation in the superior temporal gyrus (STG), superior temporal sulcus (STS), and right

insula. When asked to compensate for the FAF signal, enhanced activity in the ACC, STS, insula, putamen, pre-SMA, and IPL was observed in singers. Zarate and Zatorre suggested that the additional recruitment of the STG and STS when singers heard the FAF may be the result of increased perceptual analysis of auditory feedback and that enhanced activity observed in the ACC and insula when singers attempted to compensate indicates that these regions are tied to “audiovocal integration”. Thus, it appears the singers’ extensive vocal practice resulted in the recruitment of additional cortical areas that allow more proficient vocal pitch control. Perhaps these additional cortical regions form part of the network that instantiates a singer’s internal model.

A striking difference in the current study compared to many other studies using a FAF paradigm was the degree of compensation observed. Most previous investigations have failed to show perfect or even close to perfect compensation for F0 perturbations (e.g., Larson et al. 2000; Natke et al. 2003; cf. Hain et al. 2000). In fact, response magnitudes of half a semitone (50 cents) or less are most common regardless of how large the perturbation. Indeed, Larson et al. (2000) proposed a closed-loop mathematical model of F0 control and suggested that a complete model would include a filter with a limiting nonlinearity that prevents responses over 50 cents. The bulk of research has only addressed the integration of F0 feedback for speech. As previously mentioned, pitch targets in speech are relative to a speaker’s own productions and are not absolute like they are for singing: notes provide singers with an external reference. To date, there has only been one other systematic study investigating responses to FAF during singing. Natke et al. (2003) found that compensatory responses were greater during the singing condition (66 cents) than the speaking condition (47 cents). Moreover, the compensatory response lasted longer during the singing task than speaking task, persisting into the following trial. The increased compensation and its persistence during singing suggest that singing invokes more vigilant monitoring and integration of auditory feedback. This tighter control may be the result of the particular task constraints of singing such as the requirement of matching an absolute pitch value (and perhaps the availability of an external reference to which to match) (Burnett et al. 1997; Natke et al. 2003).

Although Burnett et al. (1997) report exposing a single singer to FAF and finding perfect compensation, the results of the present study are extraordinary when considering the larger body of work using the FAF paradigm, including Natke et al. (2003) study comparing speech and singing. However, there are other factors related to our experimental procedure other than the singing task that may further account for the near perfect compensation we

observed. Probably the most important difference between our paradigm and that of others stems from the fact that one of our aims was to look at aftereffects. This meant that participants were exposed to repeated and consecutive trials with FAF. Other studies investigating FAF responses have been primarily interested in compensation mechanisms, and as such, exposed participants to FAF trials randomly (cf. Bauer and Larson 2003). As can be seen in Fig. 2, compensations were smaller in singers after initial exposure to the FAF compared to trials later in the shift phase. Thus, compensation responses appear to get stronger with increasing exposure to FAF. We interpret this increase in the magnitude of compensation to result from recalibration of an internal model for F0 control based on error signals derived from comparing the expected outcome of vocal commands and auditory feedback.

Ultimately, singing offers a unique window in which to study the formation of internal models for vocal production. This work adds to the paucity of research conducted on the role of auditory feedback during singing. Future work should continue to address how singers and nonsingers utilize internal models and sensory feedback to regulate F0 while singing and speaking.

Acknowledgments Research supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the National Institute on Deafness and Other Communication Disorders, and a Research Fellowship from Wilfrid Laurier University.

References

- Bauer JJ, Larson CR (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
- Bauer JJ, Mittal J, Larson CR, Hain TC (2006) Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. *J Acoust Soc Am* 119:2363–2371
- Boersma P (2001) Praat, a system for doing phonetics by computer. *Glott Int* 5:341–345
- Borden GJ (1979) An interpretation of research on feedback interruption in speech. *Brain Lang* 7:307–319
- Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103:3153–3161
- Burnett TA, Senner JE, Larson CR (1997) Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. *J Voice* 11:202–211
- Cowie R, Douglas-Cowie E (1992) Postlingually acquired deafness. *Trends in linguistics, studies and monographs*. Mouton de Gruyter, New York
- Desmurget M, Grafton S (2000) Forward modeling allows feedback control for fast reaching movements. *Trends Cogn Sci* 4:423–431
- Elliot L, Niemoeller A (1970) The role of hearing in controlling voice fundamental frequency. *Int Audiol* 9:47–52
- Elman JL (1981) Effects of frequency-shifted feedback on the pitch of vocal productions. *J Acoust Soc Am* 70:45–50

- Fairbanks G (1954) Systematic research in experimental phonetics. I. A theory of the speech mechanism as a servosystem. *J Speech Hear Dis* 19:133–139
- Flanagan JR, Wing AM (1993) Modulation of grip force with load force during point-to-point arm movements. *Exp Brain Res* 95:131–143
- Garber SR, Moller KT (1979) The effects of feedback filtering on nasalization in normal and hypernasal speakers. *J Speech Hear Res* 22:321–333
- Guenther FH (1994) A neural network model of speech acquisition and motor equivalent speech production. *Biol Cybern* 72:43–53
- Guenther FH, Perkell JS (2004) A neural model of speech production and its application to studies of the role of auditory feedback in speech. In: Maassen B, Kent R, Peters H, Van Lieshout P, Hulstijn W (eds) *Speech motor control in normal and disordered speech*. Oxford University Press, Oxford, pp 29–49
- Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK (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
- Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. *Science* 279:1213–1216
- Jeffries KJ, Fritz JB, Braun AR (2003) Words in melody: an H(2)15O PET study of brain activation during singing and speaking. *Neuroreport* 14:749–754
- Jones JA, Munhall KG (2000) Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am* 108:1246–1251
- Jones JA, Munhall KG (2002) The role of auditory feedback during phonation: studies of Mandarin tone production. *J Phon* 30:303–320
- Jones JA, Munhall KG (2003) Learning to produce speech with an altered vocal tract: the role of auditory feedback. *J Acoust Soc Am* 113:532–543
- Jones JA, Munhall KG (2005) Remapping auditory-motor representations in voice production. *Curr Biol* 15:1768–1772
- Kawahara H (1998) Hearing voice: transformed auditory feedback effects on voice pitch control. In: Rosenthal DF, Okuno HG (eds) *Computational auditory scene analysis*. Lawrence Erlbaum Associates Publishers, Mahwah, pp 335–349
- Kawahara H, Masuda-Katsuse I, de Cheveigne A (1999) Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Commun* 27:187–207
- Kirchner JA, Wyke BD (1965) Articular reflex mechanisms in the larynx. *Ann Otol Rhinol Laryngol* 74:749–768
- Lane H, Tranel B (1971) The Lombard sign and the role of hearing in speech. *J Speech Hear Res* 14:677–709
- Larson CR, Burnett TA, Kiran S, Hain TC (2000) Effects of pitch-shift velocity on voice F0 responses. *J Acoust Soc Am* 107:559–564
- Lee BS (1950) Effects of delayed speech feedback. *J Acoust Soc Am* 22:824–826
- Mürbe D, Pabst F, Hofmann G, Sundberg J (2002) Significance of auditory and kinesthetic feedback to singers' pitch control. *J Voice* 16:44–51
- Murry T (1990) Pitch-matching accuracy in singers and nonsingers. *J Voice* 4:317–321
- Natke U, Donath TM, Kalveram KT (2003) Control of voice fundamental frequency in speaking versus singing. *J Acoust Soc Am* 113:1587–1593
- Oller DK, Eilers RE (1988) The role of audition in infant babbling. *Child Dev* 59:441–449
- Perkell J, Matthies M, Lane H, Guenther F, Wilhelms-Tricarico R, Wozniak J, Guiod P (1997) Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. *Speech communication special issue: speech production. Models Data* 22:227–250
- Perry DW, Zatorre RJ, Petrides M, Alivisatos B, Meyer E, Evans AC (1999) Localization of cerebral activity during simple singing. *Neuroreport* 10:3979–3984
- Purcell DW, Munhall KG (2006a) Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J Acoust Soc Am* 120:966–977
- Purcell DW, Munhall KG (2006b) Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 119:2288–2297
- Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W (2000) Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport* 11:1997–2000
- Sapir S, McClean MD, Larson CR (1983) Human laryngeal responses to auditory stimulation. *J Acoust Soc Am* 73:315–321
- Shadmehr R, Mussa-Ivaldi FA (1994) Adaptive representation of dynamics during learning of a motor task. *J Neurosci* 14:3208–3224
- Tin C, Poon CS (2005) Internal models in sensorimotor integration: perspectives from adaptive control theory. *J Neural Eng* 2:S147–S163
- Ward WD, Burns EM (1978) Singing without auditory feedback. *J Res Singing* 1:24–44
- Watts C, Murphy J, Barnes-Burroughs K (2003) Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *J Voice* 17:185–194
- Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. *Science* 269:1880–1882
- Wyke BD (1974) Laryngeal neuromuscular control systems in singing. A review of current concepts. *Folia Phoniatr* 26:295–306
- Yip M (1995) Tone in east Asian languages. In: Goldsmith JA (ed) *The handbook of phonological theory*. Blackwell, Cambridge
- Yoshida Y, Saito T, Tanaka Y, Hirano M, Morimoto M, Kanaseki T (1989) Laryngeal sensory innervation: origins of sensory nerve fibers in the nodose ganglion of the cat. *J Voice* 3:314–320
- Zarate JM, Zatorre RJ (2005) Neural substrates governing audiovisual integration for vocal pitch regulation in singing. *Ann NY Acad Sci* 1060:404–408
- Zarate JM, Zatorre RJ (2008) Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage* 40:1871–1887
- Zemlin WR (1981) *Speech and hearing science: anatomy and physiology*. Prentice-Hall, Englewood Cliffs