Published in final edited form as:
Exp Brain Res. 2008 June ; 187(4): 613–621. doi:10.1007/s00221-008-1330-z.

Interactions between auditory and somatosensory feedback for voice F_0 control

Charles R. Larson,
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA clarson@northwestern.edu

Kenneth W. Altman,
Department of Otolaryngology, Mount Sinai School of Medicine, One Gustave L. Levy Pl., Box 1189, New York, NY 10029, USA

Hanjun Liu, and
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA

Timothy C. Hain
Departments of Neurology, Otolaryngology, and Physical Therapy/Human Movement Sciences, Northwestern University, 645 N. Michigan, Suite 1100, Chicago, IL 60611, USA

Abstract

Previous studies have demonstrated the importance of both kinesthetic and auditory feedback for control of voice fundamental frequency (F_0). In the present study, a possible interaction between auditory feedback and kinesthetic feedback for control of voice F_0 was tested by administering local anesthetic to the vocal folds in the presence of perturbations in voice pitch feedback. Responses to pitch-shifted voice feedback were larger when the vocal fold mucosa was anesthetized than during normal kinesthesia. A mathematical model incorporating a linear combination of kinesthesia and pitch feedback simulated the main aspects of our experimental results. This model indicates that a feasible explanation for the increase in response magnitude with vocal fold anesthesia is that the vocal motor system uses both pitch and kinesthesia to stabilize voice F_0 shortly after a perturbation of voice pitch feedback has been perceived.

Keywords

Vocalization; Auditory feedback; Larynx; Somatosensory; Anesthetization

Introduction

The role of sensory feedback for the control of voice fundamental frequency (F_0) has been extensively studied for many years. Somatosensory receptors located in and near the larynx provide the CNS with information on the frequency of vocal fold vibration, status of the vocal folds related to quality and effort of voice production, muscle contraction, joint movements and voice intensity (Eyzaguirre et al. 1966; Gozaine and Clark 2005; Shiba et al. 1999; Wyke 1974, 1983). Stimulation of the superior laryngeal nerve, laryngeal mucosa or cartilages results in reflexive contraction of laryngeal muscles (Andreatta et al. 2002; Eyzaguirre et al. 1966;

Correspondence to: Charles R. Larson.

Kirchner and Suzuki 1968; Ludlow et al. 1992; Sasaki and Masafumi 1976; Suzuki and Sasaki 1977; Wyke 1974, 1983). Although the functional significance of these reflexes remain uncertain, some of them may be important for control of voice F_0.

Support for the role of laryngeal somatosensory receptors in the control of voice F_0 has been provided through anesthetization procedures in humans. Application of local anesthetic to the vocal folds or superior laryngeal nerve leads to an increase in voice perturbation measures, a decrease in Wne control of voice F_0 and a deterioration in rapid adjustments of F_0 in a tone-matching task (Leonard and Ringel 1979; Sorensen et al. 1980; Sundberg et al. 1993; Tanabe et al. 1975). It is unlikely that the application of the anesthetic agent to the vocal fold mucosa could affect the muscles since there are no such documented effects of which we are aware. Although these findings suggest that laryngeal mucosal receptors are involved in voice control, the subjects were still able to perform the vocal tasks, suggesting other sensory receptors or auditory feedback is involved in the control of voice F_0.

Many studies have documented a clear role for auditory feedback to control F_0. Pre and postlingually deaf speakers have abnormal levels of F_0 and variations in F_0 (Binnie et al. 1982; Leder et al. 1987; Osberger and Hesketh 1988). Other studies demonstrated that masking of auditory feedback led to a deterioration in fine control of F_0 and inaccuracy in singing musical notes (Elliott and Niemoeller 1970; Mallard et al. 1978; Mürbe et al. 2002). Presentation of sudden perturbations in voice auditory pitch feedback while subjects are speaking or sustaining a vowel sound have demonstrated compensatory changes in voice F_0 production (Burnett et al. 1998; Chen et al. 2007; Donath et al. 2002; Hain et al. 2000; Jones and Munhall 2002; Kawahara 1995; Larson et al. 2001; Natke and Kalveram 2001; Sapir et al. 1983; Xu et al. 2004).

Given that both somatosensation and auditory feedback act together to stabilize F_0, the next logical questions that arise are how relatively important are the two mechanisms, and how do they interact. In this regard, Mallard et al. 1978 disrupted both kinesthesia and auditory feedback and measured voice F_0 while subjects attempted to match specific tones and while they were reading sentences with specific intonation contours. In four experimental conditions: normal/control condition, auditory masking, anesthetization of the vocal fold mucosa and masking combined with anesthesia; it was found that the greatest disruption occurred with masking alone. It was concluded that when disruptions occur in both modalities at the same time, the subjects may change their strategy and attend more closely to whatever minimal auditory cues are present.

While the explanation of Mallard and associates is feasible, it is difficult to confirm, as changes in “strategy”—nonlinear processing—are generally capable of explaining nearly any experimental result. We instead hypothesized that both auditory and kinesthetic feedback combine in a simple linear way (as in Fig. 7). This hypothesis leads to several straightforward predictions permitting experimental verification regarding interactions between audition and somatosensation.

First consider the case of our previous work where voice auditory pitch feedback was experimentally shifted while kinesthesia was left intact (Burnett et al. 1998; Hain et al. 2000; Larson et al. 2001; Xu et al. 2004). To the extent that kinesthetic feedback contributes to stabilization of F_0, experimental inferences that neglect kinesthetic feedback should result in an incorrectly low estimate of the coupling between auditory error and F_0 error (F_{gain} and F_{Tc} in Fig. 7). This occurs because we simulated and fit the response to this external auditory perturbation by adjusting the gain and time constant (F_{gain} and F_{Tc} in Fig. 7) relating auditory-error and F_0-error. Both a gain and time constant are necessary to fit the response over time. In this paradigm, because kinesthesia is unaffected by the experimental auditory perturbation,

and because the experimental perturbation drives vocal F_0 away from intended F_0, kinesthetic feedback would recognize the perturbation as an error and would oppose it. Thus kinesthetic feedback should reduce the response of laryngeal F_0 to an external auditory perturbation.

Second, we considered the case that would occur if pitch-shift stimuli were presented (F_0 error) after kinesthetic feedback was removed. For this situation, responses to the stimuli should be larger because there is a loss of the stabilizing influence of kinesthetic input. Furthermore, the F_{gain} and F_{Tc} derived by fitting to this condition should reflect the true coupling between auditory input and output, as they are not contaminated by the opposing effects of kinesthesia as in case 1.

The third and last case is the normal condition where both F_0 and kinesthetic error change together. An example of this situation might be a change in F_0 arising from a mechanical change in the vocal cord tension associated with fluctuations in muscle activity. In this situation, because auditory and kinesthetic feedback signals both change in the same direction, one would expect a larger overall gain (the sum of F_{gain} and KF_{gain}) and a larger compensatory F_0 response than with either of the two experimental conditions outlined above. This condition is much more difficult to test experimentally as it would require an endogenous stimulus, such as the sudden fluctuation of voice F_0 of a pubescent young man.

The present study was conducted to test the prediction made in the second case above—that eliminating kinesthesia would be associated with a greater response to an external perturbation of auditory F_0 than the response measured with kinesthesia intact. We measured responses to pitch-shifted voice feedback before and during temporary anesthetization of the vocal fold mucosa. In this situation, as outlined above, because auditory feedback and kinesthetic feedback should be opposed to each other, we hypothesized that removal of kinesthesia would result in stronger responses to auditory perturbations. Results confirmed our hypothesis, and responses to pitch-shifted voice feedback were greater when the vocal folds were anesthetized compared to the non-anesthetized condition. We successfully simulated this data by extending a previously postulated mathematical model of auditory feedback control of F_0 to include kinesthesia.

Methods

Nineteen normal, native English speaking adults (9 males, 10 females, age 22–59) with no history of neurological or speech disorders served as subjects. Data from two subjects (2 females) were later discarded because in one there were early signs of vocal nodules, and the other one had an abnormally rough voice. This study was approved by the Northwestern University Institutional Review Board Human Subjects Committee.

Subjects were seated in a medical examination chair with AKG boom-set headphones and attached microphone (model K 270 H/C) placed on their head. The microphone was located 1 in from their lips. The microphone signal was amplified (Grace Design, model 101 mic amplifier) and processed for pitch shifting through an Eventide Eclipse Harmonizer. The pitch-shifted signal was amplified (PreSonus HP4 amplifier) to a gain of 10 dB greater than voice amplitude and fed back to the subject over the headphones (Fig. 1). Acoustic calibrations were made with a B&K 2250 sound level meter and model 4100 in-ear microphones. A laboratory computer running MIDI software (Max/MSP; Cycling 74) was used to control the harmonizer. Microphone, headphone and control signals were low pass Wltered at 5 kHz and digitized (12 bit) at 10 kHz and recorded on a second computer with Chart software (AD Instruments).

Subjects were Wrst instructed to produce 5-s vocalizations of the /u/ vowel at a comfortable pitch and amplitude level. The exact F_0 level differed for each subject, but by having them vocalize at their customary conversational level, the relative contraction levels of laryngeal

muscles was relatively constant across individuals. Also, since during data analysis (see below), the F_0 contours were converted to the cent scale, comparisons between different F_0 levels in absolute frequency (i.e., Hz) can be made. During each of eight vocalizations, the feedback signal to the subject was shifted up or down Wve times in a randomized sequence. The interval between successive stimuli within a set was randomized between 400 and 1,000 ms. Subjects paused between each vocalization to take a breath. In the Wrst set of vocalizations, the pitch-shift stimulus was 100 cents (100 cents = 1 semitone), 200 ms duration and in the second sequence, 50 cents, 200 ms duration. The stimulus output of the harmonizer was calibrated on the “cent” interval because this is a logarithmic scale related to a 12-tone musical scale and is not locked to specific frequencies. The 50 and 100-cent stimulus magnitudes were chosen to allow comparison with previous studies in our laboratory (Burnett et al. 1998).

After the first sets of trials were completed, the subjects’ vocal folds were anesthetized. For the first two subjects, subjects were instructed to lean forward, tilt their head upward, open their mouth and stick out their tongue. The physician then held the tongue with gauze and inserted a curved tube into the oral pharynx aiming at the larynx and administered a 2-s spray of a local anesthetic—14% benzocaine, 2% butyl aminobenzoate, 2% tetracaine (Topical Cetacaine, Cetylite Industries Inc., Pennsauken, NJ, USA). Following this, the physician inserted a fiberoptic scope through the nose and by touching the vocal folds with the end of the scope without eliciting a cough, confirmed that the vocal folds were anesthetized.

After testing the first two subjects, we became concerned that the application of the anesthetic by the spray technique may have been anesthetizing mucosa throughout the laryngeal/pharyngeal region and not restricted to the vocal folds. We therefore switched to a new technique for the remainder of the subjects. These subjects had their vocal folds anesthetized by spraying 2 cc of xylocaine 4% topical solution (Lidocaine, AstraZeneca Pharmaceuticals, Wilmington, DE, USA) directly onto the surface of the vocal folds. This was done by first inserting the fiberoptic scope through a sleeve (Endosheath, Vision Sciences, Natick, MA, USA), which also contained a side-tube for the anesthetic agent. The subject's nasal epithelium was first sprayed with pontocaine to reduce discomfort from insertion of the fiberoptic scope. A 3 cc syringe containing lidocaine was then attached to the side-tube of the sleeve, and the scope with sleeve was gently inserted through the nose to the larynx. With the scope tip in position above the vocal folds, the subject was instructed to produce the vowel /i/ at a high pitch. During this production, the physician applied 2 cc of lidocaine directly on to the upper surface of the vocal folds. After 30 s, the physician touched the vocal folds with the fiberoptic scope and by failing to elicit a cough or other gagging type of response, confirmed the anesthetization of the vocal folds in all subjects. Within 2 min of the anesthetization of the vocal folds, the sequence of pitch-shift testing was repeated. Four subjects were again tested 15 min after anesthetization of the vocal folds. Scheduling conflicts prohibited testing of the other 14 subjects a third time. Two of the subjects who had received the first form of anesthesia delivery were tested again with the second method.

Data were analyzed by transferring the digitized files to a computer running Igor (Wavemetrics Inc.) software. The vocal signal was then transferred to a program that extracted pulses corresponding to each cycle of vocal fold vibration (Praat). The pulses were then transferred back to the Igor program and converted to a waveform in which voltage corresponded to voice F_0 measured in cents. This waveform, an F_0 contour, was then displayed on a computer screen. Figure 2a provides a short sample of an F_0 contour of the feedback to the subject and the subject's vocal output. In this sample, one upward and three downward pitch-shift stimuli are seen in the feedback trace. For the purpose of obtaining the average F_0 response across several stimulus presentations, F_0 contours with a 200 ms pre- and 500 ms post-stimulus window were time-aligned with a stimulus pulse on the computer screen. Figure 2b illustrates a waterfall display of the F_0 contours for each of the trials for one subject in one condition. Inspection of

the waterfalls was done prior to averaging the signals to eliminate trials in which there was an unusually large deflection related to an error in the extraction of the F_0 contours or if there was some type of vocal interruption such as a cough during one of the trials. Averaged F_0 contours based on the trials in the waterfall display, representing the F_0 response to the pitch-shifted voice feedback, were calculated for each subject and for each of the separate conditions; 50 or 100 cent pitch-shift in the up or down direction (Fig. 2c).

Following the averaging process, the mean and standard deviation of the F_0 contour during the pre-stimulus period was calculated. The criteria for defining a response were a change in the F_0 contour that exceeded 2 SDs of the pre-stimulus F_0 that occurred at least 60 ms following the stimulus onset (response latency) and lasting at least 50 ms before returning to a value within the 2 SD boundaries of the pre-stimulus F_0 (see Fig. 2c). Response magnitude was defined as the greatest value of the averaged F_0 contour following the response onset. Measures of response magnitude and latency were taken from the averaged waveforms and submitted to significance testing with a repeated measures ANOVA using SPSS (v. 11.0). Assumptions of compound symmetry and circularity for a repeated measures ANOVA were met. We did not test for a difference between the 100 and 50-cent stimuli because we always tested 100 cent first, and a difference between them could be attributed to an order effect.

As a test of a possible effect of changes in voice F_0 level between the pre- and post-anesthesia conditions on response magnitude, the average F_0 level for all experimental conditions was measured from an FFT of the vocal waveform for the entire set of vocalizations from all the subjects. A Pearson product correlation coefficient was calculated between this difference in the average F_0 level (average post-anesthesia F_0 level—average pre-anesthesia F_0 level) and the pitch-shift response measures.

We extended a linear model of F_0 control (Hain et al. 2000) using auditory feedback to include kinesthetic input. This model topology is structured in a similar way to the speech production model of Guenther et al. (2006), having both a feedforward and feedback pathway, and auditory and kinesthetic feedback. The model of Guenther and associates is a neural network model and due to the array of neural weights intrinsic to such constructs, it contains many nonlinearities and free parameters corresponding to individual simulated neurons. Our model contains only a small number of parameters, but they are sufficient to simulate a wide variety of experimental paradigms involving perturbed auditory feedback. The feasibility of our extended model was established by simulating our data.

Results

All the subjects reacted to the pitch-shift perturbation with a change in voice F_0. Of the 136 possible responses, 121 were in the direction opposite to that of the stimulus and were categorized as opposing responses. There were ten responses in the same direction as the stimulus (“following” responses) and five that did not meet our criteria of a response (non-response). Four of the non-responses were produced under the anesthesia condition.

Figures 3 and 4 display exemplar responses for one subject in the control and anesthesia conditions for both upwards and downwards stimulus directions for 50 (Fig. 3) and 100 cent (Fig. 4) perturbations. Although there are some differences in the form of the responses, all are in the opposing direction and have latencies between 100 and 200 ms. Figure 5 displays the averaged responses across all subjects and conditions along with the 95 confidence intervals (shaded). Average response curves in the “anesthetize” condition are slightly larger than in the “normal” condition. For the 50-cent stimuli, responses were less variable than with the 100-cent stimuli. Much of the variability for responses with the 100-cent down stimuli is attributed to variations in response magnitude and timing of the responses (not shown). Figure 6 displays

boxplots of response magnitude for the normal and anesthetic conditions for both upwards and downwards stimuli (opposing responses only). There was a significant main effect for response magnitude between the pre and post anesthetic condition [$F(1,16) = 10.453$, $P = 0.005$] but not for stimulus direction [$F(1,16) = 0.024$, $P = 0.878$]. The overall mean of response magnitude in the anesthetized condition (25 ± 15 cents) was larger than for the non-anesthetized condition (20 ± 17 cents). There was no significant interaction between stimulus direction and anesthesia state [$F(1,16) = 0.144$, $P = 0.709$]. Response magnitudes in the four subjects that were re-tested after anesthetization procedures, returned to the pre-anesthetic levels with the final testing.

There was a significant main effect for stimulus direction on response latency [$F(1,16) = 20.249$, $P = 0.000$], in which longer latencies were associated with the upward stimuli (147 ± 72 ms) compared to the downward stimuli (104 ± 49 ms). There were no significant effects for the anesthetic condition [$F(1,16) = 0.804$, $P = 0.383$] on response latency or interactions between anesthetic condition and stimulus direction [$F(1,16) = 0.534$, $P = 0.476$].

Results of the correlation analysis between the F_0 level differences of the pre- and post-anesthesia conditions and the mean reflex magnitudes yielded an $r < 0.0001$. Thus, there was no correlation between the overall voice F_0 level and the magnitude of the response to pitch-shifted feedback.

In an attempt to provide a quantitative hypothesis, we extended a previously presented model of F_0 control incorporating auditory feedback (Hain et al. 2000), to include kinesthetic feedback. With our model we simulated our data and showed feasibility for a simple linear feedback implementation.

In our previous model, we used an auditory negative feedback loop, incorporating suitable delays to simulate auditory feedback data, which has a delay of about 100 ms. In this adaptation, shown in Fig. 7, we added a kinesthetic feedback loop. The kinesthetic feedback loop was designed as follows. We assumed that the organization was similar to that of the auditory feedback loop, but that the gain and delay might differ. Although there are substantial differences in auditory and kinesthetic signal processing, nevertheless in both instances, for feedback to be accomplished an F_0 error signal must be developed and used to change the drive to F_0. We chose the simplest possible model design—feedback through a delay and simply sought to establish feasibility.

Our earlier auditory feedback model incorporated an auditory processing delay, a low pass filter, and a feedback gain. In the extended model designed to model the present data set, we assumed that the filter was part of the output pathway and shared between the auditory and kinesthetic feedback loops. The delay and gain of the kinesthetic pathway, KF_{delay} and KF_{gain} in Fig. 7, were the unknowns that were identified.

Using the optimization package of Matlab (Nantick, MA, USA), we first fit our previous model of F_0, without any kinesthetic input, to data collected under anesthesia. This model has only two free parameters, a gain (F_{gain}) and time constant (F_{Tc}), as anesthesia eliminates the kinesthetic feedback loop. A gain of 0.273, and a time constant of 0.271 s provided the best fits. Next, with these two parameters set to their optimal values, we optimized the kinesthetic part of the model to experimental data in which kinesthesia was present. We allowed the optimizer to adjust the gain of kinesthetic feedback, KF_{gain}, and kinesthetic delay, KF_{delay}, as mentioned above. A low pass filter identical to that used for F_0 was also applied to kinesthetic error. The data were best fit by $KF_{gain} = 0.80$, and $KF_{delay} = 0.020$. This emergent result for a kinesthetic delay of only 20 ms, much shorter than the 100 ms delay needed to simulate the results of auditory perturbations (Hain et al. 2000), is consistent with known neurophysiology. The delay between sensory stimulation of the larynx and laryngeal muscle responses is on the order of 18–25 ms (Ludlow et al. 1992). The combination of a higher gain and shorter delay

for kinesthesia as opposed to auditory feedback parameters, suggests that over this short time frame, kinesthetic error is weighted more heavily than auditory error for controlling F_0.

Figures 3 and 4 show experimental data for the 50 and 100 cents stimuli overlaid with simulations (dashed lines). For this experimental paradigm where kinesthesia and auditory input are in opposition, the simulated responses with kinesthesia absent, are 1.65 times larger than those where kinesthesia is present.

Discussion

The present study has demonstrated that temporary anesthetization of the mucosa of the vocal folds results in a larger response to pitch-shifted voice auditory feedback compared with the pre-anesthetic condition. The findings suggest that both kinesthesia and auditory feedback are used for the control of vocalization. Furthermore, our mathematical simulations suggest that early in the response, kinesthesia alone provides feedback control, but after about 100 ms, auditory feedback also participates.

There have been several previous studies that have also studied the effects of anesthesia of laryngeal nerves or mucosal tissues on voice F_0 (Mallard et al. 1978; Sorensen et al. 1980; Sundberg et al. 1995; Tanabe et al. 1975; Yang and Chen 2005). The main effects of anesthesia were an increase in F_0 variability and/or a decrease in accuracy of producing a specific F_0. None of the study designs were directly comparable to the present study, but their results are compatible with our proposed kinesthetic feedback loop.

To explain our results, we proposed a simple linear extension of a previous mathematical model of F_0 control (Hain et al. 2000). Our model is structured in a similar way to the speech production model of Guenther et al (2006), having both a feedforward and feedback pathways, but our previous implementation did not include kinesthetic input. To simulate this set of experimental data, we added a kinesthetic negative feedback loop, configured to stabilize F_0 against external perturbations. When auditory input is perturbed, but kinesthesia remains unchanged, kinesthetic input and auditory input are in opposition to each other, and the stabilizing response is small due to their interference. When kinesthetic input is removed, the response to auditory input is not reduced by kinesthesia, and therefore it is larger. We call this the “linear interaction hypothesis”. Our modeling demonstrates the feasibility of this particular topology and set of parameters but it does not exclude the possibility that other topologies or parameter sets might be sufficient to fit the same data.

An alternative and/or complementary hypothesis that might also explain our results is a central non-linear interaction. Our data documented larger magnitude responses to pitch-shifted feedback in the presence of vocal fold anesthesia suggesting greater reliance on auditory feedback. While the simplest explanation is that there simply was a decrease in the interfering kinesthetic input, it is also possible that there was an additional increase in auditory gain. In the non-linear interaction hypothesis, the system increases auditory feedback gain to control voice F_0 in response to a disruption of kinesthetic feedback. Changes of gain are intrinsically non-linear operations. In the context of our model, this could be implemented by an increase in the gain of the F_0 auditory feedback loop, F_{gain}. To the extent that this additional non-linear response was present, our simulations would not require as large a contribution from kinesthesia to explain our results.

While the more complex non-linear control mechanism is not necessary to explain our results, our data do not exclude it. Although generally in modeling, simpler explanations are preferred, this idea is worth considering for several reasons. Previous work with other senses has shown that it is common for different sensory modalities to interact non-linearly (Horak et al. 2001; Li et al. 1999; Shimojo and Shams 2001). In a system where multiple sensory modalities are

involved in control, if one modality is lost, disrupted or unreliable, remaining functional modalities can be reweighted (Horak et al. 1994). These same types of interactions in the phonatory control system could allow the weighting of different sensory modalities to be modified to obtain optimal regulation of voice F_0. A similar non-linear interaction involving upweighting of kinesthetic input could explain the ability of singers to hold a precise note even in the presence of high levels of competing auditory feedback from other voices or instruments (Sundberg 1987).

The main argument against the non-linear interaction hypothesis is that it is not necessary. Our simulations show that the simpler linear interaction hypothesis is feasible by itself. Nevertheless, a combination of both of these explanations is also viable. One might be able to further test the non-linear interaction hypothesis by performing similar experiments to those of Mallard et al. (1978) using contemporary methodology.

Either explanation requires one to accept that there is a kinesthetic signal encoding F_0. Previous work suggests that kinesthetic information is indeed available from mucosal mechanoreceptors sensitive to laryngeal vibration and movement of the cartilages (Gozaine and Clark 2005; Shiba et al. 1999). Discharge from these receptors could provide the speaker with information related to the state of muscle contraction, position of the vocal folds and a direct estimate of the voice F_0.

If we accept that both auditory and kinesthetic feedback are used to control F_0, then the question arises naturally of their relative importance. Auditory feedback is clearly very important—the onset of deafness leads to a rapid deterioration in the ability to control voice F_0 and amplitude, and much greater effects on F_0 level control and variations in F_0 have been observed in deaf speakers (Binnie et al. 1982; Leder et al. 1987; Osberger and Hesketh 1988; Svirsky et al. 1992) than the results previously discussed for anesthesia. On the other hand, our results suggest that kinesthesia is even more important than auditory feedback, as the best fit for the gain for kinesthetic error (0.803) is more than twice that for auditory error (0.273). Although kinesthesia may be relatively more important than auditory input, nevertheless it is part of a redundant negative feedback control system. As long as correct auditory input is available, partial loss of the kinesthetic loop gain should not cause critical disruptions to pitch control. Redundant control loops are common in critical biological systems, as they make the overall system more robust.

Furthermore, the importance of auditory and kinesthetic feedback may vary with respect to the time from onset of phonation. Mallard et al. (1978) suggested that auditory and kinesthetic control may be used to a different extent as well as at differing times with respect to the onset of phonation and for differing frequency ranges. Supporting this idea, unlike auditory information, kinesthetic information could allow the speaker to know if the cartilages were in the correct position, and the vocal folds had the correct length and stiffness prior to the onset of vocalization. Such information may provide the necessary substrate for “prephonatory tuning” proposed by Wyke (1974, 1983). Prephonatory tuning refers to the preparation undertaken by a singer prior to vocalization. It involves contraction of muscles to correctly position the laryngeal cartilages, and to adjust vocal fold tension to produce the desired note.

In contrast, auditory feedback can provide feedback on voice F_0, quality and loudness, which the laryngeal mucosal receptors could not. Thus, auditory and somatosensory provide complimentary, but in some cases different types of feedback for the control of the voice. Moreover, auditory feedback, being in the acoustic domain, could also be integrated with other acoustical variables related to the environment.

Consistent with this idea, our data and modeling suggested that kinesthetic error acts with a shorter latency (20 ms) compared to auditory error (100 ms). Logically, the more rapidly

available kinesthetic feedback should be more heavily weighted over the short term, while auditory feedback should eventually dominate. In fact, because auditory feedback requires 100 ms, for the first 100 ms of phonation, all the control should logically be kinesthetic.

In conclusion, we have shown that anesthesia of the vocal folds increases the response to an externally imposed auditory perturbation. We have also shown that a simple modification of a linear feedback model of F_0 control, incorporating kinesthetic feedback, is a feasible explanation for this effect. Although our simulations demonstrate that a completely linear explanation is feasible, a plausible alternative or additional explanation of this effect is a non-linear interaction where auditory feedback is upweighted, when kinesthetic input is unavailable or simply less reliable. Overall, combining our data and the literature, it seems likely that there are different roles for these two sensory channels: auditory feedback may be used for gross control of F_0 while kinesthesia is used when auditory feedback is not available (i.e. prephonatory tuning or singing in a very noisy environment), and for fine rapid control of F_0.

Acknowledgment

This study was supported by a grant from NIH Grant No. DC006243-01A1. We thank Dr. David Conley for his assistance in administering anesthetic and Mr. Chun Liang Chan for computer programming.

References

Andreatta RD, Mann EA, Poletto CJ, Ludlow CL. Mucosal afferents mediate laryngeal adductor responses in the cat. J Appl Physiol 2002;93:1622–1629. [PubMed: 12381746]

Binnie CA, Daniloff RG, Buckingham HW. Phonetic disintegration in a five-year-old following sudden hearing loss. J Speech Hear Disord 1982;47:181–189. [PubMed: 7176596]

Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F_0 responses to manipulations in pitch feedback. J Acoust Soc Am 1998;103:3153–3161. [PubMed: 9637026]

Chen SH, Liu H, Xu Y, Larson CR. Voice F_0 responses to pitch-shifted voice feedback during English speech. J Acoust Soc Am 2007;121:1157–1163. [PubMed: 17348536]

Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables. J Acoust Soc Am 2002;111:357–366. [PubMed: 11831808]

Elliott L, Niemoeller A. The role of hearing in controlling voice fundamental frequency. Int Audiol 1970;IX:47–52.

Eyzaguirre, C.; Sampson, S.; Taylor, JR. The motor control of intrinsic laryngeal muscles in the cat.. In: Granit, R., editor. Nobel symposium I: muscular afferents and motor control. Wiley; New York: 1966. p. 209-225.

Gozaine TC, Clark KF. Function of the laryngeal mechanoreceptors during vocalization. Laryngoscope 2005;115:81–88. [PubMed: 15630372]

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 2006;96:280–301. [PubMed: 16040108]

Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. Exp Brain Res 2000;130:133–141. [PubMed: 10672466]

Horak FB, Earhart GM, Dietz V. Postural responses to combinations of head and body displacements: vestibular–somatosensory interactions. Exp Brain Res 2001;141:410–414. [PubMed: 11715086]

Horak FB, Shupert CL, Dietz V, Horstmann G. Vestibular and somatosensory contributions to responses to head and body displacements in stance. Exp Brain Res 1994;100:93–106. [PubMed: 7813657]

Jones JA, Munhall KG. The role of auditory feedback during phonation: studies of Mandarin tone production. J Phon 2002;30:303–320.

Kawahara, H. Hearing voice: transformed auditory feedback effects on voice pitch control.. Computational auditory scene analysis and International joint conference on artificial intelligence; Montreal. 1995.

Kirchner, JA.; Suzuki, M. Annals of New York Academy of Sciences. 1968. Laryngeal reflexes and voice production.; p. 98-109.

Larson CR, Burnett TA, Bauer JJ, Kiran S, Hain TC. Comparisons of voice F_0 responses to pitch-shift onset and offset conditions. J Acoust Soc Am 2001;110:2845–2848. [PubMed: 11785786]

Leder SB, Spitzer JB, Kirchner JC. Speaking fundamental frequency of postlingually profoundly deaf adult men. Ann Otol Rhinol Laryngol 1987;96:322–324. [PubMed: 3605956]

Leonard RJ, Ringel RL. Vocal shadowing under conditions of normal and altered laryngeal sensation. J Speech Hear Res 1979;22:794–817. [PubMed: 513688]

Li Z, Morris KF, Baekey DM, Shannon R, Lindsey BG. Responses of simultaneously recorded respiratory-related medullary neurons to stimulation of multiple sensory modalities. J Neurophysiol 1999;82:176–187. [PubMed: 10400946]

Ludlow C, Van Pelt F, Koda J. Characteristics of late responses to superior laryngeal nerve stimulation in humans. Ann Otol Rhinol Laryngol 1992;101:127–134. [PubMed: 1739256]

Mallard AR, Ringel RL, Horii Y. Sensory contributions to control of fundamental frequency of phonation. Folia Phoniatr 1978;30:199–213.

Mürbe D, Pabst F, Hofmann G, Sundberg J. Significance of auditory and kinesthetic feedback to singers' pitch control. J Voice 2002;16:44–51. [PubMed: 12002886]

Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. J Speech Lang Hear Res 2001;44:577–584. [PubMed: 11407562]

Osberger, MJ.; Hesketh, LJ. Speech and language disorders related to hearing impairment.. In: Lass, NJ.; McReynolds, LV.; Northern, JL.; Yoder, DE., editors. Handbood of speech-language pathology and audiology. B.C. Decker; Philadelphia: 1988. p. 858-886.

Sapir S, McClean M, Luschei ES. Effects of frequency-modulated auditory tones on the voice fundamental frequency in humans. J Acoust Soc Am 1983;73:1070–1073. [PubMed: 6841812]

Sasaki C, Masafumi S. Laryngeal reflexes in cat, dog, and man. Arch Otolaryngol 1976;102:400–402. [PubMed: 938318]

Shiba K, Miura T, Yuza J, Sakamoto T, Nakajima Y. Laryngeal afferent inputs during vocalization in the cat. Neuroreport 1999;10:987–991. [PubMed: 10321472]

Shimojo S, Shams L. Sensory modalities are not separate modalities: plasticity and interactions. Curr Opin Neurobiol 2001;11:505–509. [PubMed: 11502399]

Sorensen D, Horii Y, Leonard R. Effects of laryngeal topical anesthesia on voice fundamental frequency perturbation. J Speech Hear Res 1980;23:274–283. [PubMed: 7442190]

Sundberg, J. The science of the singing voice. Northern Illinois University Press; Dekalb: 1987.

Sundberg, J.; Iwarsson, J.; Billström, A-MH. Significance of mechanoreceptors in the subglottal mucosa for subglottal pressure control in singers.. 22nd annual symposium care of the professional voice; Philadelphia. 1993.

Sundberg J, Iwarsson J, Billstrom AH. Significance of mechanoreceptors in the subglottal mucosa for subglottal pressure control in singers. J Voice 1995;9:20–26. [PubMed: 7757146]

Suzuki M, Sasaki C. Effect of various sensory stimuli on reflex laryngeal adduction. J Otol Rhinol Laryngol 1977;86:30.

Svirsky MA, Lane H, Perkell JS, Wozniak J. Effects of short-term auditory deprivation on speech production in adult cochlear implant users. J Acoust Soc Am 1992;92:1284–1300. [PubMed: 1401516]

Tanabe M, Kitajima K, Gould W. Laryngeal phonatory reflex. The effect of anesthetization of the internal branch of the superior laryngeal nerve: acoustic aspects. Ann Otol Rhinol Laryngol 1975;84:206–212. [PubMed: 1124909]

Wyke B. Laryngeal myotatic reflexes and phonation. Folia Phoniatr 1974;26:249–264.

Wyke, B. Neuromuscular control systems in voice production.. In: Bless, DM.; Abbs, JH., editors. Vocal fold physiology. College-Hill; San Diego: 1983. p. 71-76.

Xu Y, Larson C, Bauer J, Hain T. Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. J Acoust Soc Am 2004;116:1168–1178. [PubMed: 15376682]

Yang CC, Chen SH. Impact of topical anesthesia on acoustic characteristics of voice during laryngeal telescopic examination. Otolaryngol Head Neck Surg 2005;132:110–114. [PubMed: 15632920]

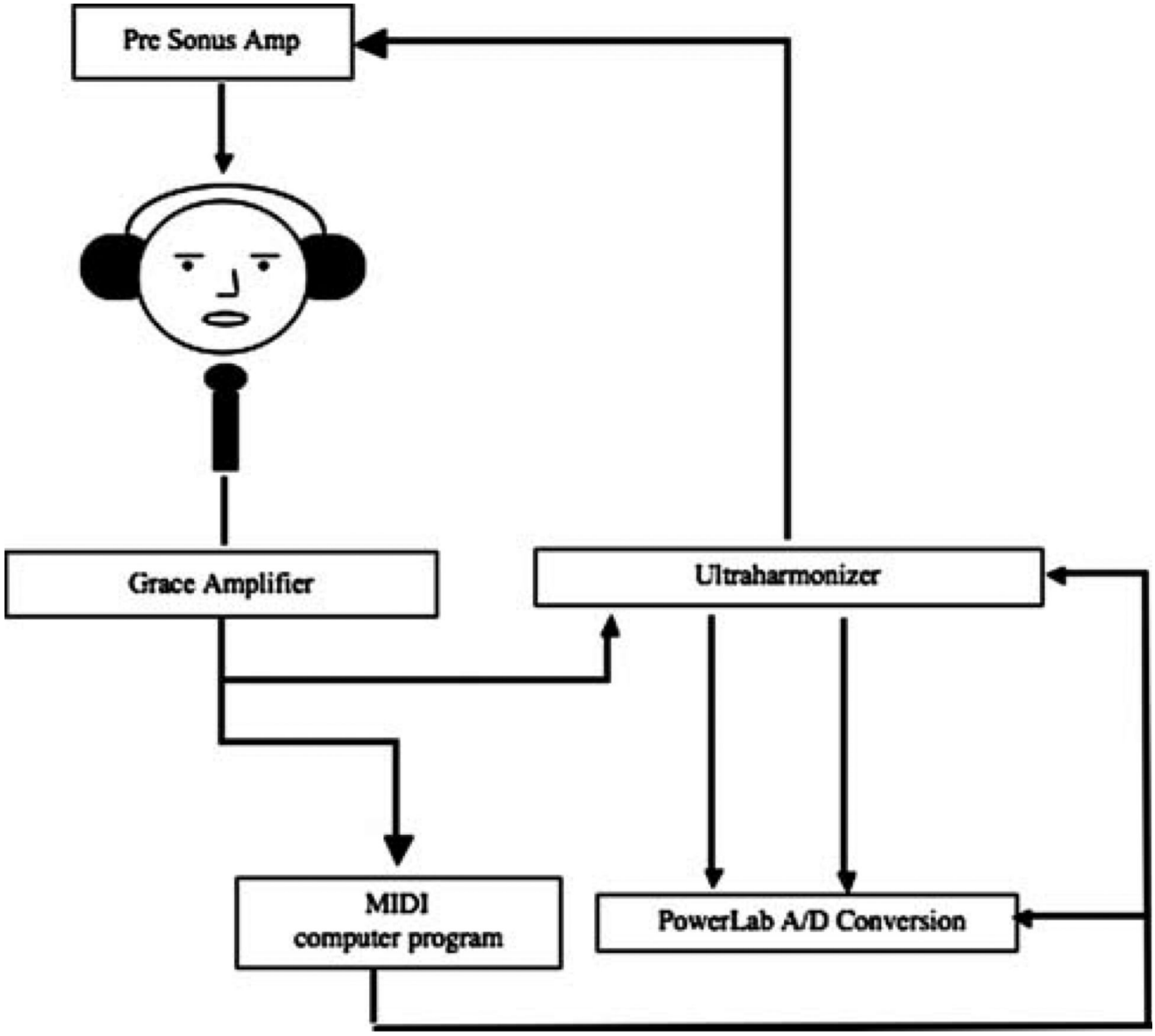

Fig. 1.
Schematic illustration of apparatus used for the pitch-shifting of voice auditory feedback

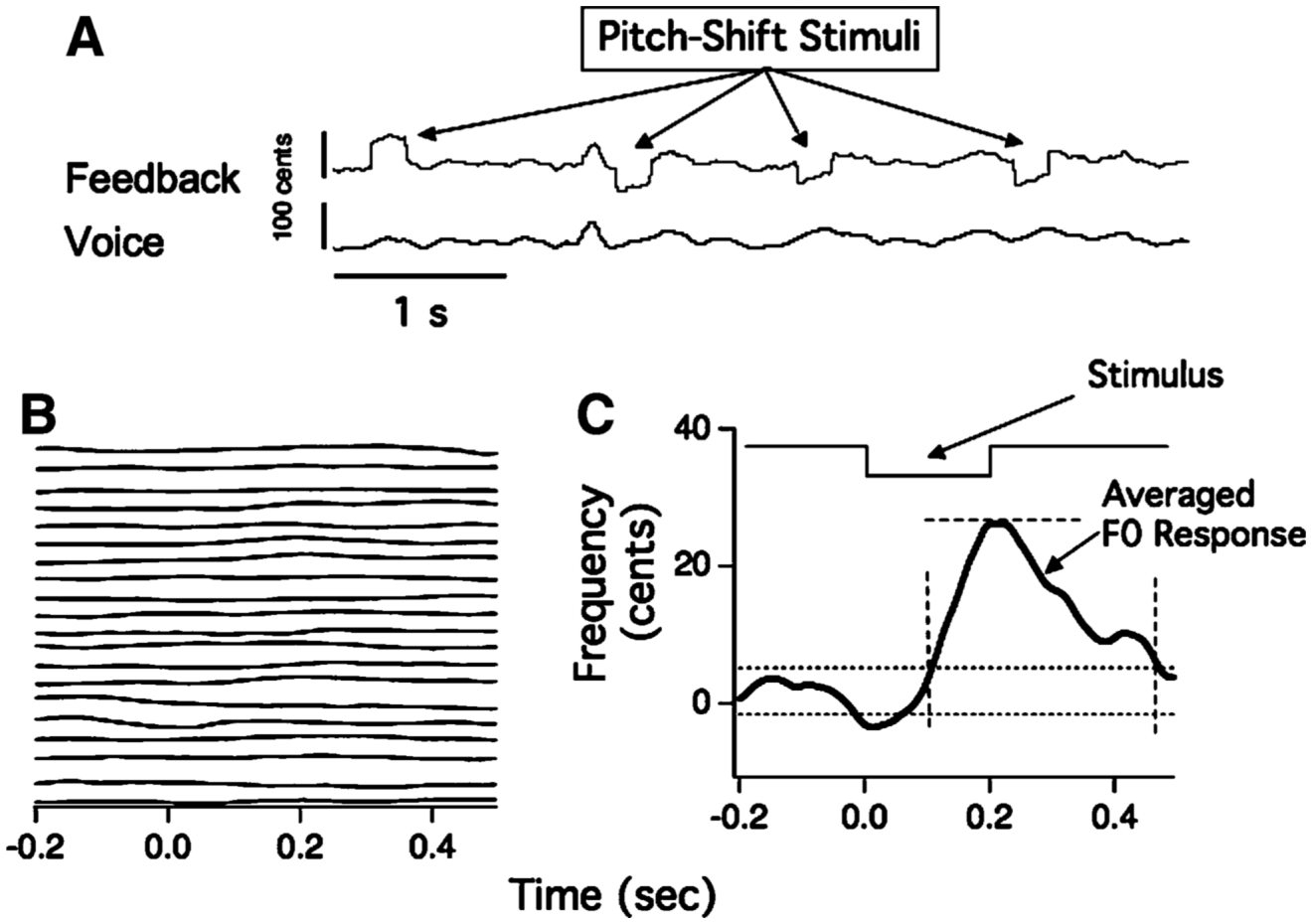

Fig. 2.

a F_0 contour (*bottom trace*) and auditory feedback (*top trace*). Pitch-shift perturbations are illustrated in feedback trace. *b* F_0 contours for many individual trials. The pitch-shift stimulus for each trace (not shown) were aligned at time = 0, as shown on scale at bottom. **c** A single averaged F_0 response. *Horizontal dotted lines* indicate ±2SDs of the prestimulus mean averaged F_0. *Vertical dashed lines* affixed to averaged F_0 response indicate onset and offset times of the response. *Horizontal dashed line* indicates response magnitude. Time and direction of the pitch-shift stimulus is indicated by *square trace* at top

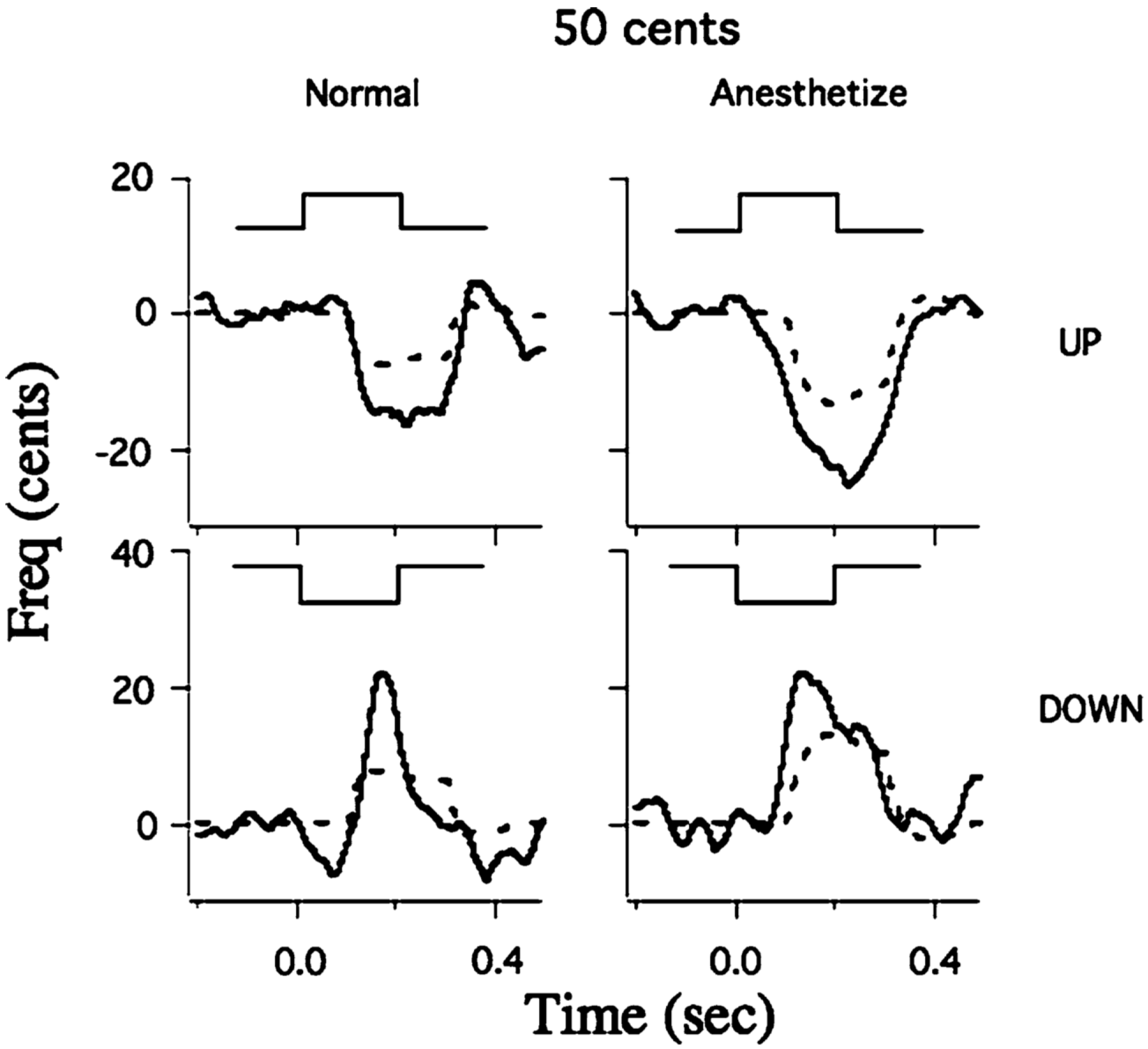

Fig. 3.

Averaged voice F_0 contours in normal condition (*left*) and with anesthesia (*right*) following 50 cent pitch-shift stimuli. *Solid lines* represent experimental F_0 responses. *Dotted lines* are simulations using the mathematical model of Fig. 7. *Square brackets* represent timing and direction of pitch-shifted feedback. Stimulus onset is at time 0.0

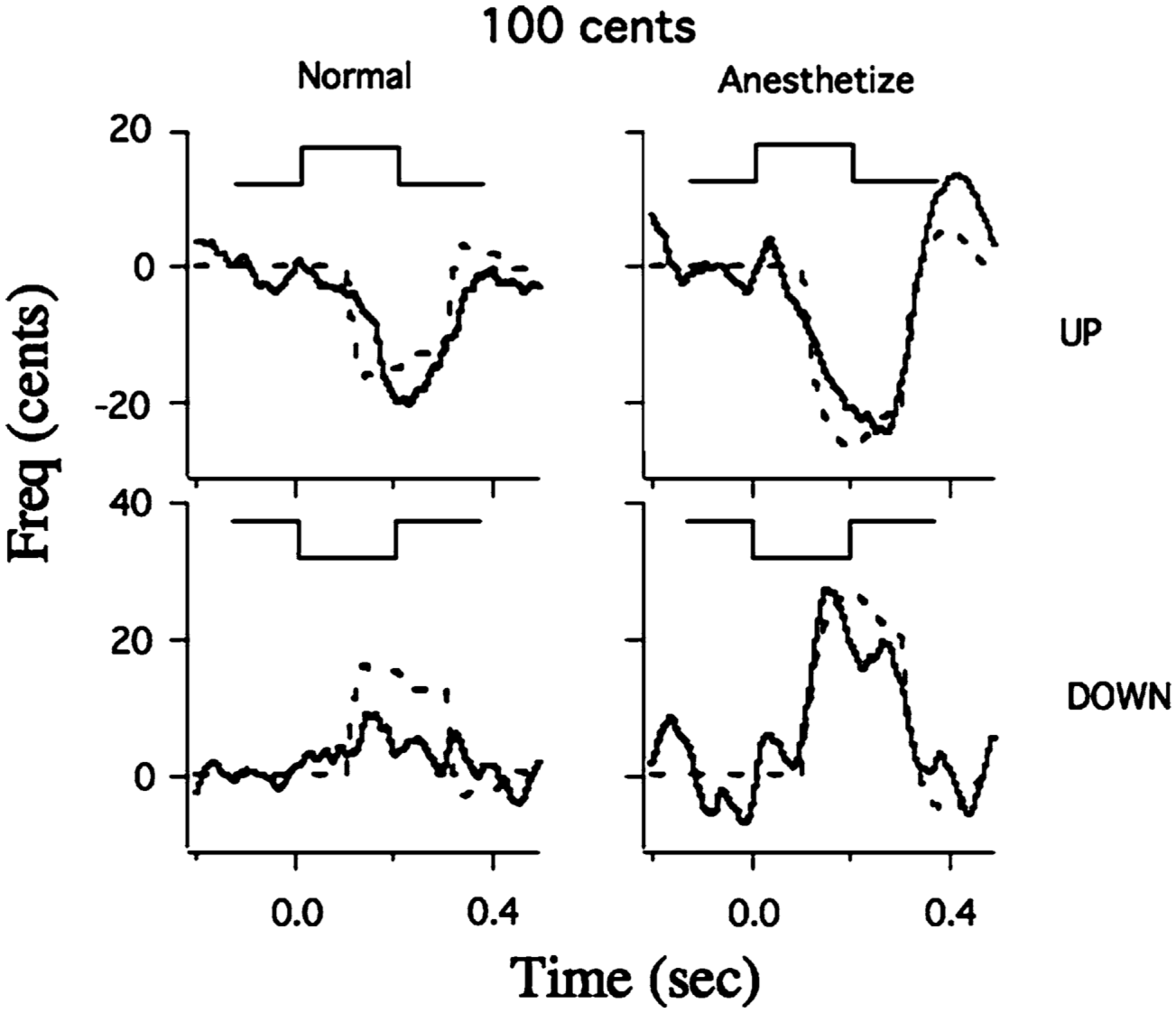

Fig. 4.
Voice F_0 contours in normal and anesthetic condition following 100 cent pitch-shift stimuli. Figure is otherwise formatted as in Fig. 3

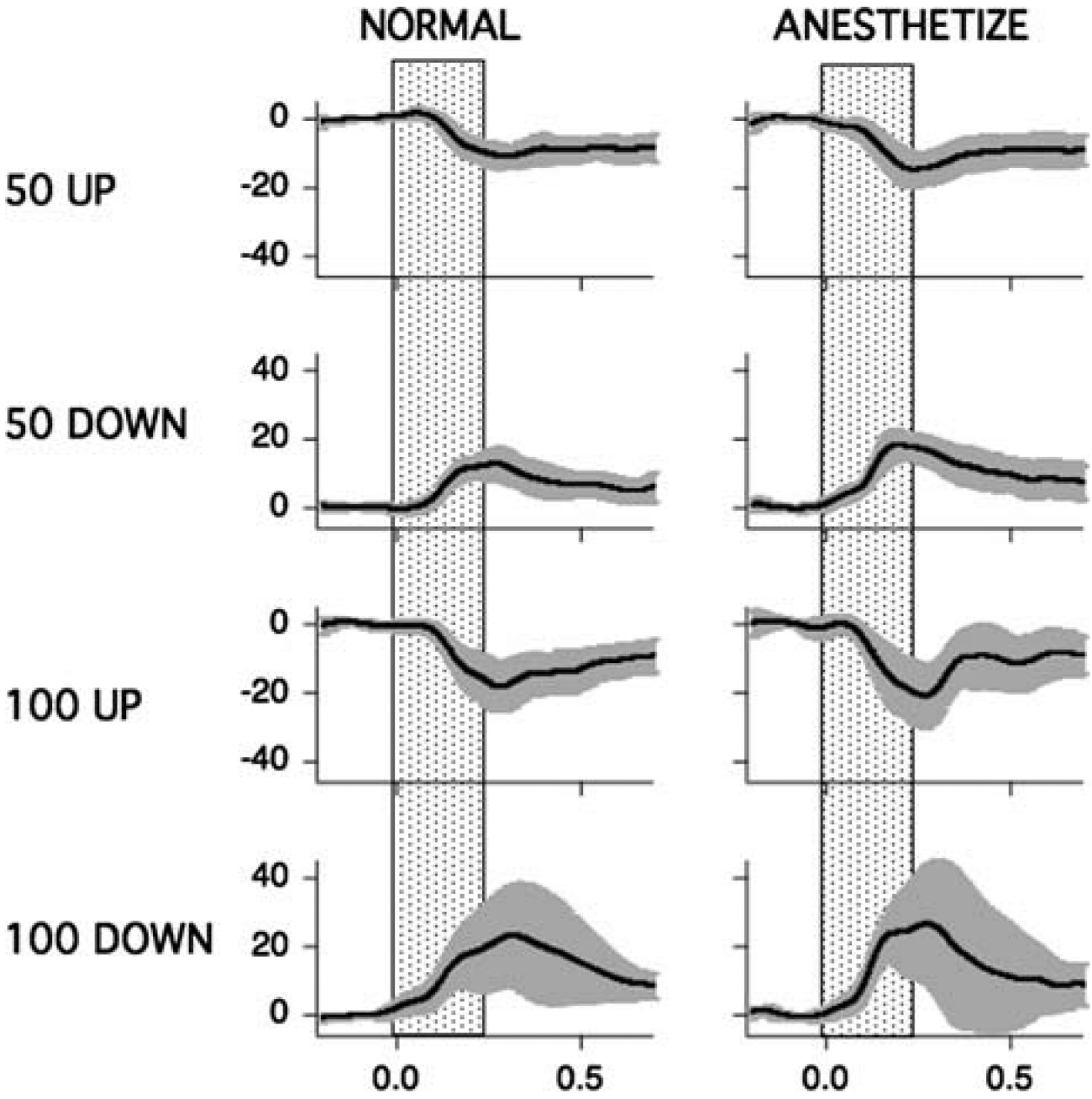

Fig. 5.

Composite averages of all responses for all subjects and conditions; 50 cent up, 50 cent down, 100 cent up, 100 cent down, normal and anesthetize. *Solid lines* represent average F_0 contour. *Gray shading* represents ±95% confidence intervals of averaged traces. *Vertical shaded rectangles* represent time of the pitch-shift stimulus. Horizontal axis is time in seconds, and vertical axis is frequency in cents

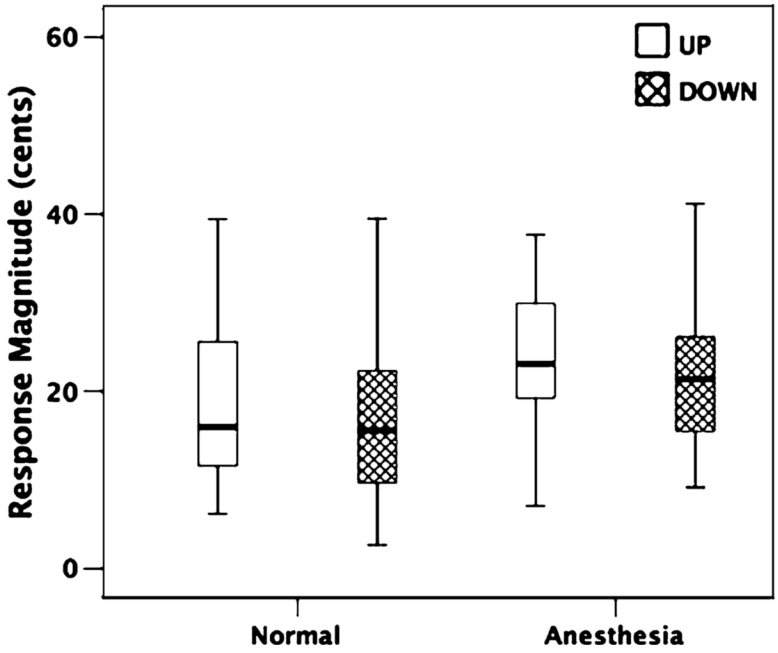

Fig. 6.

Boxplots illustrating response magnitudes with and without anesthesia of the vocal folds. *Hatched bars* are for downward stimuli and *open bars* for upward stimuli

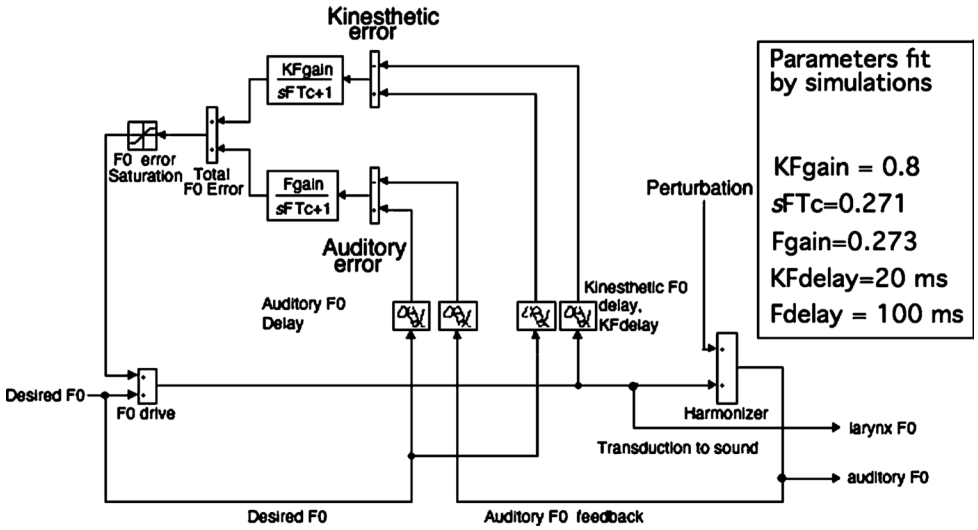

Fig. 7.

Model of audio-vocal system incorporating auditory and kinesthetic feedback. "Desired F_0" is converted through a "*black box*" representing the entire central vocal production system (summing junction labeled F_0 drive) into two F_0 signals. "Larynx F_0" is the F_0 perceived through kinesthesia. "Auditory F_0" is the F_0 perceived by the auditory system. An efferent copy of desired F_0 is the reference signal for both auditory and kinesthetic feedback loops. Separate delay elements process inputs for auditory and kinesthetic input. Kinesthetic and auditory error is computed from the delayed differences between desired F_0 and perceived larynx or auditory F_0 as appropriate. Kinesthetic and auditory error are low-pass filtered and scaled for both loops, limited to a maximum of 50 cents, and added into the F_0 drive signal