Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F_0 and amplitude control[a]

Charles R. Larson[b] and Jean Sun
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

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

(Received 20 July 2006; revised 15 February 2007; accepted 20 February 2007)

Perturbations in either voice pitch or loudness feedback lead to changes in a speaker's voice fundamental frequency (F_0) or amplitude. Voice pitch or loudness perturbations were presented individually (either pitch or loudness shift stimuli) or simultaneously (pitch combined with loudness shift stimuli) to subjects sustaining a vowel to test the hypothesis that the mechanisms for these two response types are independent. For simultaneous perturbations, pitch and loudness both changed in the same direction or in opposite directions. Results showed that subjects responded with voice F_0 or amplitude responses that opposed the direction of the respective pitch- or loudness shift stimuli. Thus, depending on the stimulus direction, both responses could either change in the same direction or in the opposite direction to each other. F_0 response magnitudes were greatest with pitch-shift only stimuli (18 cents), smallest for loudness shift stimuli (10 cents) and intermediate with pitch combined with loudness shift stimuli (13 and 16 cents). Amplitude responses were largest with +3 dB stimuli (0.96 dB) and smallest with −3 dB stimuli (0.49 dB) but were not affected by the addition of pitch-shift stimuli. Results suggest the F_0 and amplitude response mechanisms may be independent but interact in some conditions. © *2007 Acoustical Society of America.*

[DOI: 10.1121/1.2715657]

PACS number(s): 43.70.Bk, 43.66.Hg, 43.64.Bt, 43.66.Fe, 43.70.Mn [AL] Pages: 2862–2872

I. INTRODUCTION

The importance of auditory feedback for voice control in speech and singing has been recognized for centuries, primarily from observations on the speech of deaf people. Prelingually deaf people acquire the ability to speak with great difficulty and then their speech is generally considered to be quite abnormal. Postlingually deaf people experience deterioration of voice F_0 and amplitude control shortly after the onset of hearing loss (Binnie *et al.*, 1982), while control of articulatory dynamics declines more slowly.

There have been several different experimental approaches towards the study of the role of auditory feedback on voice control; the Lombard effect, side-tone amplification (Lane and Tranel, 1971), noise masking (Elliott and Niemoeller, 1970; Ward and Burns, 1978) and perturbation (Sapir *et al.*, 1983). Recent studies have demonstrated that vocalizing subjects compensate for a perturbation in voice pitch feedback by changing their voice F_0 in the opposite direction to the change in feedback. The compensatory nature of the responses suggest that they are an attempt to correct for an error between voice pitch feedback and the note the subject was attempting to produce. Such responses have been observed during sustained vowel sounds, glissandos, speech, and singing (Bauer, 2004; Bauer and Larson, 2003; Burnett *et al.*, 1998; Burnett and Larson, 2002; Donath *et al.*, 2002; Hain *et al.*, 2000; Hain *et al.*, 2001; Jones and Munhall, 2000, 2002; Kawahara and Williams, 1996; Kiran and Larson, 2001; Larson, 1998; Larson *et al.*, 2001; Larson *et al.*, 2000; Natke *et al.*, 2003; Natke and Kalveram, 2001; Sivasankar *et al.*, 2005; Xu *et al.*, 2004). In approximately 15% of the trials, subjects produced responses that changed in the same direction as the stimulus, e.g., an upward change in voice F_0 in response to an upward perturbation in voice pitch feedback. These have been termed "following" responses.

We described and modeled the system responsible for generating responses to pitch-shifted voice feedback as a negative feedback control system (Hain *et al.*, 2000). Elements of this system must have an internal referent of desired voice F_0, access to feedback pitch and the ability to correct for errors between the pitch of the feedback signal and the referent. It has also been demonstrated that perturbation of voice loudness feedback results in compensatory responses in voice amplitude (Bauer *et al.*, 2006; Heinks-Maldonado and Houde, 2005). These voice amplitude responses have the same approximate latency as F_0 responses to pitch-shifted feedback and are compensatory in nature. We have also presented a model and described the system responsible for the

---

[a] Portions of this work were presented at the meeting of the Acoustical Society of America, May, 2005, Vancouver, Canada, and at the Speech Motor Control Conference, March, 2006, Austin, Texas.

[b] Electronic mail: clarson@northwestern.edu

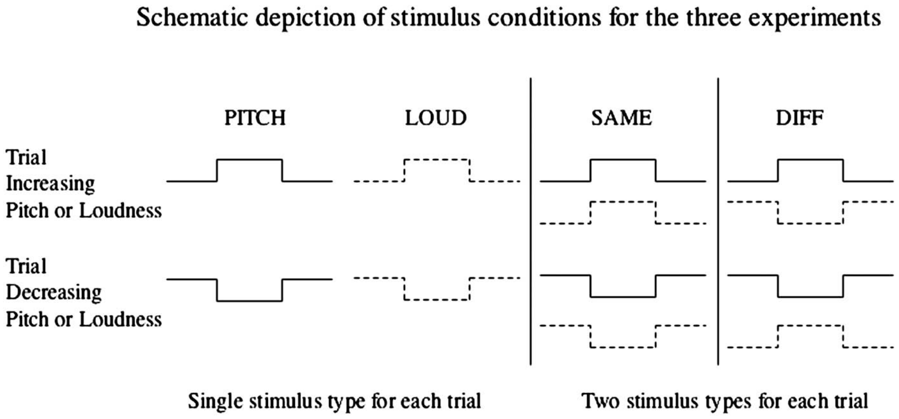

FIG. 1. Schematic representation of the stimuli used in the four experimental conditions. Square brackets depict the direction and relative timing of pitch-shifted (solid lines) or loudness-shifted (dashed lines) feedback.

loudness-shift responses as a negative feedback control system with properties similar to those of the voice F_0 control circuitry.

Although it appears that both pitch and loudness perturbations elicit similar responses, it is unknown whether the two responses are independent of each other, and thus represent different mechanisms, or whether they are part of the same responding system. We tested the hypothesis that the two responses are independent of each other by presenting to subjects vocalizing a vowel sound, simultaneous changes in pitch and loudness feedback, where both of the stimuli changed in the same direction or in the opposite direction. We predicted that if there are two independent mechanisms, then a stimulus composed of both pitch- and loudness-shifted voice feedback should elicit two independent responses, an F_0 response that changed in the opposite direction to the pitch-shift stimulus and an amplitude response that changed in the opposite direction to the loudness-shift stimulus. As control studies, subjects were also tested with pitch- and loudness-shift stimuli alone.

Results of the experiments revealed the presence of both F_0 and amplitude responses to simultaneous pitch- and loudness-shifted voice feedback that generally changed in opposite directions to the respective stimuli. Thus, the results suggest that there are two separate systems controlling F_0 and amplitude responses to pitch- and loudness-shifted stimuli and that the mechanisms underlying voice F_0 and amplitude control based on pitch and loudness feedback are independent. Additional observations demonstrate that with some stimuli there are interactions between the two response mechanisms.

II. METHODS

A. Subjects

Twenty-four undergraduate students at Northwestern University (2 male, 22 female, ages 18–22 years) served as subjects. Similar studies conducted in our laboratory have not revealed significant differences in responses as a function of sex, and hence we did not attempt to recruit equal numbers of male and female subjects. All subjects passed a hearing screening, and none reported a history of neurological or communication disorders. All subjects signed informed consent approved by the Northwestern Institutional Review Board. After preliminary data analysis, data from one subject were excluded because of an incomplete data set. Final results are based on data from 23 subjects (1 male, 22 female).

B. Apparatus

Subjects were seated in a sound-treated room and wore Sennheiser headphones with attached microphone (Model No. HMD 280). The vocal signal from the microphone was amplified with a Mackie mixer (Model No. 1202), processed for pitch and loudness shifting with an Eventide Eclipse Harmonizer, mixed with 40 dB SPL pink masking noise with a Mackie mixer (Model No. 1202-VLZ), further amplified with a Crown D75 amplifier and HP dB attenuators, and then sent back to the headphones. The harmonizer was controlled with MIDI software (Max/MSP v4.5 by Cycling '74) from a laboratory computer. Acoustic calibrations were made with a B&K 2250 sound level meter and model 4100 in-ear microphones. There was a gain of 10 dB SPL between the subject's voice amplitude, measured 2.5 cm from the mouth, and the feedback loudness measured at the input to the ear canal. The delay from MIDI input to the harmonizer to the shift in pitch was 14 ms, whereas there was no delay for the loudness shift stimulus. We considered the 14 ms difference in processing speed of the two feedback sources to be too small to affect the results. Preliminary testing revealed no perceptible difference in the timing of the pitch- and loudness-shifted feedback. The voice output signal, feedback and control pulses were digitized at 10 kHz, low-pass filtered at 5 kHz and recorded on a laboratory computer utilizing Chart software (ADInstruments, Colorado Springs, CO). Data were analyzed using event-related averaging techniques in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR). Subjects monitored their voice loudness from a Dorrough Loudness Monitor placed 0.5 m in front of them. This monitor provided the subjects with visual feedback of their voice amplitude and helped them to maintain a relatively constant level throughout the testing.

C. Procedures

Subjects produced /u/ vowel sounds into a microphone while hearing their voice modulated in pitch or loudness over headphones in near real-time (Bauer and Larson, 2003; Bauer *et al.*, 2006). Each vocalization was ≈5 s, in duration and produced at a level of ≈70 dB. Voice feedback loudness was amplified to ≈80 dB SPL. During each trial (vocaliza-

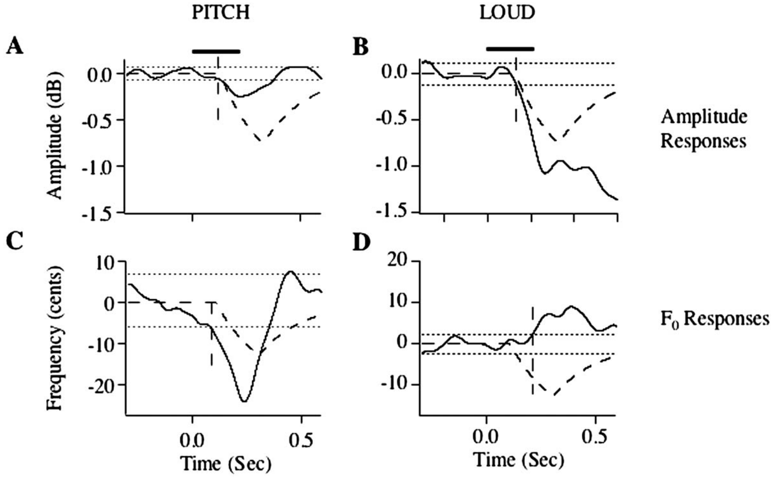

FIG. 2. Averaged F_0 and amplitude contours for the PITCH and LOUD conditions. Top row, responses to upward directed stimuli with voice amplitude responses. Bottom row, responses to upward directed stimuli with voice F_0 responses. Horizontal dashed lines represent ±2 SDs of the prestimulus mean. Stimulus onsets were at time 0.0, indicated by horizontal bar above the plots. Short vertical dashed lines indicate onset of response. Curved dashed traces are simulations from the model.

tion), five stimuli were presented at randomized intervals of 0.7–1.0 s. For each condition consisting of 8 trials, a total of 20 increasing and 20 decreasing stimuli were presented to the subjects. In the PITCH condition, randomized ±50 cent pitch-shift stimuli were presented, and in the LOUD condition, randomized ±3 dB loudness-shifted stimuli were presented. For any given set of 8 trials, the stimulus type (pitch or loudness) was held constant. In the SAME condition, simultaneous pitch and loudness-shifted stimuli were presented, and they both either increased or decreased. In the DIFF condition, simultaneous pitch- and loudness-shifted stimuli were presented, and they changed in different directions. For example, a +50 cent pitch shift was combined with a −3 dB loudness shift. Figure 1 diagrammatically illustrates the stimuli for the three conditions.

From the digitized signals, one wave representing the F_0 contour and one representing the voice amplitude contour were generated (Bauer and Larson, 2003; Bauer *et al.*, 2006). Event-related averages were generated for each subject for each experimental condition by time aligning the voice F_0 and amplitude contours with the pitch-shift or loudness-shift stimulus onset (TTL control pulse). For each type of stimulus, event-related averages were calculated for both F_0 and amplitude contours. Each averaged response consisted of a minimum of 15 trials, with a 200 ms prestimulus baseline and a 500 ms poststimulus response window. Valid responses were identified according to the following criteria: a deviation in the averaged trace (F_0 or amplitude) with a magnitude >2 standard deviations (SDs) of the pre-stimulus baseline for a minimum duration of 50 ms, and a latency $\geqslant 60$ ms after stimulus onset (see Fig. 2). Response latency was defined as the time point where the averaged trace crossed the 2 SD line following the stimulus. Response magnitude was measured as the greatest magnitude of the averaged F_0 or amplitude trace from the baseline mean. In the SAME and DIFF conditions, since each stimulus consisted of a combined pitch and loudness component, both voice amplitude and F_0 trajectories were averaged for each stimulus combination. Statistical analyses were done only on compensating responses because "following" responses may represent errors in responding, as explained in the following. Response latencies and magnitudes were submitted to statistical testing with one-way ANOVAs (Data Desk; Data Description). Excessive numbers (>8) of missing responses in the −3 dB conditions precluded a repeated-measures design. Response directions (compensating or "following") were tabulated by condition and stimulus direction.

III. RESULTS

Subjects responded in both PITCH and LOUD conditions with changes in F_0 and voice amplitude. Although most responses were compensatory, many "following" responses were also produced, primarily with cross-dimensional stimuli. Table I displays the number of F_0 and amplitude "compensatory," "following," and nonresponses (NR) in the PITCH and LOUD conditions. Most subjects produced a change in F_0 or amplitude in response to either a perturbation in pitch or loudness feedback, however, there were differences in the types and number of responses across conditions. In the PITCH condition, 87% of F_0 responses were compensatory, whereas 50% of amplitude responses were compensatory. In the LOUD condition, 52% of F_0 responses were compensatory while 67% of amplitude responses were compensatory.

Figure 2 displays illustrative examples of averaged F_0 and voice amplitude responses for both PITCH and LOUD conditions from two different subjects. As can be seen, all responses are in the compensatory direction except for Fig. 2(D), which is a "following" response. The cross-dimensional responses, i.e., amplitude response to a pitch-shift stimulus [Fig. 2(A)] and F_0 response to a loudness-shift

TABLE I. Numbers of compensatory (COMP), "following" (FOL) and nonresponses (NR) for Pitch and Loud conditions and stimulus dimension type[a].

	PITCH		LOUD		
Response type	F_0	Amplitude	F_0	Amplitude	Total
COMP	40 (87%)	23 (50%)	24 (52%)	31 (60%)	118 (64%)
FOL	5 (11%)	15 (33%)	17 (37%)	9 (20%)	46 (25%)
NR	1 (2%)	8 (17%)	5 (11%)	6 (13%)	20 (11%)
Total	46	46	46	46	184

[a] Chi square=13.21, df=2, $p<0.002$.

TABLE II. Counts of types of voice amplitude responses across conditions. “Following” (FOL), nonresponses (NR), and compensatory (COMP) responses are tabulated by simultaneous pitch- and loudness-shifted feedback with the stimuli changing in the same (SAME) or different (DIFF) directions. Responses are further organized with each specific stimulus combination of either + of −50 cent and + or −3 dB stimuli[a].

	SAME		DIFF		
	−50 c −3 dB	+50 c +3 dB	−50 c +3dB	+50c−3dB	Total
COMP	12 (52%)	20 (87%)	20 (87%)	12 (52%)	64 (70%)
FOL	8 (35%)	0	2 (9%)	7 (30%)	17 (18%)
NR	3 (13%)	3 (13%)	1 (4%)	4 (17%)	11 (12%)
Total	23	23	23	23	92

[a] Chi-square=16.26, df=6, $p=0.0124$.

TABLE III. Counts of types of F_0 responses across conditions. “Following” (FOL), nonresponses (NR), and compensatory (COMP) responses are tabulated by simultaneous pitch- and loudness-shifted feedback with the stimuli changing in the same (SAME) or different (DIFF) directions. Responses are further organized with each specific stimulus combination of either + of −50 cent and + or −3 dB stimuli.[a]

	SAME		DIFF		
	−50c−3dB	+50c+3dB	+50c−3dB	−50c+3dB	Total
COMP	20 (87%)	20 (87%)	16 (70%)	13 (57%)	69 (75%)
FOL	1 (4%)	2 (9%)	6 (26%)	8 (35%)	17 (18%)
NR	2 (9%)	1 (4%)	1 (4%)	2 (9%)	6 (7%)
Total	23	23	23	23	92

[a] Chi-square=10.39, df=6, $p=0.1093$.

stimulus [Fig. 2(D)] are smaller than the within-dimensional responses (also see Fig. 5). There were also instances where a double or triple response was observed. In such cases, we measured the first compensatory response for statistical analysis. If compensatory responses were not made, we measured the first “following” response. Responses to downward pitch- and loudness-shifted stimuli were similar to those illustrated in figure 2 but the directions were reversed.

Tables II and III provide counts of compensating, “following,” and nonresponses for amplitude (Table II) and F_0 (Table III) responses for the SAME and DIFF conditions. For these conditions, as all stimuli consisted of both a pitch and loudness component, we defined compensating and “following” F_0 and amplitude responses with respect to stimuli that were in the same acoustical dimension. For example, an upward voice amplitude response to a stimulus combination of +50 cent −3 dB, would be classified as a compensating response because its direction is opposite to the downward 3 dB component of the stimulus. For the amplitude responses, the types of responses differed significantly across the four stimulus combinations comprising the SAME and DIFF conditions. There was a disproportionately large number of amplitude “following” responses in the stimulus combinations that included a −3 dB stimulus (35% for SAME and 30% for DIFF). Otherwise, response types were rather evenly distributed across the stimulus combinations. For the F_0 responses, although not significantly different, there were more “following” responses in the DIFF (30%) than in the SAME (7 %) condition. Altogether, 18% of responses in the SAME and DIFF conditions were of the “following” type, compared with 25% for the PITCH and LOUD conditions. Compensating responses constituted 72% of the SAME and DIFF responses.

Figure 3 illustrates representative amplitude (top) and F_0 responses (bottom) to simultaneous stimuli changing in the SAME direction. Upward stimuli on the left and downward stimuli on the right led to opposing directed responses in each case. The amplitude response to a stimulus combination including a −3 dB stimulus (B) was much smaller than that to a combination including a +3 dB stimulus. The F_0 responses are roughly of the same magnitude for each stimulus. Response latencies in all of these cases are approximately 100 ms.

Figure 4 shows representative responses in the DIFF condition. On the left, the stimulus (+50 cent −3 dB) led to an upward amplitude response (A), which opposed the −3 dB stimulus, and a downward F_0 response (C) that opposed the +50 cent stimulus. On the right, with the −50 cent +3 dB stimulus, again both the amplitude (B) and F_0 responses (D) oppose the direction of the stimulus that was in

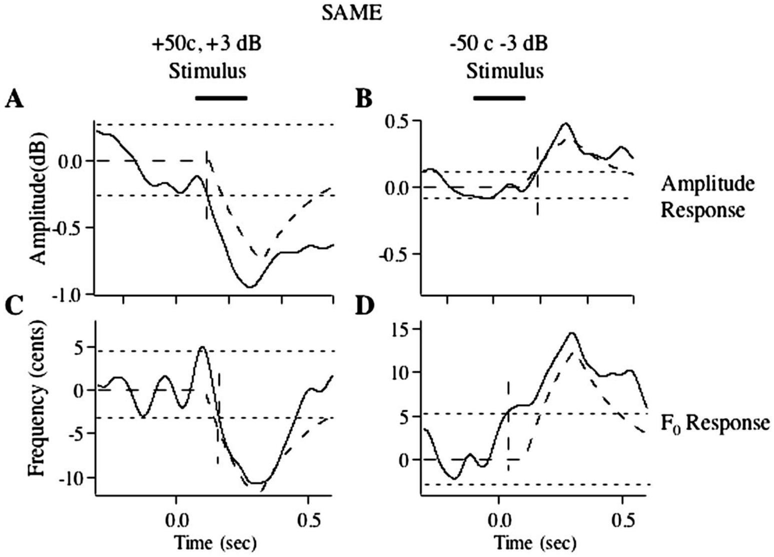

FIG. 3. Averaged F_0 and amplitude contours for the SAME condition. Top row shows contours of amplitude and bottom row of F_0 responses. Left column shows responses to stimuli composed of +50 cent and +3 dB stimulation. Right column shows responses to stimuli composed of −50 cent and −3 dB stimulation. Curved dashed lines are simulations from the model.

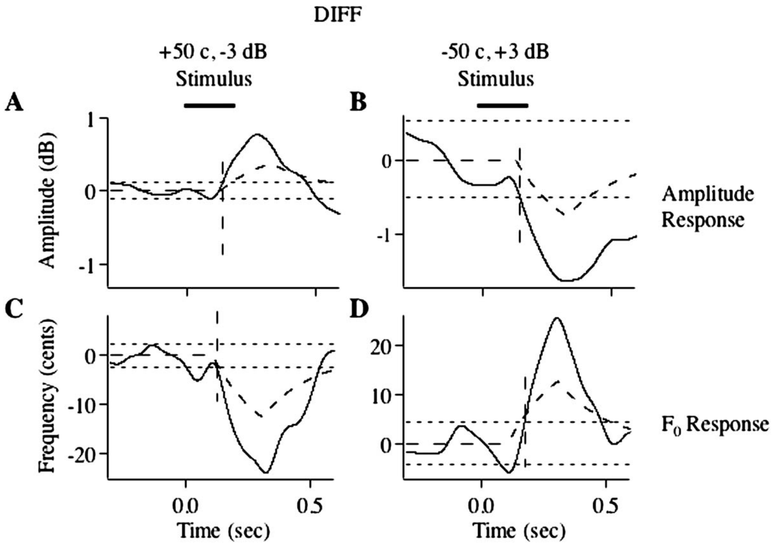

FIG. 4. Averaged F_0 and amplitude contours for the DIFF condition. Top row shows contours of amplitude and bottom row of F_0 responses. Left column shows responses to stimuli composed of +50 cent and −3 dB stimulation. Right column shows responses to stimuli composed of −50 cent and +3 dB stimulation. Curved dashed lines are simulations from the model.

the same acoustical dimension. There were also occasions when a small apparent "following" response occurred prior to a much larger compensating response [Figs. 4(C) and 4(D)]. It is unclear what these responses represent; whether they just represent 'noise' in the wave forms or a bona fide response. It is noteworthy that the amplitude response to the +50 cent −3 dB stimulus is much smaller than the response to the −50 cent +3 dB stimulus. This difference is similar to the comparison of amplitude responses in Fig. 3.

Figure 5 displays the overall mean response magnitudes for both voice amplitude and F_0 responses across all four experimental conditions, PITCH, LOUD, SAME, DIFF, and the two types of stimuli within each condition. Statistical testing across all four conditions revealed an effect for voice F_0 magnitude ($F=8.77$, df=3, 128, $p < 0.0001$). Posthoc testing with a Bonferroni correction indicated the mean response magnitude for the PITCH condition (18 cents) was significantly larger than for the SAME (13 cents; $p < 0.007$) and the LOUD condition (10 cents; $p < 0.0001$) but not different from the DIFF condition (16 cents). There was no significant effect for pitch-shift direction. For magnitude of voice amplitude responses, there was no significant difference across conditions, however, the downward loudness-shift stimulus (−3 dB) elicited significantly smaller responses (mean 0.54 ± 0.23 dB) than the upward stimuli (mean 0.91 ± 0.35 dB; $F=30.11$, df=1,93, $p < 0.0001$). Although it would be desirable to compare F_0 response magnitudes with amplitude magnitudes, their differing dimensions preclude this.

For voice amplitude magnitudes across all eight conditions, there was an overall effect ($F=6.16$, df=7,110, $p < 0.0001$). Posthoc testing with Bonferroni corrections showed that loudness-shift stimuli of −3 dB led to smaller responses than all of the following stimulus combinations:

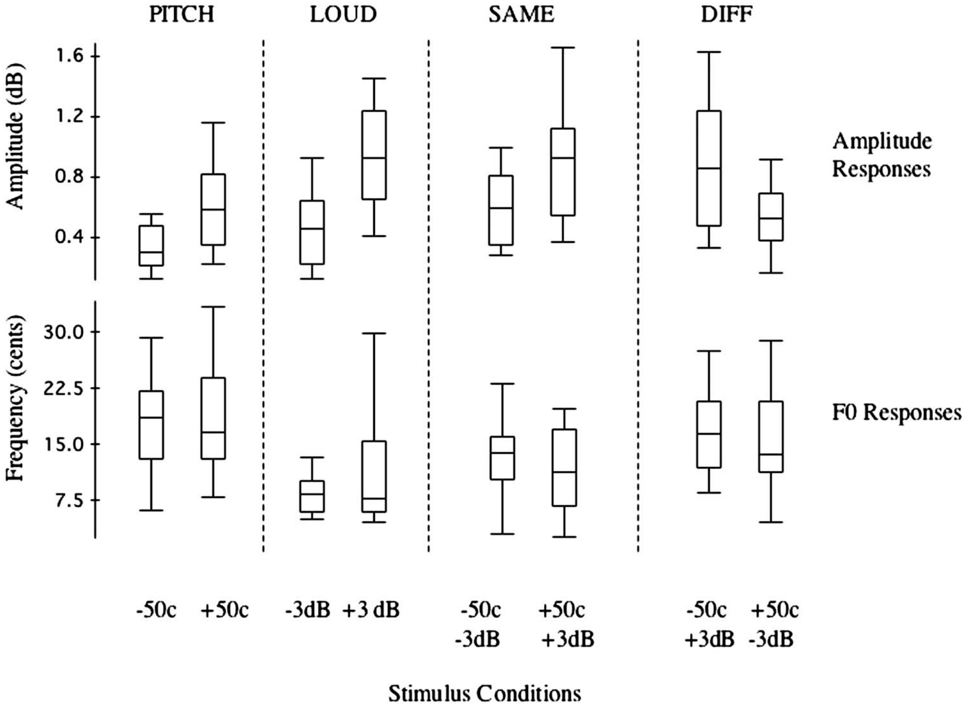

FIG. 5. Boxplots of response magnitudes for all experimental conditions. Top row shows amplitude responses and bottom row F_0 responses. Box definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge − low hinge), and low hinge −1.5 (high hinge − low hinge) (data desk; data description).

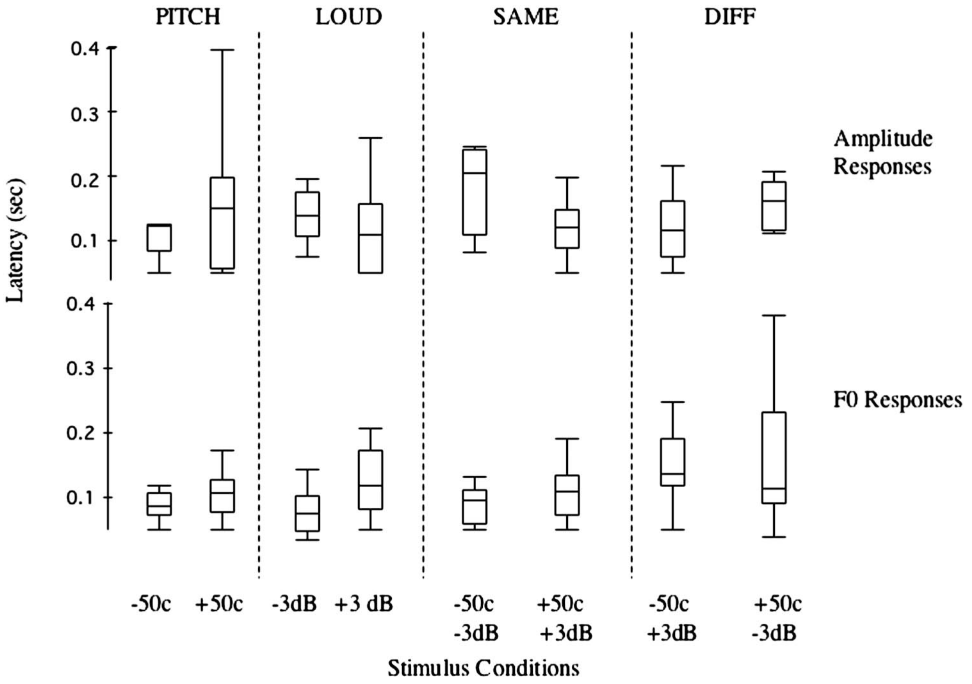

FIG. 6. Boxplots of response latencies for all experimental conditions. Top row shows amplitude responses and bottom row F_0 responses.

−50 cent +3 dB ($p < 0.04$), +3 dB ($p < 0.004$), and +50 cent +3 dB ($p < 0.04$). Stimuli of −50 cents also produced smaller responses than the following stimulus combinations: −50 cent +3 dB ($p < 0.02$), +3 dB ($p < 0.004$), and +50 cent +3 dB ($p < 0.02$). Finally, the stimulus combination of +50 cent −3 dB produced a smaller response than +3 dB stimuli ($p < 0.008$).

Figure 6 displays the mean latencies for responses across all four experimental conditions and stimuli. Statistical testing revealed an effect of the testing condition for F_0 responses ($F = 4.13$, df=3,128, df< 0.008). Posthoc testing with a Bonferroni correction showed that the latencies of the DIFF condition (mean 152±82 ms) were significantly longer than those for the PITCH (mean 106±64 ms; $p < 0.02$) and the SAME conditions (mean 104±42 ms; $p < 0.02$). There were no significant differences for F_0 response latencies as a function of direction of the pitch-shift stimulus.

There was no significant difference in voice amplitude response latencies as a function of condition, however, the downward loudness-shift stimulus produced significantly longer latencies (mean 163±49 ms) than did the upward stimulus (mean 125±63 ms; $F = 9.68$, df=1,93, $p < 0.003$). Mean latencies of F_0 responses (mean 128±81 ms) were significantly shorter than the amplitude latencies (mean 155±91 ms; $F = 8.7$, df=1,328, $p < 0.005$).

Statistical testing of response latencies across all conditions shown in Fig. 6 showed a main effect for F_0 latencies ($F = 2.56$, df=7,124, $p < 0.02$), but posthoc testing with a Bonferroni correction did not reveal significant differences between any two stimulus combinations. There was no significant difference for voice amplitude response latencies as a function of the stimulus combinations shown in Fig. 6.

We simulated these responses, as shown in Figs. 2–4 (curved dashed lines), using the negative feedback model shown in Fig. 7. This model, implemented with the SIMULINK toolbox of MATLAB (Mathworks, Nantick, MA) is a combination of two previously reported feedback models, one of stabilization of F_0 (Hain *et al.*, 2000), and the other of stabilization of loudness (Bauer *et al.*, 2006). Desired loudness and F_0 are compared to perceived loudness and F_0., and an error signal is computed through subtraction. Next, the error signal is "demultiplexed," meaning it is converted from a tonotopic representation to scalers representing F_0 and loudness error. Then, in the "pitch error feedback" and "loudness error feedback" sections, error is delayed and then used to adjust voice drive. Thus there are two negative feedback models, linked together as necessitated by known physiology, by multiplexers and demultiplexers for F_0 and loudness. The goal of the aggregate model was to establish feasibility and to provide a quantitative hypothesis for future work. We portray simulations obtained from a set of generic parameters optimized to fit the data shown in figures 2–4.

The previous individual models of F_0 and loudness were combined by explicitly modeling the computation of F_0 and loudness from auditory input (Bauer *et al.*, 2006; Hain *et al.*, 2000). In previous models we simply assumed that these signals were available. In the present model we consider F_0 and loudness as being "multiplexed" together in the cochlea. We made no assumption concerning whether the encoding is via a tonotopic representation, phase-locking, or a combination of the two. We then "demultiplex" the central auditory signal into F_0 and its loudness in a process that takes about 100 ms ("demux delay" in Fig. 7). In physiological terms, this would mean that the representation of F_0 and the loudness of F_0 are not separated from each other until a lengthy decoding step occurs, presumably in auditory cortex. A "matrix gain" cross-coupling matrix was included to allow for the possibility of interactions between F_0 and loudness, but it was configured without cross coupling as this simple configuration fit our data reasonably well.

Although we used a lumped demux delay of 100 ms, the true delay associated with the conversion of the central rep-

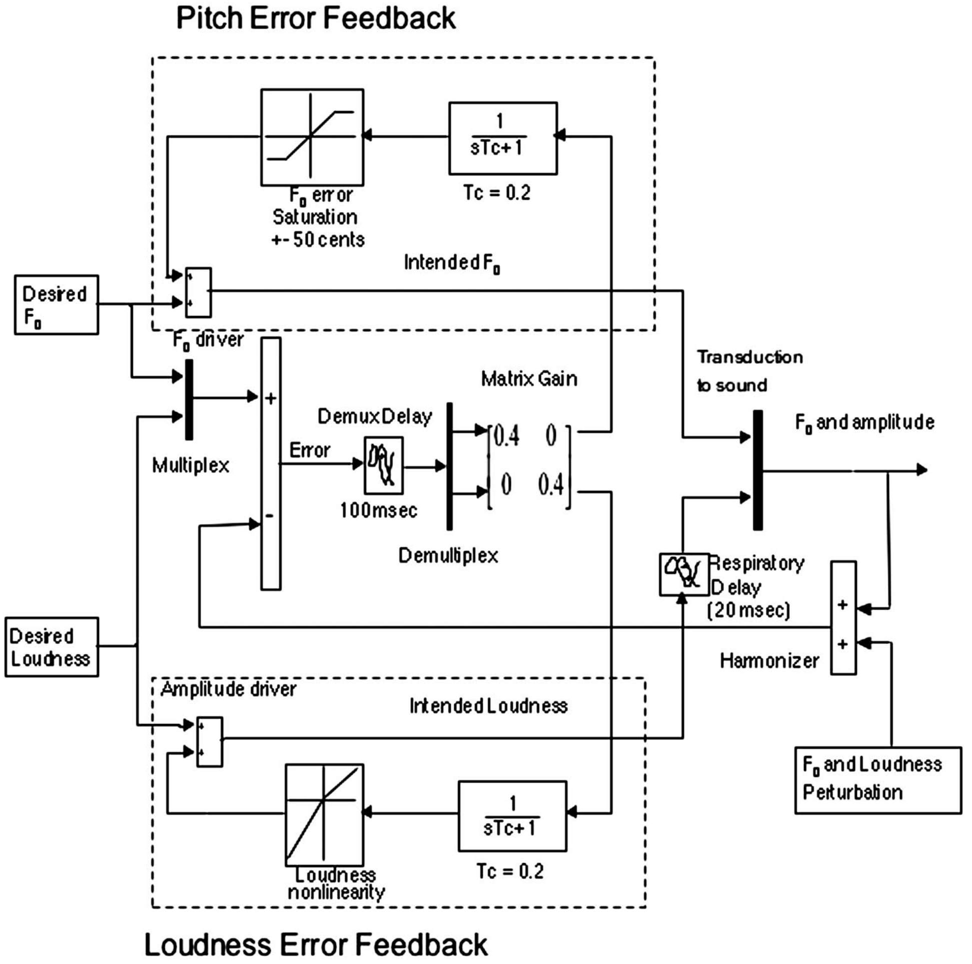

FIG. 7. Model of negative feedback system for control of voice F_0 and amplitude responses to pitch and loudness-shifted voice feedback. See the text for further details.

resentation into the scalars of F_0 and loudness error must be less, as the 100 ms includes time to get the signal to the cortex as well as sending it back to the brainstem. The 100 ms delay of the demux delay could be made shorter without affecting the simulations by shifting some of this delay to these pathways, but as our simulations would be unaffected by the shifts, for simplicity we have lumped them together. Physiological data suggests that there is likely a delay of about 20 ms between auditory input and auditory cortex (Howard *et al.*, 1996; Steinschneider *et al.*, 1999) and that the latency between cortical stimulation and activation of laryngeal and respiratory muscles is on the order of 13 ms (Gandevia and Rothwell, 1987; Ludlow and Lou, 1996). Thus, the anatomical demux delay should be about 66 ms.

We also attempted to model the longer latency observed for perturbations of loudness. Because the loudness signal is the loudness of F_0, only after F_0 has been determined can the loudness of F_0 be computed. This is most easily seen by an example—if overall loudness were used for negative feedback, then speech in noise would be softer than speech in quiet, because the brain would not be able to distinguish loudness of noise from speech. However, we know that the opposite is true from the Lombard effect literature (Lane and Tranel, 1971). It follows that because the relevant signal is the loudness of F_0, not loudness in general, and that the loudness feedback step must await the F_0 determination step. We implemented the additional loudness delay as "respiratory delay" in the loudness output pathway.

In order to simulate our observations that responses to upwards perturbations of loudness were stronger than downward perturbations, we added a nonlinearity to the loudness portion of the model, with a higher gain for upward perturbations. This nonlinearity has a slope of 0.5 for downward perturbations and 1.0 for upward perturbations, to create the roughly 2:1 gain asymmetry observed experimentally.

A model sensitivity analysis was performed to ascertain whether there was an optimal set of gain and time constant parameters as well as whether or not model performance was critically dependent on one or the other. The analysis revealed that best performance in terms of the model in a typical subset of the experimental data, was found with the gain and time constant parameters set to 0.4, and 0.2, respectively. This was the case for *both* the F_0 and loudness subsystems, and accounted for about 70% and 50% of the variance, respectively. Very poor fits occurred for the combination of large gains (e.g., 2) and short time constants (e.g., 0.2–0.6). The similarity between optimal fits for the F_0 and loudness stabilization subsystems suggests that they share neural circuitry.

IV. DISCUSSION

In this study, changes in voice F_0 and amplitude were measured in response to either pitch- or loudness-shifted voice feedback when the stimuli were presented alone or when both pitch- and loudness-shifted feedback were presented simultaneously. Other studies have shown that with dynamically changing pure tones, there may be interactions between the pitch and loudness stimuli, and in some cases these are associated with perceptual illusions (McBeath and Neuhoff, 2002; Moore and Sek, 1998; Neuhoff *et al.*, 1999). To our knowledge, no other studies have examined these relationships with respect to voice feedback. The present study was conducted to answer the question of whether or not the voice F_0 and amplitude responses to pitch- and loudness-shifted stimuli represent the same, or two independent mechanisms. By first testing the subjects with either pitch-or loudness-shifted stimuli (PITCH and LOUD), it was demonstrated that the basic properties of the F_0 and amplitude responses to stimuli in the same acoustical dimensions are very similar in form and are optimally simulated by identical mathematical constructs. This suggests that the two mechanisms share neural circuitry. Nevertheless, our results from experiments where both stimuli were present simultaneously, show that the two systems can respond largely independently of each other, albeit with small nonlinear interactions and minor cross coupling. Thus, overall it appears that the two stabilization mechanisms are predominantly independent but share circuitry and interact to a minor extent.

A. Theoretical basis for interactions between pitch and loudness processing

The responses to pitch- or loudness-shift stimuli result from a three-step process involving conversion of the stimuli into neural firing patterns, central processing and motor output. Given the fact that cross-dimensional responses were produced, a mixing together between loudness and pitch processing must occur at one of these three steps. The "mixing together" might consist of a linear combination of the two signals, which we call "cross coupling," or a nonlinear influence of one signal on the other, which we call "nonlinear interactions." In general, mixing together between two sensory modalities causes confusion between sensory streams and is undesirable. Thus, the work reported here is essentially looking for design flaws in the auditory system.

To determine where the mixing might occur requires a more detailed consideration of how the brain processes auditory signals and in particular, how it detects changes in pitch or loudness. Sound is not transmitted from the cochlea to auditory cortex as separate streams of "F_0" and "loudness," but rather it is tonotopically encoded in the cochlea. In other words, F_0 and loudness are multiplexed together. There is also processing of F_0 and loudness in the brainstem. As the signal most relevant to F_0, "phase locking," is not found in auditory cortex above about 100 Hz(Palmer, 1995), cortical processing probably uses tonotopic signals. If we accept that F_0 is necessarily computed in cortex, the loudness of F_0 cannot be computed until the cortical step has finished, and thus, although the brainstem could potentially compute overall loudness as well as the loudness of spectral components, it is highly unlikely to be the location for computation of the loudness of F_0. It is generally agreed that perception of sound depends upon the representation of individual tonotopic components being "reassembled" at a later stage in the auditory system (Moore, 1995). A lengthy reassembly (demultiplexing) step is plausible if one considers that converting from a tonotopic array representing cochlear input into separate signals encoding F_0 and loudness of F_0 intrinsically requires first F_0 identification, followed by computation of the loudness of F_0. F_0 identification is necessary because the cochlea cannot determine, without the help of the cortex, which tonotopic signal of an incoming frequency spectrum is the fundamental frequency.

Although a simple method of finding the largest component of the spectrum would work perfectly well for a pure sinusoid, there is overwhelming evidence that this is not the usual method used by the auditory system. It is well known that F_0 can be "heard" even if it is physically absent, as long as there are associated higher harmonics (Schouten, 1940). Also, two fundamentals can be separated out from other, potentially louder pitches, as for example, when we listen to a chord being played by a group of instruments (Houtsma, 1995). These observations teach us that simple algorithms such as finding the loudest component of a spectrum cannot be those used to derive F_0, and also that the cortex must participate in the process.

Previous central models of complex pitch detection, such as the "optimum processor model" of Goldstein, postulated that F_0 is identified via template matching (Goldstein, 1973). Template matching, being a high-level process, would have to be implemented in auditory cortex as well as be relatively slow. Once F_0 is determined, then the loudness of F_0 might be computed by selecting the tonotopic input of F_0, which would be relatively fast. Alternatively, considering Schouten's residue theory (Schouten, 1940), F_0 determination might emerge from the template matching process itself. Either method would seem likely to be a much quicker process than template matching alone. Thus, it seems likely that F_0 and loudness processing relevant to feedback control should share considerable common neural circuitry, up through a lengthy step involving F_0 identification and demultiplexing.

Nevertheless, our data provides compelling evidence that the F_0 and loudness systems are predominantly independent. The fact that the responses were able to respond appropriately to within-dimensional stimulation and change in a direction opposite to the within-dimensional stimulus, and opposite in direction to each other, provides strong evidence that the mechanisms controlling the F_0 and amplitude responses *can* function independently. Depending on the stimulus combinations, both responses can change in the same direction or in opposite directions.

B. Linear interaction between loudness and F_0 processing

As mentioned previously, using engineering formalism, the mixing together between F_0 and loudness—a sensory or motor confusion—can be thought of as two types of basic processes—linear combination (dimensional cross coupling) and nonlinear interactions. Dimensional cross coupling is portrayed in our model as a 2×2 matrix located after error has been computed and the tonotopic representation of F_0 and loudness have been demultiplexed. A matrix with large elements on the diagonals compared to the off-diagonals, would correspond to a system without cross coupling. A matrix with large off-diagonal elements, corresponds to a system with cross coupling. Our data suggests that there is a small off-diagonal element between F_0 and voice amplitude (element 1,2 of the gain matrix), and that there is no off-diagonal coupling between loudness and F_0 (element 2,1). Further, our simulations with the model configured with no cross coupling or nonlinear interactions accounted for most of the variance in the experimental data. Thus, our results suggest that representation of F_0 and a loudness stabilization system as predominantly independent is feasible.

Logically, given that auditory input is transmitted to the cortex using tonotopic encoding which keeps F_0 and amplitude combined, mixing could occur at any point—input, central processing, or motor output. Previous literature supports the suggestion that there may be a mixing together, which is appreciated at the perception stage (McBeath and Neuhoff, 2002; Moore and Sek, 1998; Neuhoff *et al.*, 1999). Perception includes both input and central processing. Thus, in the context of our model, these observations support the idea that a part or all of the mixing together does not occur in the output circuitry.

C. Nonlinear interactions between F_0 and loudness processing

The other potential type of mixing together between modalities is nonlinear interactions. Examples of nonlinear interactions might be an influence of one sensory modality on the timing of another, or a multiplicative type action of one sensory modality on the other. Like cross coupling, nonlinearity is generally avoided in system design. However, there was evidence for nonlinear interactions in our data.

The primary finding of the amplitude responses was that all stimuli that included a −3 dB component were smaller and slower than those produced to other stimulus combinations. The small responses to the −3 dB stimulus, represented by the "loudness nonlinearity" in our model, may be related to the observation that for equal differences in an increase or a decrease in the intensity of sounds, subjects perceive an increase to be greater than a decrease (Neuhoff *et al.*, 2002). It is also known that a decrease in stimulus intensity leads to longer reaction times (Jaskowski *et al.*, 1994; Seitz and Rakerd, 1997), which occurred with the −3 dB stimuli. Along this line, it is instructive that the amplitude responses were also much smaller for a −50 cent pitch-shift stimulus than a +50 cent stimulus. The most parsimonious explanation is that the same nonlinearity that is responsible for smaller responses to decreases in intensity, also acts on the cross-coupled response to F_0.

There was additional evidence for nonlinear interactions between the response mechanisms. When a loudness-shift stimulus was combined with a pitch-shift stimulus in the same direction (SAME), F_0 response magnitudes were smaller than those made to a pitch-shift stimulus alone (PITCH). In this condition there was a change in both pitch and loudness, and they changed in the same direction. Yet, despite the fact that both pitch and loudness were dropping, the F_0 responses were much smaller than those to a downward pitch-shift stimulus alone. In other words, there is a nonlinear interaction between the direction of change of inputs and the size of the response. It might be argued that with both a decrease in pitch and loudness, the overall perceptual impression would have been that of a decrease in vocal effort, such as a drop in voice amplitude and frequency that occurs in a nonfocused syllable (Alain, 1993; Cooper *et al.*, 1985; Eady and Cooper, 1986), and that such a "natural" event might be intended and thus should not trigger a response.

Observations from the DIFF conditions also document nonlinear interaction between the two mechanisms. In the DIFF condition, F_0 responses were somewhat smaller than in the PITCH condition, but much slower than in the other conditions.

The observations of "following" responses also suggest nonlinear interactions between the two response mechanisms. We have previously suggested that "following" responses occur due to a self-selected tracking mode where the subjects consider auditory feedback to be externally generated (as it is), rather than self-generated (Hain *et al.*, 2000). Most "following" voice amplitude responses occurred whenever the stimuli included the −3 dB component, and most "following" F_0 responses occurred when a pitch- and loudness-shift stimulus were presented in conflict with each other (DIFF). These observations suggest that the −3 dB stimulus may have been near threshold of detection of loudness, and the responses were almost random. "Following" F_0 responses were most prevalent when a pitch and loudness shift stimuli were combined in opposite directions. This stimulus combination may have made stimulus identification difficult; the responses were slower, smaller and more easily confused with "noise" in the system. This behavior may occur when the stimulus sounds unlike the subject's own voice, which could have been caused by some of the stimulus combinations in the present study. Regardless of the cause of "following" responses, the fact that they occurred under different conditions for amplitude and F_0 dimensions, suggests that the control circuitry for these two response mechanisms must not be completely combined.

D. Evidence relating to separation of F_0 and loudness control circuitry

Other observations from this study suggest there are differences in these two response mechanisms, not having to do with interactions that would support the idea that there is at least partial separation of the control circuitry. The finding that F_0 latencies (mean 116 ms) were significantly shorter than amplitude latencies (mean 150 ms) is one such example. In the calculation of these latencies, the difference in timing of the actual pitch- or loudness-shifted feedback that was caused by the harmonizer was taken into account. The latency differences cannot be entirely due to differences in transmission delays from the cortex to the respiratory or laryngeal muscles, since these delays are both about 13 ms (Gandevia and Rothwell, 1987; Ludlow and Lou, 1996). The latency differences may be due to differences in the perceptual saliency of the two stimuli. Both the ±3 dB and ±50 cent stimuli were readily apparent to the investigators, but no perceptual testing was done to determine if one or the other stimulus was more or less apparent to the subjects in the study. It is known that increases in stimulus intensity or perceptual salience can reduce the latency of responses to acoustical stimuli (Jaskowski *et al.*, 1994; Klug *et al.*, 2000), and if the subjects perceived the ± 50 cent stimuli to be more salient than the ±3 dB stimuli, this could account for the shorter latencies of the pitch-shift stimuli.

These latency differences could also be due to the motoric mechanisms controlling the responses. It is likely that the F_0 responses were controlled by laryngeal muscle contractions; the cricothyroid muscles are primarily responsible for regulating voice F_0 (Titze, 1994). Moreover EMG studies have shown the cricothyroid muscle to respond to pitch shift stimuli (Kawahara *et al.*, 1993). Voice amplitude responses could be due to respiratory muscle contractions and/or laryngeal muscle contractions. If they are due to contractions of muscles such as the intercostals, there may be a significant time delay between muscle contraction, thoracic volume reduction and compression of the broncho-tracheo air column. On the other hand, if the amplitude responses are due to changes in laryngeal impedance or stiffness, which is controlled primarily by intrinsic laryngeal muscles (Titze, 1994), then there may be no differences in response latency based on muscle contractions. Additional studies are necessary to determine which muscles are activated by pitch- and loudness-shift stimuli. It is also possible that the latency differences could be due to cortical processing delays, which are largely unknown at this time. We implemented the longer latency for loudness shifts in our model of the response as a separate delay (respiratory delay) following F_0 extraction.

E. Implications for our understanding of complex vocal control

Aside from providing information about mechanisms involved in generating responses to complex acoustical stimuli, the present study may provide important information relative to vocal control based on auditory feedback. In previous studies using the auditory perturbation paradigm, a single acoustical dimension was altered. However, in normal audio-vocal situations such as speech and singing, perturbations in both dimensions occur. Many of these are done purposefully such as for suprasegmental contrasts in speech. During normal speech, emphasized syllables generally have a higher F_0, longer duration and greater amplitude (Alain, 1993; Cooper *et al.*, 1985; Eady and Cooper, 1986), although the amplitude changes are generally not as strong as the frequency or durational modifications (Xu and Xu, 2005). Nevertheless, some degree of positive correlation between the changes in these two dimensions is common. However, to our knowledge there is no information on the correlation of unplanned perturbations in these two dimensions during speech. Certainly, vocal jitter and shimmer seem to occur during all vocalizations, but whether or not they are temporally correlated has not been examined. Similarly, slower perturbations as in tremor are likely to co-occur. If such perturbations co-occur during speech and singing, results of the present study suggest that the audio-vocal system is capable of simultaneous stabilization in both acoustical dimensions. This is true if the perturbations are positively or negatively correlated. The only caveat to this interpretation is that there may be greater numbers of errors (“following” responses) in the stabilization process when the two types of perturbations are in conflict with each other. Such errors in vocal stabilization could lead to greater variability in vocal control and the perception of an unsteady voice in some people.

V. CONCLUSION

The results of this study show that when people sustain vowel sounds, neural mechanisms respond to fluctuations in their voice pitch and loudness feedback by producing compensatory responses in voice F_0 and amplitude. These responses are part of a negative feedback loop that functions to stabilize F_0 and amplitude around an actual or intended goal. This study shows that the mechanisms for voice F_0 and amplitude stabilization are predominantly independent. Namely, the system is capable of stabilizing F_0 regardless of whether or not there is a change in loudness feedback, and vice versa. Further, a simple mathematical model using negative feedback, without any dimensional cross coupling or nonlinear interaction, reproduces the main features of our experimental data. In some cases, however, responses are cross-coupled, delayed or exhibit reduced magnitudes if there is more than one type of feedback perturbation, or if the directions of the multiple feedback dimensions are in conflict with each other. These observations suggest that some elements of either the perceptual or compensatory mechanisms interact to a small extent.

ACKNOWLEDGMENTS

This study was supported by a grant from NIH, Grant No. DC006243-01A1. The authors thank Chun Liang Chan for his help with computer programming.

Alain, C. (**1993**). “The relation among fundamental frequency, intensity, and duration varies with accentuation,” J. Acoust. Soc. Am. **94**, 2434–2436.

Bauer, J. J. (**2004**). “Task dependent modulation of voice F0 responses elicited by perturbations in pitch of auditory feedback during English speech and sustained vowels.” Ph.D. dissertation, Northwestern University, Evanston, IL.

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

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (**2006**). “Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude,” J. Acoust. Soc. Am. **119**, 2363–2371.

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

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.

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

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). “Acoustical aspects of contrastive stress in question-answer contexts,” J. Acoust. Soc. Am. **77**, 2142–2156.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). “Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Eady, S. J., and Cooper, W. E. (**1986**). "Speech intonation and focus location in matched statements and questions," J. Acoust. Soc. Am. **80**, 402–416.

Elliott, L., and Niemoeller, A. (**1970**). "The role of hearing in controlling voice fundamental frequency," Int. Aud. **IX**, 47–52.

Gandevia, S. C., and Rothwell, J. C. (**1987**). "Short-latency cortical projection to human phrenic motoneurons," in *Respiratory Muscles and their Neuromotor Control*, edited by G. C. Sieck, S. C. Gandevia, and W. E. Cameron (Alan R. Liss, New York), pp. 115–119.

Goldstein, J. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

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

Hain, T. C., Burnett, T. A., Larson, C. R., and Kiran, S. (**2001**). "Effects of delayed auditory feedback (DAF) on the pitch-shift reflex," J. Acoust. Soc. Am. **109**, 2146–2152.

Heinks-Maldonado, T. H., and Houde, J. F. (**2005**). "Compensatory responses to brief perturbations of speech amplitude," ARLO **6**, 131–137.

Houtsma, A. (**1995**). "Pitch perception," in *Hearing: Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego), pp. 267–291.

Howard, M. A., Volkov, I. O., Abbas, P. J., Damasio, H., Ollendieck, M. C., and Granner, M. A. (**1996**). "A chronic microelectrode investigation of the tonotopic organization of human auditory cortex," Brain Res. **724**, 260–264.

Jaskowski, P., Rybarczyk, K., and Jaroszyk, F. (**1994**). "The relationship between latency of auditory evoked potentials, simple reaction time, and stimulus intensity," Psychol. Res. **56**, 59–65.

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.

Kawahara, H., Hirai, T., and Honda, K. (**1993**). "Laryngeal muscular control under transformed auditory feedback with pitch perturbation," Report No. SP93–39, The Institute of Electronics, Information and Communication Engineers, Japan.

Kawahara, H., and Williams, J. C. (**1996**). "Effects of auditory feedback on voice pitch trajectories: Characteristic responses to pitch perturbations," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, Sydney), pp. 263–278.

Kiran, S., and Larson, C. R. (**2001**). "Effect of duration of pitch-shifted feedback on vocal responses in Parkinson's Disease patients and normal controls," J. Speech Lang. Hear. Res. **44**, 975–987.

Klug, A., Khan, A., Burger, R. M., Bauer, E. E., Hurley, L. M., Yang, L., Grothe, B., Halvorsen, M. B., and Park, T. J. (**2000**). "Latency as a function of intensity in auditory neurons: influences of central processing," Hear. Res. **148**, 107–123.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

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

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparisons of voice F_0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**, 2845–2848.

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

Ludlow, C. L., and Lou, G. (**1996**). "Observations on human laryngeal muscle control," in *Vocal Fold Physiology: Conrolling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 201–218.

McBeath, M. K., and Neuhoff, J. G. (**2002**). "The Doppler effect is not what you think it is: dramatic pitch change due to dynamic intensity change," Psychon. Bull. Rev. **9**, 306–313.

Moore, B. C. J. (**1995**). "Frequency analysis and Masking," in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego), pp. 161–205.

Moore, B. C. J., and Sek, A. (**1998**). "Discrimination of frequency glides with superimposed random glides in level," J. Acoust. Soc. Am. **104**, 411–421.

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

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

Neuhoff, J. G., Kramer, G., and Wayand, J. (**2002**). "Pitch and loudness interact in auditory displays: Can the data get lost in the map?," J. Exp. Psychol., Appl. **8**, 17–25.

Neuhoff, J. G., McBeath, M. K., and Wanzie, W. C. (**1999**). "Dynamic frequency change influences loudness perception: a central, analytic process," J. Exp. Psychol. Hum. Percept. Perform. **25**, 1050–1059.

Palmer, A. R. (**1995**). "Neural signal processing," in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego), pp. 75–113.

Sapir, S., McClean, M. D., and Larson, C. R. (**1983**). "Human laryngeal responses to auditory stimulation," J. Acoust. Soc. Am. **73**, 315–321.

Schouten, J. (**1940**). "The residue and the mechanism of hearing," *Proceedings of the Koninklijke Akademie van Wetenschap*.

Seitz, P. F., and Rakerd, B. (**1997**). "Auditory stimulus intensity and reaction time in listeners with longstanding sensorineural hearing loss," Ear Hear. **18**, 502–512.

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.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A., III. (**1999**). "Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex," J. Neurophysiol. **82**, 2346–2357.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ).

Ward, D., and Burns, E. (**1978**). "Singing without auditory feedback," J. Res. Sing. **1**, 4–44.

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

Xu, Y., and Xu, C. X. (**2005**). "Phonetic realization of focus in English declarative intonation," J. Phonetics **33**, 159–197.