Comparison of voice F_0 responses to pitch-shift onset and offset conditions (L)

Charles R. Larson[a], **Theresa A. Burnett**[b], **Jay J. Bauer**, and **Swathi Kiran**[c]
Department of Communication Sciences and Disorders, 2299 North Campus Drive, Northwestern University, Evanston, Illinois 60208

Timothy C. Hain
Departments of Otolaryngology, Neurology and Physical Therapy, Northwestern University Medical School, Chicago, Illinois 60611

Abstract

In order to maintain a steady voice fundamental frequency (F_0), it is assumed that people compare their auditory feedback pitch with an internal (memory) or external (acoustic) referent. In the present study we examined whether the internal referent is fixed or variable by comparing voice F_0 responses to incorrect auditory feedback in two timing conditions. In one condition, the incorrect pitch was introduced during vocalization (ON condition). In the second, the incorrect auditory feedback pitch was presented before vocal onset and then removed during vocalization (OFF condition). These conditions were examined with pitch-shift stimuli of ±25, 100, and 200 cents. There were no differences in response latency or magnitude between the two timing conditions, indicating that for a sustained-pitch vocalization task, the internal referent is not fixed. Several alternative types of referencing are discussed, which include use of a pitch relative to that which existed at the onset of vocalization (a sample and hold strategy) and pitch velocity referencing.

I. INTRODUCTION

One issue of concern in the study of voice fundamental frequency (F_0) control is how people maintain a steady voice F_0 level. Trained singers can hold their voice F_0 at a desired level in a steady manner, or they can deliberately modulate it around a desired level, such as in vibrato (Sundberg, 1987). It is also well known that auditory feedback is important for the accurate control of voice F_0 (Elliott and Niemoeller, 1970; Elman, 1981; Jones and Munhall, 2000; Ternström *et al.*, 1988) and that pitch memory is important in enabling singers to sing a specific note (Takeuchi and Hulse, 1993). We have proposed a model of voice F_0 control in which perceived pitch of auditory feedback is compared either with an internal referent (e.g., memory) or an external referent (e.g., piano note) (Hain *et al.*, 2000). A related issue is whether an internal referent is fixed or variable. It is well known that some people have "perfect" or absolute pitch, implying a fixed reference, while others are unable to reliably produce an accurate pitch without an external reference.

In the present study, we utilized the pitch-shifting paradigm in nontrained singers to determine whether the internal voice F_0 reference for comparison with auditory feedback is fixed or variable. Voice F_0 responses to altered voice pitch feedback were studied under two timing conditions. In the onset condition (ON), the pitch-shift stimulus was unexpectedly turned on shortly after the start of vocalization. In the offset condition (OFF), the pitch-shift processor

[a]Electronic mail: clarson@northwestern.edu.
[b]Laryngeal and speech section, NINDS, Bldg. 10, Rm 5D38, Bethesda, MD 20892-1416
[c]Department of Communication Sciences and Disorders, University of Texas, Austin, Texas 78712

was turned on prior to the initiation of vocalization and then unexpectedly turned off during vocalization. Thus under the OFF condition, the feedback pitch was returned to normal for the remainder of the vocalization. It was hypothesized that if the internal referent is fixed, subjects would respond to the onset of the pitch-shift stimulus (ON)but not the offset (OFF condition). The reason they would not respond in the OFF condition is because according to the *fixed* referent hypothesis, they would recognize the feedback as not being their own voice, and hence it would be irrelevant to their own production. However, if the internal referent is variable, subjects would respond to both conditions; any sudden change in voice pitch feedback would be recognized as an error, and the audio-vocal system would attempt to negate it. To test whether or not any effect depended on the magnitude of the pitch-shift stimulus or direction of pitch-shift modulation, six different stimulus magnitudes were employed across both timing conditions.

II. METHODS

A. Subjects

Thirty-three undergraduate students (26 females, 7 males, ages 18–22 years(served as subjects. All subjects passed a hearing screening at 20 dB HL (500–8 kHz), none reported any neurological or speech abnormalities, and none was trained as a professional singer or claimed to have perfect pitch.

B. Apparatus and procedures

The subjects' voices were transduced with an AKG boom-set microphone, amplified with a Mackie Mixer (model 1202), processed through an Eventide Ultraharmonizer (model H3000 SE) for pitch-shifting, mixed with 70 dB SPL pink masking noise and fed back over AKG earphones (model K 270 H/C). Subjects vocalized a vowel (ah) at a habitual pitch and at an intensity of 70 dB SPL, aided by observing a Dorrough Loudness Monitor, resulting in voice feedback loudness of about 80 dB SPL at the headphones. Subjects sat in a sound-treated room and were instructed to vocalize for 5 s, pause for a breath and repeat. They were instructed to hold voice pitch as steady as possible and to ignore any auditory feedback variations. At a random time (500–2500 ms) after vocalization onset, the pitch of the feedback signal was altered by a preset amount. Thirty consecutive vocalizations were recorded during each experimental block. During each block, 15 upward and 15 downward (pseudorandomly mixed) pitch shifts were presented. (For a more detailed description of the methodology see Burnett *et al.*, 1998; Hain *et al.*, 2001; Larson *et al.*, 2000.)

In this study two *timing*, three *magnitude*, and two stimulus *direction* experimental conditions were compared across six blocks of trials. In the ON timing condition, the pitch-shift stimulus (PSS) was presented during the vocalization. In the OFF timing condition, the pitch-shift stimulus was turned on before vocal onset and then removed during the vocalization. Thus, in the OFF condition, when subjects began vocalizing, they heard their voice pitch already shifted, and when it was removed, they heard their normal, unperturbed pitch feedback. Three pitch-shift stimulus magnitudes (25, 100, or 200 cents) (100 cents=one semitone) and two pitch-shift stimulus directions (upward and downward) were examined in addition to the timing conditions. The change in pitch feedback that occurred during the vocalization was maintained for the duration of the vocalization. Sixteen subjects were tested with PSS of ±25 cents (13 females, 3 males), while the other 17 subjects were tested with ± 100 and ±200 cent PSS (13 females, 4 males).

During the experiment, the voice signal, the feedback signal, and a TTL pulse (indicating time of change in the feedback signal) were digitized at 2 kHz on a laboratory computer. In offline analysis, a software algorithm was used to generate signals where voltage is proportional to

the F_0 of the subject's voice (F_0 analog) and the feedback signals. These signals were then time aligned to the TTL pulse for each subject for each experimental condition and event-related averages were computed. From these averaged signals, the preshift mean F_0 was calculated. In the period following the pitch shift, a response was measured whenever the voltage of the averaged F_0 signal differed by more than two standard deviations from the preshift mean. Only deviations beginning at least 60 ms after onset of the pitch-shift stimulus and lasting for at least 60 ms were considered valid. Latency and amplitude measures from the averaged responses (hereafter, "responses" refers to averaged responses) were recorded (Burnett *et al.*, 1998; Hain *et al.*, 2000; Larson *et al.*, 2000).

Responses following the change in pitch feedback were tested with a modified 3×(2×2) repeated measures factorial MANOVA. Response latency and response magnitude were the dependent variable measurements. The within subjects (repeated) factors were *timing* and stimulus *direction* conditions, each with two levels. The between-subjects factor of PSS *magnitude* was modified to fit three levels, although it was derived from two separate subject groups. The first group (*N*=16) was tested only on PSS magnitude of 25 cents, while the second group (*N*=17) was tested on both PSS magnitudes of 100 and 200 cents. Differences were considered significant for *p* values less than 0.01. Significance of incidence of compensating versus "following" responses was done with a chi-square test.

III. RESULTS

A total of 200 averaged responses were measured out of a total of 200 possible responses. For the 16 subjects receiving PSS of ±25 cents, there were a total of 64 responses (1 magnitude×2 direction×2 timing conditions). For the 17 subjects receiving PSS of ±100 and ±200 cents, there were 136 responses (2 magnitude×2 direction×2 timing conditions). These totals translate to 100 ON and 100 OFF responses. Therefore responses were observed for all subjects under each experimental condition.

Representative data from one subject in Fig. 1 show traces representing voice F_0 and feedback pitch for the ON and OFF conditions. The left half of Fig. 1 illustrates a subject's responses to upward and downward pitch-shift stimuli for the ON condition. Before the onset of the PSS (vertical dashed line), the voice F_0 and feedback pitch are identical. After the onset of the PSS, there is a discrepancy (error) of 25 cents between the voice F_0 and the feedback pitch (shaded area) that is maintained throughout the remainder of the vocalization. A gradual change in voice F_0 (dark line) is apparent, which after about 500 ms, results in the voice F_0 reaching an asymptote and the feedback pitch (light line) approaching the baseline level that existed prior to the stimulus onset. The right-hand side of Fig. 1 shows responses to upward and downward pitch stimuli in the OFF condition. In this case there is a discrepancy between feedback pitch and voice F_0 at the onset of vocalization (shaded area). When the stimulus is turned off (vertical dashed line), the feedback pitch (light line) matches that of voice F_0 (dark line). Subjects respond with a gradual change in F_0, which after approximately 500 ms, results in the feedback signal approaching the level that existed prior to the pitch shift. In both the ON and OFF conditions and regardless of pitch-shift direction, subjects respond to the change during vocalization such that the feedback pitch approaches the level that existed before the pitch shift was presented (compensatory response). It should be noted that at the end of the record, the feedback pitch does not always match the level that existed prior to the pitch-shift stimulus in either the ON or OFF condition.

There were no significant differences in response latency or magnitude as a function of the ON and OFF conditions (Wilk's Lambda 0.91, *df* =2, 46, *p*=0.12). Therefore subjects responded equally to ON and OFF timing conditions. The mean latency of the ON responses was 155 ms (SD 128) and that of the OFF responses was 170 ms (SD 109). The mean magnitude of the ON

responses was 42 (SD 36) and the OFF responses was 33 (SD 30) cents. There was a main effect for the pitch-shift magnitude (Wilk's Lambda 0.432, df =4, 92, p<0.001). Post hoc Sheffé testing revealed that the latencies for the 25-cent pitch-shift stimuli (mean 246, SD 156 ms) were significantly longer than that for the 100 cent (mean 117, SD 60 ms) and 200 cent stimuli (mean 130, SD 80 ms). Although not significant, the response magnitudes for the 25-cent pitch-shift stimuli (mean 26, SD 8.4) were less than the 100-cent (mean 48, SD 39) and the 200-cent (mean 39, SD 39) pitch-shift stimuli. Furthermore, there were nonsignificant differences in latency and magnitude measures as a function of pitch-shift direction (up or down), as well as nonsignificant interactions.

One hundred eighty four responses were identified as compensatory and 16 as "following" responses. Of these 16 "following" responses 12 were elicited from downward stimuli, 11 from OFF stimuli, and 8 from DOWN-OFF experimental conditions. Chi-square comparisons of opposing versus "following" prevalence failed to reach significance (p>0.05). However, compensating responses were statistically larger in magnitude (39.7 cents) compared with "following" responses (16.7 cents) (F=7.1, df =1, p<0.005).

IV. DISCUSSION

Results of this study show that regardless of whether a pitch-shift stimulus is presented before or after vocalization onset, subjects respond equivalently. In both the ON and OFF conditions, a pitch-shift stimulus elicits a "compensatory" F_0 response that brings the feedback signal back to- ward the level (approximately) that existed before the stimulus. Thus, the comparison between auditory feedback and voice F_0 is not fixed, but is variable. If the comparison were fixed, then the vocalist would know that the altered feedback pitch in the OFF condition was not normal, and would not respond to the change when it was turned off during vocalization. Use of a variable, or relative reference implies one of several different strategies that will be discussed in the following.

The simplest possible organization of a system that maintains a steady F_0 is to compare the auditory feedback signal with a referent. We call this scheme *absolute* F_0 referencing. The referent might be external such as a piano note, or internal such as an efference copy of motor output, proprioceptive memory or a memory of a pitch (Sundberg, 1987). External references are intrinsically fixed, and an economical control design might also use fixed (absolute) internal references. However, in the present study, no differences in response magnitude or latency were observed as a function of the ON and OFF conditions and indicates that a relative F_0 reference strategy is used, at least in most subjects, for stabilization where there is no external reference available.

There are two potential relative (variable) referencing organizations that are consistent with our observations. First, subjects may be using changes in auditory feedback velocity to compensate for and null changes in F_0 (termed *velocity control*), rather than keeping F_0 set to a fixed internal reference. Velocity control could also stabilize a glissando by comparing intended and perceived F_0 velocity. An alternative to velocity control is a "*sample and hold*" strategy based on using a memory of initial perception of F_0 as a reference. Disparities between perceived F_0 and memory would invoke a corrective response similar to those observed in this experiment. However sample/hold would likely result in a longer latency compared to the velocity control strategy as it requires first storage of an auditory input followed by a comparison with the referent. In addition, sample/hold could not stabilize a glissando because there is no stable referent, but it might be more accurate than velocity control for stabilizing a steady-state F_0.

These several types of feedback referencing are not mutually exclusive and might all be used simultaneously. Certainly, some degree of absolute F_0 referencing is necessary for accurate tracking of an external reference. Individuals that have reliable internal absolute F_0 references (i.e., "absolute or perfect pitch") might select absolute internal F_0 referencing, velocity control, or sample/hold. Those without absolute pitch might use either velocity control or sample/hold relative referencing. Weightings for these control strategies that vary with behavioral goals, individual abilities, and the availability of references from the environment would likely be the most effective strategy.

These data also indicate that the mechanism for *detecting* the error is the same regardless of the magnitude of the F_0 disparity (25–200 cents). That is, subjects produce a compensatory response to both the ON and OFF stimuli for each stimulus magnitude. However, the ability of the system to *correct* for errors depends on the stimulus magnitude. In this and previous studies (Burnett *et al.*, 1998), it was shown that response magnitude rarely exceeded 50 cents even in the presence of stimulus magnitudes of 100 and 200 cents. This limitation in response magnitude may be present to prevent this reflexive low-level stabilization mechanism from interfering with or destabilizing higher level, more sophisticated tracking mechanisms, possibly using F_0 position references.

Analysis of numbers of "compensatory" and "following" responses show there was no significant difference in the number of "following" responses with the ON or OFF stimuli, which further demonstrates that subjects did not try to match the feedback signal. Had there been more "following" responses with the OFF stimuli, it would have suggested that subjects treated it as an external reference. We have previously postulated that "compensatory" responses are made by comparing the feedback signal with an internal referent, and "following" responses made by comparison with an external referent (i.e., the feedback signal itself or musical accompaniment) (Burnett *et al.*, 1998; Hain *et al.*, 2000). Hence these conditions do not appear to influence whether a person would use an internal versus an external reference.

References

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]

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

Elman JL. Effects of frequency-shifted feedback on the pitch of vocal productions. J Acoust Soc Am 1981;70:45–50. [PubMed: 7264071]

Hain TC, Burnett TA, Larson CR, Kiran S. Effects of delayed auditory feedback (daf) on the pitch-shift reflex. J Acoust Soc Am 2001;109:2146–2152. [PubMed: 11386566]

Hain TC, Larson CR, Burnett TA, Kiran S, Singh S. Instructing participants to make a voluntary response reveals the presence of two vocal responses to pitch-shift stimuli. Exp Brain Res 2000;130:133–141. [PubMed: 10672466]

Jones JA, Munhall KG. Perceptual calibration of *f*0 production: Evidence from feedback perturbation. J Acoust Soc Am 2000;108:1246–1251. [PubMed: 11008824]

Larson CR, Burnett TA, Kiran S, Hain TC. Effects of pitch-shift onset velocity on voice *f*0 responses. J Acoust Soc Am 2000;107:559–564. [PubMed: 10641664]

Sundberg, J. The Science of the Singing Voice. Northern Illinois University Press; Dekalb, IL: 1987.

Takeuchi AH, Hulse SH. Absolute pitch. Psychol Bull 1993;113:345–361. [PubMed: 8451339]

Ternström S, Sundberg J, Colldén A. Articulatory f0 perturbations and auditory feedback. J Speech Hear Res 1988;31:187–192. [PubMed: 3398492]
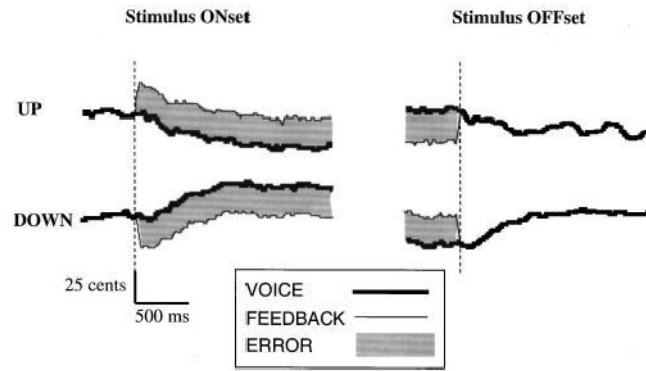
FIG. 1.
Averaged response traces from a single subject illustrating typical behavior to 25-cent pitch-shift stimulus. In each set, the heavy line is voice F_0 and the light line is feedback pitch. The vertical dashed line indicates onset of pitch-shift stimulus (left) and offset of pitch shift stimulus (right). The top row shows responses to upward pitch-shifts, while the bottom row displays downward pitch shifts. Shaded areas represent error between feedback pitch and output F_0.