

Voice fundamental frequency modulates vocal response to pitch perturbations during English speech

Hanjun Liu

Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University,
Guangzhou 510080, People's Republic of China
lhanjun@mail.sysu.edu.cn

James Auger and Charles R. Larson

Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive,
Evanston, Illinois 60208
jamesauger2007@u.northwestern.edu, clarson@northwestern.edu

Abstract: Previous research has demonstrated task-dependent vocal responses to pitch perturbations during speech production. The present study investigated the effect of voice fundamental frequency (F_0) on the modulation of vocal responses during English speech. Randomized pitch shifts of ± 100 or 200 cents during speaking were presented to English speakers. Results indicated larger vocal responses and shorter latencies at a high voice F_0 than at a low voice F_0 , but no significant differences were observed for stimulus magnitude or direction. These findings suggest that the pitch-shift reflex during speech can be modulated as a function of voice F_0 .

© 2010 Acoustical Society of America

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Gr, 43.70.Jt [AL]

Date Received: March 10, 2009 **Date Accepted:** October 15, 2009

1. Introduction

Control of voice fundamental frequency (F_0) plays an important role in vocal communication. Through the regulation of voice F_0 , humans can convey general information about the affective state, speech inflection, artistic purpose, etc. Auditory feedback has been demonstrated to be important for the on-line control of voice F_0 during sustained vowels (Hain *et al.*, 2000), in which subjects compensated for the pitch feedback perturbation by changing their voice F_0 in the direction opposite to the stimulus. Such compensatory mechanisms were also observed during speech production (Chen *et al.*, 2007; Liu *et al.*, 2009). These findings indicate that the audio-vocal system can be modulated in a task-dependent manner to correct the discrepancy between auditory feedback and vocal output.

During most of the previous pitch-shift studies, subjects were asked to vocalize a vowel sound or speak nonsense or meaningful syllables at their habitual pitch (Xu *et al.*, 2004; Chen *et al.*, 2007). A recent pitch-shift study during vowel phonation, however, demonstrated larger response magnitudes and shorter latencies when subjects vocalized at a high F_0 compared to a low F_0 (Liu and Larson, 2007). This study suggested that the sensitivity of the audio-vocal system to voice feedback perturbation might vary as a function of voice F_0 during sustained vowels. However, no research has been conducted on whether the same is true for speech production. Hypothetically, a reflexive input to motor neurons that are discharging at a high rate may lead to a greater level of muscle contraction than an equal input to neurons discharging at a lower rate. The greater degree of muscle contraction associated with vocalizing at a higher F_0 (Hirano *et al.*, 1970) may cause an increase in magnitudes and a decrease in latencies of voice F_0 responses to voice pitch-shifted feedback.

Therefore, the purpose of the present study was to investigate if vocal responses to pitch perturbation in auditory feedback can be modulated as a function of voice F_0 during English speech. We hypothesized that, similar to the sustained vowels, the response magnitudes

to pitch-shifted voice feedback would be larger and response latencies would be shorter at a high F_0 compared to a low F_0 during speech production.

2. Methods

2.1 Subjects

Fifteen female Northwestern University students (18–22 years old) participated in the experiment and produced data that met our criteria of acceptable responses. All of the subjects passed a hearing screening for a 25 dB hearing level bilaterally at 250, 500, 1000, 2000, and 4000 Hz. None of the subjects reported a history of neurological or communication disorders and they all signed informed consent approved by the Northwestern University Institutional Review Board.

2.2 Apparatus

During the testing, subjects were seated in a sound-treated room and wore Sennheiser headphones with attached microphone (model HMD 280). The vocal signal from the microphone was amplified with a Mackie mixer (model 1202) and shifted in pitch with an Eventide Eclipse Harmonizer, and then amplified with a Crown D75 amplifier and HP 350 dB attenuators at 80 dB sound pressure level (SPL). MIDI software (MAX/MSP v4.6 by Cycling 74) was used to control the harmonizer. A Brüel and Kjar sound level meter (model 2250) and in-ear microphone and headphones were used for calibration to make sure that there was a gain in amplitude of 10 dB SPL between the subject's voice amplitude and the feedback loudness. The voice output, feedback, and transistor transistor logic (TTL) control pulses were digitized at 10 kHz, low-pass filtered at 5 kHz, and recorded using CHART software (AD Instruments, Castle Hill, New South Wales, Australia).

2.3 Procedure

Prior to the experiment, a phrase “You know Nina?” was spoken by one experimenter at a high and a low voice F_0 and recorded. Then these two voice samples were processed in PRAAT (Boersma, 2001) so that the average F_0 values during the voice period before the rise of the final syllable (i.e., “You know Ni”) were maintained at around 300 and 200 Hz, which were regarded as reasonably high and low voice F_0 for female subjects according to pilot testing results. Those subjects who were unable to produce the phrase at either of these two voice F_0 s were excluded. Subjects were then instructed that they would hear the phrase sample at either a high or a low voice F_0 , and then they should repeat the phrase within 1 s in exactly the same manner as that of the sample. After a brief period of training, subjects were asked to repeat the phrase immediately following its presentation over the headphones 60 times in each of four conditions: high voice F_0 , low voice F_0 , and 100 or 200 cent pitch-shift magnitude. For each vocalization, the pitch feedback was increased, decreased, or held constant (no stimulus) in a randomized sequence for a total of 60 trials. The duration of each stimulus was 200 ms and the magnitude was held constant at ± 100 or ± 200 cents. The pitch-shift stimulus was presented 200 ms following onset of vocalization. Data were analyzed using event-related averaging techniques (Chen *et al.*, 2007) in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR). A statistical method was used to determine if responses to pitch-shifted feedback differed significantly from control trials (see Xu *et al.*, 2004). SPSS (v. 16.0) was used to test for significant differences in the response magnitude and latency across all the conditions.

3. Results

The period between vocal onset and stimulus onset was extracted to measure the mean F_0 values for each condition. Statistical results showed that on average, subjects spoke at 275 Hz for the high voice F_0 condition and 209 Hz for the low voice F_0 condition [$F(1, 58) = 35.437$, $p < 0.0001$]. From 15 subjects across two voice F_0 s, two stimulus magnitudes, and two stimulus directions, there were 120 responses ($15 \times 2 \times 2 \times 2$). 93 responses opposed the stimulus direction, and 22 responses followed the direction of stimulus. 5 of 120 responses did not meet our criteria of validity and were declared to be non-responses. A chi-square test revealed a statistically greater

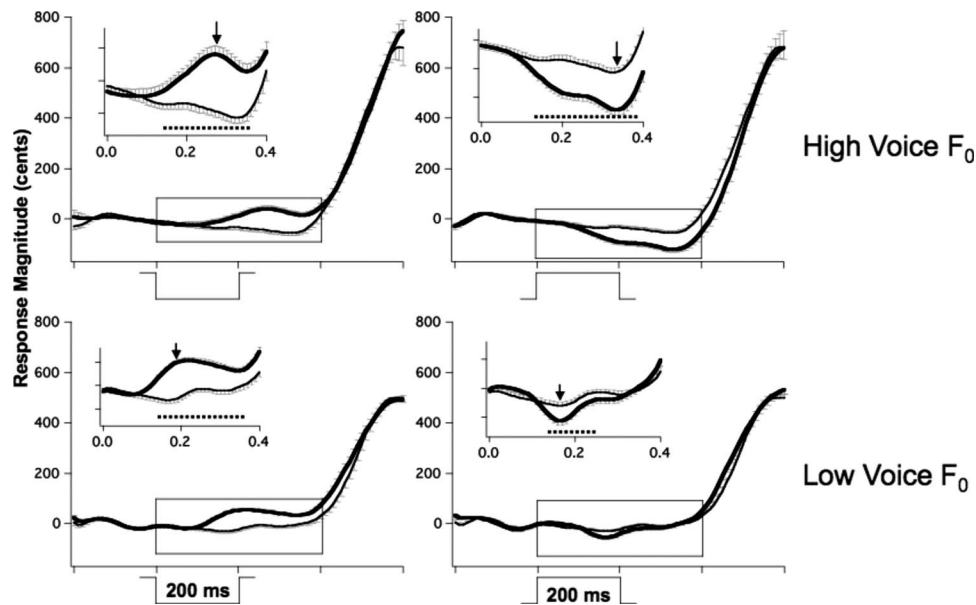


Fig. 1. Representative vocal responses to 200 cent pitch-shifted stimuli as a function of stimulus direction at a high (top) and a low (bottom) voice F_0 , respectively. Thick lines represent the averaged F_0 contours of responses to pitch-shifted feedback, and thin lines represent contours for control trials. The solid vertical arrow indicates time where the response magnitude was measured. The dashed horizontal line represents the onset and offset of the response, and response latency is indicated by the start of this line. The inset shows an expanded portion of average waves. Error bars attached to the contours represent the standard error of the mean for a single direction. Boxes at the bottom indicate the time and the direction of the stimulus.

number of non-response and “following” responses in the low voice F_0 compared to the high voice F_0 condition ($\chi^2=6.259$, d.f. = 1, and $p=0.012$). The distribution of opposing, following, and non-responses was even across stimulus direction and stimulus magnitude.

Figure 1 shows the representative vocal responses to pitch perturbations as a function of stimulus direction at a high and a low voice F_0 . As shown in this figure, larger vocal responses occurred when subjects spoke the phrase at a high voice F_0 compared to a low voice F_0 . Table 1 presents the average and standard deviations (SDs) of the response magnitudes and latencies across all conditions. Although these data lend themselves to repeated-measures ANOVAs (analysis of variance), factorial ANOVAs without repeated-measures were used in the present study to account for missing data and unequal cell size. A three-way ANOVA performed on the response magnitude indicated a significant main effect for voice F_0 [$F(1,81) = 11.817$, $p=0.001$] but not stimulus magnitude [$F(1,81)=0.916$, $p=0.341$] or stimulus direc-

Table 1. Averaged response magnitude SD in cents and response latency SD in ms as a function of voice F_0 , stimulus magnitude, and stimulus direction.

| | | Response magnitude | | Response latency | |
|------------------|------|--------------------|-----------------|------------------|-----------------|
| | | High voice F_0 | Low voice F_0 | High voice F_0 | Low voice F_0 |
| 100 cent stimuli | Up | 27 (18) | 19 (13) | 138 (61) | 165 (77) |
| | Down | 37 (24) | 28 (18) | 117 (48) | 152 (66) |
| 200 cent stimuli | Up | 40 (25) | 22 (11) | 106 (38) | 124 (67) |
| | Down | 48 (33) | 23 (22) | 115 (58) | 140 (63) |

tion [$F(1, 81)=2.062$, $p=0.155$]. The high voice F_0 produced significantly larger response magnitudes (38 ± 26 cents) than the low voice F_0 (23 ± 16 cents). No significant interactions were found across all conditions.

A three-way factorial ANOVA on the response latency revealed a significant main effect for voice F_0 [$F(1, 81)=4.251$, $p=0.042$] but not for stimulus magnitude [$F(1, 81)=2.967$, $p=0.89$] or stimulus direction [$F(1, 81)=0.029$, $p=0.866$]. Faster response latencies were generated at the high voice F_0 (119 ± 52 ms) than the low voice F_0 (146 ± 68 ms). No significant interactions were found in the response latency across all conditions.

4. Summary and discussion

Liu and Larson (2007) reported that voice F_0 has effects on the vocal responses to pitch feedback perturbations during sustained vowels. The present study showed that vocal responses during speech production were also modulated as a function of voice F_0 . As we hypothesized, the higher voice F_0 led to larger response magnitudes and shorter latencies. The finding that larger response magnitudes occurred at the high voice F_0 compared to the low voice F_0 once again demonstrates that vocal response can be modulated in a task-dependent manner, which is consistent with previous research during speech production (Xu *et al.*, 2004; Liu *et al.*, 2009). One possible explanation for the greater magnitudes with higher F_0 is that speaking a sentence at a high voice F_0 may require a greater reliance on auditory feedback than at a low voice F_0 . Thus, auditory feedback is closely monitored and pitch-shifted errors can be corrected with a greater degree of accuracy at a high voice F_0 than at a low voice F_0 .

It was also found that response latencies during speech production varied as a function of voice F_0 , which is consistent with Liu and Larson's (2007) study. Furthermore, this study provides supportive evidence that the timing of vocal responses to the pitch-shifted feedback can be adjusted according to the variations in speech contexts (i.e., voice F_0). Previous research has demonstrated that increases in voice F_0 are accompanied by increases in the activity of the tensor muscles of the larynx as reflected by the greater magnitudes of electromyographic signals (Hirano *et al.*, 1970; Gay *et al.*, 1972). The greater laryngeal motor responsiveness may also lead to the reduction in vocal response latencies when subjects were vocalizing at a high voice F_0 .

The paradigm of the present study was similar to Chen *et al.* (2007), and the only difference is that we instructed the subjects to say "You know Nina?" at either a high or a low voice F_0 while Chen *et al.* (2007) just used a habitual pitch. Nevertheless, the two studies are comparable in that mean response magnitudes were 38 and 23 cents for the high and low voice F_0 in the present study and 31.5 cents for Chen *et al.*'s (2007) study. On the other hand, Chen *et al.* (2007) reported greater response magnitudes for downward stimulus direction than upward direction, which was not observed in the present study. As suggested by Chen *et al.* (2007), the directional effect was observed because downward stimuli may sound to subjects that their voice F_0 was changing in the wrong direction during the production of the phrase with a rise of F_0 on the final syllable. In contrast, the manipulation of voice F_0 in the present study required precise control of laryngeal muscles and may have weakened this interaction between the supra-segmental production and stimulus direction, leading to the absence of the directional effect on the response.

Another difference between the present and the Chen *et al.* (2007) study was that more following responses were found in the present study compared to the Chen *et al.* (2007) study (20% vs 7.5%). It should be noted that 67% of following responses occurred in the low pitch condition and 33% in the high pitch condition, which is consistent with Liu and Larson's (2007) finding during vowel phonation at a low and high F_0 . One possible reason for this difference is that an increase in voice F_0 resulted in greater accuracy of detecting the correct direction of the pitch-shift stimulus and a greater percentage of compensating responses during the high F_0 condition in the present study. In a previous study, it was suggested that accuracy in perception of the direction of a pitch-shift stimulus may be a factor contributing to the direction of voice F_0 responses (Larson *et al.*, 2007). In addition, it was reported that some subjects were better than

others at perceiving the direction of the pitch changes (Semal and Demany, 2006). So individual differences might also contribute to the greater number of following responses and less sensitivity to the stimulus direction in the present study compared to Chen *et al.* (2007).

Acknowledgments

This work was supported by NIH Grant No. 1R01DC006243. The authors thank Chun Liang Chan for programming assistance.

References and links

- Boersma, P. (2001). "Praat, a system for doing phonetics by computer," *Glott International*, **5**, 341–345.
- Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (2007). "Voice F0 responses to pitch-shifted voice feedback during English speech," *J. Acoust. Soc. Am.* **121**, 1157–1163.
- Gay, T., Hirose, H., Strome, M., and Sawashima, M. (1972). "Electromyography of the intrinsic laryngeal muscles during phonation," *Ann. Otol. Rhinol. Laryngol.* **81**, 401–409.
- 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.
- Hirano, M., Vennard, W., and Ohala, J. (1970). "Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles," *Folia Phoniatr Logop* **22**, 1–20.
- Larson, C. R., Sun, J., and Hain, T. C. (2007). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control," *J. Acoust. Soc. Am.* **121**, 2862–2872.
- Liu, H., and Larson, C. R. (2007). "Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex," *J. Acoust. Soc. Am.* **122**, 3671–3677.
- Liu, H., Xu, Y., and Larson, C. R. (2009). "Attenuation of vocal responses to pitch perturbations during Mandarin speech," *J. Acoust. Soc. Am.* **125**, 2299–2306.
- Semal, C., and Demany, L. (2006). "Individual differences in the sensitivity to pitch direction," *J. Acoust. Soc. Am.* **120**, 3907–3915.
- 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.