VOCAL RESPONSES TO FREQUENCY MODULATED COMPOSITE SINEWAVES VIA AUDITORY AND VIBROTACTILE PATHWAYS

Xiaozhen WANG[1], Kiyoshi HONDA[1], Jianwu DANG[1, 2], Jianguo WEI[1]

[1]Tianjin Key Laboratory of Cognitive Computation & its Applications, Tianjin University, Tianjin, China

[2]School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan

ABSTRACT

Feedback control mechanisms for speaking have been examined using the transformed auditory feedback (TAF) technique. Previous studies have shown that speakers demonstrate fundamental frequency (F0) changes when they monitor their voice with artificial alterations of F0. However, those studies underestimate the role of vibrotactile information involved in feedback F0 control. This pilot study aims at exploring whether and how vibrotactile information from the larynx influences vowel F0. Participants in our experiment were asked to sustain vowel with their F0 adjusted to composite sinewave stimuli, which were given via auditory and vibrotactile channels using a headset on the ears or a bone-conduction transducer on the larynx. Results revealed the greater compensatory responses to combined vibrotactile-auditory stimuli than to the responses to auditory-only stimuli. The effect of vibrotactile stimuli on feedback F0 adjustment was also observed with the shorter latency of the responses.

Index Terms— vocal feedback response, voice frequency control, composite sinewave frequency modulation, vibrotactile feedback, auditory feedback.

1. INTRODUCTION

Speech production is under certain feedback controls involving multiple sensory pathways. This notion has been examined so far by studies to explore the linkage between perception and production using a real-time feedback procedure called the transformed auditory feedback (TAF) or frequency-shifted auditory feedback. Common findings from those studies are compensatory responses. For example, participants for pitch-shifted feedback experiments tend to change their vocal fundamental frequency (F0) to the opposite direction to the real-time shift in F0 of the monitored TAF signals [1-5]. This observation supports the role of auditory feedback mechanisms of F0 control in voice production, and the latencies of responses also suggest the nature of the feedback system involved in such control mechanisms.

The TAF procedure was proposed to evaluate contributions of auditory feedback to F0 trajectory generation under natural speech conditions by Kawahara et al [2]. Using this technique, Burnett et al. revealed the shorter latencies (192 ms) for the opposing responses than those for the following response (327 ms) in cases with small unexpected interferences [3]. Although responses to vocal intensity and F0 in pitch-shifted feedback have been studied frequently, no previous investigation is reported to date on the impact of the vibration of the laryngeal wall that may be involve in feedback vocal control for speech production. Thus, a new attempt was made in this study to examine the role of vibrotactile information as described in what follows.

We focus on the fact that speakers do not only hear their own voice but also sense vibration produced by vocal-fold oscillation. Vibrotactile sensory apparatus is abundant in the laryngeal area, and they send afferent signals to the somatosensory area mainly via the sensory branch of the superior laryngeal nerve. It has been reported that laryngeal reflexes are elicited by stimulation of the internal superior laryngeal nerve, inferring its role in different functions such as vocalization, breathing, swallowing, and coughing [6]. In the TAF condition, the subject's auditory system responds to altered frequency feedback, while the vibrotactile sensory system receives no such changes in wall vibration accompanied by vocal-fold vibration. The purpose of this pilot study is to examine the role of vibrotactile feedback in F0 control in voice production using a bone-conduction transducer to stimulate the laryngeal wall and ears separately or simultaneously.

2. METHODS

The task used in this experiment is simply to produce a sustained vowel while receiving composite sinusoidal signals that shift upward or downward in frequency. Subjects maintain their vocal pitch adjusted to initial F0 of the heard signals. The signals were given to each subject through a bone conduction transducer on the neck or a headset on the ears. We call this task Composite-sinewave Frequency Modulation (CFM).

2.1. Participants

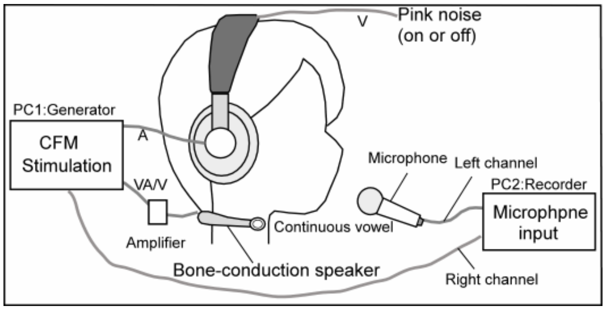


Fig.1. Experimental setup for auditory and vibrotactile stimulation. A paired bone-conduction transducer was fixed near the larynx. PC1 is used to generate composite sinewaves with frequency shifts, and PC2 is used to record produced sounds and stimuli.

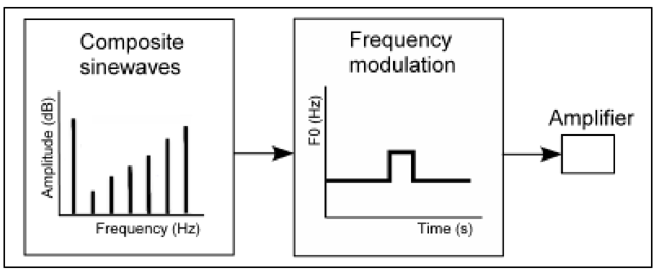


Fig.2. Generating frequency-modulated composite-sinewave stimuli. Seven harmonics were combined to generate composite sinewave signals, which were modulated for fundamental frequency perturbation.

Sixteen subjects of 19 to 26 years (6 female, and 10 male subjects) participated in this study. All the subjects were native Chinese speakers who reported no history of speech and hearing dysfunctions.

2.2. Auditory and vibrotactile stimuli

Our procedure for the CFM experiment is shown in Figure 1. A set of bone conduction sport headphones (Aftershokz) was fixed bilaterally on the neck wall near the larynx using an elastic band. An audio headset (Sennheiser) with a microphone was used to output CFM signals for listening or masking. A flat-sheet pressure sensor (FSR) was used to monitor the force on the bone-conduction transducer by the elastic band to keep the tightness of the band about the same across subjects. The harmonics pattern shown in Figure 2 was chosen as the most effective of the composite sinewave for vibration as among several test patterns including that with a pattern of -6 dB/oct slope.

The subjects produced sustained vowel /a/ with the pitch adjusted to the initial F0 of the CFM signals, repeating 60 times for each type of stimuli over three conditions. The first condition is CFM vibrotactile-auditory (VA) stimulation using the bone-conduction transducer that sends

Table 1. The value for fundamental frequency (F0) with different shift magnitudes.

Gender	F0	Upwards 10%	Upwards 20%	Downwards 10%	Downwards 20%
male	130Hz	143Hz	156Hz	117Hz	104Hz
female	230Hz	253Hz	276Hz	207Hz	184Hz

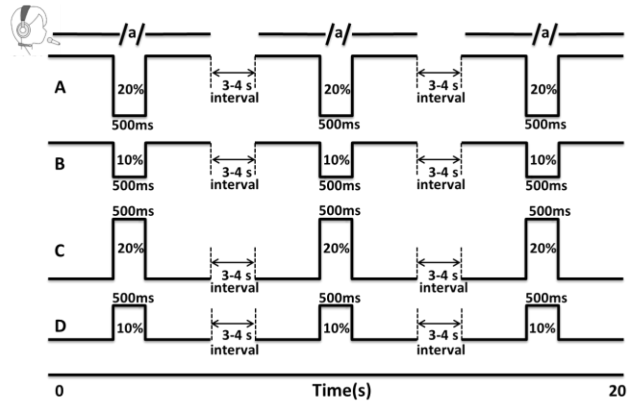
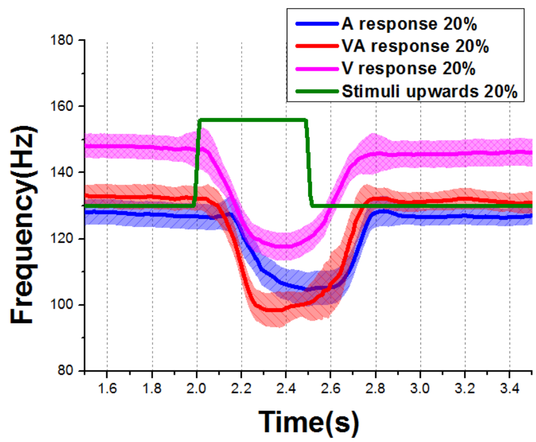


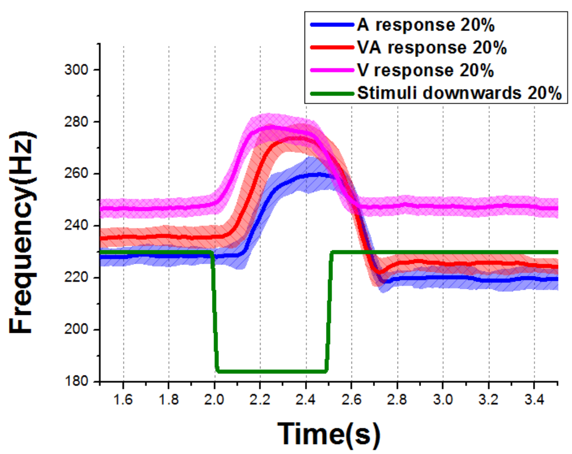
Fig.3. Diagram of the stimulus sequences used in the experiments. Each stimulus contains a 10% or 20% frequency-shift downwards and upwards for a shift duration of 500 ms in the time sequence from 0 s to 20 s with an inter-stimulus interval of 3-4 s.

both sound to the ears and vibration to the laryngeal wall, while the signal frequencies shift for a short period as shown in Figure 2. The second condition is CFM auditory-only (A) stimulation through a headset, and subjects receive sounds as in the natural hearing condition. The third condition is CFM vibrotactile-only (V) stimulation with 70-dB pink noise added via a headset to elicit the response to the laryngeal vibration with no effects on hearing. The session for the CFM-V condition with auditory masking was conducted after that for the CFM-VA and CFM-A conditions. This is to familiarize the subjects with the CFM stimuli so as to facilitate subjects' responses to non-audible signals. In the three conditions, the F0 base value of the CFM signals for male subjects is set at 130 Hz, and that for female subjects is set at 230Hz. All the subjects conducted a hearing training at the 70 dB sound pressure level (SPL) through the bone-conduction transducer and audio headset to test whether they can distinguish different F0 signals.

All the stimuli in the experiments were prepared to keep the nearly identical intensity according to subjective judgement. Upward or downward frequency shift was either 10% or 20% as shown in Table 1. Each F0 modulation takes place with a fixed duration of 200 ms or 500 ms, including rise and fall ramps of 5 ms. The stimuli were presented with an inter-stimulus interval of 3~4 s to let the subject take a break after each vocalization for 5 s. Each condition contained 8 trials (upwards 10% and 20% with shift durations of 200 ms and 500 ms; downwards 10% and



(a)



(b)

Fig.4. F0 response curves in the three conditions. The averaged F0 trajectories with standard deviation are plotted for the responses in the conditions of vibrotactile-auditory (VA) (red), auditory-only (A) (blue) and vibrotactile-only (V) (purple), during (a) 20% upward F0 shift for all male subjects, and (b) 20% downward F0 shift for all female subjects.

20% with shift durations of 200 ms and 500 ms). Each trial includes 60 repetitions (taking about 8 minutes) as shown in Figure 3. While the subjects sustained the vowel, they listened the frequency modulation after 2 s from the stimulus initiation. The responses in the first and last 300 ms of each session were discarded from analysis to minimize potential initiation and termination effects.

All the subjects participated in all the trails (8 trials) under the three conditions, and the half of all the subjects received the vibrotactile-auditory (VA) stimulation after the auditory-only (A) stimulation with a break of 5 minutes between conditions. The other half of the subjects received the vibrotactile- auditory (VA) stimulation followed by the auditory-only (A) stimulation.

Table 2. Summary of subjects' response directions to each condition.

Response	Opposite	Follow	No response
Auditory(A)	92.82%	3.89%	3.28%
Vibrotactile-Auditory(VA)	94.36%	3.13%	2.51%
Vibrotactile(V)	80.95%	6.24%	12.81%

3. RESULTS

The data acquired from the experiment were analyzed to report subjects' responses to the composite-sinewave frequency modulation (CFM) stimuli regarding the direction and magnitude of F0 changes as well as the latency for initiation of those changes (on-responses only). F0 response curves were obtained using Praat and averaged for each condition for each subject.

3.1. Responses to CFM Stimuli

The subjects in this experiment responded to the CFM stimuli in most of the three conditions, and the response patterns resembled with what were expected from TAF experiments [7-8]. The direction of responses was mostly opposite to that of CFM stimulus changes, and the compensatory responses were observed in 94.36% for the vibrotactile-auditory (VA) and 92.82% for auditory-only (A) conditions as shown in Table 2. This result supports that the CFM procedure is applicable to examine speakers' vocal responses to F0 modulation stimuli. Also, it is found that vocal responses can be obtained by vibrotactile-only (V) stimulation to the larynx.

3.2 F0 Response Curves

F0 response curves were calculated from the audio data recorded during the experiments for the three conditions. The results show a consistent tendency: the deeper the F0 modulation the larger the vocal response in F0, being consistent with previous TAF studies [9-11]. Figure 4 shows two representative data for averaged F0 response curves obtained from all the male and female speakers. In the figure, two sets of representative data are shown for the responses to 20% upward F0 shift in Figure 4(a) and to 20% downward F0 shift in Figure 4(b). Comparing the on-responses across conditions, the vibrotactile-auditory (VA) responses are larger than the auditory-only (A) responses. The vibrotactile-only (V) responses under auditory masking are varied in F0 patterns for on- and off-responses, and the on-responses tend to start earlier than those in other conditions. In the detail, about 19.05% of the V responses showed no changes or changes to the direction same to the modulation in stimuli (i.e., following responses).

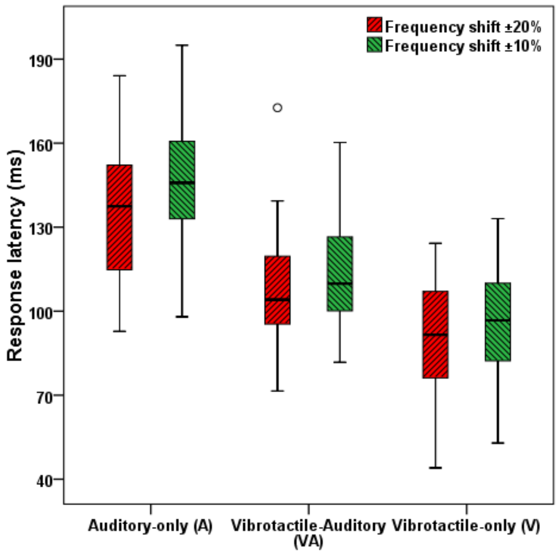


Fig.5. Averaged latency values for the three conditions with frequency shifts upwards or downwards (20% or 10%).

3.3. Response Latencies

The latency of the responses was analyzed by measuring the time interval between the initial changes of the stimuli and responses. Measured latency values were averaged over the subjects in each condition and plotted in Figure 5. As the general trend, the latency is shorter in the VA condition than in the A condition, and it is shortest in the V condition. Also, the latency is slightly shorter in the responses to 20% modulation than those to 10% modulation. Those observations suggest a possible role of vibrotactile signals from the larynx to facilitate vocal feedback responses.

4. DISCUSSION

The transformed auditory feedback (TAF) has been the tool to explore neural mechanisms in vocal feedback control when subjects with a headset hear their own F0 shifted upward or downward during phonation. In this study, we considered another factor that may be involved in the feedback control of F0: it is vibrotactile information generated at the larynx by vocal-fold vibration during phonation. To examine the effects of auditory and vibrotactile information on vocal responses, we employed the composite-sinewave frequency modulation (CFM) procedure in place of the common TAF technique. We found that CFM stimuli also cause compensatory responses in the form of negative feedback adjustment to regulate vocal F0.

It was observed that F0 changes in response to the vibrotactile-auditory (VA) stimulation are larger than those to the auditory-only (A) stimulation. This observation suggests that the vibrotactile feedback system is involved in F0 control, and vibration of the larynx wall can be active afferent information for F0 control. Auditory feedback causes certain responses to audible signals, while combined vibrotactile-auditory feedback integrates both pathways to facilitate overall responses. This is what was speculated by the F0 response curves obtained in this study.

The most interesting observation is seen in the result from the vibrotactile-only condition: the subjects respond to the vibrotactile-only (V) stimuli. The responses to the vibrotactile-only (V) stimuli always showed the higher F0 curves than those to the auditory-only (A) and vibrotactile-auditory (VA) stimuli as shown in Figure 4. One of the probable reasons is that the subjects were unable to adjust their vocal pitch to the initial F0 of the stimuli under the masking noise via a headset to suppress the auditory signals from the bone-conduction transducer. Since auditory masking causes uncertainty in detecting target F0 from vibrotactile stimuli in this study, different modes of vibrotactile-only stimulation may need to be considered in the next study.

The averaged latency values were compared across the three CFM conditions, and the responses to the vibrotactile-only stimulation were the fastest. This suggests that vibrotactile signals may have a shorter feedback loop and that they are available to be monitored before auditory signals activate the feedback control mechanism. It has been known as a general tendency that the greater the stimulus magnitude the shorter the latency. Attention will be paid in the next study to maintain the same intensity of auditory signals between the auditory-only and vibrotactile-auditory conditions.

A remaining issue is that the subjects may predict the timing of F0 changes with the constant lead time before each F0 change in the stimulus. The randomization of the lead time will be considered in the next study.

To summarize, vocal control in speech does not only involve auditory feedback but also utilizes vibrotactile feedback from the larynx caused by vocal-fold vibration. However, our work is only preliminary to learn how integrated vibrotactile-auditory information influences speech perception and production. Further studies are needed to explore the nature of overall feedback mechanisms with multiple pathways.

5. ACKNOWLEDGEMENT

The research is supported by the National 1000-Plan Project of China (WQ20111200010), the National Basic Research Program of China (No. 2013CB329301), and the National Natural Science Foundation of China (No. 61175016, No. 61303109 and Key Program No. 61233009). The authors are thankful to subjects who participated in the study for 3 hours data collection.

6. REFERENCES

[1] Elman JL, “Effects of frequency-shifted feedback on the pitch of vocal productions,” *J Acoust Soc Am,* 70, 45-50, 1981.

[2] Kawahara H, “Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies,” *J. Acoust Soc Jpn (E)*, 15, 201-202, 1994.

[3] Burnett TA, Freedland MB, and Larson CR, “Voice F0 responses to manipulation in pitch feedback,” *J Acoust Soc Am,* 103, 3153–3161, 1998.

[4] Honda M, Fujino A, Kaburag T, “ Compensatory responses of articulators to unexpected perturbation of the palate shape,” *J Phonetics,* 30, 281-302, 2002.

[5] Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK, “Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex,” *Experimental Brain Research*, 130, 133-141, 2000.

[6] Ludlow CL, “Central nervous system control of the laryngeal muscles in humans,” *Respir Physiol Neurobiol*, 205-222, 2005.

[7] Xu Y, Larson CR, Bauer J, and Hain T, “ Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences,” *J Acoust Soc Am*, 116, 1168–1178,2004.

[8] Sivasankar M, Bauer JJ, Babu T, Larson CR, “Voice responses to changes in pitch of voice or tone auditory feedback,” *J Acoust Soc Am,* 117, 850-857, 2005.

[9] Larson CR, “The role of auditory feedback for the control of voice fundamental frequency and amplitude,” *SIG 5 Perspectives on Speech Science and Orofacial Disorders,* 9-17, 2009.

[10] Patel S, Nishimura C, Lodhavia A, Korzyukov O, Parkinson A, Robin DA, Larson CR, “Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback,” *J Acoust Soc Am,* 135, 3036-3041, 2014.

[11] Chen SH, Liu H, Xu Y, Larson CR, “Voice F0 responses to pitch-shifted voice feedback during English speech,” *J Acoust Soc Am,* 121, 1157-1163, 2006.