Influences of Auditory and Vibrotactile Information on Vocal F0 Responses

Xiaozhen Wang*, Kiyoshi Honda*, Jianwu Dang*, †, Hongcui Wang* and Jianguo Wei*
*Tianjin Key Laboratory of Cognitive Computation & its Applications, Tianjin University, Tianjin, China
†School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan
E-mail: wxzzhen1990@126.com Tel: +86-13821056162
E-mail: khonda@sannet.ne.jp

Abstract— **Feedback mechanisms for fundamental frequency (F0) control have been explored using the transformed auditory feedback (TAF) technique. However, those studies have underestimated the fact that the vibrotactile information from the laryngeal cavity wall during vocal-fold vibration is also involved in F0 control in speech. Our previous study examined the role of auditory and somatosensory information in vocalization. The present study further investigates how the vibrotactile information from the larynx influences vocal response parameters (F0, latency, and magnitude). Subjects participated in this experiment were instructed to sustain vowel /a/ while receiving the stimuli called the composite sine-wave modulation (CFM) of tone or vibration that shifts in frequency for a short period. The CFM signals were delivered through a bone-conduction speaker fixed on the neck surface near the larynx. Results demonstrated the larger magnitude, shorter latency and deeper peak F0 modulation in compensatory responses to the combined vibrotactile-auditory stimuli than to the responses to auditory-only stimuli. These findings strongly suggest that the vibrotactile afferent information is utilized in feedback control of F0 with the shorter delay to adjust vocal responses.**

I. INTRODUCTION

Vocal control in speech involves feedback mechanisms to minimize the error between intended and realized vocal fundamental frequencies (F0). The main feedback loop for the control system is thought to be auditory, but the somatosensory feedback also plays a role because vibration of the laryngeal cavity wall during vocal-fold oscillation is sensed by the mechanosensory apparatus in the laryngeal mucosa and monitored by the higher sensory system. The contribution of the auditory channel to activating the feedback mechanism has been examined in vivo using a technique for real-time frequency modulation of a speaker's voice, called transformed auditory feedback (TAF) or frequency-shifted auditory feedback [1-3]. In such experiments, speakers respond to the frequency-modulated auditory stimuli mostly to maintain the intended F0 by shifting their vocal F0 in the direction opposite to the F0 shift in the stimuli [4-6]. The TAF technique to perturb the auditory system has been successful at evaluating the role of auditory feedback in F0 trajectory generation, but the procedure is somewhat unnatural because the vocal feedback control is multisensory in nature. In addition to auditory monitoring of F0, vibration of the vocal folds and laryngeal cavity wall is also monitored by the sensory system. It has been noted that a subtle shift of vowel sounds in the artificial auditory feedback may cause a small mismatch between the auditory and somatosensory feedback signals [7]. Indeed, vocal-fold vibration during speech activates not only the auditory mechanism but also other sensory systems of many modalities simultaneously. Mechanosensory modulation in the larynx also takes part as one of the underlying neural mechanisms of vocal control [8].

In our previous study [9], an attempt was made to examine the role of auditory and vibrotactile feedback signals in F0 control using frequency-modulated auditory and vibratory stimuli, and the result revealed both similarity and difference with those from the TAF studies. The main findings from our study included the compensatory responses to vibrotactile-only stimulation and the larger response magnitude in combined vibrotactile-auditory stimulation than in auditory-only stimulation. Although the result supported the contribution of vibrotactile stimuli in vocal feedback control, a few problems were noticed that may result in inaccuracy. The lead time between stimulus initiation and modulation onset was constant (not randomized), which may facilitate subjects' responses by prediction of stimulus onset timing. Also, the auditory and vibrotactile stimuli were delivered from different transducers, which resulted in different magnitudes of auditory stimuli between the auditory-only and vibrotactile-auditory conditions. Thus, the present study was planned to replicate the previous experiment with the renewed procedures toward the same goal of examining the role of auditory and somatosensory feedback in controlling vocal F0.

II. METHODS

The subjects in this experiment were instructed to produce a sustained vowel /a/, while receiving the composite sine-wave frequency modulation (CFM) stimuli that shifted upward or downward in frequency. They adjusted their vocal pitch to the base frequency of the CFM stimuli that were delivered through a bone-conduction speaker placed on the neck surface. The experiments were conducted in three sessions with auditory (A), vibrotactile-auditory (VA) and vibrotactile (V) conditions.

A. Participants

Eighteen subjects of 21 to 26 years (6 female, and 12 male subjects) participated in this study. All the subjects were

native Chinese speakers who reported no history of speaking, hearing, or motor disorders.

B. *Experiment procedure*

The setup for the experiment is shown in Figure 1. The subjects wore a bone-conduction speaker with bilateral transducer units (Aftershokz) on the neck near the larynx to receive airborne and/or vibratory signals. The stimulus signals synthesized with a PC (PC1) were harmonic signals with a short period of frequency modulation as shown in Figure 1. This signal was called the composite frequency modulation (CFM) in this study. In one condition, a headset was used to mask airborne signals from the bone-conduction speaker by pink noise. In another condition, plates of visco-elastic material were placed between the transducer units and the neck surface to damp the vibration on the neck skin. The experiment was conducted in a quiet room to record the subject's vowel production with a condenser microphone (Sennheiser) plugged into an USB audio interface (Creative) with another PC (PC2).

Different modalities of stimulation were employed in three conditions. The first was the vibrotactile-auditory (VA) condition. The subjects received both stimuli of sound to the ears and vibration on the neck surface. The second was the auditory-only (A) condition. The subject listened to auditory stimuli from the speaker, while the vibration on the neck was heavily damped by the visco-elastic plates. The third was the vibrotactile-only (V) condition. The subjects listened to pink noise via a noise-cancelling headset (ATH-ANC29) to mask airborne sound that generated by the bone-conduction speaker. In this condition, they received vibration on the neck with no effective auditory stimuli. Using the above procedure to deliver effective stimuli in all three conditions, the identical amplitude and F0 were maintained in all the conditions in the experiment. The bone-conduction speaker units were supported by an elastic band, with a flat-sheet pressure sensor (FSR) to monitor the force of the elastic band on the transducers. The effect of the noise-cancelling headset was found to be minimal for attenuating the airborne stimuli to the ears, but its effect on altering the effective auditory stimuli was also negligible.

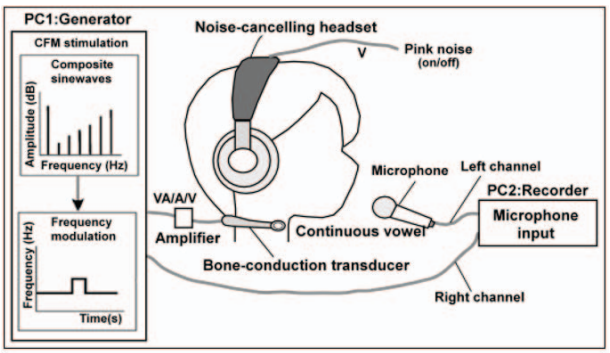

Fig. 1 Experimental setup for auditory and vibrotactile stimulation using the bone-conduction speaker fixed on the larynx. Seven harmonics were combined to generate composite sine-wave signals, which were frequency-modulated for 10% or 20% either in upward or downward directions.

The task for the subjects was to produce vowel /a/ with their vocal pitch adjusted to the base frequency of the stimuli. All the subjects participated in the trial sessions with four types of stimuli under the three conditions. The V condition with auditory masking was conducted after the VA and A conditions so that the subjects were familiarized with the CFM stimuli when they responded to non-audible stimuli in the V condition.

C. *Stimuli used*

The choice of the type of the stimuli was considered first among the saw-tooth wave, vowel-like harmonics pattern (with a -6 dB slope), and other artificial patterns in order to confirm the effectiveness for vibration and the naturalness for hearing. Among those, the harmonics pattern shown in Figure 1 was found most effective for the subjects to sense the modulation delivered to the larynx, and it was chosen as the CFM stimulus for the experiment.

The depth of F0 modulation in the CFM signals was either 10% or 20% with the base frequency of 130 Hz for male subjects and of 230 Hz for female subjects, and the directions of modulation was upward (+) or downward (-). The control stimuli with no modulation were also included. The F0 modulation in each stimulus was set at a fixed duration of 500 ms with rise and fall ramps of 5 ms and the randomized lead time of 1-2 sec. Each stimulus was delivered with the randomized inter-stimulus interval of 3-4 sec. Each condition contained four trials, where the F0 modulation of the CFM stimuli was 10% or 20% in depth and upward or downward in direction. The order of the stimuli to be presented was altered between those with 10% and 20% modulation in every two trials. Each type of stimuli included 40 repetitions (nearly 6 minutes) as shown in Figure 2.

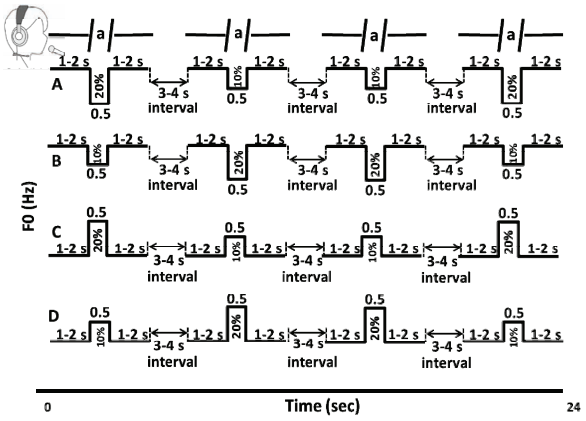

Fig. 2 Schematic patterns of the four types of stimuli. Each CFM stimulus of 4.5-sec duration was presented to the subjects. F0 modulation of 500-ms duration occurred with randomized lead time of 1-2 sec. Presentation of the four types of modulation was also randomized for the depth of modulation (10% and 20%) and direction of modulation (downward and upward), where downward (A) 20% or (B) 10% modulation occurred first, and upward (C) 20% or (D) 10% modulation occurred first.

D. Data analysis

The vocal responses from the subjects were segmented into individual records of 4.5-sec duration for each stimulus type. Short-time energy was calculated as in (1).

$$E_n = \sum_{m=n-N+1}^{n} [x(m)w(n-m)]^2 \quad (1)$$

where $x(n)$ is the speech signal, $w(n)$ is the window function, and N is time-window length. Then voice F0 was extracted from each vocal response using the autocorrelation algorithm, as in (2).

$$R_n(\tau) = \sum_{m=0}^{N-\tau-1} [x(n+m)w(m)][x(n+m+\tau)w(m+\tau)] \quad (2)$$

where τ is the delay of speech signal $x(n)$. The F0 curves were averaged across all the individual records having the same stimulus type in the same condition. Then, the response latency, peak F0 value, and maximum F0 response magnitude were automatically measured. In order to examine the effects of auditory and vibrotactile stimulation on vocal F0 control, statistical analyses were conducted on the response latency and magnitude of the A-V-VA complex using One-way ANOVA on SPSS with the significance level with the p-values less than or equal to 0.05. For statistically significant effects, a post hoc Tukey–Kramer analysis was performed with the alpha set at 0.05.

III. Results

All the subjects responded to the CFM stimuli in the three conditions. The latency of the response was analyzed by measuring the time interval between the initial F0 change of the stimuli and the moment when the amplitude of the on-response signals exceeded the threshold value set at the twice of the standard variation. The magnitude of responses was defined by the difference between the mean F0 value before the vocal F0 shift and the peak F0 value during the shift in the response.

A. F0 response curves

Figure 3 showed averaged F0 response curves obtained among the compensatory responses in the audio data recorded from all the male subjects. The panels in the figure indicate characteristic response patterns in the three conditions. The frequency of occurrence of the compensatory responses (opposite to the direction of F0 modulation in the stimulus) was 91.53% for the vibrotactile-auditory (VA) and 94.21% for the auditory-only (A) conditions. In contrast, the compensatory responses were observed in 80.26% for the vibrotactile-only (V) condition. The result for the response magnitude showed a consistent tendency: the deeper the F0 modulation, the larger the vocal response, being consistent with previous results from TAF-based studies [10-13].

B. Vibrotactile-only response

From the total of 2880 responses in the V condition, 2312 compensatory responses (80.26%) were measured and analyzed for the latency and magnitude as shown in Figure 3. Among all the subjects, 7.24% responses were in the same direction with the stimulus modulation (following responses), and 12.50% did not meet the criteria of a response (no-response). The larger response magnitude was obtained from the stimuli with upward 20% modulation than those with 10%.

C. Response magnitudes and latencies

Figure 4 showed box plots for quantitative measures of the response latency and F0 magnitude. Pairwise comparisons among the A, V and VA conditions indicated that the stimuli in the VA condition elicited the significantly larger response magnitudes ($F(1,22)=7.158$, $p<0.001$) with the shorter latencies ($F(1,34)=11.565$, $p<0.001$) than those in the A condition. The subjects tended to show the shortest latency ($F(2,51)=10.175$, $p<0.001$) in the V condition. The response magnitude was significantly greater in upward 20% modulations in the V condition by comparing among all other stimulus conditions. The data from the female subjects demonstrated the same trend with the male data in the latency and magnitude. It was also found that the latency for upward responses was shorter than that for downward responses in the compensatory responses.

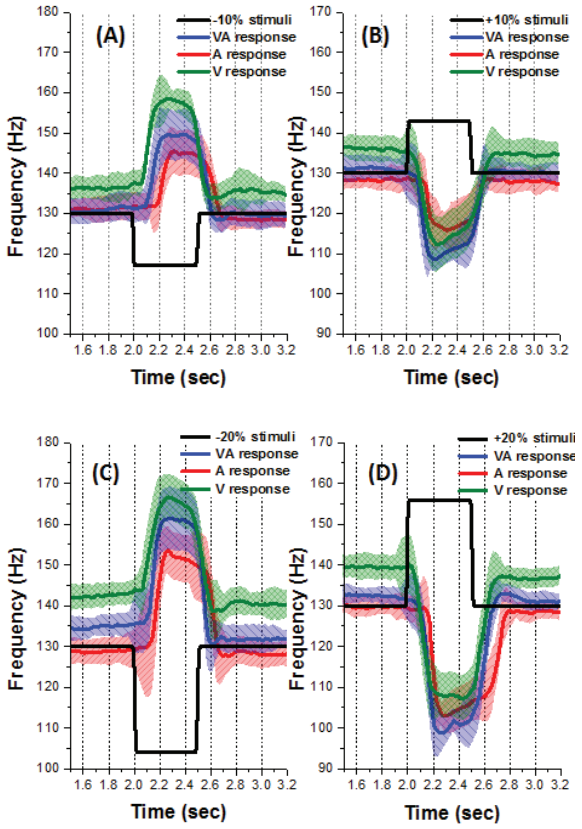

Fig. 3 Averaged F0 trajectories with standard deviations for all the male subjects. Responses are plotted for VA (blue), A (red), and V (green) conditions with (A) 10% and (C) 20% downward (-) modulations and (B) 10% and (D) 20% upward (+) modulations. The shaded regions with different colors indicate the standard deviations.

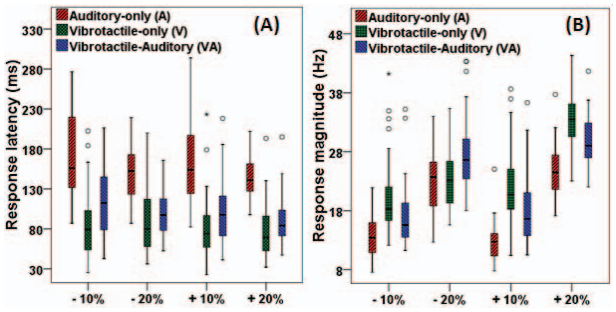

Fig. 4 Box plots showing (A) response latencies and (B) response magnitudes in the three conditions. The data were shown for the compensatory responses only from all the male subjects. Points depicted as "o" represent F0 value beyond the upper and lower limits (±1.5 high/low hinge) of main body of data (mild outliers); Points shown as "*" are those exceeding the above limits of ±3.0 high/low hinge (extreme outliers).

D. Response peak F0

The peak F0 was analyzed for the compensatory on-responses by measuring the maximum F0 in the downward modulation and the minimum F0 in the upward modulation. The impacts of the VA, A and V conditions on vocal F0 responses are characterized by the larger of the peak F0 in the downward modulation and the smaller of the peak F0 in the upward modulation. Figure 5 showed the averaged peak F0 measured from the compensatory responses for all subjects. Comparing the peak F0 values across the three conditions, the peak F0 values were more extreme in the VA condition than in the A condition regardless the depths of the modulation. In contrast, the peak F0 values were dependent on the directions of modulation: In the V condition, the peak F0 values were the largest in the downward modulation, while they were between those in the A and VA conditions in the upward modulation.

IV. DISCUSSIONS

In this study, we re-examined the effect of the vibrotactile information at the larynx on the feedback control of F0. To do so, the composite-sinewave frequency modulation (CFM) was used to explore the influence of auditory and somatosensory information on vocal control using improved procedures. It was re-confirmed that the CFM stimuli caused compensatory responses to maintain vocal F0. The responses to the vibrotactile-auditory (VA) stimuli occurred with the shorter latency and larger magnitude than that to the auditory-only (A) stimuli (Figure 4). This observation suggests that the vibrotactile feedback system plays a certain effective role in F0 control. The auditory feedback induces responses to audible sounds, while the combined vibrotactile-auditory stimuli integrate the sensory signals in the afferent pathways to elicit significantly larger vocal responses than other conditions.

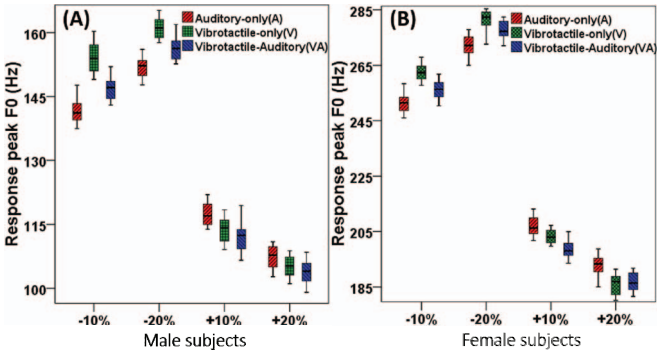

Fig. 5 Box plots showing averaged peak F0 values for all the subjects in the male (A) and female (B) groups. The data were plotted for the compensatory responses in the VA (blue), A (red) and V (green) conditions during downward (-) and upward (+) modulations with 10% and 20% modulation depths.

In our previous study, 7.1% of the compensatory responses showed negative latency values and were judged to be the responses by subjects' prediction of the onset timing of F0 modulation. This was probably due to the previous CFM stimuli with the constant lead time (the interval between stimulus initiation and modulation onset). The renewed CFM stimuli used in the present experiment included randomization of the lead time, and the responses with the prediction of onset timing were only 1.9%. Therefore, the results essentially agree with those from the previous study, further supporting the confidence of the results.

The responses to the vibrotactile-only (V) stimulation result in the shortest latency in all the conditions (Figure 4(A)), which suggests that the vibrotactile sensory pathway may have the shorter feedback loop and that the vibrotactile signals are possibly monitored before the auditory signals activate the feedback control mechanism.

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 (Figure 3). Two reasons can be considered for the shift of F0 trajectories in the presence of masking noise: (1) difficulty in adjusting vocal F0 to the inaudible stimuli, and (2) augmentation of vocal effort due to the masking noise similar to the Lombard effect [14].

The response magnitude reveals that the deeper the F0 modulation, the larger the response magnitude in the VA and A conditions, whereas no clear tendency is observed in the V condition (Figure 4(B)). This difference remains as a question partly because of the uncertainty of the frequency characteristics of the somatosensory system and the bone-conduction speaker when placed on the skin surface.

The peripheral somatosensory system involved in the vibrotactile feedback from the larynx is the sensory branch of the superior laryngeal nerve (SLN) that innervates the laryngeal mucosa lining the cavity wall. According to Ludlow et al. [15], the SLN is involved in feedback control of various functions of the larynx. Laryngeal muscle reflexes are elicited by stimulation of the SLN in animals and humans, and these reflexes are thought to play a role in the sensorimotor control of voice. Although the direct evidence lacks in the literature, it has been reported that anesthesia of the laryngeal mucosa affects feedback F0 responses in humans [16]. This finding

supports that compound action potentials in the SLN caused by vocal-fold vibration convey afferent information of vocal F0. Therefore, it is reasonable to conjecture that the responses to the external vibration to the larynx observed in this study are mediated by the feedback loop via the SLN to monitor the vibrotactile stimuli applied to the larynx.

V. CONCLUSION

Our study demonstrates that the auditory and somatosensory information interacts to influence the vocal F0 responses. The vocal F0 responses are more sensitive to the combined vibrotactile-auditory stimuli judging from the shorter latency and the larger magnitude of peak F0 than those to the auditory-only stimuli. The responses to the vibrotactile-only stimuli exhibit the shortest response latency among all the conditions. These results support our assumptions that the vibrotactile information contributes to fast adjustment of F0 in voicing and that the feedback control of F0 is multisensory in nature.

ACKNOWLEDGMENT

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 the subjects who participated in the study for 2 hours data collection.

REFERENCES

[1] J. L. Elman, "Effects of frequency-shifted feedback on the pitch of vocal productions," *J. Acoust. Soc. Am.* vol. 70, no. 1, pp. 45–50, 1981.

[2] H. Kawahara, "Effects of Natural Auditory Feedback on Fundamental Frequency Control," *Proc. 3th ICSLP,* Yokohama, pp. 1399–1402, 1994.

[3] H. Kawahara, H. Kato, and J. C. Williams, "Effects of Auditory Feedback on F0 Trajectory Generation," *Proc. 4th ICSLP*, Philadelphia, pp. 287–290, 1996.

[4] S. H. Chen, H. Liu, Y. Xu, and C. R. Larson, "Voice F0 responses to pitch-shifted voice feedback during English speech," *J. Acoust. Soc. Am.* vol. 121, no. 2, pp. 1157–1163, 2006.

[5] T. A. Burnett, M. B. Freedland, and C. R. Larson, "Voice F0 responses to manipulation in pitch feedback," *J. Acoust. Soc. Am.* vol. 103, no. 6, pp. 3153–3161, 1998.

[6] M. Honda, A. Fujino, and T. Kaburag, "Compensatory responses of articulators to unexpected perturbation of the palate shape," *J. Phonetics.* vol. 30, no. 3, pp. 281–302, 2002.

[7] S. Katseff, and J. F. Houde, "The role of auditory feedback in speech production," *Proc. 11th LabPhon,* Wellington, 2008.

[8] M. J. Hammer, and M. A. Krueger, "Voice-related modulation of mechanosensory detection thresholds in the human larynx," *J. Experimental Brain Research.* vol. 232, no. 1, pp. 13–20, 2014.

[9] X. Z. Wang, K. Honda, J. W. Dang and J. G. Wei, "Vocal responses to frequency modulated composite sinewaves via auditory and vibrotactile pathways," *Proc. 40th ICASSP 2015*, Brisbane, pp. 4355–4359, 2015.

[10] Y. Xu, C. R. Larson, J. Bauer, and T. C. Hain, "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," *J. Acoust. Soc.Am.* vol. 116, no.2, pp. 1168–1178, 2004.

[11] M. Sivasankar, J. J. Bauer, T. Babu, C. R. Larson, "Voice responses to changes in pitch of voice or tone auditory feedback," *J. Acoust. Soc.Am*. vol. 117, no. 2, pp. 850–857, 2005.

[12] C. R. Larson, "The role of auditory feedback for the control of voice fundamental frequency and amplitude," *SIG 5 Perspectives on Speech Science and Orofacial Disorders,* pp. 9–17, 2008.

[13] S. Patel, C. Nishimura, A. Lodhavia, O. Korzyukov, A. Parkinson, D. A. Robin, C. R. Larson, "Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback," *J. Acoust. Soc. Am.* vol. 135, no. 5, pp. 3036–3041, 2014.

[14] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," *J. Speech Language and Hearing Research.* vol. 14, no. 4, pp. 677–709, 1971.

[15] C. L. Ludlow, "Central nervous system control of the laryngeal muscles in humans," *J. Respir Physiol & Neurobiol,* vol. 147, no. 2-3, pp.205–222, 2005.

[16] C. R. Larson, K. W. Altman, H. J. Liu, and T. C. Hain, "Interactions between auditory and somatosensory feedback for voice F0 control," *J. Exp Brain Res,* vol. 187, no. 4, pp.613–621, 2008.