Principal component analysis reveals differential attentional modulation of the vocal response to pitch perturbation

Shuntaro Okazaki,[a] Koichi Mori,[b] and Minae Okada
Research Institute of National Rehabilitation Center for Persons with Disabilities, 4-1 Namiki, Tokorozawa, Saitama, 359-8555, Japan

(Received 12 March 2013; revised 26 April 2014; accepted 13 May 2014)

The auditory-vocal system modifies voice fundamental frequency (F0) with auditory feedback. The responses to F0 changes in auditory feedback are known to depend on the task. The hypothesis explored in this study is that the task dependency is the result of multiple components of the F0 responses differently modulated with different tasks. Attention to audition was manipulated by task condition by the instruction to ignore or to count the number of the F0 shifts heard during vocalization. A synthetic voice with pitch shifts was used as auditory pseudo-feedback. The upward and downward shifts evoked very similar vocal F0 response patterns with polarity reversal. Attention to the auditory feedback caused a reduction in the grand-average response amplitude. By decomposing the F0 responses with principal component analysis (PCA), three principal components (PCs) with different peak latencies were found to have contributions above the criterion of 5%, totaling to 74%. All three PCs contributed to a compensatory response under the "ignore" condition. The slowest PC changed its polarity and the intermediate PC was reduced to almost zero under the "count" condition. Thus, the task-dependency of the F0 response to auditory feedback can be described in terms of different sensitivities of components to attention. © *2014 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4881921]

PACS number(s): 43.70.Aj, 43.66.Hg, 43.70.Bk, 43.70.Mn [CHS] Pages: 334–340

I. INTRODUCTION

Auditory feedback is important for the control of vocal loudness and fundamental frequency (F0) (Jones and Munhall, 2000). When an upward pitch shift is introduced via auditory feedback during production of a sustained vowel, speaking, or singing, the auditory-vocal system responds with a downward shift of F0 to compensate for it, and vice versa (e.g., Elman, 1981; Kawahara, 1993; Burnett *et al.*, 1998; Hain *et al.*, 2000; Burnett and Larson, 2002; Larson *et al.*, 2007). This compensatory response is considered to be reflexive and is often termed a "pitch-shift-reflex (PSR)" (e.g., Burnett *et al.*, 1998; Burnett and Larson, 2002; Chen *et al.*, 2007; Hain *et al.*, 2000; Kiran and Larson, 2001; Sivasankar *et al.*, 2005) because the latency is between 100 and 200 ms after the onset of the introduced pitch shift. The PSR serves to stabilize voice F0 by counteracting unintended F0 changes in the vocalization (Burnett and Larson, 2002).

Although the PSR has been considered to be an automatic and reflexive response, it has been argued that the amplitude of the PSR can be modulated depending on the vocal task. The PSR was shown to be enhanced during speech (Chen *et al.*, 2007; Xu *et al.*, 2004) and singing (Natke *et al.*, 2003) compared with that during sustained vowel vocalization. On the other hand, the PSR during dynamic voice pitch control was found to be smaller than that during sustained vowel vocalization (Burnett and Larson, 2002) although this is not compatible with the previous finding of Mandarin tone sequence production (Xu *et al.*, 2004). Similarly, pitch perturbations presented just prior to a planned pitch increment induced a smaller PSR than when it was presented earlier or when there was no planned F0 change (Liu *et al.*, 2009).

Previous studies have also proposed two or more complex processes in auditory-vocal F0 control. In contrast to the compensating PSR, a following F0 response (in which F0 changes in the same direction as the stimulus) was observed in approximately 15% of the trials (Larson *et al.*, 2007), and even more frequently when the pitch shift was larger (300 cents, Burnett *et al.*, 1998) or when the timing of the pitch-shift stimulus was anticipated ("ideomotor response," Burnett *et al.*, 2008). Furthermore, a second response with a latency of approximately 300 ms was observed in addition to, or as a part of, the PSR (Kawahara 1993; Burnett *et al.*, 1998; Hain *et al.*, 2000). The second response occurred more frequently when the duration of the pitch shift stimulus was greater than 100 ms (Burnett *et al.*, 1998). Hain *et al.* (2000) have proposed a conceptual model that explains the compensating and tracking (following) modes of voice F0 control. When subjects are told to ignore an external input, the compensating mode is facilitated. On the other hand, when they are told to respond to the input, the tracking mode is facilitated. However, the effects of the intention to respond voluntarily and the attention to the external stimuli on the F0 responses are not fully elucidated.

Although most of the previous studies have tried to isolate the PSR from other components of the responses by using pitch-shift stimuli with a relatively short duration (e.g., 100 ms in Burnett *et al.*, 2008), the phasic property of the PSR and the second response may be shaped by the response reaction in a closed-loop auditory-vocal system (cf. Leydon

[a]Current affiliation: Division of Cerebral Integration, Department of Cerebral Research, National Institute for Physiological Sciences, 38, Nishigonaka, Myodaiji, Okazaki, Aichi 444-8585, Japan.
[b]Author to whom correspondence should be addressed. Electronic mail: mori-koichi@rehab.go.jp

 0001-4966/2014/136(1)/334/7/$30.00

et al., 2003). The complex responses in the auditory-vocal system could only be fully characterized in an open-loop measurement system. Hain *et al.* (2001) have proposed that the duration of the pitch shift responses lengthens by up to several hundred milliseconds in the open-loop auditory-vocal system, which may suggest that the multiple components of the pitch shift response are temporally overlapped and thus cannot be separated from each other even in an open-loop system. We hypothesized that the more voluntary components of the response to the pitch shift may be distinct from the PSR, and modulated by the attention to the external stimuli. Since the two modes of responses (Hain *et al.*, 2000) are stochastic rather than deterministic (Larson *et al.*, 2007; Burnett *et al.*, 1998), we tested these hypotheses by decomposing the pitch shift response based on the variability of the response across conditions (with or without pitch shift and ignoring or attending), trials, and participants using principal component analysis (PCA).

II. MATERIALS AND METHODS

A. Participants

Ten healthy adults (2 males, 8 females, age 22–29) with no history of neurological or speech disorders participated. None of the participants had had any formal vocal training or experience as a singer (by self-report). All participants signed informed consent approved by the Internal Ethical Review Board of National Rehabilitation Center for Persons with Disabilities.

B. Apparatus

Participants were seated in a sound-attenuated chamber with a headset (HMD 25-1, Sennheiser Electronic KG, Germany) placed over the head (Fig. 1). The microphone was located approximately 3 cm from the lips. The nominal frequency response ranges of the headphones and the microphone were 20–16 000 Hz and 100–16 000 Hz, respectively. A monitor (LP2275w, Hewlett-Packard, Palo Alto, CA) was placed in front of the participant. A vowel /a/ was made from a "small choir" of a GM2 instrument sound sources (General MIDI System Level 2, Roland, Osaka, Japan). First, the "small choir" was played for 5 s from the GM2 instrument at C3 for male and at C4 for female participants, respectively and was recorded at a sampling rate of 48 kHz. Second, the F0 contours of the recorded sounds were replaced with those of a constant 120 Hz for males and 240 Hz for females by the "Time-Domain Pitch-Synchronous Overlap-and-Add" (TD-PSOLA) method (Moulines and Charpentier, 1990) in Praat software (Boersma, 2001). This method changes F0 without significantly altering the overall spectral shape, or formant information, of the original vowel (see spectra in Fig. 2). For pitch-shift stimuli, the F0 contours additionally included intermittent upward and downward shifts as described in the next section. The auditory stimuli were delivered at a sampling rate of 48 kHz. To measure the open-loop characteristics of the voice F0 control, the synthesized vowel sound was presented as a pseudo-feedback voice at 82 dB sound pressure level (SPL), mixed with white noise at 80 dB SPL from a

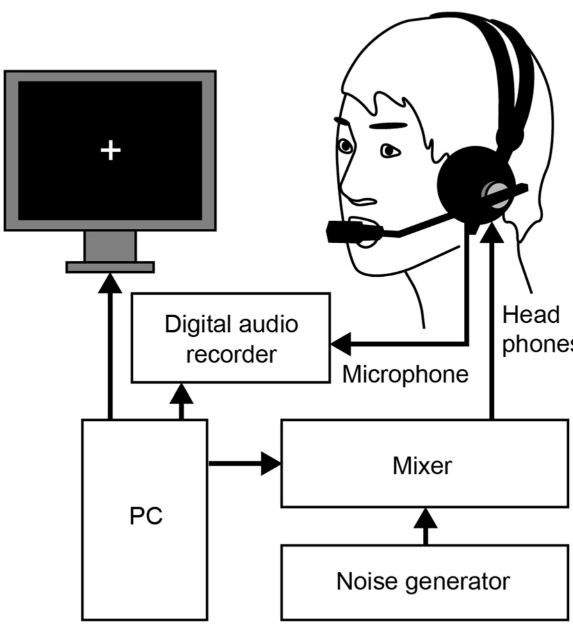

FIG. 1. Schematic diagram of the experimental system. Participants were sitting in front of the monitor and wore a headset. Synthesized voice, mixed with masking noise, was presented through the headphones. The uttered voices and unmixed pseudo-feedback were recorded on a digital recorder.

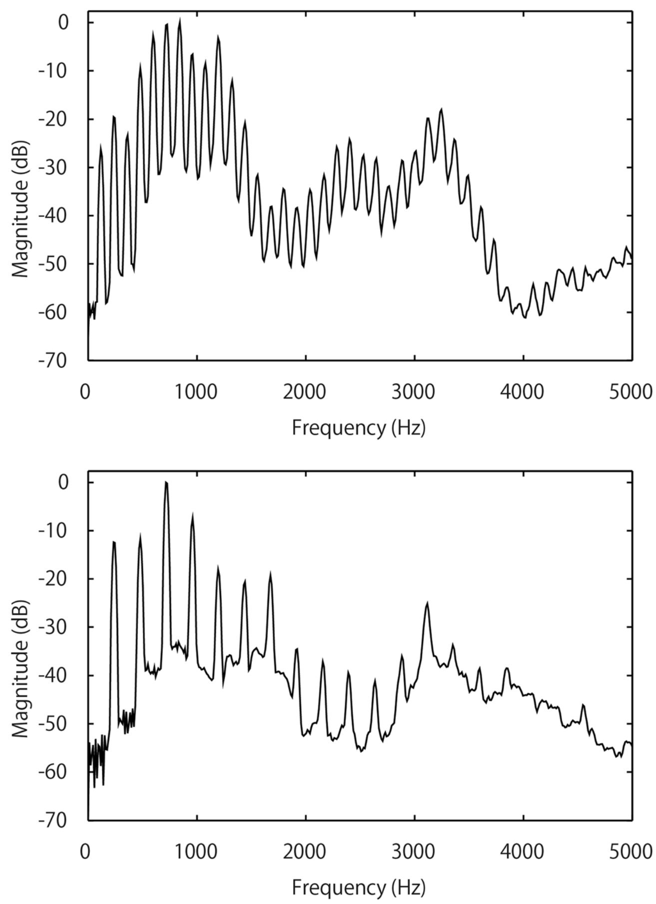

FIG. 2. Spectrum density of the pseudo-feedback sound stimuli for males (upper panel) and females (lower panel). The spectra are computed with Welch's periodogram method (analysis window: 50 ms) after pre-emphasis of high frequencies with a slope of 6 dB/oct. The fundamental frequencies are transposed with the TD-PSOLA method from C3 to 120 Hz for males and from C4 to 240 Hz for females.

noise generator (type 1405, Brüel & Kjær, Nærum, Denmark), as shown in Fig. 1. The white noise was presented to mask the subject's air- and bone-conducted voice. The subject's voice itself was not played back into the headphones. Acoustic level calibration was made with a sound level meter (type 1405, Brüel & Kjær, Nærum, Denmark) and a probe microphone (ER-7C, Etymotic Research, Elkgrove Village, IL). The experiment was run using Presentation software (Neurobehavioral Systems, Berkeley, CA), and both auditory and visual stimuli were presented by a computer (HPZ200 Workstation, DELL, Austin, TX). Participants' utterances and unmixed pseudo-feedback were recorded on a digital recorder (EDIROL R-4, Roland, Osaka, Japan) for off-line F0 analysis.

C. Experimental task

The experiment was performed in a sound attenuated room and the participant was cued by visual and auditory stimuli. In each trial, a visual red "+" cue (500 ms in duration) centered in the display and a diotic auditory cue of the synthesized voice /a/ (500 ms in duration) were synchronously presented twice with an interval of 0.5 s. At 2 s after the onset of the first cue, a visual green "+" cue of 5 s appeared where the red cue had been presented. Participants were instructed to vocalize the vowel /a/ at the same pitch as the auditory cue. They were to start vocalizing at the onset of the visual green "+" cue and to maintain it as long as the green cue was on. Concurrently with the green cue, a diotic pseudo-feedback voice was presented with the noise that masked the subject's voice. When a visual blue "+" cue (500 ms in duration) replaced the green cue, the pseudo-feedback voice was terminated and participants were to stop the vocalization. The next trial began 9 s after the onset of the previous trial (Fig. 3).

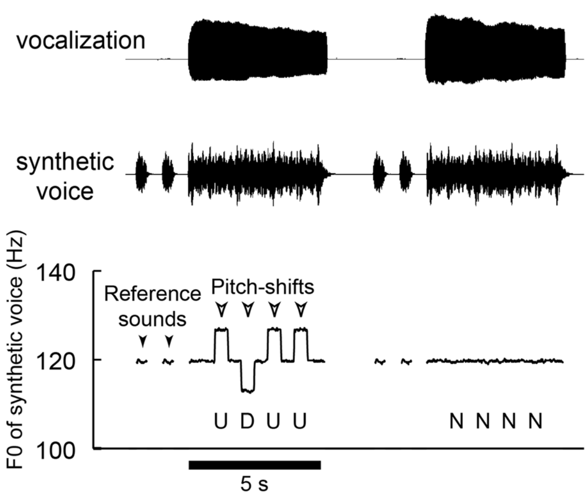

FIG. 3. Time chart of the experiment. The envelopes of the recorded vocalization (upper trace), and the reference and pitch-shifted pseudo-feedback (second trace) are shown. The F0 contours of the stimulation are shown in the bottom graph. Its left half shows the F0 trace of an example pitch-shift trial with upward (U) and downward (D) pitch shifts in a random order with a stimulus-onset asynchrony of 1 s. The right half shows that of a control trial without pitch-shift (N).

The pitch of the synthesized pseudo-feedback voice was shifted upward (U) or downward (D) four times in a random order during each vocalization trial. The pitch shift started 1 s after the onset of pseudo-feedback. The duration of each pitch-shift was 500 ms (Fig. 3, left half). The amount of the pitch-shift was 1 semitone (100 cents), separated from the next by an interval of 0.5 s. One in every three trials had no pitch-shift (N) added to the stimulus voice (Fig. 3, right half).

Participants performed two experimental tasks. In the first session, they were instructed to ignore the feedback and to keep their pitch constant ("ignore" condition). In the second task, they were instructed to keep their pitch constant and, at the same time, to count the numbers of upward and downward pitch-shifts in the pseudo-feedback voice, and to press one of three keys on a keyboard ("count" condition). They were required to press (1) the right arrow key when they heard more upward than downward pitch-shifts, (2) the left arrow key when they heard more downward than upward pitch-shifts, and (3) the downward arrow key when the upward and downward pitch-shifts were heard an equal number of times or there was no pitch-shift. Before each experimental session, participants were trained to vocalize together with synthesized pseudo-feedback voice, and, only before the second session, to count the number of pitch shifts in the same manner as during the experiment, using a shorter version of the experimental procedure.

D. Method of analysis

The F0 contours of both recorded utterances and pseudo-feedback voices were extracted using an analysis window of 25 ms with a step size of 10 ms using WaveSurfer software (Sjölander and Beskow, 2000). The segments from 100 ms before to 690 ms after each of the four onsets of the pitch-shifts in a given trial were subjected to analysis. The F0 contours were converted to cent values of a half-tone interval relative to the F0 mean in the 100-ms pre-stimulus period. After this processing, the F0 responses to downward pitch-shift were inverted to be compared to those to the upward shift in terms of the response amplitude: The time-series data of F0 responses were averaged across 64 trials for each stimulus direction per each task condition of attention in each subject. The averaged time-series data were analyzed by a three-way repeated measures analysis of variance (ANOVA) [Time (80 time points from −100 ms to 690 ms relative to the stimulus pitch shift onset) × stimulus directions (upward or downward) × task conditions of attention ("ignore" or "count")]. Since this analysis did not reveal a significant difference in the response amplitude relative to the pre-stimulus average between the opposite stimulus directions (see Sec. III), a two-way repeated measures ANOVA (Time × task conditions of attention) was performed on the dataset of the averaged upward and downward shift responses. *Post hoc* multiple comparisons with a Bonferroni correction were performed at the corresponding time points across the attentional conditions (80 time points in total, aligned to the stimulus pitch shift onset timing) to address the time profile of the attention effect on the F0 response.

Finally, the F0 responses of all participants in all trials under all conditions were analyzed using Principal Component Analysis (PCA, Jolliffe, 2002) (Fig. 5). The F0 responses to no pitch-shift were also included in the PCA analysis, which were the F0 contours during the corresponding periods of the analysis for the pitch shift responses, i.e., the segments of 900–1700 ms, 1900–2700 ms, 2900–3700 ms, and 3900–4700 ms from the onset of the visual green "+" cue. The PCA is hypothesized to reveal the differential effects of attention on the orthogonal response components in the auditory-vocal system. The factor loadings for each principal component (PC) whose contribution ratio exceeded 5% was subjected to further statistical analysis in order to determine the effects of the attention ("ignore" and "count"). Since the polarity of PCA components is arbitrary, we chose the polarity so that the earliest deflection after time 0 (stimulus pitch shift onset timing) was positive-going for each component. The factor loadings for each experimental condition (with and without pitch shift in "ignore" condition, and with and without pitch shift in "count" condition) were statistically evaluated against zero by a one-sample *t*-test with a Bonferroni correction (N = 4).

III. RESULTS

In both "ignore" and "count" tasks (48 trials in each), participants vocalized a sustained vowel /a/ while hearing the auditory pseudo-feedback with four pitch shifts in two thirds of the trials (32 of 48 trials, upward and downward pitch shifts were presented with the same probability) and with no pitch shift in the remaining 16 trials. Thus, we obtained and analyzed 3840 responses (10 participants × 2 task conditions of attention × 3 stimulus directions (upward, downward, and no pitch shift) × 16 trials × 4 repetitions). Figure 4 shows the respective averaged F0 responses of all participants to the upward (upper left) and downward pitch shifts (lower left). The pitch difference between the voice and the reference sound (at 120 Hz for males and 240 Hz for females) during the 100 ms window before the first pitch shift stimulus was 2.1 ± 1.1 Hz (average ± standard deviation). To apply an ANOVA, we first calculated the mean values of the F0 responses for each subject. The mean F0 responses to the upward and downward pitch-shift stimuli, with the responses to the latter reversed in polarity, were compared by a three-way repeated measures ANOVA (80 time points × 2 stimulus directions × 2 task conditions of attention). We found a significant main effect of the attentional condition [$F(1,9) = 23.495$, $p = 0.001$], but not of time [$F(79\,711) = 1.796$, $\varepsilon_{GG} = 0.019$, $p = 0.205$] or stimulus direction [$F(1,9) = 0.248$, $p = 0.630$]. This indicates that the F0 responses to the upward and downward frequency shifts were different in direction but not in amplitude or time course, nor did they show an interaction with the two experimental conditions. Consequently, we averaged the responses to upward and downward stimuli (Fig. 4, right) and analyzed them with a two-way repeated measures ANOVA (2 task conditions of attention × 2 stimulus directions × 80 time points). As expected, the main effect of the attentional condition [$F(1,9) = 30.218$, $p < 0.001$] and the interaction between the attentional condition and the time [$F(79,1501) = 16.660$, $p < 0.001$] were significant. The *post hoc* analysis of the F0 at each time point revealed that the F0 response was significantly decreased by attention manipulation in the latency range between 180 and 470 ms after the stimulus onset ($p < 0.05$, corrected by a Bonferroni method, thick bar in the right figure in Fig. 4).

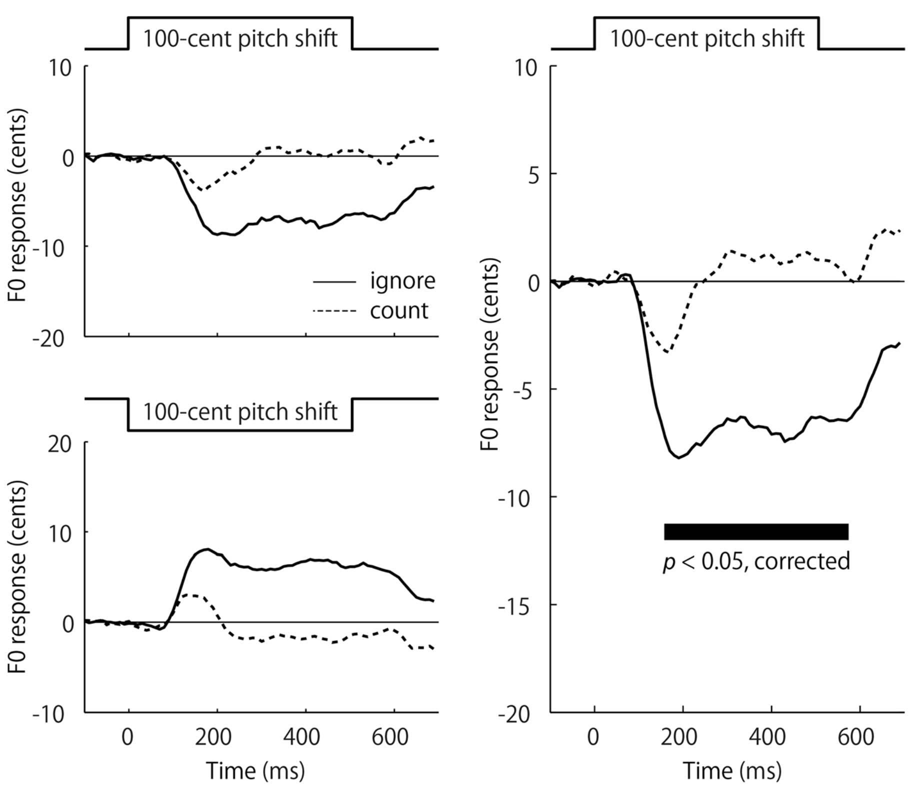

FIG. 4. Grand average F0 responses across subjects to the upward and downward pitch-shifted pseudo-feedback under "ignore" and "count" conditions. The left upper and lower panels show the F0 responses to upward and downward pitch shifts, respectively. The right panel shows their averages, with those to the downward pitch shifts reversed in polarity. Solid and dashed lines indicate the F0 responses under "ignore" and "count" conditions, respectively. The segments of F0 responses significantly different between the two conditions are marked by a thick line ($p < 0.05$ corrected by a Bonferroni procedure).

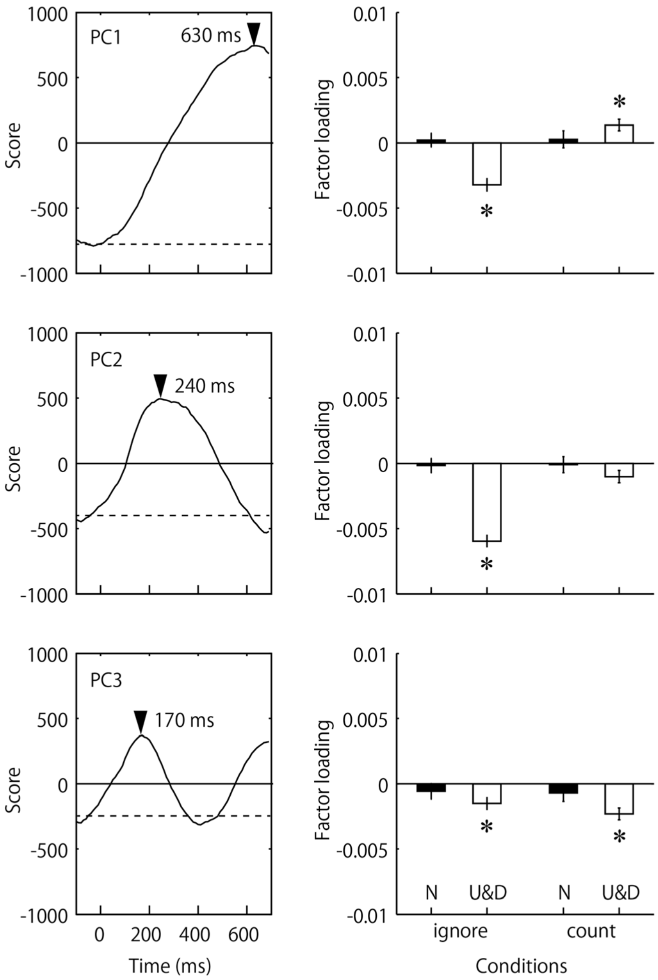

FIG. 5. The three largest principal components and factor loadings. The three largest principal components (PC1, PC2, and PC3, left panels) of the F0 response to the pseudo-feedback with and without pitch shifts showed different loadings depending on the task and pitch-shift conditions (right panels). Horizontal broken lines indicate the averaged pre-stimulus levels of the respective components. Arrowheads indicate the peak and the latency of the principal components. Black and white bars indicate factor loadings in the F0 responses to pitch shift (U&D) and those to no pitch shift (N), respectively. Asterisks indicate significant factor loadings against 0 by one-sample t-test ($p < 0.05$).

The averaged F0 response in the "ignore" condition showed a peak of 8.2 cents at 190 ms and a plateau afterward that ended more than 100 ms after the stimulus offset. On the other hand, the averaged F0 response in the "count" condition had two phases with the transient peak of 3.2 cents at 170 ms, and a smaller and slower following response peaking at 300 ms after the stimulus onset. All of these peaks are within the period of the significant decrease of the F0 response with attention ("count" task, Fig. 4, right). However, the two peaks, or local minima, are both discernible in the "ignore" and in the "count" conditions with different amplitudes, which suggests that similar sub-processes but with different compositional weights may underlie the condition-dependent results. Since the slopes of the two peaks overlapped, we next tried to decompose the F0 responses using PCA.

All the responses obtained in this study were subjected to principal component analysis. Three orthogonal principal components (factors) that exceeded 5% in contribution were adopted for further evaluation (Fig. 5, left column, PC1, PC2, PC3). The integrated contribution of the three factors was 74% (48%, 18%, and 8%, respectively). Since no averaged factor loadings of these components for the no pitch-shift control were different from zero, all three factors are considered as the constituents of the pitch shift response. The first factor (PC1) was the slowest response peaking at 630 ms from the stimulus onset. Its factor loadings were evaluated across two conditions of attention and with or without pitch shift stimuli by Bonferroni-corrected one-sample t-test (Fig. 5). The multiple comparisons showed a significant factor loading opposite to the stimulus pitch-shift (i.e., "compensatory") in the "ignore" condition, and a significant factor loading to the same direction as the stimuli (i.e., "following") in the "count" condition ($p < 0.05$, corrected). The second principal component (PC2) indicated a faster response with the peak latency of 240 ms. In contrast to the PC1, a significant factor loading, which was in the opposite direction to the pitch-shift stimulation, was found only in the "ignore" condition but not in the "count" condition ($p < 0.05$, corrected). The third principal component (PC3) was the fastest of the three, with peak latency of 170 ms. The PC3 exhibited significant opposing response to the stimulation in both "ignore" and "count" conditions ($p < 0.05$, corrected). The F0 responses were reconstructed only from the major three PC's multiplied by the respective factor loadings, with the responses during the baseline period (-100 to 0 ms) zeroed (Fig. 6).

IV. DISCUSSIONS

In this study, we investigated the open-loop characteristics of the auditory-vocal F0 control using F0 step stimuli of a synthetic voice as auditory pseudo-feedback, and quantitatively evaluated the effects of auditory attention on the control. The auditory synthetic voice stimuli with pitch shifts

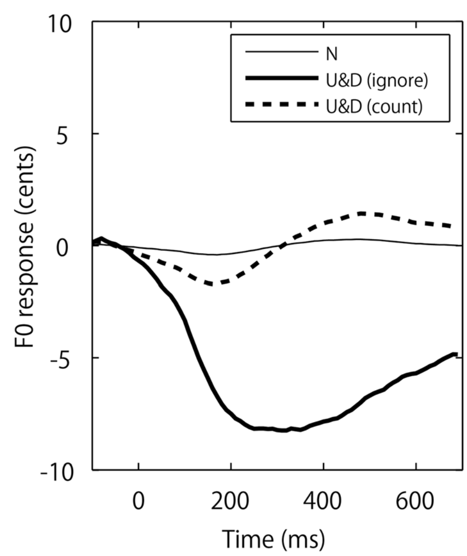

FIG. 6. The reconstructed F0 responses from the major principal components. The F0 response is reconstructed for each condition of pitch shift and attention from the three largest principal components (PC1, PC2, and PC3) whose contributions were larger than 5%. Thin, thick, and dashed lines indicate the reconstructed F0 responses to no pitch shift (N), those to the ignored pitch shift [U&D (ignore)] and those to the counted pitch shift [U&D (count)], respectively.

induced pitch shift responses in the vocalization of the listener, just as tonal stimuli do (Lee *et al.*, 2008; Sivasankar *et al.*, 2005). As proposed in a previous study (Sivasankar *et al.*, 2005), the amplitude of the F0 response in this study (8.3 cents under the "ignore" condition in the grand average) may be slightly smaller than that of the real auditory feedback of one's own voice.

Although PCA has been extensively applied to the time-domain data, such as EEG signals (e.g., Donchin and Heffley, 1979), the physiological and functional implications of its results should be carefully interpreted. In this study, we observed (1) that 74% of total variance was explained using only three components, (2) that the three components (PC1, PC2, and PC3) have positive earliest peaks at 630, 240, and 170 ms, respectively, and (3) that the value difference of each component's loading factor between different tasks and conditions was statistically significant. Experimental manipulation of the pitch shifts and the attention to the stimuli seemed to be suitable for the PCA decomposition in the present study. We observed a rapid component in the pitch-shift responses with earlier peak latencies of 170 ms (for the grand average F0 response in "count" condition and for the third principal component) and 190 ms (for the grand average F0 response in "ignore" condition), which are comparable to the previously proposed PSR (e.g., Burnett *et al.*, 1998; Burnett and Larson, 2002; Chen *et al.*, 2007; Hain *et al.*, 2000; Kiran and Larson, 2001; Sivasankar *et al.*, 2005). The grand average F0 responses (Fig. 4) showed a plateau that mirrored the stimulus step in the "ignore" condition, indicating an inverting low-pass filter characteristic of the auditory-vocal F0 control. Hain *et al.* (2001) examined the open-loop characteristics of auditory-vocal system by inserting a delay into the feedback loop and reported a similar plateau characteristic of the PSR. A previous study has proposed a second response with a longer peak latency and greater magnitude that temporally overlaps the first PSR (Burnett *et al.*, 1998). Although these responses have been hard to isolate from each other, we have been able to separate them using PCA, and we have shown the individually different effects that attention has on them.

Since the earliest response component with a peak latency less than 200 ms (PC3) was opposite both to the unattended and attended pitch shifts, it can be interpreted to reflect a pre-attentive process, and consistent with the PSR that is always produced in the opposing direction to the pitch shift stimulus, regardless of how accurately an individual overtly identifies the pitch shift's direction (Burnett and Madonia, 2006; Burnett *et al.*, 2008). As previous studies have proposed that the PSR attenuates differently depending on task (Burnett and Larson, 2002; Liu *et al.*, 2009), the earliest peak of the F0 response in this study was also attenuated by attending to the pitch shift stimuli as shown in Figs. 4 and 6. However, this phenomenon could be analytically explained by the different combinations of the rapid and pre-attentive response component (PC3) and the slower but overlapping, attention-dependent components (PC1 and PC2) in this study.

The previously proposed second response with the peak latency of approximately 300 ms (Kawahara, 1993; Burnett *et al.*, 1998; Hain *et al.*, 2000) is apparently compatible with the second principal component of the pitch shift response in this study. However, this response has also been reported to occasionally follow the pitch shift stimuli (Burnett *et al.*, 1998; Burnett *et al.*, 2008; Larson *et al.*, 2007), which could be explained by the polarity reversal of the loading of the first principal component and the significant reduction of the loading of the second component of the response in the opposite direction to the stimulation in the attended condition. Actually, a small following response was observed in the grand average as well (Fig. 4) under the "count" condition in the present study.

As demonstrated in the present study, the amplitude of the PSR has been modulated depending on the vocal task (Burnett and Larson, 2002; Chen *et al.*, 2007; Donath *et al.*, 2002; Liu *et al.*, 2009; Natke *et al.*, 2003; Xu *et al.*, 2004), although there were some conflicting results in the previous studies (i.e., amplitude of the PSR in sustained vowel phonation vs other tasks). The previous findings that the PSR amplitude was decreased during the dynamical vocal control (Burnett and Larson, 2002; Liu *et al.*, 2009) can be partly explained by the higher demand of attention to the auditory feedback in these tasks. The other findings about PSR in speaking and singing (Chen *et al.*, 2007; Natke *et al.*, 2003; Xu *et al.*, 2004) could not be explained solely by the attention to the auditory feedback. Some other factors, such as the F0 dependent characteristics of the PSR (Liu and Larson, 2007; Okazaki *et al.*, 2009), should be considered as well. In the present study, the slow response component to the pitch shift stimuli (PC2) was attenuated and the slowest component (PC1) exhibited following-like characteristics. The attention-dependent modulation of these responses is compatible with the previously proposed model of the compensating and tracking (following) modes of the voice F0 control (Hain *et al.*, 2000).

The findings of the neural substrate for the pitch shift response have suggested that the auditory cortex is involved (Eliades and Wang, 2008; Hain *et al.*, 2000; Masuda *et al.*, 2008; Tourville *et al.*, 2008; Toyomura *et al.*, 2007). However, the relationships between the variable characteristics of pitch shift responses and the auditory cortical processing still remain unclear. In neurophysiological (Creutzfeldt *et al.*, 1989) and MEG (Hari *et al.*, 1989; Numminen *et al.*, 1999; Curio *et al.*, 2000; Houde *et al.*, 2002) studies, the neuronal response to one's own voice in the temporal lobe (or auditory cortex) was significantly lower during its production than that during listening without speaking. In addition, by counting heard words, auditory attention is induced, which in turn enhances the activity in the auditory cortex (Hari, 1991; Grady *et al.*, 1997). Taken together with previous findings, the slower components (PC1 and PC2) rather than the pre-attentive component (PC3) in the present study may well be regulated in the auditory cortex. Although it is not clear whether attention-induced higher activation in the auditory cortex attenuates the pitch shift response or induces the following type of pitch shift responses, the finding that the ERP components in response to feedback perturbation were significantly larger during vocalization than listening (Behroozmand *et al.*, 2009), lends more support for the latter. The following response in the auditory feedback control may contribute to the speech coordination and auditory attention in aligning one's own voice with the external one, possibly through a mirror neuron system (Kalinowski and Saltuklaroglu, 2003; Garrod and Pickering, 2009).

V. CONCLUSIONS

Using PCA, we have quantitatively and objectively identified the attention independent PSR, and two types of more complex response components that were modulated by auditory attention. Particularly, one type of the complex components (the first principal component) had a long latency and could explain the previously proposed following response under an attended condition. The other response (the second principal component) had a moderate latency, overlapped the fast PSR component in time, and could explain the task dependent PSR attenuation by auditory attention. These findings provide a collective view for the variable response components in the auditory feedback voice F0 control. PCA is an essential method of analysis for extracting and understanding the nature of the individual components under different attentional conditions.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI 22700569 to S.O., 19591999 and 23592543 to K.M., and 2370330 to M.O.

Behroozmand, R., Karvelis, L., Liu, H., and Larson, C. R. (**2009**). "Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation," Clin. Neurophys. **120**, 1303–1312.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**, 341–345.

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.

Burnett, T. A., and Madonia, D. A. (**2006**). "Voice F0 responses to actual versus consciously perceived auditory feedback pitch shifts," in *Meeting of the Acoustical Society of America*, Honolulu, HI, p. 3090.

Burnett, T. A., McCurdy, K. E., and Bright, J. C. (**2008**). "Reflexive and volitional voice fundamental frequency responses to an anticipated feedback pitch error," Exp. Brain Res. **191**, 341–351.

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.

Creutzfeldt, O., Ojemann, G., and Lettich, E. (**1989**). "Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice," Exp. Brain Res. **77**, 476–489.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., and Hari, R. (**2000**). "Speaking modifies voice-evoked activity in the human auditory cortex," Hum. Brain Mapp. **9**, 183–191.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice F[sub 0] contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Donchin, E., and Heffley, E. (**1979**). "Multivariate analysis of event-related potential data: A tutorial review," in *Multidisciplinary Perspectives in Event-Related Potential Research*, EPA600/9-77-043, edited by D. Otto (U.S. Government Printing Office, Washington, DC), pp. 555–572.

Eliades, S. J., and Wang, X. (**2008**). "Neural substrates of vocalization feedback monitoring in primate auditory cortex," Nature **453**, 1102–1106.

Elman, J. L. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**, 45–50.

Garrod, S., and Pickering, M. J. (**2009**). "Joint action, interactive alignment, and dialog," Topics Cognit. Sci. **1**, 292–304.

Grady, C. L., Meter, J. W. Van, Maisog, J. M., Pietrini, P., Krasuski, J., and Rauschecker, J. P. (**1997**). "Attention-related modulation of activity in primary and secondary auditory cortex," Neuroreport **8**, 2511–2516.

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.

Hari, R. (**1991**). "Activation of the human auditory cortex by speech sounds," Acta Otolaryngologica Suppl. **491**, 132–137.

Hari, R., Hämäläinen, M., Kaukoranta, E., Mäkelä, S.-L., and Tiihonen, J. (**1989**). "Selective listening modifies activity of the human auditory cortex," Exp. Brain Res. **74**, 463–470.

Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (**2002**). "Modulation of the auditory cortex during speech: An MEG study," J. Cognit. Neurosci. **14**, 1125–1138.

Jolliffe, I. T. (**2002**). *Principal Component Analysis, Encyclopedia of Statistics in Behavioral Science*, 2nd ed., edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg, I. Olkin, N. Wermuth, and S. Zeger (Springer-Verlag, New York), Vol. 30, p. 487.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Kalinowski, J., and Saltuklaroglu, T. (**2003**). "Speaking with a mirror: Engagement of mirror neurons via choral speech and its derivatives induces stuttering inhibition," Med. Hypotheses **60**, 538–543.

Kawahara, H. (**1993**). "Transformed auditory feedback: Effects of fundamental frequency perturbation," J. Acoust. Soc. Am. **94**, 1883–1884.

Kiran, S., and Larson, C. R. (**2001**). "Effect of duration of pitch-shifted feedback on vocal responses in patients with Parkinson's disease," J. Speech Language Hearing Res. **44**, 975–987.

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.

Lee, S.-H., Lee, K.-C., Ban, J.-H., Lee, N.-H., and Jin, S.-M. (**2008**). "Vocal analysis related to changes in frequency of pure tone auditory feedback," Yonsei Med. J. **49**, 610–614.

Leydon, C., Bauer, J. J., and Larson, C. R. (**2003**). "The role of auditory feedback in sustaining vocal vibrato," J. Acoust. Soc. Am. **114**, 1575–1581.

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.

Masuda, S., Okazaki, S., Mizokami, N., Kamatani, D., Mori, K., and Nakajima, Y. (**2008**). "The neural basis subserving unconscious pitch control revealed by transformed auditory feedback: A magnetic resonance imaging study," report, Society for Neuroscience, Washington, DC.

Moulines, E., and Charpenticr, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Communication. **9**, 453–467.

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.

Numminen, J., Salmelin, R., and Hari, R. (**1999**). "Subject's own speech reduces reactivity of the human auditory cortex," Neurosci. Lett. **265**, 119–122.

Okazaki, S., Mori, K., and Cai, C. (**2009**). "Effect of the fundamental frequency and vocal register on the voice pitch compensation," in *Meeting of the Acoustical Society of America*, San Antonio, TX, p. 2242.

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.

Sjölander K., and Beskow J. (**2000**). "WaveSurfer—An open source speech tool," ICSLP 2000 **4**, 464–467. Available at: http://www.speech.kth.se/wavesurfer/.

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (**2008**). "Neural mechanisms underlying auditory feedback control of speech," NeuroImage **39**, 1429–1443.

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., and Kuriki, S. (**2007**). "Neural correlates of auditory feedback control in human," Neuroscience **146**, 499–503.

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