Time-dependent Neural Processing of Auditory Feedback during Voice Pitch Error Detection

Roozbeh Behroozmand[1], Hanjun Liu[2], and Charles R. Larson[1]

Abstract

■ The neural responses to sensory consequences of a self-produced motor act are suppressed compared with those in response to a similar but externally generated stimulus. Previous studies in the somatosensory and auditory systems have shown that the motor-induced suppression of the sensory mechanisms is sensitive to delays between the motor act and the onset of the stimulus. The present study investigated time-dependent neural processing of auditory feedback in response to self-produced vocalizations. ERPs were recorded in response to normal and pitch-shifted voice auditory feedback during active vocalization and passive listening to the playback of the same vocalizations. The pitch-shifted stimulus was delivered to the subjects' auditory feedback after a randomly chosen time delay between the vocal onset and the stimulus presentation. Results showed that the neural responses to delayed feedback perturbations were significantly larger than those in response to the pitch-shifted stimulus occurring at vocal onset. Active vocalization was shown to enhance neural responsiveness to feedback alterations only for nonzero delays compared with passive listening to the playback. These findings indicated that the neural mechanisms of auditory feedback processing are sensitive to timing between the vocal motor commands and the incoming auditory feedback. Time-dependent neural processing of auditory feedback may be an important feature of the audio-vocal integration system that helps to improve the feedback-based monitoring and control of voice structure through vocal error detection and correction. ■

INTRODUCTION

Studies of the neural mechanisms underlying voice control in humans and animals have led to the identification of key components that are necessary for efficient vocal communication. It is now known that the robust control of voice is carried out through the integration of the sensory-motor systems that allows for the feedback-based monitoring and control of the self-produced voices. The enhanced ability of the system for voice control is thought to be driven by the vocal motor system that issues an efference copy of the motor commands (corollary discharges) to predict the sensory consequences associated with the intended vocal output (Guenther, 2006). The comparison between the predicted and the incoming sensory feedback input from self-vocalizations enables the system to detect and correct for unexpected alterations in feedback. Although the role of the sensory-motor integration has been studied during vocal production and control, its underlying neural mechanisms remain to be elucidated.

The interaction between the vocal motor and the auditory feedback mechanisms is among the most widely studied aspects of the sensory-motor integration for vocal production and control. Several studies have provided supporting evidence for the fact that applying auditory feedback perturbations to voice fundamental frequency (F0) (Chen, Liu, Xu, & Larson, 2007; Sivasankar, Bauer, Babu, & Larson, 2005; Xu, Larson, Bauer, & Hain, 2004; Bauer &

Larson, 2003; Burnett & Larson, 2002; Donath, Natke, & Kalveram, 2002; Jones & Munhall, 2002; Natke & Kalveram, 2001; Hain et al., 2000; Burnett, Freedland, Larson, & Hain, 1998; Larson, 1998; Burnett, Senner, & Larson, 1997; Elman, 1981), formant frequencies (Villacorta, Perkell, & Guenther, 2007; Purcell & Munhall, 2006a, 2006b; Houde & Jordan, 1998, 2002), or intensity (Liu, Zhang, Xu, & Larson, 2007; Bauer, Mittal, Larson, & Hain, 2006; Heinks-Maldonado & Houde, 2005) elicits compensatory vocal responses that tend to maintain the structure of the voice against the disruptive effect of feedback alterations. The prolonged exposure to the altered auditory feedback has also been shown to result in adaptation, meaning that the output of the vocal-motor system gradually adjusts to the altered feedback and remains so for some period after the alteration is removed (Villacorta et al., 2007). The adaptation is hypothesized to occur by a mechanism that tends to minimize mismatch (error) between the internal representation of mappings between the vocal motor output and its sensory feedback.

Another aspect of audio-vocal integration has been investigated in studies in which it has been shown that the activity of the vocal motor system may modulate the processing of auditory feedback during self-vocalization. The evidence for this effect comes from early studies in humans showing that ERP components elicited by self-triggered tones were substantially smaller than those elicited by machine-triggered stimuli (Martikainen, Kaneko, & Hari, 2005; McCarthy & Donchin, 1976; Schafer & Marcus, 1973). Studies in animals such as nonhuman primates have shown that the cortical auditory neurons are suppressed in

[1]Northwestern University, [2]Sun Yat-sen University, Guangzhou, P.R. China

response to electrically stimulated (Muller-Preuss & Ploog, 1981; Müller-Preuss, Newman, & Jürgens, 1980) or self-generated (Eliades & Wang, 2003, 2005) vocalizations. Work on crickets has also shown that neural responses to the auditory feedback of self-generated sounds are suppressed when the animal engages its singing motor network (Poulet & Hedwig, 2002, 2006, 2007). Intracranial recordings from the temporal lobe (Creutzfeldt, Ojemann, & Lettich, 1989) and scalp-recorded auditory-evoked responses using EEG and MEG (Ford, Gray, Faustman, Roach, & Mathalon, 2007; Ford, Roach, Faustman, & Mathalon, 2007; Heinks-Maldonado, Nagarajan, & Houde, 2006; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Ford et al., 2001; Curio, Neuloh, Numminen, Jousmaki, & Hari, 2000; Numminen, Salmelin, & Hari, 1999) have shown that human auditory cortex is less responsive to self-voice during vocalization or speaking compared with passive listening. This suppression has been shown to be highly specific to normal auditory feedback (NAF) of self-vocalizations and diminished in response to altered or modified voice feedback (Heinks-Maldonado et al., 2005, 2006). These findings suggest that the vocalization-induced suppression is sensitive to the extent to which the sensory consequences of self-produced motor acts match the predicted sensory input. Greater suppression during normal voice feedback is thought to occur because the auditory system is less responsive to feedback that is accurately predicted by the efference copies from the vocal motor system (Heinks-Maldonado et al., 2005). The minimal feedback mismatch (error) during normal voice feedback can translate to smaller neural responses than those during altered or modified feedback. The neural mechanisms underlying mismatch detection in the auditory feedback can possibly enable the audio-vocal system to distinguish between self- and externally generated sounds.

However, a recent study in primates suggested that vocalization-induced suppression might enhance the neural sensitivity of the auditory neurons to alterations in voice auditory feedback (Eliades & Wang, 2008). This effect has been reported in vocalizing primates, wherein a majority of the cortical auditory neurons that were suppressed in response to normal voice auditory feedback became highly sensitive to voice pitch feedback alterations during vocalization. The fact that this effect was not observed during vocalization for those neurons that were excited in response to normal feedback and also that it was absent for both groups of neurons (suppressed and excited) during passive listening suggests that vocalization-induced suppression may play an important role in enhancing the sensitivity of the auditory neurons to alterations in voice feedback. This phenomenon implies a sophisticated neural mechanism that involves internal modulation (reafference projections) as well as responses to changes in feedback for feedback-based monitoring of the self-produced vocalizations. Although the neural mechanisms of such processes are still unclear, one possible hypothesis is that active vocalization may change the tuning properties of auditory neurons in such a way as to increase their sensitivity to alterations in voice auditory feedback. The enhanced sensitivity to feedback alterations during active vocalization may help the audio-vocal system control the structure of the self-produced vocal outputs through feedback-based error detection and correction.

Although the vocalization-induced auditory suppression has previously been reported in humans (Heinks-Maldonado et al., 2006; Houde et al., 2002; Numminen et al., 1999) and animals (Eliades & Wang, 2003), the effect of active vocal production on neural sensitivity enhancement has only been demonstrated in primates (Eliades & Wang, 2008), and previous studies in humans showed no sign of this effect. One possible explanation for the lack of any clue regarding the enhancement of the neural responses during vocal production is that responses to feedback pitch perturbation in humans were recorded at the onset of vocalization. The disadvantage of obtaining the neural responses at vocal onset is that they reflect brain activities arising from two independent but time-overlapping neural processes; these are the suppression at vocal onset versus feedback processing at the onset of pitch perturbation. Therefore, the temporal overlap between these two neural events may result in the masking of enhanced neural responsiveness by suppression at vocal onset. Interestingly, findings of studies in the somatosensory (Blakemore, Wolpert, & Frith, 1998, 2000) and auditory modalities (Aliu, Houde, & Nagarajan, 2009) support this idea by showing that the motor-induced suppression (MIS) of sensory feedback develops for zero delays between the onset of the motor act and its sensory feedback, but it does not generalize to nonzero delays. In the somatosensory system, this effect was suggested to decrease the sensation of tickliness as the stimulus delivery delays were reduced because the suppression of sensory neural responses to self-generated tactile stimulation were greater for zero stimulus delays (Blakemore et al., 2000; Blakemore, Goodbody, & Wolpert, 1998; Blakemore, Wolpert, et al., 1998). A similar effect has been reported in the auditory system by showing that the MIS of the auditory cortex in response to a simple tone triggered by a button press develops for zero delays but does not generalize to nonzero delays (Aliu et al., 2009). These findings support the theory of an internal forward model that accommodates intrinsic sensory delays and provides time-dependent suppression of the predicted sensory consequences of a self-produced action.

The present study investigated whether introducing a time delay between the vocal onset and the onset of auditory feedback perturbation can reveal the enhanced neural responsiveness to feedback alterations during vocal production in humans, which was previously reported in primates (Eliades & Wang, 2008). The properties of this effect was characterized by examining the P_1–N_1–P_2 components of the auditory-evoked potentials in response to +200 cents[1] voice pitch feedback perturbations that were delivered at a randomly chosen time delay following the vocal onset. The experimental setup for this task is schematized in Figure 1. The main hypothesis in the present study was that the temporal separation of the stimulus onset from vocal onset

would result in the elimination of overlap between the time course of the neural responses arising from these two neural events (vocal onset vs. stimulus) and, consequently, would eliminate the masking effect of suppression from the responses to pitch-shifted feedback at the onset of vocalization. On the basis of this hypothesis, we predicted that vocalization would enhance neural responses to feedback pitch perturbation during active vocalization compared with passive listening for nonzero delays between the vocal onset and the stimulus. However, to verify consistency with previous findings regarding the suppression effect at vocal onset, the neural responses were also obtained in response to normal and pitch-shifted auditory feedback at the onset of vocalization (zero delay). The investigation of time-dependent neural processing of auditory feedback in humans is possibly a key to improving our knowledge about the mechanisms that underlie neural sensitivity tuning to enhance feedback mismatch detection for vocal error detection and correction during speaking.

METHODS

Subjects

Seventeen right-handed native speakers of American English (12 women and 5 men, 18–29 years of age) with a mean age of 22.24 years ($SD = 3.85$) participated in this study. All subjects passed a bilateral pure-tone hearing screening test and reported no history of neurological disorders or voice training. All study procedures, including recruitment, data acquisition, and informed consent, were approved by the Northwestern University institutional review board, and subjects were monetarily compensated for their participation.

Procedure

ERPs were recorded in response to normal and pitch-shifted auditory feedback while subjects maintained a short utterance of the vowel sound /a/ for approximately 3–4 sec. During each utterance, subjects' voice auditory feedback was pitch shifted by +200 cents (upward) for 600 msec. The onset of the pitch-shift stimulus (PSS) was randomly chosen at 0, 200, 500, or 1000 msec following the onset of the vocalization (Figure 1). Subjects were asked to take a short break (2–3 sec) between successive utterances and repeat this vocal task 200 times during the active vocalization condition. Each active vocalization condition was immediately followed by a condition during which subjects passively listened to the playback of the auditory feedback of their self-produced vocalizations. The experiment consisted of three active vocalization (200 utterances each) and three passive listening conditions, giving a total number of 600 utterances for the vocalization and listening tasks separately. ERPs in response to NAF at vocal onset (zero delay) were obtained by averaging the recorded neural responses at the onset of vocalization in utterances for which the PSS was delivered after 1000 msec. ERPs in response to pitch-shifted feedback were obtained by averaging the neural responses to the onset of PSS at randomly chosen time delays (zero and nonzero delays at 200, 500, and 1000 msec). The extent of vocalization-induced modulation of the neural responsiveness to normal and pitch-shifted voice feedback at different time delays with respect to vocal onset was characterized by comparing the neural peak amplitudes of the P_1–N_1–P_2 ERP components during active vocalization and passive listening conditions.

Instrumentation

Subjects were seated in a sound-treated room in which their voices were recorded with an AKG boomset microphone (model C420; AKG Co., Vienna), amplified with a Mackie mixer (model 1202; LOUD Technologies, Woodinville, WA), and pitch shifted through an Eventide Eclipse Harmonizer (Eventide, Inc., Little Ferry, NJ). The pitch-shifted

Figure 1. Schematic of the experimental setup for introducing time delays between the vocal onset and the onset of auditory feedback pitch perturbation. The vocal onset, detected by a voice onset detector (VOD), triggered the pitch-shift stimulus (PSS) after a randomly chosen time delay (Δt). The PSS onset triggered a TTL pulse that was recorded on the computer and used for averaging the neural responses with respect to the onset of the stimulus.
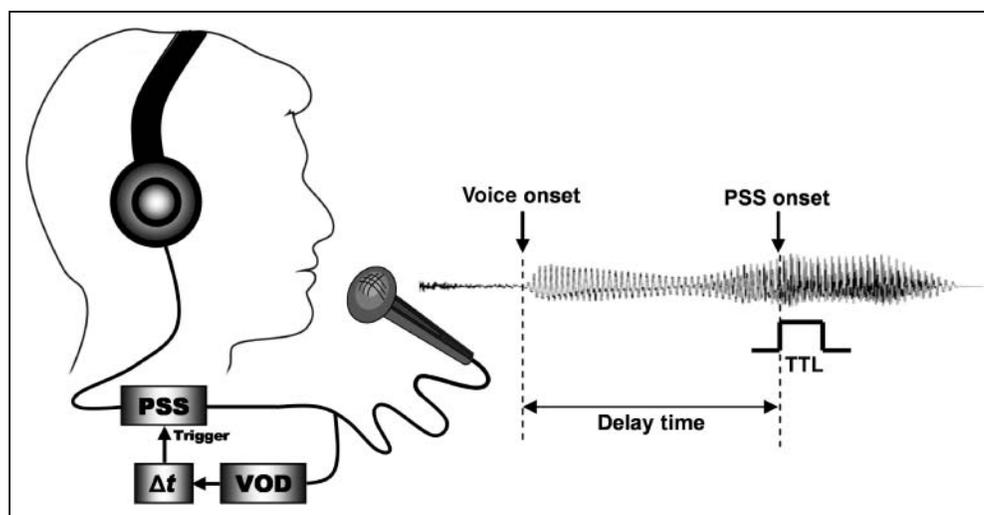
auditory feedback was heard through Etymotic earphones (model ER1-14A; Etymotic Research, Inc., ElkGrove Village, IL) inserted into the subject's ear canals. The gain between the subject's voice and the feedback was further manipulated with a Crown amplifier (D75; Crown Audio Inc., Elkhart, IN) and HP350 dB-attenuators to +10 dB SPL, calibrated with a Zwislocki coupler and a Brüel & Kjær sound level meter (model 2250; Brüel & Kjaer Sound & Vibration Measurement A/S, Denmark). The 10-dB gain between the voice and the feedback channels allowed the recording of ERPs in response to pitch-shifted auditory feedback by partially masking the normal voice feedback through bone and airborne conduction.

Following each vocalization condition, the feedback channel was converted to a sound file to be played back during the passive listening condition. Two objective and one subjective methods were used to calibrate the gain during the passive listening condition with respect to active vocalization. The objective methods included using the Brüel & Kjær sound level meter and Zwislocki coupler to ensure the sound pressure level (dB, SPL) in the output of the insert earphones during passive listening was nearly identical to the earphone output level during vocalization. Furthermore, because the feedback channel was recorded on Chart recorder software (AD Instruments, Castle Hill, Australia) during vocalization and listening, we verified that the voltage driving the earphones was identical during both conditions. Lastly, we asked subjects to verify that the sound intensity during vocalization and listening conditions was nearly identical. At conversional levels, subjects maintained their voice loudness at about 70–75 dB, which was delivered to the feedback channel (earphones) at 80–85 dB resulting in a 10-dB gain between voice (microphone) and feedback (earphones) channels.

All parameters of the pitch-shifted auditory feedback stimulus, such as duration, magnitude, and time delays, were controlled by MIDI software (Max/MSP v.4.1. by Cycling 74, San Francisco, CA). The MIDI software also generated a TTL pulse to mark the onset of each stimulus for synchronized averaging of the recorded ERPs. Voice, feedback, and TTL pulses were sampled at 10 kHz using PowerLab A/D Converter (model ML880, AD Instruments) and recorded on a laboratory computer utilizing Chart software (AD Instruments).

Recording

The EEG signals were recorded from 13 sites on the subject's scalp (C_Z, C_3, C_4, T_3, T_4, F_Z, F_3, F_4, F_7, F_8, P_Z, P_3, and P_4) using an Ag-AgCl electrode EEG cap (10-20 system). Scalp-recorded brain potentials were amplified with 13 Grass amplifiers (Grass P511 AC amplifier; Astro-Med, Inc., West Warwick, RI), sampled at 10 kHz (PowerLab A/D Converter), and recorded using Chart software. All amplifiers were calibrated according to the instructions from the manufacturers. The gain of the EEG amplifiers was set to 10k, and the cut-off frequencies of their on-line high-pass and low-

pass filters were set to 0.1 Hz and 10 kHz, respectively. All recorded EEG channels were referenced to linked earlobes, and their impedances were measured using a Grass impedance meter (Model: EZM-5AB) and maintained below 5 kΩ. The effect of visual and muscle artifacts on the recorded brain potentials was reduced by asking the subjects to close their eyes and relax their muscles throughout the course of the experiment.

ERP Signal Analysis

Auditory-evoked responses to normal and pitch-shifted feedback were obtained by averaging the recorded EEG signals with respect to the onset of a TTL pulse that marked the onset of the vocalization for the normal (0 delay) condition and the onset of the PSS for pitch-shifted feedback at different delay times. Before averaging, the EEG signals from all channels were subjected to off-line filtering, using a band-pass filter with cut-off frequencies set to 1 and 30 Hz. The auditory neural responses to normal and pitch-shifted feedback were obtained by averaging the filtered EEGs, which were cut into epochs ranging from −100 to 500 msec with respect to the onset of the TTL pulse. Artifact rejection was carried out before baseline correction by excluding those trials that exceeded ±50 μV in amplitude. The baseline of the remaining trials was corrected by calculating the prestimulus mean amplitude in the 100-msec prestimulus time window and then subtracting it from all samples within the individual epochs ranging from 100 msec prestimulus to 500 msec poststimulus. ERPs were obtained by averaging trials for conditions with a minimum number of 100 epochs. The latency and the amplitude of the P_1–N_1–P_2 complex were extracted from the averaged neural responses by finding the most prominent peaks in 50-msec-long time windows centered at 50, 100, and 200 msec.

Topographical Distribution Maps

The surface distribution maps of measures of brain activity in response to voice pitch feedback perturbation were created using the peak amplitudes of the neural responses for 13 electrode sites (C_Z, C_3, C_4, T_3, T_4, F_Z, F_3, F_4, F_7, F_8, P_Z, P_3, and P_4) over the surface of the scalp. These topographical distribution maps of the neural activity were created by color coding the amplitudes of the ERP components using the interpolation method between adjacent electrodes to obtain a map of electrical activity distribution.

Statistical Analysis

The extracted latencies and amplitudes of the P_1, N_1, and P_2 peaks in the averaged neural responses were separately analyzed for the recording sites. The SPSS software (v.16.0) was used to perform a two-way repeated measures ANOVA (Rm-ANOVA) to examine main effects of condition (active vocalization, passive listening) and electrode

position (central: C_z; left centro-medial: C_3; right centro-medial: C_4; left temporal: T_3; right temporal: T_4; fronto-central: F_z; left fronto-medial: F_3; right fronto-medial: F_4; left fronto-lateral: F_7; right fronto-lateral: F_8; parieto-central: P_z; left parieto-medial: P_3; and right parieto-medial: P_4) factors and their interactions on the extracted P_1–N_1–P_2 neural peak amplitudes and latencies of the ERPs in response to NAF at vocal onset (zero delay). The amplitudes and the latencies of the P_1–N_1–P_2 responses to +200 cents pitch-shifted voice feedback were subjected to a three-way Rm-ANOVA to examine main effects of condition (active vocalization, passive listening), stimulus delay (0, 200, 500, and 1000 msec), and electrode position (same as for NAF) factors and their interactions.

RESULTS

Figure 2A–E shows the grand-averaged (over 17 subjects) auditory neural responses to NAF at vocal onset (0-msec delay) and responses to pitch-shifted feedback (PSS) at 0-, 200-, 500-, and 1000-msec time delays after vocal onset, respectively. For each individual, the P_1–N_1–P_2 cortical components were extracted from the averaged neural responses across all conditions and electrode positions and were subjected to statistical analysis. The bar plots in Figure 3 show the difference between neural peak amplitudes of P_1–N_1–P_2 ERP components during active vocalization versus passive listening conditions for normal (0 delay) and pitch-shifted voice feedback at different delay times. The positive and the negative numbers on the *y*-axis in Figure 3 demonstrate the extent of enhancement and suppression of neural responsiveness during active vocalization compared with passive listening, respectively.

P_1 Neural Responses

Rm-ANOVAs on the amplitude of P_1 responses to NAF of self-vocalizations revealed no significant main effect. Rm-ANOVAs on P_1 responses to PSS revealed a significant main effect of stimulus onset delay, $F(3, 48) = 4.13$, $p = .011$, electrode position, $F(12, 192) = 10.02$, $p < .001$, and Stimulus Onset Delay × Position, $F(36, 576) = 1.60$, $p = .015$, and Condition × Position, $F(12, 192) = 3.82$, $p < .001$, interactions. Post hoc tests revealed that the significant main effect of stimulus onset delay was due to a significant difference between P_1 amplitudes for 0- versus 1000-msec delay times ($p = .014$), indicating that the 1000-msec stimulus onset delays elicited P_1 responses that were larger than those for 0 msec. Post hoc tests for the electrode position main effect revealed significant positivity in the fronto-central (F_z), left fronto-medial (F_3), and right fronto-medial (F_4) regions for the P_1 amplitudes ($p < .001$) compared with other electrode sites. No significant main effect of electrode position was revealed for electrode pairs on left versus right hemispheres, indicating no laterality effect. Figure 4A shows scalp distribution of the P_1 amplitudes for NAF and pitch-shifted auditory feedback (PSS) at different stimulus onset

delays during active vocalization and passive listening conditions separately. Significant Stimulus Onset Delay × Position and Condition × Position interactions indicated that the scalp distribution pattern of the P_1 amplitudes was different across different stimulus onset delays and conditions (active vocalization vs. passive listening). Analysis of the P_1 latencies revealed no significant effects. The overall mean latencies of the P_1 components were 67.08 msec ($SD = 18.32$ msec).

N_1 Neural Responses

Rm-ANOVAs on the amplitude of N_1 responses to NAF of self-vocalizations revealed a significant main effect of the condition factor, $F(1, 16) = 7.21$, $p = .016$, electrode position, $F(12, 192) = 15.62$, $p < .001$, and Condition × Position interaction, $F(12, 192) = 7.43$, $p < .001$. The significant main effect of the condition factor indicated that the N_1 responses to the NAF were suppressed (less negative) during active vocalization compared with passive listening (see Figures 2A and 3). Post hoc tests for the electrode position main effect revealed significant negativity in the fronto-central (F_z), left fronto-medial (F_3), and right fronto-medial (F_4) regions for the N_1 amplitudes ($p < .001$) compared with other electrode sites. No significant main effect of electrode position was revealed for electrode pairs on left versus right hemispheres. Figure 4B shows scalp distributions of the N_1 amplitudes for NAF and pitch-shifted auditory feedback (PSS) at different stimulus onset delays during active vocalization and passive listening conditions separately. The significant Condition × Position interaction indicated that the scalp distribution pattern of the N_1 amplitudes was different across different conditions (active vocalization vs. passive listening).

Rm-ANOVAs on N_1 responses to PSS revealed a significant main effect of electrode position, $F(12, 192) = 23.55$, $p < .001$, as well as a Stimulus Onset Delay × Position interaction, $F(36, 576) = 1.95$, $p = .001$, and a Condition × Position interaction, $F(12, 192) = 2.49$, $p = .005$. Post hoc tests for the electrode position main effect revealed significant negativity in the fronto-central (F_z), left fronto-medial (F_3), and right fronto-medial (F_4) regions for the N_1 amplitudes ($p < .001$) compared with other electrode sites (see Figure 4B). Results of the pairwise comparison showed no significant main effect of electrode position on left versus right hemispheres. Significant Stimulus Onset Delay × Position and Condition × Position interactions indicated that the scalp distribution pattern of the N_1 amplitudes was different across different stimulus onset delays and conditions (active vocalization vs. passive listening). Analysis of the N_1 latencies revealed no significant effects. The overall mean latencies of the N_1 components were 124.72 msec ($SD = 22.62$ msec).

P_2 Neural Responses

Rm-ANOVAs on P_2 responses to NAF of self-vocalizations revealed a significant main effect of electrode position, $F(12,$

Figure 2. Time course of the neural responses to (A) normal auditory feedback (NAF) occurring at vocal onset (delay = 0 msec) and to PSS for (A) 0-msec, (B) 200-msec, (C) 500-msec, and (D) 1000-msec delays between the vocal onset and the onset of the stimulus. The neural responses for each condition are separately plotted for all recording sites during active vocalization (solid) and passive listening (dashed) conditions. The dashed vertical lines mark the onset of normal (A) and pitch-shifted (B–E) feedback in each subplot.
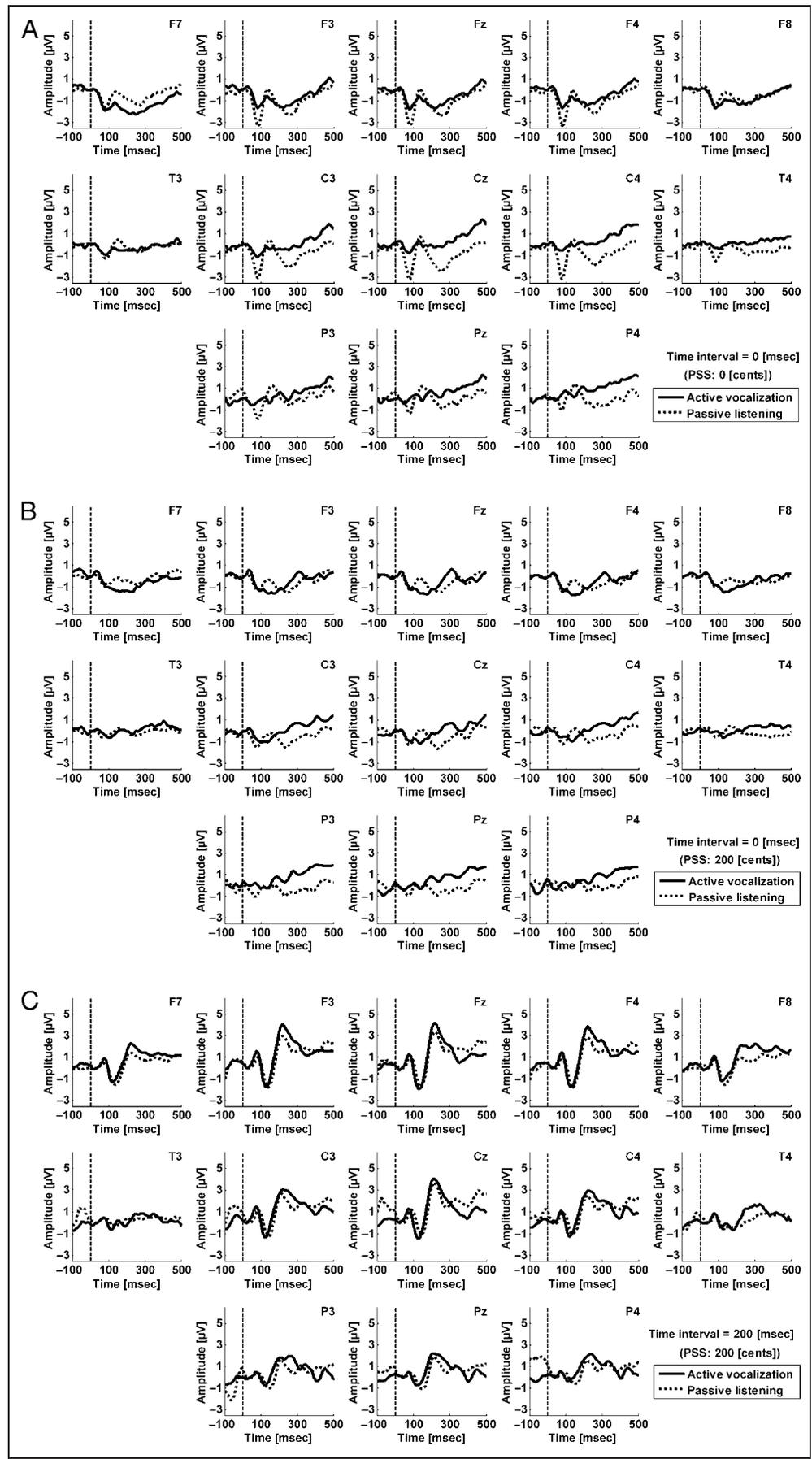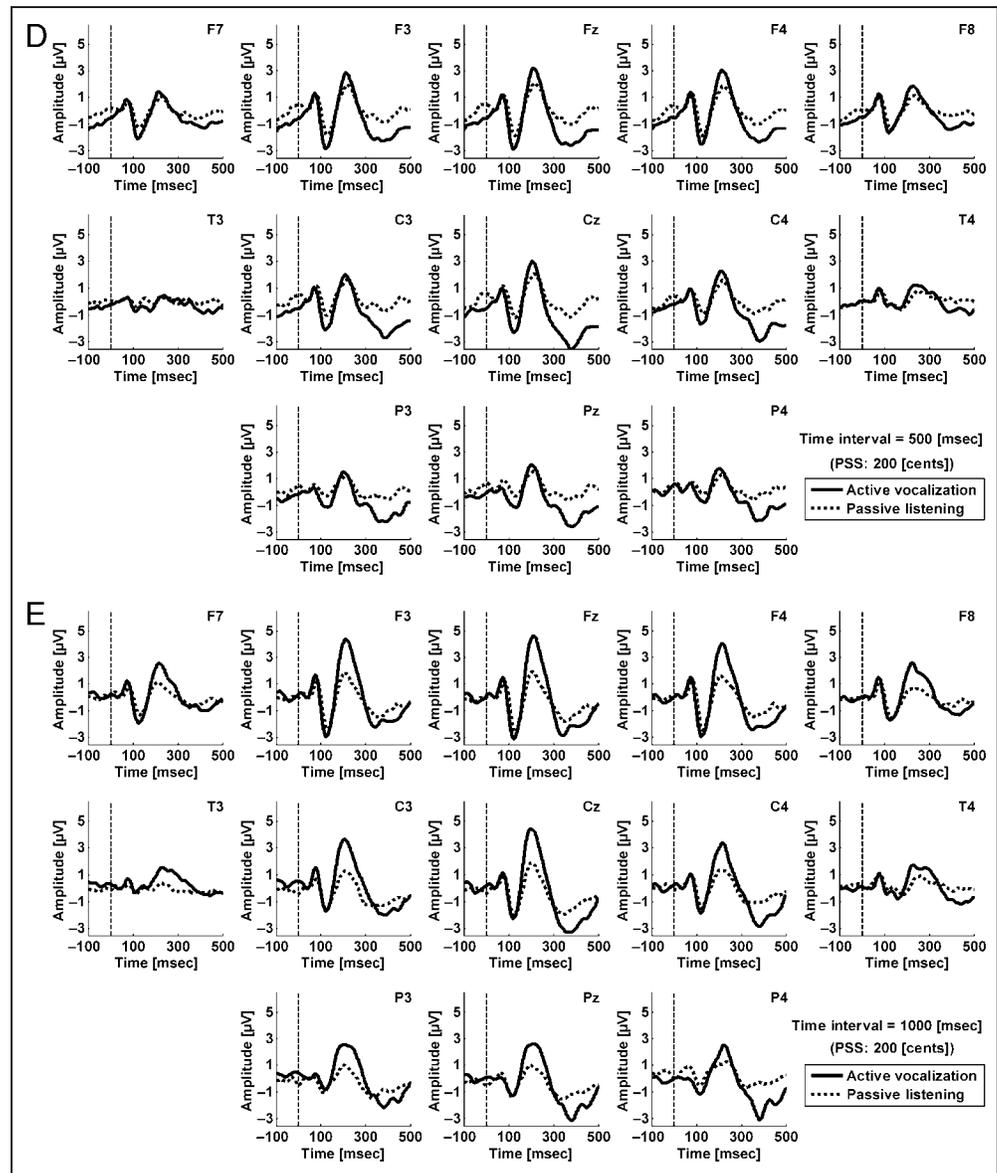
Figure 2. (*continued*)

Figure 3. Bar plots of the difference between P_1–N_1–P_2 neural peak amplitudes for active vocalization versus passive listening conditions in response to normal auditory feedback (NAF) and +200 cents pitch-shift stimulus (PSS) occurring at different delay times after vocal onset. The negative amplitudes on the y-axis indicate neural suppression during active vocalization compared with passive listening, and the positive amplitudes indicate vocalization-induced enhancement of neural responsiveness to +200 cents PSS. The asterisks (*) indicate a significant amplitude difference ($p < .05$) between vocalization and listening conditions.
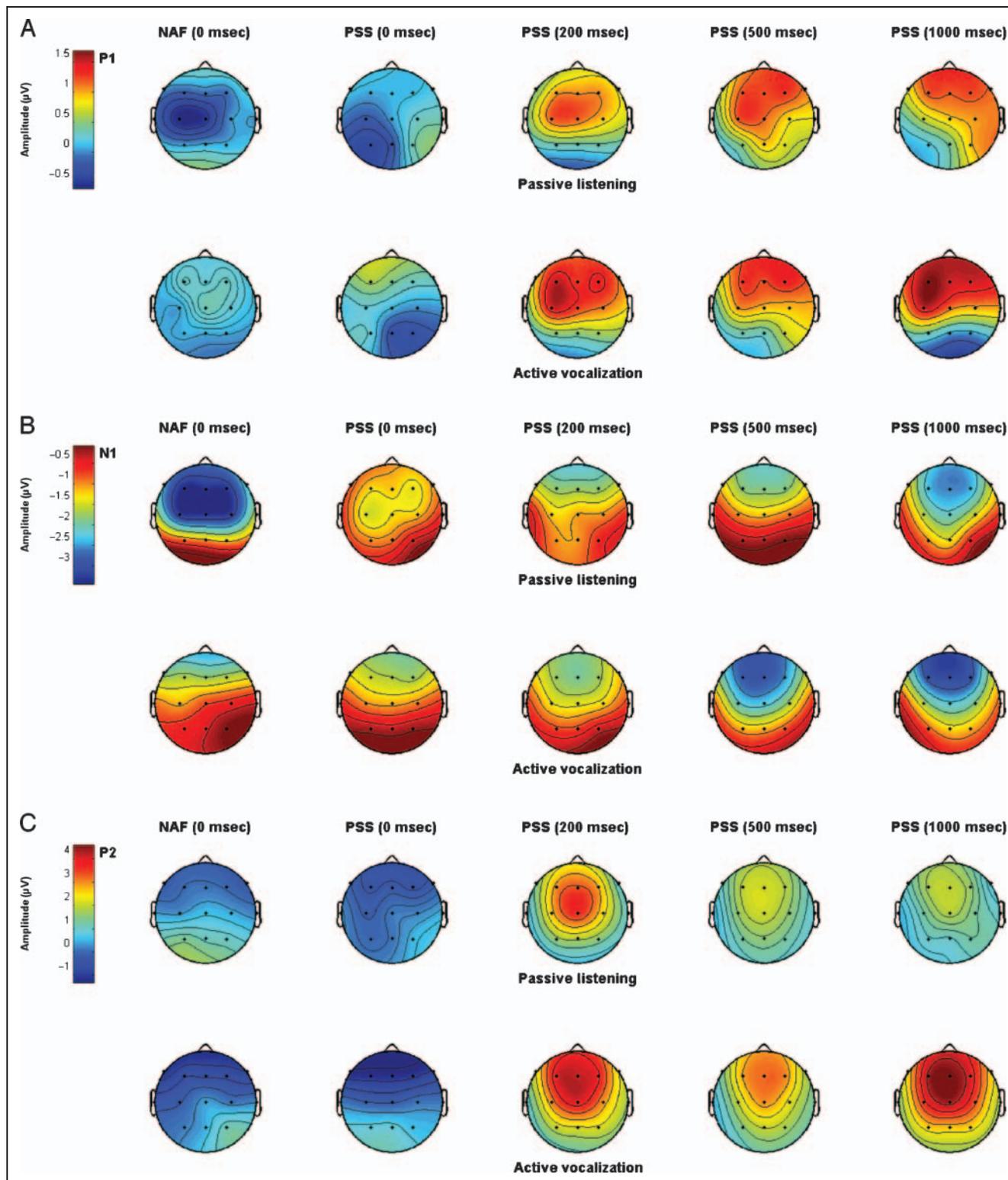
Figure 4. Topographical scalp distributions of the averaged neural responses to normal (first column from left) and pitch-shifted (second to fifth column from left) auditory feedback across 17 subjects. The maps are calculated for 13 recording sites on the surface of the scalp (C_Z, C_3, C_4, T_3, T_4, F_Z, F_3, F_4, F_7, F_8, P_Z, P_3, and P_4) for (A) P_1 (latency = 67.08 msec), (B) N_1 (latency = 124.72 msec), and (C) P_2 (latency = 213.02 msec) peak amplitudes. The top and the bottom rows in each plot show the scalp distributions during passive listening and active vocalization conditions, respectively.

192) $= 10.73$, $p < .001$, as well as a Condition × Position interaction, $F(12, 192) = 2.48$, $p = .005$. Post hoc tests for the electrode position main effect revealed significant positivity in the parieto-central (P_z), left parieto-medial (P_3), and right parieto-medial (P_4) regions for the P_2 amplitudes ($p = .002$) compared with other electrode sites. No significant laterality effect was revealed during pairwise comparison of electrode sites on left and right hemispheres. Figure 4C shows scalp distribution of the P_2 amplitudes for NAF and pitch-shifted auditory feedback (PSS) at different stimulus onset delays during active vocalization and passive listening conditions separately. A significant Condition × Position interaction indicated that the scalp distribution pattern of the P_2 amplitudes was different across different conditions (active vocalization vs. passive listening).

Rm-ANOVAs on P_2 responses to PSS revealed a significant main effect of stimulus onset delay, $F(3, 48) = 40.90$, $p < .001$, condition, $F(1, 16) = 26.62$, $p < .001$, and electrode position, $F(12, 192) = 19.24$, $p < .001$, as well as Stimulus Onset Delay × Condition, $F(3, 48) = 12.65$, $p < .001$, Stimulus Onset Delay × Position, $F(36, 576) = 15.22$, $p < .001$, and Condition × Position, $F(12, 192) = 2.00$, $p = .025$, interactions. Post hoc tests revealed that the significant main effect of stimulus onset delay was due to significant differences between P_2 amplitudes for 0- versus 200-msec, 0- versus 500-msec, and 0- versus 1000-msec delay times ($p < .001$ for all), indicating that all delayed PSS onsets elicited P_2 responses that were larger than those for 0 msec (see Figure 2B–E). Post hoc tests for the electrode position main effect revealed significant positivity in the central (C_z) region for the P_2 amplitudes ($p < .001$) compared with other electrode sites (see Figure 4C). No significant main effect of position was found for electrode sites on left versus right. Significant Stimulus Onset Delay × Position and Condition × Position interactions indicated that the scalp distribution pattern of the P_2 amplitudes was different across different stimulus onset delays and conditions (active vocalization vs. passive listening). Post hoc tests for significant Stimulus Onset Delay × Condition interaction revealed a significant main effect of condition factor for P_2 peak amplitudes for 200 msec, $F(1, 16) = 6.78$, $p = .019$, 500 msec, $F(1, 16) = 6.76$, $p = .019$, and 1000 msec, $F(1, 16) = 87.52$, $p < .001$, PSS onset delays but no significant effect of condition for 0 msec, indicating that P_2 amplitudes were larger during active vocalization compared with passive listening for all nonzero delays but no difference for zero PSS onsets (Figures 2 and 3). Analysis of the P_2 latencies revealed no significant effects. The overall mean latencies of the P_2 components were 213.02 msec ($SD = 33.68$ msec).

DISCUSSION

The significant main effect of the condition factor on N_1 responses to NAF confirmed the previously reported suppression effect during self-vocalization in humans (Heinks-Maldonado et al., 2006; Houde et al., 2002; Numminen & Curio, 1999; Numminen et al., 1999) and primates (Eliades & Wang, 2003). However, results of the analysis did not show any significant main effect of condition on N_1 responses to PSS at different delay times, indicating that the neural responses to +200 cents pitch shifts in the auditory feedback were not significantly suppressed during active vocalization condition compared with passive listening. This finding may imply an important characteristic of the vocalization-induced cortical suppression of the auditory responses in humans and supports the hypothesis of a feedforward model in the speech production system (Blakemore, Rees, & Frith, 1998; Wolpert, 1997). The N_1 suppression seems to be feedback specific, meaning that it shows sensitivity to whether the feedback from self-produced vocalizations is normal or altered. Our results showed that N_1 neural responses are significantly suppressed to NAF, whereas there was no significant suppression observed for +200 cents pitch-shifted auditory feedback (see Figures 2A and B and 3). The notion of feedback specificity of N_1 suppression is further supported by previous studies showing that the cortical auditory neural responses to self-vocalizations are most suppressed during normal feedback compared with conditions where the feedback was either pitch shifted or modified with a voice from a different speaker (Heinks-Maldonado et al., 2005, 2006).

The feed-forward theory explains the effect of N_1 by stating that during vocalization, the auditory system is less responsive (more suppressed) to normal voice feedback that closely matches the predicted feedback that is represented by efference copies of the vocal motor commands. Smaller prediction error (no mismatch) in response to normal voice feedback may contribute to smaller neural responses and, consequently, greater suppression during normal or unaltered feedback. The same hypothesis may explain the reason for the absence of suppression during altered auditory feedback. Although the efference copies may have produced an accurate prediction of the sensory feedback associated with the intended vocal output, the auditory feedback perturbation generated larger prediction error that may have contributed to stronger neural responses and, consequently, less suppression in response to mismatch during vocalization. During the passive listening condition, the absence of the efference copy may have allowed greater auditory responsiveness to normal feedback. The process of feedback prediction based on efference copies of the vocal motor commands is possibly one way in which the audio-vocal system can distinguish self-produced sounds from those generated by an external source. This effect might be an important characteristic of the vocal production system for feedback-based monitoring of self-vocalizations that helps to detect and correct for errors during vocal production.

Results of our analysis revealed a significant main effect of condition on P_2 responses to pitch-shifted auditory feedback, indicating that the auditory system was more responsive to feedback perturbations during active vocalization than passive listening (Figure 2B–E). The increase in P_2

peak amplitudes during vocalization suggests that the proposed feed-forward model (Blakemore, Rees, et al., 1998; Wolpert, 1997) can enhance auditory responsiveness to feedback perturbations during speaking. This effect might be an important characteristic of the speech motor control system in humans that allows for accurate detection and correction of unintended changes in the vocal output during speaking. A similar effect observed in primates (Eliades & Wang, 2008) was that the vocalization-induced suppression of cortical auditory neurons during normal feedback led to their increased sensitivity to feedback alterations. Although the neural mechanisms of such a phenomenon are poorly understood, the enhanced neural sensitivity to changes in the auditory feedback is thought to be due to changes in the tuning properties of the auditory neurons (e.g., increasing their dynamic range) as a result of vocal motor system activity (Eliades & Wang, 2008). Vocalization-induced enhancement of the neural responsiveness to feedback alteration may support the internal forward model theory by explaining that the hypothesized reafference projection might be involved in fine tuning the auditory cortex to improve feedback-based monitoring and control of voice.

Despite the significant main effect of condition on P_2 peak amplitudes, results of post hoc tests on Stimulus Onset Delay × Condition interaction revealed that the P_2 responses to pitch-shifted feedback were significantly larger during vocalization than listening only for non-zero delay times (Figures 2B–E and 3). The absence of vocalization-induced enhancement of neural responsiveness to feedback alteration for zero delays suggests that the hypothetical motor-induced sensitivity tuning of the auditory neurons may be a time-dependent process that is mostly effective for mismatch detection after the onset of vocalization. One possible consequence of this effect is that the temporal separation of the vocal and PSS onsets might eliminate the masking effect of the cortical auditory neuron suppression at the vocal onset and therefore lead to larger neural responses to delayed changes in feedback. Previous studies on the somatosensory system have demonstrated that the MIS of the sensory consequences of an internally generated motor act are sensitive to delays between the motor act and the onset of the stimulus (Blakemore et al., 2000; Blakemore, Wolpert, et al., 1998). A similar study on the auditory system has suggested that MIS of the auditory cortex in response to a simple tone triggered by a button press develops for zero delays but not to nonzero delays (Aliu et al., 2009). The absence of enhanced neural responsiveness at vocal onset may mean that the scalp-recorded potentials in response to altered feedback may be masked by the suppression of the auditory neurons at the onset of vocalization. However, for nonzero delay times, the diminished masking effect may allow the recording of neural components purely in response to PSS that indicate enhanced neural responsiveness to alterations in voice pitch feedback.

The main effect of stimulus onset delay on P_2 neural responses to altered feedback was revealed to be significant for 0- versus 200-msec, 0- versus 500-msec, and 0- versus 1000-msec delay times, indicating that extended time intervals between vocalization and PSS onsets elicited responses that were greater in amplitude than those at vocal onset (zero delay). This finding implies that the cortical processing of feedback alterations may be a time-dependent process that is sensitive to the time delay between the vocal onset and the onset of feedback mismatches. Larger neural responses to delayed PSS compared with those in response to PSS occurring at vocal onset during passive listening may indicate that the sensory memory from the incoming feedback can help the auditory system be more responsive to feedback alterations by enabling it to compare the pitch-shifted feedback with the normal feedback before the onset of the stimulus (see Figure 3). However, further enhancement of the P_2 peak amplitudes in response to the delayed PSS during vocalization compared with passive listening suggests that the activity of the vocal motor system may further sharply tune the auditory system to respond more vigorously to alterations in voice pitch feedback. The enhancement of the neural responsiveness to feedback alterations during vocalization may imply a complex neural mechanism that functions to fine tune the auditory neurons to detect alterations in the feedback of self-generated vocal outputs to enhance the ability of the audio-vocal system for feedback-based monitoring and control of the voice structure. A similar significant main effect of the stimulus onset time was observed for the P_1 response to PSS, indicating that the P_1 responses were larger for PSS that were delayed by 1000 msec compared with those occurring at vocal onset (zero delay). Modulation of the P_1 and P_2 peak amplitudes by the stimulus onset timing suggests that sensory discrimination of the feedback mismatches is enhanced for prolonged experience of normal feedback from self-produced voice.

Results also revealed a significant main effect of the electrode position factor on the neural peak amplitudes for P_1–N_1–P_2 complex. This effect was observed as a significant positivity that started in the frontal and fronto-medial region for P_1 during both normal and pitch-shifted feedback followed by a significant negativity for N_1 in about the same region that eventually moved posteriorly toward the central region during pitch-shifted and toward parietal region for NAF as a significant positivity for the P_2 peak amplitudes (Figure 4). The dynamic flow of the potential distribution on the surface of the scalp over time suggested several stages of feedback information processing in different auditory related areas in the brain.

Previous studies have suggested that several neural generators may contribute to auditory P_1–N_1–P_2 components that reflect neural processing at multiple stages. The generators of P_1 have traditionally been identified in the primary auditory cortex (specifically superior temporal gyrus and Heschl's gyrus) and are considered to

reflect neural processes that index preattentive and early cortical processing of the auditory input (Burkard, Don, & Eggermont, 2006). The N_1 component has been discussed in numerous studies to have neural generators in the higher (e.g., secondary) auditory cortical areas as well as the upper bank of the Sylvian fissure in the temporal lobe (Hari, Aittoniemi, Jarvinen, Katila, & Varpula, 1980) and in cortical frontal areas (Näätänen & Picton, 1987). The N_1 is considered an index of the preattentive auditory processing that reflects higher cortical processing of the incoming auditory stream (Burkard et al., 2006). However, the neural generators of P_2 are not as well understood as the P_1 and N_1 ERP components. P_2 appears to have generators in multiple auditory and non-auditory areas and does not appear to be a unitary potential, meaning that it is likely that there are several component generation processes occurring in the time frame of P_2. Performing different cognitive or noncognitive tasks may elicit neural responses from multiple sources that have scalp distribution similar to P_2. In a recent study of feedback-based error monitoring during musical performance, P_2 has been suggested to reflect neural mechanisms that underlie auditory mismatch detection (Katahira, Abla, Masuda, & Okanoya, 2008) during the performance of a motor task and is assumed to arise from the ACC, triggered by the basal ganglia when subjects notice their own motor error. This assumption is supported by an fMRI study that investigated the neural substrates of vocal pitch regulation during singing (Zarate & Zatorre, 2008). The P_2 component during the passive listening to the playback of self-produced sounds has been suggested to be an index of a cognitive control-related auditory component that is thought to be generated in ACC as a result of template mismatches and therefore has a scalp distribution similar to the P_2 component during motor tasks (Folstein & Van Petten, 2008).

Greater responsiveness to feedback alterations for nonzero delays between the onset of vocalization and the stimulus onset indicates that the audio-vocal integration system is sensitive not only to the predicted feedback (feedback-specificity) but also to the timing between the vocal motor commands and the incoming sensory feedback. The modulation of the neural peak amplitudes as a result of active vocalization for nonzero delays suggests that the vocalization-induced enhancement of auditory sensitivity to feedback alterations is driven by a neural process that takes into account the intrinsic sensory delays and adjusts the parameters of the auditory neurons in such a way as to increase their sensitivity to feedback alterations after the onset of the vocalization. Feedback specificity along with time-dependent neural processing of auditory feedback may be important features of the audio-vocal system that mediate feedback-based monitoring of the self-produced vocal output and help to improve vocal error detection and correction and thereby allowing robust control over the structure of the voice.

Conclusion

In the present study, the time-dependent neural processing of auditory feedback perturbation was investigated during active vocalization and passive listening conditions. Results showed that the magnitude of scalp-recorded ERPs was larger for non-zero delays between the vocal onset and the onset of the feedback perturbation compared with perturbations delivered right at vocal onset. The vocalization-induced enhancement of neural responsiveness to feedback alterations was shown to be present for the delayed feedback perturbation with respect to the onset of the vocalization. These findings suggested that the audio-vocal system accommodates intrinsic sensory delays and predicts the timing between the onset of the vocal motor act and its auditory feedback. Time-dependent neural processing of auditory feedback alterations is one possible way in which the system can perform vocal error detection and correction accurately in time on the basis of the incoming sensory feedback information.

Note

1. Cents is defined as a logarithmic measure of frequency ratios between different notes in musical scales. In general, 1200 cents is equal to 1 octave, which measures the frequency ratio between one note with itself in the next octave (frequency ratio = 2:1). The difference between two adjacent notes (e.g., C and C#) in cents scale is 100 cents, often referred to as 1 semitone.

REFERENCES

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience, 21,* 791–802.

Bauer, J. J., & Larson, C. R. (2003). Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique. *Journal of the Acoustical Society of America, 114,* 1048–1054.

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *Journal of the Acoustical Society of America, 119,* 2363–2371.

Blakemore, S. J., Goodbody, S. J., & Wolpert, D. M. (1998). Predicting the consequences of our own actions: The role of sensorimotor context estimation. *Journal of Neuroscience, 18,* 7511–7518.

Blakemore, S. J., Rees, G., & Frith, C. D. (1998). How do we predict the consequences of our actions? A functional imaging study. *Neuropsychologia, 36,* 521–529.

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport, 11,* R11–R16.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience, 1,* 635–640.

Burkard, R. F., Don, M., & Eggermont, J. J. (Eds.). (2006). *Auditory evoked potentials: Basic principles and clinical application* (1st ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America, 103,* 3153–3161.

Burnett, T. A., & Larson, C. R. (2002). Early pitch shift response is active in both steady and dynamic voice pitch control. *Journal of the Acoustical Society of America, 112,* 1058–1063.

Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F0 responses to pitch-shifted auditory feedback; a preliminary study. *Journal of Voice, 11,* 202–211.

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America, 121,* 1157–1163.

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe: II. Responses to the subjects own voice. *Experimental Brain Research, 77,* 476–489.

Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping, 9,* 183–191.

Donath, T. M., Natke, U., & Kalveram, K. T. (2002). Effects of frequency-shifted auditory feedback on voice F$_0$ contours in syllables. *Journal of the Acoustical Society of America, 111,* 357–366.

Eliades, S. J., & Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology, 89,* 2194–2207.

Eliades, S. J., & Wang, X. (2005). Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex, 15,* 1510–1523.

Eliades, S. J., & 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. *Journal of the Acoustical Society of America, 70,* 45–50.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45,* 152–170.

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology, 44,* 522–529.

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry, 158,* 2069–2071.

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. *American Journal of Psychiatry, 164,* 458–466.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39,* 350–365.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research, 130,* 133–141.

Hari, R., Aittoniemi, K., Jarvinen, M. L., Katila, T., & Varpula, T. (1980). Auditory evoked transient and sustained magnetic fields of the human-brain—Localization of neural generators. *Experimental Brain Research, 40,* 237–240.

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online, 6,* 131–137.

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology, 42,* 180–190.

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport, 17,* 1375–1379.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279,* 1213–1216.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech: I. Compensation and adaptation. *Journal of Speech, Language, and Hearing Research, 45,* 295–310.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience, 14,* 1125–1138.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics, 30,* 303–320.

Katahira, K., Abla, D., Masuda, S., & Okanoya, K. (2008). Feedback-based error monitoring processes during musical performance: An ERP study. *Neuroscience Research, 61,* 120–128.

Larson, C. R. (1998). Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control. *Journal of Communication Disorders, 31,* 489–503.

Liu, H., Zhang, Q., Xu, Y., & Larson, C. R. (2007). Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech. *Journal of the Acoustical Society of America, 122,* 2405–2412.

Martikainen, M. H., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex, 15,* 299–302.

McCarthy, G., & Donchin, E. (1976). The effects of temporal and event uncertainty in determining the waveforms of the auditory event related potential (ERP). *Psychophysiology, 13,* 581–590.

Müller-Preuss, P., Newman, J. D., & Jürgens, U. (1980). Anatomical and physiological evidence for a relationship between the "cingular" vocalization area and the auditory cortex in the squirrel monkey. *Brain Research, 202,* 307–315.

Muller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Research, 215,* 61–76.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound—A review and an analysis of the component structure. *Psychophysiology, 24,* 375–425.

Natke, U., & Kalveram, K. T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language, and Hearing Research, 44,* 577–584.

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters, 272,* 29–32.

Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters, 265,* 119–122.

Poulet, J. F., & Hedwig, B. (2002). A corollary discharge maintains auditory sensitivity during sound production. *Nature, 418,* 872–876.

Poulet, J. F., & Hedwig, B. (2006). The cellular basis of a corollary discharge. *Science, 311,* 518–522.

Poulet, J. F., & Hedwig, B. (2007). New insights into corollary discharges mediated by identified neural pathways. *Trends in Neurosciences, 30,* 14–21.

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *Journal of the Acoustical Society of America, 120,* 966–977.

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America, 119,* 2288–2297.

Schafer, E. W., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science, 181,* 175–177.

Sivasankar, M., Bauer, J. J., Babu, T., & Larson, C. R. (2005). Voice responses to changes in pitch of voice or tone auditory feedback. *Journal of the Acoustical Society of America, 117,* 850–857.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America, 122,* 2306–2319.

Wolpert, D. M. (1997). Internal models in human motor control: A computational and psychophysical perspective. *Journal of Physiology (London), 501P,* S12.

Xu, Y., Larson, C., Bauer, J., & Hain, T. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *Journal of the Acoustical Society of America, 116,* 1168–1178.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage, 40,* 1871–1887.