Published in final edited form as:

Clin Neurophysiol. 2009 July ; 120(7): 1303–1312. doi:10.1016/j.clinph.2009.04.022.

Vocalization-Induced Enhancement of the Auditory Cortex Responsiveness during Voice F_0 Feedback Perturbation

Roozbeh Behroozmand, Laura Karvelis, Hanjun Liu, and Charles R. Larson*

Speech Physiology Laboratory, Department of Communication Science and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208, USA

Abstract

Objective—The present study investigated whether self-vocalization enhances auditory neural responsiveness to voice pitch feedback perturbation and how this vocalization-induced neural modulation can be affected by the extent of the feedback deviation.

Method—Event related potentials (ERPs) were recorded in 15 subjects in response to +100, +200 and +500 cents pitch-shifted voice auditory feedback during active vocalization and passive listening to the playback of the self-produced vocalizations.

Result—The amplitude of the evoked P_1 (latency: 73.51 ms) and P_2 (latency: 199.55 ms) ERP components in response to feedback perturbation were significantly larger during vocalization than listening. The difference between P_2 peak amplitudes during vocalization vs. listening was shown to be significantly larger for +100 than +500 cents stimulus.

Conclusion—Results indicate that the human auditory cortex is more responsive to voice F_0 feedback perturbations during vocalization than passive listening. Greater vocalization-induced enhancement of the auditory responsiveness to smaller feedback perturbations may imply that the audio-vocal system detects and corrects for errors in vocal production that closely match the expected vocal output.

Significance—Findings of this study support previous suggestions regarding the enhanced auditory sensitivity to feedback alterations during self-vocalization, which may serve the purpose of feedback-based monitoring of one's voice.

Keywords

Electroencephalography; Event-Related Potentials; Audio-Vocal Integration; Speech Motor Control; Pitch-Shifted Auditory Feedback; Auditory Cortex; Vocalization

1. Introduction

The control of voice fundamental frequency (F_0) plays an important role in speech production and contributes to efficient transmission of linguistic and non-linguistic cues for human communication. Two major controlling mechanisms have been proposed to be involved in voice F_0 regulation (Guenther, 2006, Guenther et al., 2006). First, the feed-forward motor

*Corresponding Author: Charles R. Larson, E-mail: E-mail: claron@northwestern.edu, Address: Frances Searle Building, 2240 Campus Drive, Room 3-247, Evanston, IL 60208-2952, Tel: 847-491-2424.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

system adjusts the biomechanical parameters of the laryngeal muscles through previously learned motor commands. Second, sensory feedback (kinesthetic and auditory) arising from vocalization is used to minimize mismatch (error) between the intended and actual vocal output. The integration of feed-forward and feedback control mechanisms involves voice F_0 regulation through error-induced corrective commands that update the internal models of the vocal motor system.

Numerous studies have used a pitch perturbation technique to investigate the role of auditory feedback in voice F_0 control (Elman, 1981, Burnett et al., 1997, Burnett et al., 1998, Larson, 1998, Hain et al., 2000, Natke and Kalveram, 2001, Burnett and Larson, 2002, Donath et al., 2002, Jones and Munhall, 2002, Bauer and Larson, 2003, Xu et al., 2004, Sivasankar et al., 2005, Chen et al., 2007). Results of these studies have shown that applying real-time pitch shifts to the voice auditory feedback evokes compensatory vocal responses that oppose the direction of the pitch-shifted stimulus (PSS) to maintain F_0 at a desired level. Moreover, other studies have shown that the role of auditory feedback in voice regulation is not limited to pitch compensation. In a series of sensorimotor adaptation experiments it was shown that subjects compensated for shifts in formant frequency of the auditory feedback during vowel production (Houde and Jordan, 1998, Houde and Jordan, 2002, Purcell and Munhall, 2006b, 2006a, Villacorta et al., 2007). It has also been shown that applying loudness shifts to voice auditory feedback leads to vocal responses that compensate for alteration in voice intensity feedback (Heinks-Maldonado and Houde, 2005, Bauer et al., 2006, Liu et al., 2007). Evidence from all these studies suggests that auditory feedback plays an important role in online regulation of voice and for updating an internal representation of the mapping between auditory feedback and the motor control system.

In efforts to understand the neural mechanisms underlying voice control in humans, magneto-encephalographic (MEG) (Numminen and Curio, 1999, Numminen et al., 1999, Houde et al., 2002, Heinks-Maldonado et al., 2006) and electro-encephalographic (EEG) (Heinks-Maldonado et al., 2005) recordings of the brain activity have been obtained during vocalization. Results of these studies showed that the auditory cortex responses to normal and pitch-shifted voice feedback were attenuated during speaking compared to listening, a phenomenon often referred to as motor-induced suppression (MIS). It was suggested that MIS dampens sensory input from self-initiated actions in order to help distinguish between self and externally generated sounds. The “re-afference hypothesis” has been proposed to explain this phenomenon through integration of the internal feed-forward speech production mechanism (Wolpert, 1997, Blakemore et al., 1998) and auditory feedback system (Houde et al., 2002, Heinks-Maldonado et al., 2005, Guenther et al., 2006). This theory suggests that, during speaking, re-afference projections from motor cortex transmit the neural representation of the intended vocal output to the auditory cortical areas. The comparison between the efference copy (corollary discharge) and the sensory consequences of a self-produced vocalization in auditory error cells has been proposed to provide a mechanism to detect and correct for vocal errors in order to stabilize voice F_0 . Therefore, it was thought that MIS during speaking might result from attenuated auditory responsiveness to the incoming sensory feedback that is represented by the re-afference projections (Heinks-Maldonado et al., 2005). The notion of suppression due to subtractive comparison between the incoming auditory feedback information and the efference copies of the intended vocal motor commands is further supported by showing that the cortical auditory neural responses to self-vocalizations are most suppressed during normal feedback compared to conditions where the feedback was either pitch-shifted or modified with an alien’s voice (Heinks-Maldonado et al., 2005, Heinks-Maldonado et al., 2006). This maximum suppression is thought to occur because the auditory feedback information is maximally represented by the efference projections when the vocal output closely matches its normal auditory feedback.

With relevance to this discussion, projections from the cingular vocalization area were shown to attenuate neural responses to acoustical stimulation in the superior temporal gyrus (STG) in squirrel monkeys (Müller-Preuss et al., 1980, Muller-Preuss and Ploog, 1981). The vocalization-induced inhibition in primate auditory cortex was further reported in later studies in marmoset monkeys (Eliades and Wang, 2003, 2005). However, results of a recent study showed that during normal feedback of self-vocalization, some neurons increased and others decreased their discharge rate (Eliades and Wang, 2008). Those auditory neurons that were suppressed during normal voice feedback showed a larger increase in their firing rate in response to the PSS than those neurons that were excited. Based on this observation, it was suggested that the vocalization-induced suppression enhanced neural sensitivity to feedback perturbation. The absence of neural sensitivity enhancement during passive listening indicated that the vocalization-induced increase of the neural firing rates during feedback perturbation could not be simply explained by the frequency-specificity (tonotopic map) of the cortical auditory neurons. This evidence suggested that the vocal motor system might be involved in modulation of neural sensitivity by fine-tuning the auditory neurons (e.g. increasing their dynamic range) that respond more vigorously to the alteration in feedback.

It can be inferred from the results of the latter experiment (Eliades and Wang, 2008) that the neural response to the pitch-shifted voice feedback might be affected by two factors during active vocalization: 1) vocalization-induced suppression and 2) suppression-induced enhancement of neural responses to feedback perturbation. Therefore, for the feedback perturbation that co-occurs with the onset of the vocalization and persists during vocal production, the scalp-recorded brain potentials may not accurately reflect the enhanced neural responses to the feedback perturbation stimuli because these responses may be masked by the suppression due to vocalization onset. This masking effect might be a possible explanation for the previously reported decrease in the auditory responses to voice pitch feedback perturbation during vocalization compared to listening (Numminen and Curio, 1999, Numminen et al., 1999, Houde et al., 2002, Heinks-Maldonado et al., 2006). In the present study, we tested the hypothesis that self-vocalization enhances the auditory cortex responsiveness to unexpected, brief (200 ms duration) voice pitch feedback perturbation by presenting PSS to the subjects 500 ms following vocalization onset. This method temporally separates the effects of pitch-shifted voice feedback and vocalization onset and allows the recording of the neural responses to pitch perturbations independent of vocalization onset.

We further asked the question of whether vocalization-induced modulation of the neural responsiveness can be affected by the extent of the feedback deviation. Presenting pitch-shift stimuli at different magnitudes allowed us to investigate whether the auditory cortex is most sensitive to perturbations in feedback that closely matched the intended vocal output. It was previously shown that large magnitude pitch-shift stimuli produce more responses that follow the direction of the stimulus rather than opposing it (Burnett et al., 1998). One interpretation of this finding is that large magnitude stimuli are perceived to come from another speaker rather than self-produced vocalization, and would therefore not require a corrective response. More recently, it was shown that smaller voice F_0 shifts in feedback would evoke proportionally greater compensatory vocal responses when the vocal response magnitudes are expressed as the percentage of the feedback perturbation stimulus magnitude (Liu and Larson, 2007). This observation suggests that the audio-vocal system might be more sensitive to smaller feedback deviations during voice F_0 control. Greater sensitivity to smaller feedback deviations may allow the audio-vocal system to apply robust control and error correction over self-produced vocalization compared to the disruptive effect of external sounds during vocal production.

2. Materials and Methods

Subjects

Fifteen right-handed native speakers of American English (10 females and 5 males, 18–27 years of age; mean age=20.73 and std=2.55) with no history of speech, hearing or neurological disorders, and no voice training participated in the study. All study procedures, including recruitment and data acquisition of informed consent were approved by the Northwestern University institutional review boards and subjects were monetarily compensated for their participation.

Procedure

In this study, the neural mechanisms involved in auditory feedback-based control of voice F_0 were investigated using ERPs. The ERPs were recorded in response to the pitch-shifted auditory feedback stimulus during active production of the vowel sound /a/ and passive listening to the playback of the self-generated vocalization.

The experiment was comprised of a total of 6 different conditions (3 vocalization and 3 listening) in which the PSS magnitude was manipulated at +100 cents, +200 cents and +500 cents. For all subjects, the order of the stimulus magnitude presentation was randomly chosen and each vocalization condition was followed by the passive listening to the playback of the same utterance. During each vocalization condition, subjects were asked to vocalize the vowel sound 25 times, while maintaining a steady pitch and voice intensity. The maximum duration of each vocalization was approximately 6–7 seconds and 8 PSS were presented during each vocalization. Thus, a total of 200 ERP trials were collected and analyzed for each condition. The duration of the PSS was set to 200 ms and the inter-stimulus interval (ISI) was randomized between 500–900 ms. The ISI was randomized in order to make the subjects unable to anticipate the onset of the auditory feedback perturbation.

Instrumentation

Subjects were seated in a sound-treated room in which their voices were recorded with an AKG boomset microphone (model C420), amplified with a Mackie mixer (model 1202), and pitch-shifted through an Eventide Eclipse Harmonizer. The pitch-shifted auditory feedback was heard through Etymotic earphones (model ER1-14A) inserted into the subject's ear canal. The gain between the subject's voice and feedback was further manipulated with a Crown amplifier (D75) and HP350 dB-attenuators to +10 dB SPL, calibrated with a Zwisklocki coupler and a Brüel & Kjær sound level meter (model 2250). The 10 dB gain between the voice and feedback channels allowed the recording of ERPs in response to pitch-shifted auditory feedback by masking the normal voice feedback through bone and airborne conductions. Subjects were instructed to maintain their vocalization as long as they heard the perturbations (taking about 6–8 seconds depending on the randomized inter-stimulus-interval) in their own auditory feedback and to take short breaks (2–3 seconds) between successive vocalizations. During each vocalization condition, the pitch-shifted feedback channel was recorded and then converted to a sound file to be played back during the passive listening condition. The gain of the feedback channel during passive listening was carefully calibrated to about the same level (80–85 dB) at which subjects heard the feedback of their own voice during active vocalization.

All parameters of the pitch-shifting auditory feedback stimulus, such as duration, magnitude, and ISI were controlled by MIDI software (Max/MSP v.4.1. by Cycling 74). 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 electroencephalogram (EEG) signals were recorded from 13 sites on the subject's scalp (C_Z, C₃, C₄, T₃, T₄, F_Z, F₃, F₄, F₇, F₈, P_Z, P₃, P₄) using an Ag-AgCl electrode EEG cap (10–20 system). Scalp-recorded brain potentials were amplified with 13 Grass amplifiers (Grass P511 AC amplifier), 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 online 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 Grass impedance meter (Model: EZM-5AB) and maintained below 5 kΩ. The effect of the visual and muscle artifact on the recorded brain potentials were reduced by instructing the subjects to close their eyes and relax their muscles throughout the course of the experiment.

ERP Signal Analysis

The evoked ERPs were obtained by averaging the recorded EEG signals, synchronized to the onset of the pitch-shifted stimulus. Prior to the synchronized averaging, the EEG signals from all channels were subjected to offline filtering, using a band-pass filter with cut-off frequencies set to 1 and 30 Hz. The filtered EEGs were then cut into epochs ranging from –100 ms to 500 ms, relative to the onset of the PSS. The epochs with amplitudes exceeding $\pm 50 \mu\text{V}$ were excluded from the data analysis. The baseline (pre-stimulus mean amplitude) of the remaining epochs was removed, and the ERPs were obtained for conditions with a minimum number of 100 epochs.

The latency and amplitude of the P₁-N₁-P₂ complex were extracted from the averaged neural responses by finding the most prominent peaks in 50 ms-long time windows centered at 50 ms, 100 ms and 200 ms. In addition to the extracted peak amplitudes and latencies, the normalized difference index (NDI) was calculated as the difference between response peak amplitudes during active vocalization vs. passive listening, normalized to the sum of peak amplitudes for each individual neural component. The NDI is a measure that reflects the changes in the stimulus-evoked auditory neural responses as a result of vocal-motor system activity. In other words, NDI is an indication of vocalization-induced changes in the neural responsiveness to feedback alteration and can be calculated using the following formula:

$$NDI = \frac{|M_{voc} - M_{lis}|}{|M_{voc}| + |M_{lis}|}$$

In which M_{voc} and M_{lis} are the amplitude of the averaged evoked neural components during active vocalization and passive listening conditions, respectively.

Topographical Distribution Maps

The surface distribution maps of measures of brain activity in response to voice pitch feedback perturbation are created using the peak amplitudes of the neural responses for 13 electrode sites (C_Z, C₃, C₄, T₃, T₄, F_Z, F₃, F₄, F₇, F₈, P_Z, P₃, P₄) over the surface of the scalp. These topographical distribution maps of the neural activity are created by color coding the amplitudes of the ERP components using the interpolation method between adjacent electrodes to obtain a fine map of electrical activity distribution.

Statistical Analysis

The extracted latencies and amplitudes of the P₁, N₁ and P₂ peaks in the averaged neural responses were separately analyzed for 10 recording sites using a four-way repeated-measures analysis of variance (ANOVA). The analyzed electrode sites included C₃, C₄, T₃, T₄, F₃, F₄,

F₇, F₈, P₃ and P₄. Repeated-measures ANOVAs were conducted to examine effects of condition (active vocalization, passive listening), stimulus magnitude (100 cents, 200 cents, 500 cents), laterality (Left: C₃, T₃, F₃, F₇, P₃ vs. Right: C₄, T₄, F₄, F₈, P₄) and electrode site (Centro-Medial: C₃, C₄ – Temporal: T₃, T₄ – Fronto-Medial: F₃, F₄ – Fronto-Lateral: F₇, F₈ – Parieto-Medial: P₃, P₄), and their interactions on the latencies and amplitudes of the P₁-N₁-P₂ cortical neural components.

3. Results

Figure 1 shows the grand average (all subjects) ERP responses as a function of the stimulus magnitude during active vocalization and passive listening conditions. Figure 2 shows the overlaid ERP responses during active vocalization and passive listening conditions across three PSS magnitudes. The repeated-measures ANOVA on the P₁ amplitude revealed a significant main effect of condition, $F(1,14)=6.39$, $p=0.024$, stimulus magnitude, $F(2,28)=5.74$, $p=0.008$, electrode position, $F(4,56)=9.51$, $p=0.000$, as well as laterality $\times$ position interaction, $F(4,56)=2.55$, $p=0.049$. The significant main effect of the condition factor on the P₁ amplitude indicated that the PSS evoked larger P₁ responses during the active vocalization than the passive listening condition.

Post-hoc tests using a Bonferroni adjustment for multiple comparisons of P₁ mean amplitudes across three different PSS magnitudes revealed that the P₁ amplitudes for 500 cents significantly differed from 100 cents ($p=0.037$) and 200 cents ($p=0.017$) stimuli. These findings indicated that 500 cents pitch-shifts elicited larger P₁ responses compared to two other smaller stimuli. No significant difference between P₁ amplitudes was found for 100 cents and 200 cents stimuli. Post-hoc tests for the electrode position main effect revealed significant positivity in the frontal and frontal-medial regions for the P₁ amplitudes ($p=0.003$) compared with other electrode sites. Figure 3a illustrates the topographical distribution of the P₁ peak amplitudes over the surface of the scalp, grand averaged over three PSS magnitudes during both vocalization and passive listening conditions. Figure 4 (top row) also shows the P₁ amplitude distributions in more detail for different stimulus magnitudes and conditions separately. For the grand averaged P₁ amplitudes over three PSS magnitudes and both vocalization and listening conditions, significant laterality $\times$ position interaction indicated that the pattern of P₁ amplitude distribution was different for left vs. right hemispheres. Significant differences were observed for P₁ amplitudes between centro-medial vs. parieto-medial ($p_{Left}=0.002$, $p_{Right}=0.025$) and fronto-medial vs. parieto-medial ($p_{Left}=0.008$, $p_{Right}=0.003$) electrode sites on either side, however, the difference in the distribution pattern was due to the significant difference in P₁ amplitudes between fronto-medial vs. temporal on the left ($p=0.003$) instead of fronto-lateral vs. parieto-medial sites on the right ($p=0.043$).

The statistical test on the N₁ peak amplitude only revealed a significant main effect of electrode position, $F(4,56)=13.5$, $p<0.001$. Post-hoc tests for the electrode position main effect on the grand averaged N₁ amplitudes for all three PSS magnitudes during both vocalization and passive listening conditions revealed significant negativity in the frontal and fronto-medial regions ($p=0.001$) compared with other electrode sites. Figure 3b illustrates the topographical distribution of the N₁ peak amplitudes over the surface of the scalp, grand averaged over three PSS magnitudes during both vocalization and passive listening conditions. Figure 4 (middle row) also shows the N₁ amplitude distributions in more detail for different stimulus magnitudes and conditions separately.

For the P₂ amplitude, the repeated-measures ANOVA revealed a significant main effect of condition, $F(1,14)=5.99$, $p=0.028$, stimulus magnitude, $F(2,28)=18.04$, $p<0.001$, electrode position, $F(4,56)=12.86$, $p<0.001$, as well as magnitude $\times$ position, $F(8,112)=5.78$, $p<0.001$, and laterality $\times$ position interactions, $F(4,56)=3.83$, $p=0.008$. The significant main effect of

condition on P_2 amplitude indicated that the PSS elicited larger P_2 responses during the active vocalization than passive listening. Post-hoc tests using a Bonferroni adjustment revealed that the P_2 amplitudes significantly differed for 100 cents vs. 200 cents ($p=0.036$), 200 cents vs. 500 cents ($p=0.035$) and for 100 cents vs. 500 cents ($p=0.000$). The significant difference between P_2 amplitudes for all pairs of the stimulus magnitudes indicates that the shifts in the auditory feedback of voice F_0 elicited graded P_2 responses that become larger as the pitch-shifted magnitude increased. Post-hoc tests for the electrode position main effect revealed significant positivity at the central and centro-medial regions for the P_2 amplitudes ($p=0.001$) compared with other electrode sites. Figure 3c illustrates the topographical distribution of the P_2 peak amplitudes over the surface of the scalp, grand averaged over three PSS magnitudes during both vocalization and passive listening conditions. Figure 4 (bottom row) also shows the P_2 amplitude distributions in more detail for different stimulus magnitudes and conditions separately. For the grand averaged P_2 amplitudes over three PSS magnitudes and both vocalization and listening conditions, significant laterality $\times$ position interaction indicated that the pattern of P_2 amplitude distribution was different for left vs. right hemispheres. Significant differences were observed for P_2 amplitudes between centro-medial vs. parieto-medial ($p_{Left}=0.007$, $p_{Right}=0.013$) and temporal ($p_{Left}=0.000$, $p_{Right}=0.002$) and between fronto-medial vs. fronto-lateral ($p_{Left}=p_{Right}=0.000$) and temporal ($p_{Left}=0.001$, $p_{Right}=0.019$) electrode sites on either side, however, the P_2 amplitude was only significantly different between centro-medial vs. fronto-lateral ($p=0.003$) electrodes on the left side.

Significant magnitude $\times$ position interaction for the P_2 peak amplitudes indicated that the potential distribution pattern was different for different PSS magnitudes. Separate ANOVAs for stimulus magnitudes considering the condition, laterality and electrode position factors revealed a significant main effect of electrode position ($p=0.012$) with the largest positivity at the centro-frontal area for 100 cents stimulus, a significant main effect of electrode position ($p=0.000$) with the largest positivity at the centro-frontal area and also a significant laterality $\times$ position interaction ($p=0.015$) for 200 cents stimulus and finally a significant main effect of electrode position ($p=0.000$) with the largest positivity at the centro-frontal area for 500 cents stimulus.

Analysis of the P_1 , N_1 and P_2 latencies revealed no significant effects. The overall mean latencies of the P_1 - N_1 - P_2 complex were 73.51 ms (std=11.38 ms), 117.41 ms (std=13.43 ms) and 199.55 ms (std=23.49 ms), respectively.

Due to a significant main effect of condition on P_1 and P_2 amplitudes, the NDI was calculated and analyzed for these auditory neural components. A two-way repeated-measures ANOVA was conducted to investigate the main effects of the stimulus magnitude, laterality, and their interaction on P_1 and P_2 NDIs. Results of the statistical tests revealed no significant effects on the P_1 NDIs. Separate ANOVAs for the right and left sides also confirmed that there was no significant main effect of the stimulus magnitude on the P_1 NDIs (Figure 5a). The ANOVA for the P_2 NDIs only revealed a significant main effect of the stimulus magnitude, $F(2,28)=5.39$, $p=0.01$. Pair-wise comparison using Bonferroni adjustment showed that the main effect of the stimulus magnitude on P_2 NDI was only significant between 100 cents and 500 cents ($p=0.018$). Separate ANOVAs also showed that the P_2 NDIs significantly differed only across 100 cents and 500 cents for both the left ($F(2,28)=4.2$, $p=0.025$) and right ($F(2,28)=5.07$, $p=0.013$) sides (Figure 5b). The significant main effect of the stimulus magnitude on the P_2 NDIs indicated that, although shifts in voice F_0 during active vocalization elicited larger P_2 responses, the difference between P_2 peak amplitudes during active vocalization and passive listening conditions were smaller for larger pitch shifts.

4. Discussion

A recent study in monkeys suggested that the previously reported auditory suppression (Numminen and Curio, 1999, Numminen et al., 1999, Houde et al., 2002, Heinks-Maldonado et al., 2006) can lead to increased neural sensitivity to feedback perturbation during vocalization (Eliades and Wang, 2008). The evidence for such an effect came from the results of single neuron recordings showing that the suppressed auditory neurons during normal voice feedback exhibited a larger increase in their firing rates in response to PSS than neurons that were excited during normal voice feedback. The present study investigated this effect by recording ERPs in humans to test the hypothesis that self-vocalization enhances auditory responsiveness to voice F_0 feedback perturbation.

The suppression-induced enhancement of auditory sensitivity to alteration in voice F_0 feedback during vocalization (Eliades and Wang, 2008) implies a complex neural mechanism that involves internal modulation (re-auditory projections) as well as responses to changes in feedback. Therefore, the auditory responses to feedback perturbation can be studied accurately only if the vocalization-induced modulation (suppression) does not occur at the same time as the perturbation. In the current study we proposed that presenting the PSS after the onset of vocalization can reveal neural responsiveness to perturbations in self-vocalization without the confounding influence of the simultaneous suppression at the vocalization onset. Thus, the recorded ERPs in response to PSS that occur after a time interval with respect to the vocalization onset might reduce or eliminate the previously reported suppression effect (Heinks-Maldonado et al., 2005) from the neural responses to auditory feedback pitch perturbations.

Results of the analysis revealed that P_1 and P_2 peak amplitudes were larger during vocalization compared to listening to the same vocalizations, indicating that the auditory cortex is more responsive to feedback perturbations during active vocalization than during passive listening (Figure 2). The increase in P_1 and P_2 peak amplitudes during vocalization suggests that the proposed feed-forward model (Wolpert, 1997, Blakemore et al., 1998) can enhance auditory responsiveness to feedback perturbations during speaking. This effect might be an important characteristic of the speech motor control system that allows for accurate detection and correction of unintended changes in the vocal output during speech production. In fact, the observed vocalization-induced enhancement of neural responsiveness suggests that the hypothesized re-auditory projection might be involved in fine-tuning the auditory cortex to improve its ability to detect feedback deviations during online voice regulation. This effect was discussed by Eliades et al. (Eliades and Wang, 2008) to result from changes in the tuning properties of the auditory neurons (e.g. increasing their dynamic range) as a result of vocal motor system activity. Although the neural mechanisms of such a phenomenon are poorly understood, the vocalization-induced increase in auditory sensitivity (Eliades and Wang, 2008) and responsiveness (current study) might be an indication of enhanced neural processing of sensory information to mediate feedback-based monitoring and control of voice F_0 .

Results of the current experiment did not show any significant difference between N_1 peak amplitudes across vocalization and passive listening conditions. We suggest that this inconsistency with the previous findings (Numminen and Curio, 1999, Numminen et al., 1999, Houde et al., 2002, Heinks-Maldonado et al., 2006) may be due to differences between experimental procedures in these studies. The temporal separation of vocalization and PSS onsets in the current experiment allows for independent recording of the neural responses to the onset of the feedback perturbation, whereas in the previous experiments, the neural responses to feedback perturbation may have been masked by the suppression of the auditory cortex at the onset of vocalization. In the previous experiments, the coinciding vocalization and feedback perturbation onsets caused alteration of subject's auditory feedback as soon as vocalization was initiated. Therefore, the post-stimulus neural activity during vocalization

consisted of the suppressive effect of the vocal motor system on the auditory neurons and the changes in auditory cortex neural activity in response to the feedback alteration. In other words, the post-stimulus responses in those cases reflected the sum of changes in the neural activity as a result of the vocalization-induced inhibition (suppression) and stimulus-evoked excitation reported by Eliades et al. (Eliades and Wang, 2008).

The absence of the vocalization-induced suppression in the current experiment does not challenge or reject the previous findings regarding the suppression of auditory cortex during vocalization. Although the recorded ERPs in the current experiment might not reflect the suppression effect during vocalization onset, we propose that the previously reported auditory suppression (Numminen et al., 1999, Houde et al., 2002, Heinks-Maldonado et al., 2006) causes the enhancement of neural responsiveness observed in the current study. This proposal is based on the recent finding in primate auditory cortex that during normal feedback only the suppressed neurons exhibited increased neural activity to feedback perturbation (Eliades and Wang, 2008). Vocalization-induced suppression was hypothesized to be responsible for adjusting tuning properties of the auditory neurons in order to increase their sensitivity to unexpected feedback alteration during vocalization (Eliades and Wang, 2008).

The effect of the PSS magnitude on auditory neural responses was also investigated during active vocalization and passive listening to the playback. The results of the analysis showed that manipulation of the PSS magnitude modulates the amplitudes of the P₁ and P₂ neural components during both conditions. The increase in the amplitudes of the P₁ and P₂ peaks for larger pitch shifts indicates that larger voice F₀ feedback deviations increase the auditory cortex responsiveness during both vocalization and passive listening. This modulation suggests that the underlying neural substrates giving rise to these two peaks might be involved in detecting and processing the perturbation magnitude in voice F₀ feedback. The effect of the stimulus magnitude on neural responses was more pronounced in P₂ for which the response amplitudes were graded according to the magnitudes of the PSS (Figure 1). P₁ response magnitudes were significantly different between 500 cents stimuli and the two other stimuli, but no significant difference between the P₁ responses to 100 cents and 200 cents stimuli was found. These observations suggest that larger feedback deviations might be needed to modulate the responsiveness of the neural substrates giving rise to P₁.

The condition (vocalization vs. listening) sequence could also explain the apparent vocalization-induced enhancement of the neural responsiveness to feedback perturbation. The sequence effect arises because the subjects experienced the vocalization condition before passive listening for each PSS magnitude and therefore, this could have resulted in the stimulus being novel during the vocalization condition, but more familiar during the listening condition. However, this factor is unlikely to fully account for the observed effect because the stimulus would have been novel during vocalization only for the first stimulus magnitude but not for all three of them. Moreover, results showed the enhanced neural responsiveness for all three PSS magnitudes when the order of stimulus magnitude presentation was randomized across different subjects.

In addition to the significant main effect of the PSS magnitude and condition factors on P₁ and P₂ peaks amplitudes, results revealed that there was a significant main effect of the electrode position factor on the neural peak amplitudes for P₁-N₁-P₂ complex. This effect was observed as a significant positivity that started in the frontal and fronto-medial region for P₁ followed by a significant negativity for N₁ in about the same region that eventually moved posteriorly toward the centro-frontal region as a significant positivity for the P₂ peak amplitudes (Figures 3 and 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 might contribute

to auditory P₁-N₁-P₂ components that reflect auditory processing at multiple stages. The P₁ component has been proposed to have dominant generators in the primary auditory cortex located in the superior temporal gyrus (STG) where the early cortical detection of any change in the auditory system occurs (Burkard et al., 2006). The N₁ 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 et al., 1980) and in cortical frontal areas (Naatanen and Picton, 1987). The N₁ is considered as an index of the pre-attentive auditory processing that reflects mechanisms underlying automatic change detection based on the comparison of the incoming auditory feedback information with the memory trace of the previous sensory input to the auditory system (Naatanen, 1991). A recent study on feedback-based error monitoring during musical performance has suggested that the P₂ component might reflect the process of auditory mismatch detection (Katahira et al., 2008) during the performance of a motor task and is assumed to arise from the anterior cingulate cortex (ACC), triggered by the basal ganglia when subjects notice their own motor error. This assumption is supported by a fMRI study that investigated the neural substrates of vocal pitch regulation during singing (Zarate and Zatorre, 2008). However, the P₂ component during the passive listening condition in the present study can be interpreted as the 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 P₂ component during motor tasks (Folstein and Van Petten, 2008). Higher responsiveness to larger feedback deviations along with the modulation of the P₁ and P₂ peak amplitudes in the current study suggested that the neural substrates giving rise to these potentials are involved in detection of auditory perturbation magnitude, and their sensitivity to changes in the auditory feedback can be modulated by the vocal motor system during voice F₀ control.

Although the P₁ and P₂ peak magnitudes were shown to be modulated by PSS magnitude during both vocalization and listening conditions, the significant main effect of condition revealed that the neural responses were greater during vocalization than listening. The NDI was then calculated as the difference between response peak amplitudes for vocalization and listening to measure the vocalization-induced increase in the auditory responsiveness to feedback perturbation. In addition, NDIs were compared across different voice F₀ feedback perturbation levels to investigate how vocalization-induced enhancement of neural responsiveness can be affected by the extent of feedback deviation. Results revealed a significant main effect of the PSS magnitude on P₂ NDIs for 100 cents and 500 cents stimuli, indicating that the vocalization-induced increase in neural responsiveness is greater for the smaller feedback perturbation magnitude (Figure 5). This finding suggests that the effect of the self-vocalization on the human auditory system is greater when the deviated feedback signal is closer to one's own voice F₀. In other words, even very small deviations around a target pitch level might be significant from the system's standpoint, whereas a larger deviation might be negligible because it is recognized as not belonging to the speaker. This important characteristic of the system may enable it to apply corrective motor commands to correct for deviations only if the feedback signal is distinguished as self. Therefore, increasing feedback deviation might systematically reduce neural responsiveness to perturbations in the feedback signal because that signal does not closely match the intended output (highly deviated self or externally-generated). However, although suppression is suggested to play a major role in fine-tuning the auditory cortex, the way that motor induced suppression and increased sensitivity to perturbations can relate to each other is not clearly understood.

Larger increases in neural responsiveness to smaller feedback alterations, measured using the normalized difference between neural peak magnitudes during active vocalization and passive listening, indicated that the integration of audio-vocal modalities in humans might provide a mechanism for stabilizing the voice against small perturbations. As a result, higher neural sensitivity to feedback perturbation during active vocalization might indicate that the audio-

vocal system applies robust control over the structure of the self-produced voices against the disruptive effect of the external sounds. This suggestion is supported by a recent finding that human subjects showed proportionally greater compensation to smaller changes in voice F_0 feedback when the magnitude of the vocal responses were expressed as the percentage of the PSS stimulus magnitude (Liu and Larson, 2007). Moreover, greater relative responsiveness to smaller deviations might indicate that the speech motor control system largely corrects for the disparity between one's own voice and its feedback (error) as long as the feedback signal is recognized to be produced by the self (no or small unintended deviation). Therefore, it is possible that distinguishing between self- and externally-generated sounds would be a critical issue for the audio-vocal system during speaking.

5. Conclusion

In the present study, the effect of voice F_0 feedback perturbation was investigated on the cortical neural responses during active vocalization and passive listening conditions. Results showed that the auditory cortex is more responsive to changes in auditory feedback during vocalization than when passively listening to the playback of their own voice. This phenomenon was suggested to result from the internal modulation of auditory neurons (re-afference theory) that provides a mechanism to increase the neural sensitivity to detect and help the vocal motor system correct for deviations in self-generated voice feedback. The vocalization-induced enhancement of neural responsiveness was also shown to be more effective (greater increase) for smaller F_0 perturbations in voice feedback. This finding suggests that the auditory cortex is more sensitive to smaller feedback alterations that result from natural fluctuations in self-produced vocalizations. It is possible that the audio-vocal system carries information about the predicted vocal output that enables it to distinguish between self and externally generated sounds. This effect might be an important characteristic of the vocal production system for monitoring self-vocalizations that helps to detect and correct for vocal errors during speech production.

Acknowledgments

This research was supported by a grant from NIH, Grant No. 1R01DC006243. Authors wish to thank Chun Liang Chan for his help with the computer programming.

References

- Bauer JJ, Larson CR. 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* 2003;114:1048–1054. [PubMed: 12942983]
- Bauer JJ, Mittal J, Larson CR, Hain TC. Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *Journal of the Acoustical Society of America* 2006;119:2363–2371. [PubMed: 16642849]
- Blakemore SJ, Rees G, Frith CD. How do we predict the consequences of our actions? A functional imaging study. *Journal of Cognitive Neuroscience* 1998;58–58.
- Burkard, RF.; Don, M.; Eggermont, JJ. Auditory Evoked Potentials: Basic Principles and Clinical Application. Lippincott Williams & Wilkins; 2006.
- Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F_0 Responses to Manipulations in Pitch Feedback. *Journal of the Acoustical Society of America* 1998;103:3153–3161. [PubMed: 9637026]
- Burnett TA, Larson CR. Early pitch shift response is active in both steady and dynamic voice pitch control. *Journal of the Acoustical Society of America* 2002;112:1058–1063. [PubMed: 12243154]
- Burnett TA, Senner JE, Larson CR. Voice F_0 Responses to Pitch-Shifted Auditory Feedback; A Preliminary Study. *Journal of Voice* 1997;11:202–211. [PubMed: 9181544]
- Chen SH, Liu H, Xu Y, Larson CR. Voice F_0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America* 2007;121:1157–1163. [PubMed: 17348536]

- Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on voice F₀ contours in syllables. *Journal of the Acoustical Society of America* 2002;111:357–366. [PubMed: 11831808]
- Eliades SJ, Wang X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology* 2003;89:2194–2207. [PubMed: 12612021]
- Eliades SJ, Wang X. Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex* 2005;15:1510–1523. [PubMed: 15689521]
- Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 2008;453:1102–1106. [PubMed: 18454135]
- Elman JL. Effects of frequency-shifted feedback on the pitch of vocal productions. *Journal of the Acoustical Society of America* 1981;70:45–50. [PubMed: 7264071]
- Folstein JR, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology* 2008;45:152–170. [PubMed: 17850238]
- Guenther FH. Cortical interactions underlying the production of speech sounds. *J Commun Disord* 2006;39:350–365. [PubMed: 16887139]
- Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 2006;96:280–301. [PubMed: 16040108]
- Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research* 2000;130:133–141.
- Hari R, Aittoniemi K, Jarvinen ML, Katila T, Varpula T. Auditory Evoked Transient and Sustained Magnetic-Fields of the Human-Brain - Localization of Neural Generators. *Experimental Brain Research* 1980;40:237–240.
- Heinks-Maldonado TH, Houde JF. Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online* 2005;6:131–137.
- Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. *Psychophysiology* 2005;42:180–190. [PubMed: 15787855]
- Heinks-Maldonado TH, Nagarajan SS, Houde JF. Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport* 2006;17:1375–1379. [PubMed: 16932142]
- Houde JF, Jordan MI. Sensorimotor adaptation in speech production. *Science* 1998;279:1213–1216. [PubMed: 9469813]
- Houde JF, Jordan MI. Sensorimotor adaptation of speech I: Compensation and adaptation. *J Speech Lang Hear Res* 2002;45:295–310. [PubMed: 12003512]
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience* 2002;14:1125–1138. [PubMed: 12495520]
- Jones JA, Munhall KG. The role of auditory feedback during phonation: studies of Mandarin tone production. *Journal of Phonetics* 2002;30:303–320.
- Katahira K, Ablat D, Masuda S, Okanoya K. Feedback-based error monitoring processes during musical performance: An ERP study. *Neuroscience Research* 2008;61:120–128. [PubMed: 18359117]
- Larson CR. Cross-modality influences in speech motor control: The use of pitch shifting for the study of F₀ control. *Journal of Communication Disorders* 1998;31:489–503. [PubMed: 9836138]
- Liu H, Larson CR. Effects of perturbation magnitude and voice F₀ level on the pitch-shift reflex. *J Acoust Soc Am* 2007;122:3671–3677. [PubMed: 18247774]
- Liu H, Zhang Q, Xu Y, Larson CR. Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech. *Journal of the Acoustical Society of America* 2007;122:2405–2412. [PubMed: 17902874]
- Müller-Preuss P, Newman JD, Jürgens U. Anatomical and physiological evidence for a relationship between the ‘cingular’ vocalization area and the auditory cortex in the squirrel monkey. *Brain Res* 1980;202:307–315. [PubMed: 7437905]
- Müller-Preuss P, Ploog D. Inhibition of auditory cortical neurons during phonation. *Brain Research* 1981;215:61–76. [PubMed: 7260601]
- Naatanen R. The Role of Attention in Auditory Information-Processing as Revealed by Event-Related Potentials and Other Brain Measures of Cognitive Function. *Behavioral and Brain Sciences* 1991;14:761–761.

- Naatanen R, Picton T. The N1 Wave of the Human Electric and Magnetic Response to Sound - a Review and an Analysis of the Component Structure. *Psychophysiology* 1987;24:375–425. [PubMed: 3615753]
- Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language, and Hearing Research* 2001;44:577–584.
- Numminen J, Curio G. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neurosci Lett* 1999;272:29–32. [PubMed: 10507535]
- Numminen J, Salmelin R, Hari R. Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters* 1999;265:119–122. [PubMed: 10327183]
- Purcell DW, Munhall KG. Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *Journal of the Acoustical Society of America* 2006a;120:966–977. [PubMed: 16938984]
- Purcell DW, Munhall KG. Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America* 2006b;119:2288–2297. [PubMed: 16642842]
- Sivasankar M, Bauer JJ, Babu T, Larson CR. Voice responses to changes in pitch of voice or tone auditory feedback. *Journal of the Acoustical Society of America* 2005;117:850–857. [PubMed: 15759705]
- Villacorta VM, Perkell JS, Guenther FH. Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America* 2007;122:2306–2319. [PubMed: 17902866]
- Wolpert DM. Internal models in human motor control: A computational and psychophysical perspective. *Journal of Physiology-London* 1997;501:S12–S12.
- Xu Y, Larson C, Bauer J, Hain T. Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *Journal of the Acoustical Society of America* 2004;116:1168–1178. [PubMed: 15376682]
- Zarate JM, Zatorre RJ. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage* 2008;40:1871–1887. [PubMed: 18343163]

Figure 1. Comparison of the evoked cortical neural responses to three different pitch-shifting stimulus magnitudes during **a)** active vocalization of the vowel sound /a/ and **b)** passive listening to the playback of the self-generated vocalization. The vertical dashed lines in each subplot mark the onset of the pitch-shifting stimulus.

Figure 2.

Comparison of the evoked cortical neural responses during active vocalization and passive listening conditions for **a)** 100 cents **b)** 200 cents and **c)** 500 cents pitch-shifting stimulus magnitudes. The vertical dashed lines in each subplot mark the onset of the pitch-shifting stimulus.

Figure 3.

Topographical distribution of the grand average auditory neural responses to pitch-shifted auditory feedback perturbation for **a)** P₁ (latency: 73.51 ms), **b)** N₁ (latency: 117.41 ms) and **c)** P₂ (latency: 199.55 ms) ERP components. The maps are calculated for 13 recording sites on the surface of the scalp (C_Z, C₃, C₄, T₃, T₄, F_Z, F₃, F₄, F₇, F₈, P_Z, P₃, P₄) and the neural responses are grand averaged over 15 subjects for three stimulus magnitudes (100, 200 and 500 cents) during active vocalization and passive listening conditions.

Figure 4.

Topographical distribution of the averaged auditory neural responses to pitch-shifted auditory feedback perturbation across 15 subjects. The maps are calculated for 13 recording sites on the surface of the scalp (C_Z, C₃, C₄, T₃, T₄, F_Z, F₃, F₄, F₇, F₈, P_Z, P₃, P₄) and the distributions of the P₁ (latency: 73.51 ms), N₁ (latency: 117.41 ms) and P₂ (latency: 199.55 ms) peak amplitudes are shown in three separate rows. The distribution topographies are arranged in six columns to show the maps for three stimulus magnitudes (100, 200 and 500 cents) across active vocalization and passive listening conditions separately.

Figure 5.

Comparison of normalized difference index (NDI) across three different levels of the pitch-shifting stimulus magnitude for **a)** P₁ and **b)** P₂ peak amplitudes on the left and right hemispheres separately.