Published in final edited form as:

Brain Res. 2016 April 1; 1636: 1–12. doi:10.1016/j.brainres.2016.01.040.

A Temporal Predictive Code for Voice Motor Control: Evidence from ERP and Behavioral Responses to Pitch-shifted Auditory Feedback

Roozbeh Behroozmand¹, Stacey Sangtian¹, Oleg Korzyukov², and Charles R. Larson²

¹

²

Abstract

The predictive coding model suggests that voice motor control is regulated by a process in which the mismatch (error) between feedforward predictions and sensory feedback is detected and used to correct vocal motor behavior. In this study, we investigated how predictions about timing of pitch perturbations in voice auditory feedback would modulate ERP and behavioral responses during vocal production. We designed six counterbalanced blocks in which a +100 cents pitch-shift stimulus perturbed voice auditory feedback during vowel sound vocalizations. In three blocks, there was a fixed delay (500, 750 or 1000 ms) between voice and pitch-shift stimulus onset (predictable), whereas in the other three blocks, stimulus onset delay was randomized between 500, 750 and 1000 ms (unpredictable). We found that subjects produced compensatory (opposing) vocal responses that started at 80 ms after the onset of the unpredictable stimuli. However, for predictable stimuli, subjects initiated vocal responses at 20 ms before and followed the direction of pitch shifts in voice feedback. Analysis of ERPs showed that the amplitudes of the N1 and P2 components were significantly reduced in response to predictable compared with unpredictable stimuli. These findings indicate that predictions about temporal features of sensory feedback can modulate vocal motor behavior. In the context of the predictive coding model, temporally-predictable stimuli are learned and reinforced by the internal feedforward system, and as indexed by the ERP suppression, the sensory feedback contribution is reduced for their processing. These findings provide new insights into the neural mechanisms of vocal production and motor control.

Keywords

Voice Motor Control; Auditory Feedback; Internal Forward Model; Predictive Code; Pitch-Shift Stimulus; Event-related Potential

Corresponding Author: Roozbeh Behroozmand, Ph.D., University of South Carolina, Department of Communication Sciences and Disorders, Keenan Building Rm 356, 1224 Sumter St., Columbia, SC, 29208, r-behroozmand@sc.edu, Tel: 803-777-5055, Fax: 803-777-3081.

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.

1. Introduction

Skilled motor behavior is driven by the effective integration of feedforward and sensory feedback mechanisms to achieve the optimal goals of performed actions (Wolpert and Ghahramani, 2000). This effectiveness is determined by the relative contribution and weighting of feedforward and feedback mechanisms to generate, monitor and control our movements (Wolpert et al., 2011, 1995). Among all actions, speaking is one of the most complex goal-directed motor behaviors developed to facilitate human communication. A widely-accepted idea hypothesizes that during speech production, a copy of the motor commands known as the efference copy (Wolpert and Flanagan, 2001) is translated by an internal forward model to provide predictions about sensory consequences of self-produced speech sounds (Hickok and Poeppel, 2007; Houde and Nagarajan, 2011). This process is part of a predictive coding model in which speech errors resulting from a mismatch between the internally-predicted and actual sensory feedback are used to monitor and correct subsequent motor behavior during speech production and control (Guenther et al., 2006; Hickok, 2012; Houde and Nagarajan, 2011; Tourville et al., 2008).

In recent years, a growing number of studies have been conducted to better understand the predictive coding mechanism as it relates to vocal production and motor control. An effect associated with the predictive coding has been consistently reported by showing that the N1 component of the auditory-evoked event-related potentials (ERPs) was suppressed during vocal production of speech sounds compared with passive listening to the playback of the same self-produced speech (Curio et al., 2000; Heinks-Maldonado et al., 2006, 2005; Houde et al., 2002). It has been proposed that this motor-induced suppression effect results from the cancellation of sensory neural responses to self-produced speech by the internal feedforward predictions during vocal production. This notion was further supported by a study showing that the suppression was maximum for normal voice auditory feedback and was reduced or almost completely eliminated when a pitch-shift stimulus created mismatch between the internal predictions and the auditory feedback during vocalization (Behroozmand and Larson, 2011; Heinks-Maldonado et al., 2006). Additionally, a study by Wang et al. (Wang et al., 2014) showed that the activation of inferior frontal gyrus (IFG) at about 300 ms before speaking is associated with the suppression of N1 responses in auditory cortex at about 100 ms following speech onset. Thus, it was concluded that the transmission of predictive codes from motor-related areas such as IFG is responsible for the suppression of neural activity in the auditory cortex during speaking. These findings suggest that the motor-driven feedforward internal predictions play a key role in achieving the communication goals during vocal production and motor control.

Converging evidence from more recent studies has suggested that predictions about different aspects of sensory feedback stimuli subsequently affect behavioral and neural responses during vocal production and motor control. In the study by Scheerer & Jones (Scheerer and Jones, 2014), behavioral vocal responses to predictable and unpredictable pitch-shift stimulus magnitude were examined and they reported that the magnitude of vocal responses was significantly reduced for predictable vs. unpredictable stimuli. Behroozmand et al. (Behroozmand et al., 2012) and Korzyukov et al. (Korzyukov et al., 2012) examined the effect of pitch-shift stimulus direction predictability and found that the magnitude of

opposing (compensatory) vocal responses to unpredictable stimulus direction was significantly larger than following responses (Behroozmand et al., 2012), and there was a significantly larger number of opposing responses for unpredictable vs. predictable stimulus direction (Korzyukov et al., 2012). Korzyukov et al. (Korzyukov et al., 2012) and Scheerer & Jones (Scheerer and Jones, 2014) also reported that the amplitude of the N1 component of ERPs was significantly reduced for predictable vs. unpredictable stimulus direction and magnitude, respectively. Moreover, Scheerer & Jones (Scheerer and Jones, 2014) found that the latency of the P1 and N1 components was significantly shorter for predictable stimulus magnitude. Although behavioral vocal responses were not measured in Chen et al.'s study (Chen et al., 2012), they reported that the amplitude of the P2 ERP responses was reduced for manually-triggered temporally-predictable vs. unpredictable pitch-shift stimuli. Findings of these studies have suggested that the expectancy of the predictable stimulus eventually develops into recognition of the perturbation as being an external stimulus thereby leading to reduced vocal compensation (i.e., opposing responses) and a change in the underlying sensory-motor neural processes as indexed by modulation of the P1/N1/P2 components (Behroozmand et al., 2012; Chen et al., 2012; Korzyukov et al., 2012; Scheerer and Jones, 2014). In addition, these findings suggest that exposure to repeated presentations of predictable stimuli results in the increased contribution of feedforward mechanisms during vocal motor control. This reasoning supports the framework for predictions by the internal forward model: learned predictions result in more accurate efference copies and, consequently, a decreased mismatch in sensory feedback (Chen et al., 2012; Korzyukov et al., 2012; Scheerer and Jones, 2014; Wang et al., 2014; Wolpert and Flanagan, 2001).

Although the behavioral and neural correlates of vocalization have been examined for predictable pitch-shift stimulus magnitude and direction (Behroozmand et al., 2012; Korzyukov et al., 2012; Scheerer and Jones, 2014), research on temporal predictability effects on voice motor control is limited. Previous studies have shown that the suppression of neural responses in the auditory cortex in response to pure tones (Aliu et al., 2009) and speech (Behroozmand et al., 2011; Chen et al., 2012) develops for zero time delays but does not generalize to non-zero delays between feedforward predictions and sensory feedback perturbation. These findings indicate that the neural mechanisms of auditory feedback processing are sensitive to timing between the vocal motor commands and the incoming auditory feedback, and therefore, the observed suppression effect is not merely a movement-related non-specific effect. Further support for this notion is provided by studies showing that the degree of auditory suppression can be modulated by variations in vocal production (Sitek et al., 2013), speech targets (Ventura et al., 2009) and categorical boundaries of a spoken vowel sound (Niziolek and Guenther, 2013; Niziolek et al., 2013).

While previous voice motor control studies have had a temporal predictability element to their framework (Behroozmand et al., 2012, 2011; Burnett et al., 2008; Chen et al., 2012), no studies have been conducted focusing solely on the effects of predicting the delay of pitch perturbation in auditory feedback. Therefore, the purpose of the present study was to investigate the behavioral and neurophysiological correlates of temporal predictability of feedback pitch perturbation during vocal production and motor control. We employed the use of the auditory feedback pitch perturbation paradigm to elicit vocal behavior and ERP responses when subjects repeatedly produced and maintained a steady vocalization of a

vowel sound and received pitch shifts in their voice feedback under predictable and unpredictable onset time conditions. We hypothesized that predictions about temporal features of sensory feedback stimuli would modulate subsequent vocal motor behavior as evidenced by reduced magnitude and proportion of compensatory vocal responses. We also hypothesized that the neural responses to predictable feedback pitch perturbation stimuli would be decreased compared with those in response to non-predictable stimuli, as indexed by reduced amplitudes of the ERP components. The findings of the present study will provide new insights into the role of the temporal predictive code in sensory-motor mechanisms involved in vocal production and motor control.

2. Results

2.1. Behavioral vocal responses to pitch shift stimuli

The experimental design of the study is illustrated in Figure 1 (for more details see section 4.2). As can be seen in this figure, subjects were instructed to repeatedly produce and maintain a steady vocalization of the vowel sound /a/ while a brief (200 ms) upward pitch shift stimulus at +100 cents perturbed their voice auditory feedback. In blocks with temporally-predictable stimuli, there was a fixed delay of 500, 750 or 1000 ms between voice and pitch shift stimulus onset whereas in temporally-unpredictable blocks, the time delay between voice and stimulus onset was randomized between 500, 750 or 1000 ms. The order of all six vocalization blocks was counterbalanced across subjects and vocal and ERP responses were measured relative to the onset of pitch-shift stimuli in each condition separately.

Figure 2a shows the grand-average behavioral vocal responses to pitch shift stimuli overlaid across predictable and unpredictable stimulus onset for three different stimulus onsets at 500 ms, 750 ms and 1000 ms. As can be seen in this figure, when the stimulus onset was unpredictable, subjects produced compensatory vocal responses that opposed the direction of pitch shift in the auditory feedback. These vocal compensation responses were initiated at latencies around 80 ms (marked by arrows) after the onset of the pitch shift stimuli. However, when stimulus onset was predictable, subjects produced vocal responses that followed the direction of pitch shifts. These following responses were initiated at about 20 ms prior to the onset of pitch shift stimuli in the auditory feedback.

A repeated-measures analysis of variance (RM-ANOVA) model was used to analyze the magnitude and onset latency of the vocal responses to pitch shifts extracted for each condition (predictable and unpredictable stimulus onset) and stimulus onset time (500 ms, 750 ms and 1000 ms). The onset latency of vocal responses was calculated as the first time point at which vocal response magnitude exceeded the mean magnitude of vocal responses in a preceding 10 ms time window for more than ± 2 standard deviations. The search window for the response onset latency involved a range from 100 ms before to 500 ms after the onset of pitch-shift stimulus. The magnitude of vocal responses to temporally-predictable and unpredictable pitch-shift stimuli was measured at the first prominent peak in a time window from 0-500 ms post-stimulus for each stimulus onset delay and condition separately. The choice of these time windows was based on visual inspection of the single-subject and grand-averaged profiles of vocal responses to pitch shifts across all subjects.

Results of the analysis showed a significant main effect of condition [$F(1, 10) = 87.23, p < 0.001$] on the onset latency of the vocal responses, indicating that for the predictable pitch shift stimuli, vocal responses were initiated at a shorter latency compared with those in response to unpredictable stimuli. We also found a significant main effect of condition [$F(1, 10) = 19.14, p < 0.01$] on the magnitude of vocal responses, indicating that subjects produced larger (following) vocal responses for predictable compared with unpredictable stimulus onset. Moreover, our analysis revealed a significant main effect of stimulus onset on vocal response magnitudes only for unpredictable stimulus onset [$F(2, 20) = 5.64, p < 0.05$]. Post-hoc tests using Bonferroni's correction revealed that for the unpredictable stimulus onset, vocal response magnitudes were significantly larger for stimulus onset at 500 ms compared with 750 ms ($p < 0.01$) and 1000 ms ($p < 0.05$). The bar plots in figure 2b summarize the results of analysis on the magnitude and onset latency of vocal responses to pitch shift stimuli.

2.2. ERP responses to pitch shift stimuli

In order to analyze these data, the amplitude and latency of the P1, N1 and P2 ERP components were extracted for each condition (predictable and unpredictable stimulus onset) and stimulus onset time (500 ms, 750 ms and 1000 ms) for 24 electrode locations (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4, P3, Pz, P4, PO3, POz, PO4, FT7, T7, TP7, FT8, T8 and TP8). The amplitude of the ERPs were extracted within a 20 ms time window centered at the response peaks at 50 ms, 125 ms and 210 ms for P1, N1 and P2 components, respectively. The choice of these response peak latencies was based on visual examination of the grand-average ERP response waveforms. A RM-ANOVA model was used to analyze ERP amplitudes and latencies for condition, stimulus onset time and frontality (Frontal: F3, Fz, F4, Fronto-central: FC3, FCz, FC4, Central: C3, Cz, C4, Centro-parietal: CP3, CPz, CP4, Parietal: P3, Pz, P4, and Parieto-occipital: PO3, POz, PO4) factors and their interactions. A separate RM-ANOVA model was also used to analyze ERP amplitudes and latencies for condition, stimulus onset time and laterality (Left Temporal: FT7, T7, TP7 and Right Temporal: FT8, T8, TP8) factors and their interactions.

Results of the analysis showed significant main effects of condition [$F(1, 10) = 24.48, p < 0.01$] and frontality [$F(5,50) = 60.08, p < 0.001$] as well as a significant condition $\times$ frontality interaction [$F(5,50) = 15.27, p < 0.001$] on the N1 amplitude. We also found significant main effects of condition [$F(1, 10) = 143.73, p < 0.001$] and frontality [$F(5,50) = 4.60, p < 0.01$] as well as a significant condition $\times$ frontality interaction [$F(5,50) = 3.13, p < 0.05$] on the P2 amplitude.

Post-hoc tests using Bonferroni's correction revealed that the amplitude of the N1 component was significantly larger in response to unpredictable compared with predictable pitch shift stimulus onset at frontal ($p < 0.01$), fronto-central ($p < 0.001$), central ($p < 0.001$) and centro-parietal ($p < 0.01$) electrodes. We also found that the amplitude of the P2 component was significantly larger in response to unpredictable compared with predictable pitch shift stimulus onset at fronto-central ($p < 0.05$), central ($p < 0.001$), centro-parietal ($p < 0.001$), parietal ($p < 0.01$) and parieto-occipital ($p < 0.01$) electrodes.

Figure 3 shows examples of grand-average N1 and P2 response amplitude modulations by condition at Fz, FCz and Cz electrodes for each stimulus onset time, separately. In figure 4, grand-average ERP responses are separately shown for predictable and unpredictable pitch shift stimulus onset overlaid across all three stimulus onset times at Fz, FCz and Cz electrodes. The bar plots in figures 5a and 5b summarize the results of our findings with respect to N1 and P2 response amplitude modulation for predictable vs. unpredictable stimulus onsets at different electrode sites.

Figures 5a and 5b also show the results of laterality analysis, indicating that there was no significant difference between N1 and P2 response amplitudes in the left vs. right hemisphere. Results of our analysis on the latency of the N1 and P2 responses did not reveal any significant effect. We also did not find any significant effect on the amplitude and latency of the P1 ERP responses.

We found that the N1 component had a fronto-central distribution with the strongest responses at the Fz and FCz electrodes ($p < 0.001$) whereas the P2 component was more posterior and had a central distribution with the strongest responses at the Cz electrode ($p < 0.01$). The topographical scalp distribution maps of the N1 and P2 components are shown in Figures 6a and 6b.

3. Discussion

In the present study, we investigated the behavioral and neurophysiological correlates of the predictive coding model for voice motor control. Our study utilized an auditory feedback pitch perturbation paradigm to explore how predictability of the pitch-shift stimulus onset can affect subsequent vocal motor and neural responses during steady vocalizations of a vowel sound. Based on the findings of previous studies (Behroozmand et al., 2012; Korzyukov et al., 2012; Scheerer and Jones, 2014) and in line with the internal forward model theory (Wolpert et al., 2011, 1995), we hypothesized that predictions about the temporal aspect of perturbed auditory feedback would result in reduction of the amplitude of compensatory vocal responses and generating a greater proportion of following compared with opposing responses. We also hypothesized that pitch shifts with predictable onset time would elicit ERP components with smaller amplitudes compared with unpredictable stimuli.

Behavioral data revealed differences in direction of vocal responses to pitch perturbation with respect to temporal predictability of stimulus onset. For temporally-unpredictable stimuli, subjects produced a compensatory vocal response and changed their voice pitch in the opposite direction to pitch shifts in the auditory feedback at 80 ms post-stimulus. Conversely, for temporally-predictable stimuli, subjects demonstrated a following vocal response in the same direction as pitch shifts in the auditory feedback that was initiated at 20 ms pre-stimulus. The temporal predictability-induced modulation of vocal response direction (i.e., opposing for unpredictable stimuli and following for predictable stimuli) in the present study was consistent with results from a previous study that showed an increase in the number of following vocal responses to predictable direction of pitch shifts in voice auditory feedback (Korzyukov et al., 2012). These findings supported our hypothesis that predictions about timing of perturbations (i.e. pitch shifts) in the auditory feedback modulate

behavioral vocal responses during steady vocalization of a vowel sound. However, this effect was not reported in similar studies by Behroozmand et al. (Behroozmand et al., 2012) and Scheerer & Jones (Scheerer and Jones, 2014). These inconsistencies are partially accounted for by the difference in data analysis strategies and the type of stimulus predictability in these studies. Korzyukov et al. (Korzyukov et al., 2012) looked at the effect of stimulus direction predictability and used a pre-sorting algorithm to separate opposing and following responses before averaging them across all trials. This approach was similar to that in Behroozmand et al.'s study (Behroozmand et al., 2012) but different threshold criteria were selected for separating opposing from following responses in these two studies. Scheerer & Jones (Scheerer and Jones, 2014) investigated the effect of stimulus magnitude predictability and averaged vocal responses across all trials without pre-sorting them based on opposing vs. following responses. The differences between these methodologies urge the need for developing more uniform strategies for analysis of behavioral data to study sensory-motor mechanisms of vocal production and control in future studies.

Based on the findings of the present study, we propose the notion that the behavioral goals of feedforward motor mechanisms may change as a result of temporal predictability of sensory feedback stimuli. This notion argues that temporally-predictable sensory stimuli do not result in generation of feedback error signals that need to be corrected for by the motor system, but rather they might be processed as behaviorally-relevant sensory events that need to be further reinforced. This effect suggests that the contribution of feedforward mechanisms is increased for predictable sensory events, and the motor system learns and replicates feedback patterns that can be internally-simulated. Our data provide evidence in support of this notion by showing that the vocal motor system does not produce compensatory responses to control for temporally-predictable stimuli, but rather it produces vocal responses that follow the direction of pitch shifts in voice auditory feedback. However, when feedback changes are not temporally-predictable, the vocal motor system generates opposing (compensatory) responses that minimize mismatch (error) between vocal output and its auditory feedback. In addition, our findings indicate that the contribution of sensory feedback mechanisms is reduced for processing of temporally-predictable vs. unpredictable stimuli. This latter proposal is supported by the observation that vocal responses to predictable stimuli were initiated at approximately 20 ms prior to the onset of pitch shifts in voice auditory feedback, whereas for the unpredictable stimuli, responses started at 80 ms after the onset of pitch-shift stimuli. This effect has also been reported in vocal responses to predictable self-triggered pitch shift stimuli (Burnett et al., 2008) but not in response to predictable pitch shift magnitude or direction (Korzyukov et al., 2012; Scheerer and Jones, 2014).

Analysis of the ERP responses revealed a significant reduction in the amplitudes of the N1 and P2 components in response to predictable compared with unpredictable stimuli. These findings were consistent with results of studies on ERP responses to predictable perturbation magnitudes (Scheerer and Jones, 2014), direction (Korzyukov et al., 2012), and self-triggered stimulation (Chen et al., 2012). The suppression of the N1 and P2 components in this study supports the internal forward model and suggests that the efference copies of vocal motor commands transmit a temporal predictive code that cancels out neural responses to temporally-predictable sensory stimuli. This motor-induced suppression effect has been

previously reported in other studies by showing that the activity of the auditory cortex is suppressed during vocal production compared with passive listening to the playback of the same self-produced vocalizations (Behroozmand and Larson, 2011; Curio et al., 2000; Heinks-Maldonado et al., 2006, 2005; Houde et al., 2002). Our findings of the ERP responses combined with the behavioral results in this study further confirm the notion that temporally-predictable stimuli are reinforced and followed by the vocal motor system, and the contribution of the sensory feedback is reduced for their processing.

The knowledge about underlying mechanisms of the temporal predictive code is crucially important for diagnosis and treatment of neurological deficits in patients with speech motor disorders and/or cognitive impairments. Research has suggested that the symptoms associated with auditory hallucination in patients with schizophrenia are related to imprecise temporal synchronization between the feedforward and sensory feedback mechanisms (Ford et al., 2008, 2001; Heinks-Maldonado et al., 2007). In addition, more recent evidence has indicated that disruption of the predictive coding mechanisms by the development of neurological deficits may account for speech motor disorders in patients with Parkinson's disease (Chen et al., 2013; Liu et al., 2012; Mollaei et al., 2013) and stuttering (Cai et al., 2012; Loucks et al., 2012). These findings emphasize the importance of understanding the neural bases of the predictive coding model that serves as a binding mechanism for sensory-motor processing during vocal production, speech and other non-speech cognitive and mental tasks.

In addition to the observed effects of stimulus temporal predictability on behavioral and ERP responses, our results indicated a significant main effect of stimulus onset time on the magnitude of vocal responses only to unpredictable pitch perturbations in the auditory feedback. Our analysis revealed that the magnitude of opposing (compensatory) vocal responses to unpredictable stimuli was significantly larger at 500 ms compared with 750 and 1000 ms stimulus onset latencies. This finding suggests that temporally-unpredictable vocal production errors that occur within a shorter time delay relative to the voice onset are more effectively corrected for by the vocal motor system. In the framework of the internal forward model, such modulation of compensation magnitude as a function of stimulus onset time delay may be accounted for by differences in error detection and correction mechanisms in early vs. late parts of sustained vowel sound vocalizations. Evidence for this notion has been provided in a study by Niziolek et al. (Niziolek et al., 2013), showing that when speakers produced a vowel sound that was away from the center of its categorical boundary in the F1-F2 formant space, the motor system initiated corrective movements that changed production toward the expected median within the formant frequencies of that specific vowel sound. Niziolek et al. (Niziolek et al., 2013) showed that these corrective commands are initiated as early as 50 ms, suggesting that the motor system has a relatively short reaction time to correct for production errors in the earlier parts of speech. This finding suggests that sensory-motor mechanisms are more sensitive to production errors that occur in an earlier time frame relative to the onset of speech, because speakers are typically more likely to produce errors at the onset of their utterances.

This latter notion is supported by our findings in the present study by showing that unpredicted auditory feedback perturbations (errors) that occur at shorter time latencies from

voice onset are more effectively corrected for compared with those occurring later in time during steady phonation of vowel sounds. However, since in the pitch-shifting paradigm the ERP responses are reflective of changes in the auditory feedback processing, the absence of the N1 and P2 modulation for early (500 ms) vs. late (750 or 100 ms) stimulus onsets indicates that larger vocal compensations for shorter stimulus onsets are not driven by differences in neural processing of sensory feedback. This finding further supports our notion that the observed modulation of compensatory responses is in fact driven by feedforward mechanisms involved in vocal production and motor control. Furthermore, our results suggest that the modulation of vocal response magnitude by the stimulus onset time is an attribute of the compensatory mechanisms that stabilize voice fundamental frequency; however, following vocal responses do not share the same involvement. These findings suggest that the feedforward mechanisms may be more sensitive to detect and correct for feedback errors in the beginning of production trials that are more likely to be deviated from the desired goals of production during vocalization or speech.

The neuroanatomical correlates of sensory-motor integration during speech production and motor control have been explored in studies using fMRI recordings (Behroozmand et al., 2015; Golfinopoulos et al., 2010; Parkinson et al., 2012; Tourville et al., 2008; Zheng et al., 2010). Data from these studies have revealed that speech motor control is mediated by a complex interconnected neural network within auditory, motor and frontal cortices. Findings of these studies have shown that speech motor control involves areas within bilateral superior temporal gyrus (STG), Heschl's gyrus, insula, precentral gyrus, supplementary motor area (SMA), Anterior Cingulate Cortex (ACC), Rolandic operculum, postcentral gyrus and inferior frontal gyrus (IFG).

While the exact source location of neural activity during voice motor control cannot be pinpointed from this study, information from the topographical scalp distributions and previously conducted fMRI studies (Behroozmand et al., 2015; Parkinson et al., 2012) may suggest areas of activity for the observed ERP components in the present study. Our topographical scalp distributions revealed that the N1 responses were stronger over the fronto-central electrodes with an inverted polarity over the bilateral temporal areas. The P2 component was distributed more posteriorly with stronger activity over the central electrodes and an inversion over the bilateral temporo-parietal areas. Based on the findings from previous studies (Butler and Trainor, 2012; Wang et al., 2014), we suggest that the N1 responses arise from primary and secondary cortical auditory areas, and their modulation in this study is a neurophysiological index of sensory feedback suppression in response to temporally-predictable pitch-shift stimuli during vocal production and motor control. Less details are known about the neural generators of the P2 component, but its longer latency suggests that this neural response component reflect higher-level sensory-motor and cognitive processes and may receive contribution from multiple sources within sensory, motor and frontal cortices. Based on the findings of the fMRI studies (Behroozmand et al., 2015; Parkinson et al., 2012) we suggest that possible neural generators of the P2 may include areas within both auditory and vocal motor cortices that detect mismatch in the auditory feedback and issue corrective motor commands during voice motor control. The identification of the neural generators of the ERP components is a subject of future studies

that will help better understand the neuroanatomical correlates and function of the predictive coding mechanisms during voice production and motor control.

4. Experimental Procedure

4.1. Subjects

This study included a sample of 15 subjects (6 male and 9 female, age range: 18–23 years, mean age: 20.27 years) from Northwestern University. All subjects passed a bilateral pure-tone hearing screening test at 20 dB sound pressure level (SPL) (octave frequencies between 250 and 8000 Hz) and reported no history of neurological disorders, voice or musical 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.

4.2. Experimental Design

The experiment was conducted in a sound attenuated booth in which the subject's voice and EEG signals were recorded during steady vowel sound vocalizations. It comprised six vocalization blocks all counterbalanced across subjects in order to rule out block order effects. Subjects were verbally instructed to repeatedly produce and maintain a steady vocalization of the vowel sound /a/ at their conversational pitch and loudness and to disregard changes in their auditory feedback. No goal-oriented task was defined for the subjects. In each block, subjects were asked to produce approximately 150 vocalizations of the vowel sound for an approximate total of 900 vocalizations recorded per subject. Despite the natural trial-by-trial variability during vocal production, subjects were asked to produce vocalization as consistently as possible across trials (i.e., with relatively similar pitch and loudness). During each vocalization, subjects maintained steady vowel productions for approximately 2–3 s while taking short breaks (2–3 s) between successive trials. During each vocalization, a brief (200 ms) upward pitch shift stimulus at +100 cents perturbed voice auditory feedback in the middle of each vocalization trial. In three blocks, there was a fixed delay of 500, 750 or 1000 ms between voice and pitch shift stimulus onset (predictable timing), and in the remaining three blocks, the time delay between voice and stimulus onset was randomized between 500, 750 or 1000 ms (unpredictable timing). The experimental procedure for this task is illustrated in Figure 1. The total duration of each block was approximately 15 minutes, and each experimental session lasted approximately 2.5 hours in which each subject was prepared for voice and EEG data recording before completing all 6 vocalization blocks during the experiment.

The voice data were amplified with a Mackie mixer (model 1202-VLZ3), picked up using an AKG boomset microphone (model C420), sampled at 10 kHz using PowerLab A/D Converter (Model ML880, AD Instruments), and recorded on a laboratory computer utilizing Chart software (AD Instruments). A custom-designed program in Max/Msp (Cycling 74, v.5.0) controlled an Eventide Eclipse Harmonizer to pitch shift the voice online and feed it back to the ears using Etymotic earphones (model ER1-14A). All stimulus parameters including magnitude, direction and timing was controlled by the Max/Msp program. The Max/Msp also generated TTL pulses to accurately mark the onset of pitch-

shift stimuli in each trial. A 10 dB gain between voice and its feedback was maintained to partially mask air-born and bone-conducted voice feedback during vocalizations. A Brüel & Kjær sound level meter (model 2250) along with a Brüel & Kjær prepolarized free-field microphone (model 4189) and a Zwislocki coupler were used to calibrate the gain between voice and feedback channels.

4.3. Voice and EEG Data Acquisition

The EEG signals were recorded from 32 sites on the subject's scalp using Ag–AgCl electrodes placed on a cap (Easy-Cap GmbH, Germany) according to the standard 10-20 montage. Scalp-recorded brain potentials were low-pass filtered with a 400 Hz cut-off frequency (anti-aliasing filter), digitized at 2 kHz and recorded with common reference using a BrainVision QuickAmp amplifier (Brain Products GmbH, Germany) on a computer utilizing BrainVision Recorder software (Brain Products GmbH, Germany). Electrode impedances were kept below 5 kΩ for all channels. The electro-oculogram (EOG) signals were recorded using two pairs of bipolar electrodes placed above and below the right eye and on the lateral canthus of each eye to monitor vertical and horizontal eye movements. An electrode with unipolar shielded cable was placed on the subject's forehead and was connected to the QuickAmp dedicated slot for ground connection.

4.4. Analysis of Behavioral Vocal Responses

The pitch frequency of the recorded voice signals was extracted in Praat (Boersma and Weenik, 1996) using an autocorrelation method and then exported to MATLAB for further processing. The extracted pitch frequencies were segmented into epochs ranging from –100 ms before to 500 ms after the onset of pitch-shift stimuli. Pitch frequencies were then converted from Hertz to the Cents scale to calculate vocal compensation in response to the pitch-shift stimulus using the following formula:

$$Vocal\ Compensation\ [cents] = 1200 \times \log_2 \left(\frac{F}{F_{Baseline}} \right)$$

Here, F is the post-stimulus pitch frequency and $F_{Baseline}$ is the baseline pitch frequency from –100 to 0 ms pre-stimulus. The calculated pitch contours in Cents were averaged across all trials for unpredictable and predictable stimulus onset times at 500, 750 and 1000 ms, separately. The extracted pitch contours were then averaged across all subjects to obtain the grand-average profile of the vocal responses to pitch-shift stimulus for each condition. The onset latency of vocal responses was calculated as the first time point at which vocal response magnitude exceeded the mean magnitude of vocal responses in a preceding 10 ms time window for more than ± 2 standard deviations. The search window for the response onset latency involved a range from 100 ms before to 500 ms after the onset of pitch-shift stimulus. The vocal response peak magnitudes were extracted for the first prominent peak in a time window from 0–500 ms post-stimulus.

4.5. EEG Data Analysis

The EEGLAB toolbox (Delorme and Makeig, 2004) was used to analyze recorded EEG signals in order to calculate ERPs time-locked to the onset of upward pitch-shift stimuli with predictable and unpredictable onsets. The recorded EEG was first filtered offline using a band-pass filter with cut-off frequencies set to 1 and 30 Hz (–24 dB/oct) and then segmented into epochs ranging from –100 ms before and 500 ms after the onset of the stimulus. Following segmentation, artifact rejection was carried out by excluding epochs with EEG or EOG amplitudes exceeding $\pm 50 \mu\text{V}$. Individual epochs were then subjected to baseline correction by removing the mean amplitude of the pre-stimulus time window from –100 to 0 ms for each electrode. The extracted epochs were then averaged across all trials separately for each condition to obtain the ERP responses to pitch shift in each individual subject. A minimum number of 100 trials was used to calculate the ERP responses for each subject. The extracted ERP profiles were then averaged across all subjects to calculate the grand-average ERP responses and the amplitude of P1-N1-P2 components were extracted with a 20 ms time window centered at 60, 110 and 220 ms after the stimulus onset, respectively. These time points were calculated based on the peak amplitude of the ERP components at Cz electrode (vertex).

Acknowledgement

This research was supported by NIH Grant No. 1R01DC006243.

References

- Aliu SO, Houde JF, Nagarajan SS. Motor-induced suppression of the auditory cortex. *J. Cogn. Neurosci.* 2009; 21:791–802.10.1162/jocn.2009.21055 [PubMed: 18593265]
- Behroozmand R, Korzyukov O, Sattler L, Larson CR. Opposing and following vocal responses to pitch-shifted auditory feedback: evidence for different mechanisms of voice pitch control. *J. Acoust. Soc. Am.* 2012; 132:2468–77.10.1121/1.4746984 [PubMed: 23039441]
- Behroozmand R, Larson CR. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neurosci.* 2011; 12:54.10.1186/1471-2202-12-54 [PubMed: 21645406]
- Behroozmand R, Liu H, Larson CR. Time-dependent neural processing of auditory feedback during voice pitch error detection. *J. Cogn. Neurosci.* 2011; 23:1205–17.10.1162/jocn.2010.21447 [PubMed: 20146608]
- Behroozmand R, Shebek R, Hansen DR, Oya H, Robin DA, Howard MA, Greenlee JDW. Sensory-motor networks involved in speech production and motor control: An fMRI study. *Neuroimage.* 2015.10.1016/j.neuroimage.2015.01.040
- Boersma, P.; Weenik, D. PRAAT: a system for doing phonetics by computer. *Rep. Inst. Phonetic Sci. Univ.; Amsterdam:* 1996.
- Burnett TA, McCurdy KE, Bright JC. Reflexive and volitional voice fundamental frequency responses to an anticipated feedback pitch error. *Exp. Brain Res.* 2008; 191:341–351.10.1007/s00221-008-1529-z [PubMed: 18712372]
- Butler BE, Trainor LJ. Sequencing the cortical processing of pitch-evoking stimuli using EEG analysis and source estimation. *Front. Psychol.* 2012; 3:1–13.10.3389/fpsyg.2012.00180 [PubMed: 22279440]
- Cai S, Beal DS, Ghosh SS, Tiede MK, Guenther FH, Perkell JS. Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS One.* 2012; 7:1–14.10.1371/journal.pone.0041830

- Chen X, Zhu X, Wang EQ, Chen L, Li W, Chen Z, Liu H. Sensorimotor control of vocal pitch production in Parkinson's disease. *Brain Res.* 2013; 1527:99–107.10.1016/j.brainres.2013.06.030 [PubMed: 23820424]
- Chen Z, Chen X, Liu P, Huang D, Liu H. Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. *BMC Neurosci.* 2012; 13:55.10.1186/1471-2202-13-55 [PubMed: 22646514]
- Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. *Hum. Brain Mapp.* 2000; 9:183–191. [PubMed: 10770228]
- Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods.* 2004; 134:9–21.10.1016/j.jneumeth.2003.10.009 [PubMed: 15102499]
- Ford JM, Mathalon DH, Heinks T, Kalba S, Faustman WO, Roth WT. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *Am. J. Psychiatry.* 2001; 158:2069–2071.10.1176/appi.ajp.158.12.2069 [PubMed: 11729029]
- Ford JM, Roach BJ, Faustman WO, Mathalon DH. Out-of-Synch and Out-of-Sorts: Dysfunction of Motor-Sensory Communication in Schizophrenia. *Biol. Psychiatry.* 2008; 63:736–743.10.1016/j.biopsych.2007.09.013 [PubMed: 17981264]
- Golfinopoulos E, Tourville JA, Guenther FH. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage.* 2010; 52:862–874.10.1016/j.neuroimage.2009.10.023 [PubMed: 19837177]
- Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 2006; 96:280–301.10.1016/j.bandl.2005.06.001 [PubMed: 16040108]
- Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. *Psychophysiology.* 2005; 42:180–90.10.1111/j.1469-8986.2005.00272.x [PubMed: 15787855]
- Heinks-Maldonado TH, Mathalon DH, Houde JF, Gray M, Faustman WO, Ford JM. Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Arch. Gen. Psychiatry.* 2007; 64:286–296.10.1001/archpsyc.64.3.286 [PubMed: 17339517]
- Heinks-Maldonado TH, Nagarajan SS, Houde JF. Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport.* 2006; 17:1375–9.10.1097/01.wnr.0000233102.43526.e9 [PubMed: 16932142]
- Hickok G. Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* 2012
- Hickok G, Poeppel D. 2007; 8:393–403. processing.
- Houde JF, Nagarajan SS. Speech production as state feedback control. *Front. Hum. Neurosci.* 2011; 5:82.10.3389/fnhum.2011.00082 [PubMed: 22046152]
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* 2002; 14:1125–38.10.1162/089892902760807140 [PubMed: 12495520]
- Korzyukov O, Sattler L, Behroozmand R, Larson CR. Neuronal mechanisms of voice control are affected by implicit expectancy of externally triggered perturbations in auditory feedback. *PLoS One.* 2012; 7:10.1371/journal.pone.0041216
- Liu H, Wang EQ, Metman LV, Larson CR. Vocal responses to perturbations in voice auditory feedback in individuals with parkinson's disease. *PLoS One.* 2012; 7:10.1371/journal.pone.0033629
- Loucks T, Chon H, Han W. Audiovocal integration in adults who stutter. *Int. J. Lang. Commun. Disord.* 2012; 47:451–456.10.1111/j.1460-6984.2011.00111.x [PubMed: 22788230]
- Mollaei F, Shiller DM, Gracco VL. Sensorimotor adaptation of speech in Parkinson's disease. *Mov. Disord.* 2013; 28:1668–1674.10.1002/mds.25588 [PubMed: 23861349]
- Niziolek CA, Guenther FH. Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *J. Neurosci.* 2013; 33:12090–8.10.1523/JNEUROSCI.1008-13.2013 [PubMed: 23864694]

- Niziolek CA, Nagarajan SS, Houde JF. What does motor efference copy represent? Evidence from speech production. *J. Neurosci.* 2013; 33:16110–6.10.1523/JNEUROSCI.2137-13.2013 [PubMed: 24107944]
- Parkinson AL, Flagmeier SG, Manes JL, Larson CR, Rogers B, Robin DA. Understanding the neural mechanisms involved in sensory control of voice production. *Neuroimage.* 2012; 61:314–22.10.1016/j.neuroimage.2012.02.068 [PubMed: 22406500]
- Scheerer NE, Jones JA. The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. *Eur. J. Neurosci.* 2014; 40:3793–3806.10.1111/ejn.12734 [PubMed: 25263844]
- Sitek KR, Mathalon DH, Roach BJ, Houde JF, Niziolek CA, Ford JM. Auditory cortex processes variation in our own speech. *PLoS One.* 2013; 8:e82925.10.1371/journal.pone.0082925 [PubMed: 24349399]
- Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage.* 2008; 39:1429–43.10.1016/j.neuroimage.2007.09.054 [PubMed: 18035557]
- Ventura MI, Nagarajan SS, Houde JF. Speech target modulates speaking induced suppression in auditory cortex. *BMC Neurosci.* 2009; 10:58.10.1186/1471-2202-10-58 [PubMed: 19523234]
- Wang J, Mathalon DH, Roach BJ, Reilly J, Keedy SK, Sweeney JA, Ford JM. Action planning and predictive coding when speaking. *Neuroimage.* 2014; 91:91–8.10.1016/j.neuroimage.2014.01.003 [PubMed: 24423729]
- Wolpert DM, Diedrichsen J, Flanagan JR. Principles of sensorimotor learning. *Nat. Rev. Neurosci.* 2011; 12:739–51.10.1038/nrn3112 [PubMed: 22033537]
- Wolpert DM, Flanagan JR. Motor prediction. *Curr. Biol.* 2001; 11:R729–32. [PubMed: 11566114]
- Wolpert DM, Ghahramani Z. Computational principles of movement neuroscience. *Nat. Neurosci.* 2000; 3(Suppl):1212–1217.10.1038/81497 [PubMed: 11127840]
- Wolpert DM, Ghahramani Z, Jordan MI. An Internal Model for Sensorimotor Integration. *Science* (80-.). 1995; 269:1880–1882.
- Zheng ZZ, Munhall KG, Johnsrude IS. Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *J. Cogn. Neurosci.* 2010; 22:1770–81.10.1162/jocn.2009.21324 [PubMed: 19642886]

Highlights

- Humans use auditory feedback to control their voice during speaking
- Temporally-unpredictable changes in auditory feedback trigger opposing vocal responses
- Temporally-predictable changes in auditory feedback trigger following vocal responses
- The brain activity is suppressed for temporally-predictable changes in voice feedback

Temporal predictability of changes in auditory feedback modulates voice motor control

Figure 1.

Experimental design for measuring behavioral and neurophysiological responses to predictable and unpredictable pitch shift onsets in voice auditory feedback. Subjects repeatedly produced a steady vocalization of the vowel sound /a/ while a brief (200 ms) upward pitch shift stimulus at +100 cents perturbed voice auditory feedback in the middle of each vocalization trial. **a)** In three blocks, there was a fixed delay of 500, 750 or 1000 ms between voice and pitch shift stimulus onset (predictable timing), and **b)** in the remaining three blocks, the time delay between voice and stimulus onset was randomized between 500, 750 or 1000 ms (unpredictable timing). The order of all six vocalization blocks was counterbalanced across subjects and approximately a total of 900 vocalizations were recorded for each individual subject.

Figure 2.

a) Grand-average behavioral vocal responses to pitch shift stimuli overlaid across predictable and unpredictable stimulus onset for three different stimulus onsets at 500 ms, 750 ms and 1000 ms. **b)** Bar plot representation of the results of analysis on the onset latency and **c)** the peak magnitude of vocal responses to predictable and unpredictable pitch shift stimuli onset at 500, 750 and 1000 ms latencies.

Figure 3. The overlaid grand-average ERP responses for predictable and unpredictable pitch shift stimulus onset at Fz, FCz and Cz electrodes plotted for each stimulus onset time, separately.

Figure 4.

The grand-average ERP responses overlaid across all three stimulus onset times (500, 750 and 1000 ms) plotted separately for predictable and unpredictable pitch shift stimulus onsets at Fz, FCz and Cz electrodes.

Figure 5.

The bar plots summarizing the results of analysis for **a)** the N1 and **b)** the P2 ERP amplitudes in response to predictable vs. unpredictable pitch-shift stimulus onsets at frontal, frontocentral, central, centroparietal, parietal, parietooccipital, and the left and right temporal electrode sites.

Figures 6.

The topographical distribution maps of the scalp-recorded potentials in response to predictable and unpredictable pitch-shift stimulus onsets for **a)** the N1 and **b)** the P2 ERP components.