Behavioral/Cognitive

Top-Down Modulation of Auditory-Motor Integration during Speech Production: The Role of Working Memory

Zhiqiang Guo,[1]* Xiuqin Wu,[2]* Weifeng Li,[3] Jeffery A. Jones,[4] Nan Yan,[5] Stanley Sheft,[6] Peng Liu,[2] and Hanjun Liu[2,7]

[1]Department of Computer Science and Technology, Zhuhai College of Jilin University, Zhuhai, 519041, China, [2]Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, China, [3]Department of Emergency and Critical Care Medicine, Guangdong General Hospital, Guangdong Academy of Medical Sciences, Guangzhou, 510080, China, [4]Psychology Department and Laurier Centre for Cognitive Neuroscience, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, Canada, [5]CAS Kay Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, [6]Department of Communication Disorders and Sciences, Rush University Medical Center, Chicago, Illinois 60612, and [7]Guangdong Provincial Key Laboratory of Brain Function and Disease, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510080, China

Although working memory (WM) is considered as an emergent property of the speech perception and production systems, the role of WM in sensorimotor integration during speech processing is largely unknown. We conducted two event-related potential experiments with female and male young adults to investigate the contribution of WM to the neurobehavioural processing of altered auditory feedback during vocal production. A delayed match-to-sample task that required participants to indicate whether the pitch feedback perturbations they heard during vocalizations in test and sample sequences matched, elicited significantly larger vocal compensations, larger N1 responses in the left middle and superior temporal gyrus, and smaller P2 responses in the left middle and superior temporal gyrus, inferior parietal lobule, somatosensory cortex, right inferior frontal gyrus, and insula compared with a control task that did not require memory retention of the sequence of pitch perturbations. On the other hand, participants who underwent extensive auditory WM training produced suppressed vocal compensations that were correlated with improved auditory WM capacity, and enhanced P2 responses in the left middle frontal gyrus, inferior parietal lobule, right inferior frontal gyrus, and insula that were predicted by pretraining auditory WM capacity. These findings indicate that WM can enhance the perception of voice auditory feedback errors while inhibiting compensatory vocal behavior to prevent voice control from being excessively influenced by auditory feedback. This study provides the first evidence that auditory-motor integration for voice control can be modulated by top-down influences arising from WM, rather than modulated exclusively by bottom-up and automatic processes.

Key words: **auditory feedback; auditory-vocal integration; top-down influence; training; working memory**

Significance Statement

One outstanding question that remains unsolved in speech motor control is how the mismatch between predicted and actual voice auditory feedback is detected and corrected. The present study provides two lines of converging evidence, for the first time, that working memory cannot only enhance the perception of vocal feedback errors but also exert inhibitory control over vocal motor behavior. These findings represent a major advance in our understanding of the top-down modulatory mechanisms that support the detection and correction of prediction-feedback mismatches during sensorimotor control of speech production driven by working memory. Rather than being an exclusively bottom-up and automatic process, auditory-motor integration for voice control can be modulated by top-down influences arising from working memory.

Introduction

Speech motor control relies on the integration of sensory information, particularly auditory feedback, into the vocal motor systems. Several models of speech motor control hypothesize that a prediction of the sensory consequence of motor commands is

Received May 15, 2017; revised Aug. 6, 2017; accepted Sept. 13, 2017.

Author contributions: J.A.J. and H.L. designed research; Z.G., X.W., W.L., and P.L. performed research; N.Y. and S.S. contributed unpublished reagents/analytic tools; Z.G., X.W., W.L., and N.Y. analyzed data; Z.G., X.W., J.A.J., P.L., and H.L. wrote the paper.

*Z.G. and X.W. contributed equally to this work.

This work was supported by grants from the National Natural Science Foundation of China (31371135 and 81472154), Guangdong Natural Science Funds for Distinguished Young Scholar (S2013050014470), Guangdong Province Science and Technology Planning Project (2017A050501014), Guangzhou Science and Technology Programme (201604020115), and the Fundamental Research Funds for the Central Universities (15ykjc13b).

compared with the actual auditory feedback from the vocal output (Guenther et al., 2006; Hickok et al., 2011), and that a prediction-feedback mismatch leads to motor commands that correct for speech feedback errors (Burnett et al., 1998; Houde and Jordan, 1998). Prediction-feedback mismatches modulate the event-related potentials (ERPs) known as the N1-P2 complex (Behroozmand et al., 2009; Liu et al., 2011), which are hypothesized to respectively reflect the early pre-attentive detection and late higher-level processing of speech feedback errors (Korzyukov et al., 2012a; Chen et al., 2015). These cortical responses represent part of a complex neural network within the auditory- and motor-related cortical areas responsible for sensorimotor integration during speech (Zarate and Zatorre, 2008; Parkinson et al., 2012; Chang et al., 2013; Behroozmand et al., 2016).

Despite great strides toward understanding sensorimotor control of speech production, the neural mechanisms underlying the detection and correction of prediction-feedback mismatch remain poorly understood (Houde and Chang, 2015). Given that speakers produce rapid vocal compensations for altered auditory feedback (~100 ms; Chen et al., 2007) and cannot consciously modify their responses (Keough et al., 2013), auditory-motor integration for voice control is generally considered to be reflex-like or involuntary (Munhall et al., 2009). However, recent work has shown that this process can be modulated by attentional demands, as reflected by significantly increased vocal compensations and/or P2 responses to attended pitch perturbations (Tumber et al., 2014; Hu et al., 2015; Liu et al., 2015). Thus, top-down cognitive processing may influence the detection and/or correction of voice feedback errors.

A separate but growing literature suggests that working memory (WM) may also influence sensorimotor integration during speech. WM is considered to be an emergent property of the speech perception and production systems (Jacquemot and Scott, 2006). Research has shown that rehearsal of verbal WM information (Paulesu et al., 1993; Buchsbaum et al., 2005) and sensorimotor control of speech production (Zarate and Zatorre, 2008; Behroozmand et al., 2015) recruit overlapping neural networks that include the superior temporal, posterior parietal, and prefrontal cortical areas. More interestingly, Ranasinghe et al. (2017) found that patients with Alzheimer's disease (AD) produced abnormally enhanced vocal compensations that predicted their executive dysfunction and reduced compensation durations that predicted their memory dysfunction, which might be related to their impaired frontally-mediated inhibitory mechanisms. On the other hand, healthy participants who received auditory WM training based on a digit-span backward (DSB) paradigm exhibited smaller N1 and larger P2 responses to pitch-shifted voice auditory feedback than participants who did not receive training, suggesting that improved WM capacity can increase neural efficiency in the detection and correction of vocal feedback errors (Li et al., 2015). It is thus reasonable to hypothesize that WM exerts top-down influences on sensorimotor control of speech production: facilitating the cortical processing of auditory feedback errors to estimate mismatches between predicted and incoming auditory feedback, and inhibiting compensatory adjustment of vocal outputs to prevent vocal production from being excessively influenced by auditory feedback. Presently, there is insufficient direct evidence for this hypothesis.

To address this question, we measured vocal and cortical (N1 and P2) responses to pitch-shifted auditory feedback while participants completed two experimental tasks: a delayed matching-to-sample (DMS) task that required participants to indicate whether a test sequence of pitch perturbations matched or did not match a preceding sample sequence, and a frequency pattern reconstruction (FPR) training protocol (Sheft et al., 2014) that allowed participants to improve their auditory WM capacity. We predicted that the neurobehavioral responses to voice feedback errors would be modulated as a function of WM engagement and that improved WM capacity following training would facilitate sensorimotor processing. Our results provide strong evidence for top-down influences arising from WM on sensorimotor integration for voice control.

Materials and Methods

Subjects

Forty right-handed native-Mandarin speakers, who were students from Sun Yat-sen University of China, were recruited to participate in one of two experiments. All participants reported no history of language, neurological disorders, or musical training. Participants passed a bilateral hearing screening at ≤25 dB hearing level (HL) for octave intervals of 250–4000 Hz. Written informed consent was obtained from all participants, and the research protocol was approved by the Institutional Review Board of The First Affiliated Hospital at Sun Yat-sen University of China in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Experimental design

Experiment 1. In Experiment 1 (Fig. 1*A*), we used a frequency altered feedback (FAF) paradigm to examine top-down influences of WM on auditory-motor integration for voice control by evaluating the hypothesis that neurobehavioral responses elicited by vocal pitch perturbations would be modulated as a function of the degree of WM engagement. Fifteen participants [11 female and 4 male; age: 23 ± 3 years (mean ± SD)] took part in the DMS and control tasks. Across these two tasks, they were instructed to start vocalizing the /u/ sound at their habitual voice each time a black cross appeared on the computer monitor and to stop vocalizing when the black cross disappeared. While vocalizing, participants heard their voice randomly pitch-shifted +200 or +500 cents [duration: 200 ms; interstimulus interval (ISI): 700–900 ms] five times. For the DMS task, two consecutive vocalizations constituted one trial that lasted 17.5 s. On each trial of the DMS task, participants heard a sample sequence of five pitch perturbations during their first 6 s vocalization. This first vocalization was followed by a 2.5 s delay period to allow rehearsal. Following the delay period participants heard a test sequence of five pitch perturbations while they produced their second 6 s vocalization. Finally, an immediate recall test was performed during a 3 s response period. If the test sequence matched the sample sequence in terms of the magnitude of each pitch perturbation, participants were asked to say the word "Yes" to indicate they perceived the sample and test sequences to be the same. Otherwise, participants were asked to report the way in which the sample and test sequences differed. For example, when presented with a sample sequence of +200, +200, +500, +500, and +500 cents pitch perturbations followed by a test sequence of +500, +500, +500, +500, and +200 cents pitch perturbations, an accurate response would indicate that the first (+200 vs +500 cents), second (+200 vs +500 cents), and fifth (+500 vs +200 cents) pitch perturbations did not match. Participants were required to respond as accurately and quickly as possible before initiating the next trial.

A control task was also included, which differed from the DMS task only in instruction, such that participants were not required to remember the sequence of pitch perturbations during their vocalizations. Rather, they were instructed to maintain a steady vocalization by ignoring the perceived pitch perturbations. These instructions have been ad-

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Hanjun Liu, Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, P. R. China. E-mail: lhanjun@mail.sysu.edu.cn.

DOI:10.1523/JNEUROSCI.1329-17.2017

Copyright © 2017 the authors 0270-6474/17/3710324-11$15.00/0

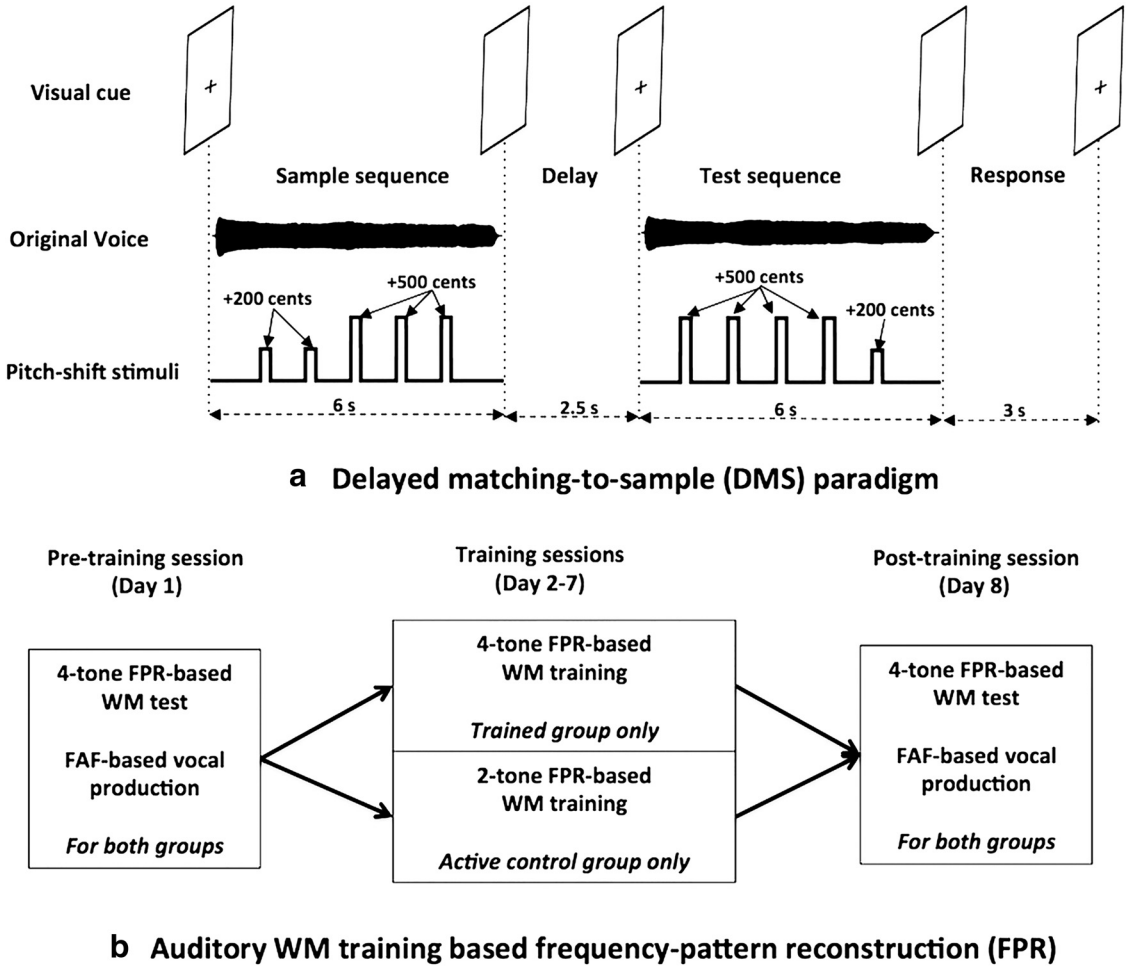

Figure 1. Schematic illustration of the experimental procedure. ***a***, A delayed matching-to-sample (DMS) paradigm was applied in Experiment 1, during which participants were required to indicate whether the pitch perturbations in the test sequence matched or did not match those in the sample sequence and specify which items were different between the two sequences. A control task was also included, which differed from the DMS task only in instruction that did not require memory retention of the sequence of pitch perturbations. ***b***, In Experiment 2, both the trained and active control groups received an 8 d auditory WM training based on a pure-tone frequency-pattern reconstruction (FPR) paradigm. Before and after the training, both groups participated in an frequency altered feedback (FAF)-based vocal production experiment. The stimulus parameters were identical to that of the control task in Experiment 1.

opted in a number of previous studies (Burnett et al., 1998; Hain et al., 2000; Zarate and Zatorre, 2008; Liu et al., 2010; Parkinson et al., 2012; Keough et al., 2013). The control task was always the first task that participants performed, ensuring that they were unaware of the WM-related task that was to follow.

To minimize the temporal predictability of feedback errors that could lead to a reduction in the proportion of compensatory vocal responses and a reduction in the magnitude of ERP N1 responses (Chen et al., 2012a; Korzyukov et al., 2012b), participants heard a randomly mixed presentation of the +200 and +500 cents pitch perturbations during each vocalization. These two perceptually salient pitch perturbations were included in the experiments because they were easily distinguished by participants, which ensured successful performance of the DMS task. The neurobehavioral responses to the +500 cents pitch perturbations, however, were not measured and submitted to statistical analyses because previous work has shown that participants perceive small pitch perturbations (e.g., <200 cents) as self-produced speech errors that result in compensatory vocal adjustments, whereas large pitch perturbations (e.g., 400 cents or more) are perceived as externally-generated and tend to elicit vocal responses that follow the perturbation direction (Burnett et al., 1998; Hain et al., 2000). In addition to these behavioral response differences, ERP N1 responses to self-produced voice are significantly suppressed relative to alien voice feedback (Heinks-Maldonado et al., 2005); this N1 suppression is eliminated when participants hear pitch perturbations of +400 cents during active vocalization versus passive listening (Behroozmand and Larson, 2011). As stated in the Introduction, the aim of the present study was to investigate the functional role of WM in perceiving and compensating for mismatches between predicted and incoming vocal feedback. Therefore, only the vocal and ERP responses to +200 cents pitch perturbations during the DMS and control tasks were measured and submitted to statistical analyses.

Participants produced 80 vocalizations during the DMS and control tasks, respectively. For each task, across the 80 vocalizations, 200 pitch-shift stimuli were +200 cents in size and 200 pitch-shift stimuli were +500 cents in size. The vocal and ERP responses to the 100 pitch-shift stimuli that were +200 cents in size in the test sequence were compared across the DMS and control tasks to examine whether auditory-motor control of vocal production can be modulated as a function of WM engagement.

Experiment 2. We conducted Experiment 2 to provide further evidence for top-down influences of auditory WM on speech motor control by evaluating the hypothesis that improved WM capacity following training would result in changes in neurobehavioral responses to pitch feedback errors during vocalization. Twenty-five naive participants were recruited and randomly assigned to the trained and active control groups. None of these participants participated in Experiment 1. The trained group included 13 participants (8 female and 5 male; age: 23 ± 2 years), and the active control group included 12 participants (8 female and 4 male; age: 21 ± 2 years). The trained and active control groups did not differ significantly in age, gender, or education.

Both the trained and active control groups participated in an 8 d auditory WM training based on a pure-tone FPR paradigm (Sheft et al., 2014; Fig. 1*B*). Participants were required to assemble a number of constituent tones in the correct order after hearing the target sequence. Before submitting the response, participants could listen to both constituent tones and their interim reconstruction of the sequence and reassemble the tones as often as they wanted. Correct-answer feedback for each sequence tone was provided after every trial. Constituent tones were shaped with a 50 ms rise/fall time and randomly selected from logarithmically scaled distributions (frequency range: 400–1750 Hz; duration: 75–600 ms). To make the sequence component discriminable, the ratios of any two component frequencies and durations were >1.2 and 1.4, respectively. Auditory WM training was implemented with MATLAB software (RRID:SCR_001622) running on a personal computer, and participants heard the sounds through headphones (Sennheiser HMD 250). Listeners were allowed to choose a preferred loudness level at the outset of the experiment, and this level was maintained throughout training.

The auditory WM training consisted of three phases: a pretraining phase (Day 1), a training phase (Days 2–7), and a post-training phase (Day 8). During the pretraining and post-training phases, both the trained and active control groups were tested using the four-tone FPR paradigm to document training-induced change in auditory WM capacity. During each day of the training phase, the trained group received auditory WM training using a four-tone version of the FPR paradigm, whereas the active control group received a similar, but low load version of the training with a two-tone FPR paradigm. Note that participants heard the target sequence only once during the pretraining and post-training phase days, whereas they were allowed multiple repetitions of the intact target sequence to correct their interim responses on each day of the training phase. On each day of the pretraining, training, and post-training phases, participants completed two 25-trial blocks, and each block lasted ~30 min.

Before and after the training, both the trained and active control groups participated in the FAF-based vocal production experiment. Similar to Experiment 1, participants in Experiment 2 were instructed to vocalize the /u/ sound at their habitual voice pitch and loudness until the black cross on the computer monitor disappeared. Each vocalization lasted 6 s, during which voice auditory feedback was randomly pitch-shifted +200 or +500 cents (200 ms long) five times with the ISI of 700–900 ms to minimize the temporal predictability of feedback errors. In contrast with Experiment 1, participants in Experiment 2 did not perform the DMS task but producing and stabilizing their vocalizations by ignoring the perceived pitch perturbations. This setup would not only provide further evidence for the relationship between WM and auditory-vocal integration but also allow us to extend our results to other studies using the same FAF paradigm (Burnett et al., 1998; Hain et al., 2000; Zarate and Zatorre, 2008; Parkinson et al., 2012). Participants produced 40 consecutive vocalizations in Experiment 2, leading to 100 stimuli that were +200 cents in size and 100 stimuli that were +500 cents in size. For the same reason mentioned in Experiment 1, we only measured the vocal and ERP responses to +200 cents pitch perturbations.

Data recording

Both Experiments 1 and 2 were conducted in a sound-treated booth. Before the experiment, the acoustic recording system was calibrated to ensure that the intensity of voice feedback was 10 dB SPL higher than that of the participant's vocal output to partially mask the air-born and bone-conducted feedback. The voice signals were picked up using a dynamic microphone (DM2200, Takstar) and amplified with a MOTU Ultralite Mk3 Firewire audio interface. The amplified voices were then pitch-shifted using an Eventide Eclipse Harmonizer that was controlled by a custom-developed MIDI software program (Max/MSP v5.0, Cycling 74). The acoustic parameters of the pitch perturbations, including the magnitude, duration, direction, and timing, were manipulated with this program. Finally, the voice signals were amplified by an ICON NeoAmp headphone amplifier and fed back to participants through insert earphones (ER1–14A, Etymotic Research). Transistor-transistor logic (TTL) pulses generated by the Max/MSP program were used to signal the onset of each pitch perturbation. The TTL pulses were also sent to the EEG recording system via a DIN cable. The original and pitch-shifted voice signals as well as the TTL pulses were digitized at 10 kHz by a PowerLab A/D converter (model ML880, AD Instruments; RRID:SCR_001620), and recorded using LabChart software v7.0 (AD Instruments).

The EEG signals were collected using a 64-electrode Geodesic Sensor Net (Electrical Geodesics Inc.), amplified by a high input-impedance Net Amps 300 amplifier ($Z_{in} \approx 200$ MΩ; Electrical Geodesics Inc.) that accepts scalp-electrode impedances up to 40–60 kΩ, and recorded using NetStation software v4.5 (Electrical Geodesics Inc.; RRID:SCR_002453) with a sampling frequency of 1000 Hz. During the online recording, the EEG signals from all channels were referenced to the vertex (Cz; Ferree et al., 2001). Individual sensors were carefully adjusted to ensure that their impedance levels were <50 kΩ (Ferree et al., 2001) throughout the recording.

Data analysis

The magnitudes and latencies of vocal responses to pitch-shifted auditory feedback were measured using event-related averaging techniques (Li et al., 2013) in a custom-designed IGOR PRO software v6.0 (Wavemetrics; RRID:SCR_000325). Voice F_0 contours in Hertz were extracted from the voice signals using Praat software (Boersma, 2001), and converted to the cents scale using the following formula: cents = 100 × (12 × $\log_2(F_0/\text{reference})$) [reference = 195.997 Hz (G4)]. The voice contours in cents were then segmented into epochs ranging from 200 ms before to 700 ms after the onset of the pitch perturbation. Following visual inspection of all individual segmented trials using a waterfall procedure, trials with vocal interruptions, signal processing errors, or highly variable baseline periods were excluded from further analyses. Overall, 78% and 72% of trials were regarded as artifact-free trials in Experiments 1 and 2, respectively. These artifact-free trials were normalized by subtracting the mean F_0 values in the baseline period (−200 to 0 ms) from the F_0 values after the onset of the pitch perturbation, and averaged to generate an overall response for each condition. The magnitude of a vocal response in cents was measured as the greatest F_0 value following the response onset. The response latency was defined as the time when the response exceeded 2 SD above or below the baseline period following the onset of the pitch perturbation.

The EEG signals were submitted to NetStation software for the off-line analysis. They were first bandpass filtered with cutoff frequencies of 1–20 Hz and then segmented into epochs ranging from −200 ms to +500 ms relative to the onset of the pitch perturbation. Segments with voltage values exceeding ±55 μv of the moving average over an 80 ms window were excluded from analysis to remove trials contaminated by excessive muscular activity, eye blinks, or eye movements from further analyses. Individual electrodes that contained artifacts in >20% of the segments were rejected, and any file that contained >10 bad channels was excluded from the averaging procedure. An additional visual inspection was performed for all individual trials to ensure that artifacts were appropriately rejected. On average, 76% and 81% of trials were retained for the following averaging procedure for Experiments 1 and 2, respectively. Finally, all channels were re-referenced to the average of the electrodes on each mastoid, and artifact-free trials were averaged and baseline-corrected to generate an overall response. The amplitude and latency of the negative peak 80–180 ms after the onset of the pitch perturbation in the averaged ERP was used to describe the N1 component. Amplitudes and latencies for the P2 component were measured using the most positive peak between 160 and 280 ms after the onset of the pitch perturbation.

The cortical distribution of current densities was calculated using standard low-resolution electromagnetic tomography (sLORETA; RRID: SCR_013829; Pascual-Marqui, 2002), a widely used method that has been validated using fMRI (Mulert et al., 2004) and intracerebral recordings (Zumsteg et al., 2006), for the N1 and P2 responses as a function of task (DMS vs control task) or phase (pretraining vs post-training). Based on a linear weighted sum of the scalp electric potentials, sLORETA calculates the standardized current density of a dense grid of 6239 voxels at a 5 mm spatial resolution in the gray matter and the hippocampus of the Montreal Neurological Institute (MNI)-reference brain. The voxel-based sLORETA images were calculated using a realistic standardized head model computed with the boundary element approach (Fuchs et al.,

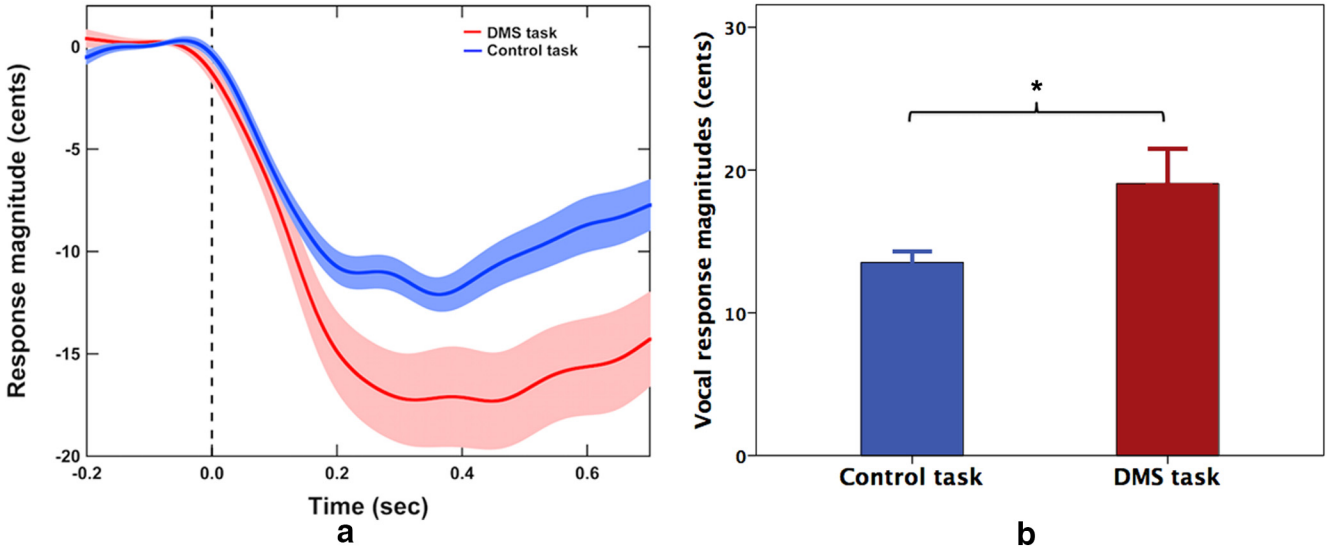

Figure 2. Behavioral results of Experiment 1. ***a***, Grand-averaged voice F_0 contours and (***b***) T-bar graphs of the absolute values of vocal responses (mean and SE) to +200 cents pitch perturbations during the DMS (red) and control (blue) tasks. Highlighted areas indicate the SE of the mean vocal responses. The asterisk indicates significantly larger magnitudes of vocal responses during the DMS task compared with the control task ($p = 0.035$).

2002) and the MNI152 template (Mazziotta et al., 2001) with the three-dimensional solution space restricted to cortical gray matter. In the present study, sLORETA images were computed at the 5 ms time windows of maximal global field power peaks within the N1 and P2 time windows. Voxel-by-voxel comparisons of the current density distributions as a function of task or phase were performed using sLORETA voxelwise randomization tests (5000 permutations) based on statistical nonparametric mapping, and corrected for multiple comparisons. The results correspond to maps of log-*F*-ratio statistics for each voxel. The voxels with significant differences (for corrected $p < 0.05$) were specified in MNI coordinates and labeled as Brodmann areas (BAs) within the EEGLAB software (Delorme and Makeig, 2004).

Statistical analyses

The values of the vocal and ERP responses to pitch-shifted auditory feedback across all the conditions met assumptions of normality and homogeneity of variance and were thus subjected to repeated-measures ANOVA (RM-ANOVA) using SPSS v16.0 (RRID:SCR_002865). For Experiment 1, the magnitudes and latencies of vocal responses were subjected to one-way RM-ANOVAs, in which task (DMS task vs control task) was the within-subject factor. The amplitudes and latencies of the N1 and P2 components were subjected to two-way RM-ANOVAs, in which task and electrode (FC1, FC2, FCz, FC3, FC4, C1, C2, Cz, C3, and C4) were the within-subject factors. Those electrodes were chosen for statistical analyses because cortical responses to pitch-shifted auditory feedback are most pronounced in the frontal and central electrodes (Hawco et al., 2009; Chen et al., 2012b). For Experiment 2, the magnitudes and latencies of vocal responses were subjected to two-way RM-ANOVAs, in which phase (pretraining vs post-training) was the within-subject factor, whereas group (trained group vs control group) was the between-subject factor. The amplitudes and latencies of the N1 and P2 components were subjected to three-way RM-ANOVAs. Phase and electrode were the within-subject factors, whereas group was the between-subject factor. Any significant higher-order interactions led to subsidiary RM-ANOVAs, and *post hoc* comparisons were performed using the Bonferroni adjustment for multiple comparisons. Probability values were corrected using Greenhouse-Geisser for multiple degrees of freedom in the case of violations of the sphericity assumption. Effect size was calculated using partial η^2 to describe the size of differences between the conditions. *P* values <0.05 and partial $\eta^2 > 0.14$ (Richardson, 2011) were required to be considered significant.

Results

Experiment 1

Figure 2 shows grand-averaged voice F_0 contours (*a*) and T-bar plots of the magnitudes of vocal responses (*b*) to pitch-shifted auditory feedback for the DMS task (red) and control task (blue). As expected, there was a significant main effect of task ($F_{(1,14)} = 5.436$, $p = 0.035$, partial $\eta^2 = 0.280$); larger vocal compensation responses were elicited when WM was engaged by the DMS task than during the control task. By contrast, the latencies of vocal responses did not differ as a function of task ($F_{(1,14)} = 0.862$, $p = 0.369$, partial $\eta^2 = 0.058$).

Figure 3*a* shows the grand-averaged ERP waveforms in response to pitch perturbations during the DMS (red) and control tasks (blue). The cortical N1 and P2 responses appeared to be differentially modulated by the task. A two-way ANOVA revealed a significant main effect of task ($F_{(1,14)} = 13.750$, $p = 0.002$, partial $\eta^2 = 0.496$), indicating that the DMS task elicited significantly larger (more negative) N1 amplitudes than the control task (Fig. 3*a–c*). The main effect of electrode ($F_{(9,126)} = 1.320$, $p = 0.278$, partial $\eta^2 = 0.086$) and the interaction between task and electrode ($F_{(9,126)} = 2.675$, $p = 0.053$, partial $\eta^2 = 0.160$) did not reach significance. For the N1 latencies, the main effects of task ($F_{(1,14)} = 0.162$, $p = 0.694$, partial $\eta^2 = 0.011$), electrode ($F_{(9,126)} = 1.314$, $p = 0.281$, partial $\eta^2 = 0.086$), and their interaction ($F_{(9,126)} = 0.324$, $p = 0.794$, partial $\eta^2 = 0.023$) also did not reach significance.

A two-way ANOVA conducted on the P2 amplitudes revealed a significant main effect of task ($F_{(1,14)} = 8.107$, $p = 0.013$, partial $\eta^2 = 0.367$), indicating that the DMS task elicited significantly smaller P2 amplitudes than the control task (Fig. 3*a–c*). There was also a significant main effect of electrode ($F_{(9,126)} = 13.430$, $p < 0.001$, partial $\eta^2 = 0.490$), which was primarily driven by larger P2 amplitudes in the frontal electrodes as compared with the central electrodes. The interaction between task and electrode, however, did not reach significance ($F_{(9,126)} = 1.410$, $p = 0.246$, partial $\eta^2 = 0.091$). For the P2 latencies, the DMS task elicited significantly faster P2 responses than the control task

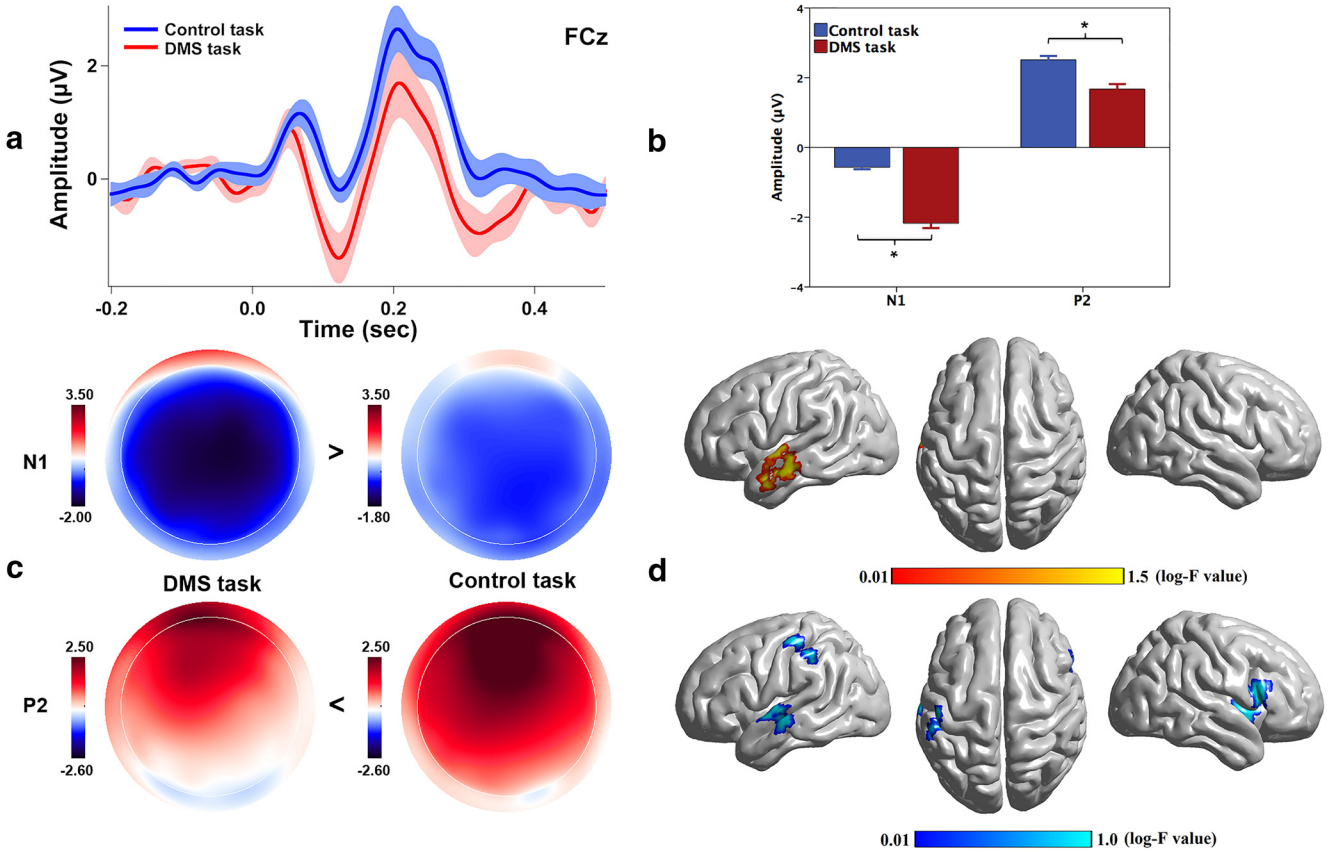

Figure 3. ERP findings of Experiment 1. ***a***, Grand-averaged ERP waveforms at FCz in response to +200 cents pitch perturbations during the DMS (red) and control (blue) tasks. Highlighted areas indicate the SE of the mean ERP responses. ***b–d***, T-bar graphs (mean and SE), topographical distributions, and source reconstructions of the amplitudes of N1 and P2 responses to +200 cents pitch perturbations. The asterisks indicate significantly larger N1 amplitudes ($p = 0.002$) and smaller P2 amplitudes ($p = 0.013$) during the DMS task as compared with the control task. Warm colors indicate enhanced brain activity in the N1 time window and cold colors indicate suppressed brain activity in the P2 time window during the DMS task relative to the control task.

Table 1. sLORTA *log*-F statistics for maximum activations obtained from comparisons between the DMS and control tasks in the N1 time window (MNI coordinates)

Condition	BA	Brain region	*Log*-F value	*X*	*Y*	*Z*
DMS vs Control	21	Left MTG	1.404	−65	−15	−10
	22	Left STG	1.361	−45	−10	−5

Table 2. sLORTA *log*-F statistics for maximum activations obtained from comparisons between the DMS and control tasks in the P2 time window (MNI coordinates)

Condition	BA	Brain region	*Log*-F value	*X*	*Y*	*Z*
DMS vs Control	21	Left MTG	1.404	−65	−15	−10
	22	Left STG	1.361	−60	−15	0
	40	Left IPL	−0.740	−50	−40	40
	45	Right IFG	−0.907	60	20	10
	1	Left S1	−0.754	−50	−25	55
	13	Right insula	−0.726	45	10	5

($F_{(1,14)} = 6.764$, $p = 0.021$, partial $\eta^2 = 0.326$). The main effect of electrode ($F_{(9,126)} = 1.283$, $p = 0.293$, partial $\eta^2 = 0.084$) was not significant nor was the interaction between task and electrode ($F_{(9,126)} = 0.983$, $p = 0.413$, partial $\eta^2 = 0.066$).

The results thus far show the modulatory effects of WM on the processing of pitch feedback errors; however, they do not shed light on the neural substrates involved in the observed WM-driven modulation of vocal pitch regulation. To clarify this issue, we performed source reconstruction analysis of the N1 and P2 responses with sLORETA to localize the neural resources that support the WM-related changes in vocal pitch regulation. The results revealed that the enhanced N1 responses during the DMS task were a result of increased brain activity in the left middle temporal gyrus (MTG; BA 21) and left superior temporal gyrus (STG; BA 22; Table 1; Fig. 3*d*). Both regions have been identified as areas involved in the detection of feedback errors during vocal output monitoring (Parkinson et al., 2012; Huang et al., 2016a). On the other hand, the suppressed P2 responses were found to receive contributions from a complex network, including in the right inferior frontal gyrus (IFG; BA 45), left STG (BA 22), left MTG (BA 21), left inferior parietal lobule (IPL; BA 40), left somatosensory cortex (S1; BA 1), and right insula (BA 13; Table 2; Fig. 3*d*). All regions have been found previously to be involved in the auditory-motor integration for vocal pitch regulation (Zarate and Zatorre, 2008; Behroozmand et al., 2015). Enhancement and suppression of activity in this fronto-temporo-parietal network are indicative of the differential modulatory effects of WM on the cortical processing of feedback errors during speech production.

Experiment 2

For the trained group, the FPR scores that index auditory WM capacity significantly increased from 43 ± 15% to 74 ± 15% after the training ($t_{(12)} = -10.356$, $p < 0.001$). Therefore, participants' auditory WM capacity was significantly improved following training. By contrast, the FPR scores did not differ as a function of

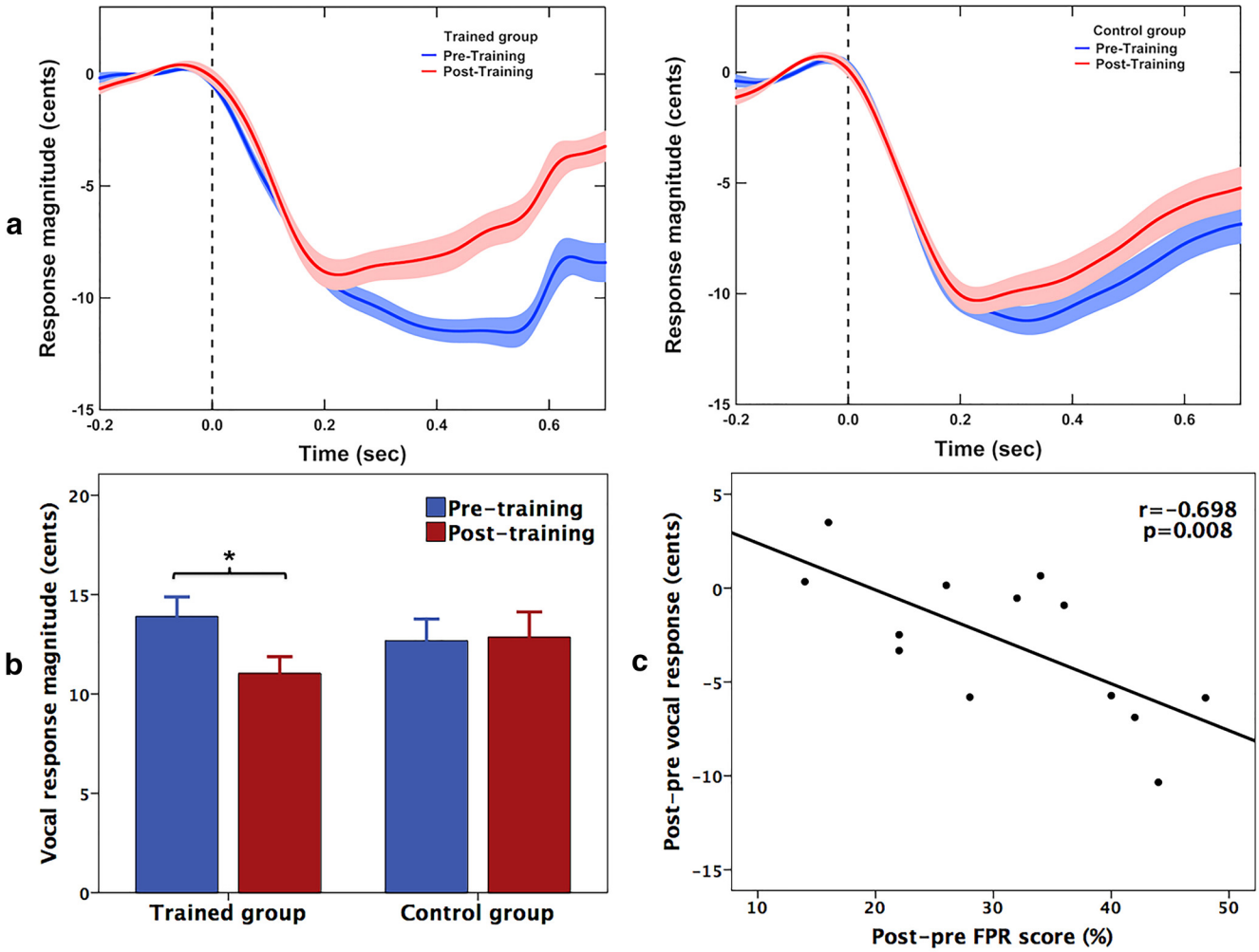

Figure 4. Behavioral results of Experiment 2. ***a***, Grand-averaged voice F_0 contours and (***b***) T-bar graphs of the absolute values of vocal responses (mean and SE) to +200 cents pitch perturbations during the pre- (blue) and post-training (red) phases for the trained and active control group, respectively. Highlighted areas indicate the SE of the mean vocal responses. The asterisk indicates significantly larger vocal responses during the pre- versus post-training phase ($p = 0.020$) for the trained group. ***c***, The trained group exhibited a significant negative correlation between the post-pre vocal responses and the post-pre WM FPR scores ($r = -0.698$, $p = 0.008$).

training for the active control group ($t_{(11)} = -0.544$, $p = 0.597$; 48 ± 23% vs 50 ± 22%).

Figures 4, *a* and *b*, show the grand-averaged voice F_0 contours and T-bar plots of the magnitudes of vocal responses to pitch perturbations before and after WM training for the trained and active control groups. A two-way RM-ANOVA conducted on the magnitudes of vocal responses revealed no significant main effects of phase ($F_{(1,23)} = 3.737$, $p = 0.066$, partial $\eta^2 = 0.140$) and group ($F_{(1,23)} = 0.053$, $p = 0.821$, partial $\eta^2 = 0.002$), whereas there was a significant interaction between phase and group ($F_{(1,23)} = 4.798$, $p = 0.039$, partial $\eta^2 = 0.173$). Following-up analyses revealed that the trained group exhibited a significant decrease of vocal response magnitudes after WM training ($F_{(1,12)} = 7.129$, $p = 0.020$, partial $\eta^2 = 0.373$; Fig. 4*a*,*b*). Moreover, the post-pre FPR scores were negatively correlated with the post-pre vocal response magnitudes ($r = -0.698$, $p = 0.008$; Fig. 4*c*), indicating that training-related WM capacity gains caused significantly decreased vocal compensations for pitch feedback errors. For the active control group, however, the magnitudes of vocal responses did not differ as a function of phase ($F_{(1,11)} = 0.043$, $p = 0.839$, partial $\eta^2 = 0.004$). For the latencies of vocal responses, there were no significant main effects of phase ($F_{(1,23)} = 0.418$, $p = 0.525$, partial $\eta^2 = 0.018$) and group ($F_{(1,23)} = 2.537$, $p = 0.125$, partial $\eta^2 = 0.099$), nor was there a significant interaction between these factors ($F_{(1,23)} = 0.091$, $p = 0.766$, partial $\eta^2 = 0.004$).

Figure 5*a* shows the grand-averaged ERP waveforms in response to pitch perturbations before and after WM training for the trained (left) and active control groups (right). Relative to the active control group, the trained group exhibited more prominent training effects on the cortical ERP responses. A three-way RM-ANOVA conducted on the N1 amplitudes revealed no significant main effects of phase ($F_{(1,23)} = 0.138$, $p = 0.714$, partial $\eta^2 = 0.006$), electrode ($F_{(9,207)} = 0.904$, $p = 0.453$, partial $\eta^2 = 0.038$), and group ($F_{(1,23)} = 2.091$, $p = 0.162$, partial $\eta^2 = 0.083$). Interactions between these variables did not reach significance either ($p > 0.05$ for all factors). Analyses of the N1 latencies revealed that N1 responses were significantly faster during the post-training phase than during the pretraining phase ($F_{(1,23)} = 14.256$, $p = 0.001$, partial $\eta^2 = 0.383$). However, the main effects of electrode ($F_{(9,207)} = 2.140$, $p = 0.109$, partial $\eta^2 = 0.085$) and group ($F_{(1,23)} = 0.168$, $p = 0.686$, partial $\eta^2 = 0.007$), and the interactions between these variables did not reach significance ($p > 0.05$).

By contrast, a three-way RM-ANOVA conducted on the P2 amplitudes revealed a significant main effect of electrode ($F_{(9,207)} = 21.086$, $p < 0.001$, partial $\eta^2 = 0.478$), which was primarily

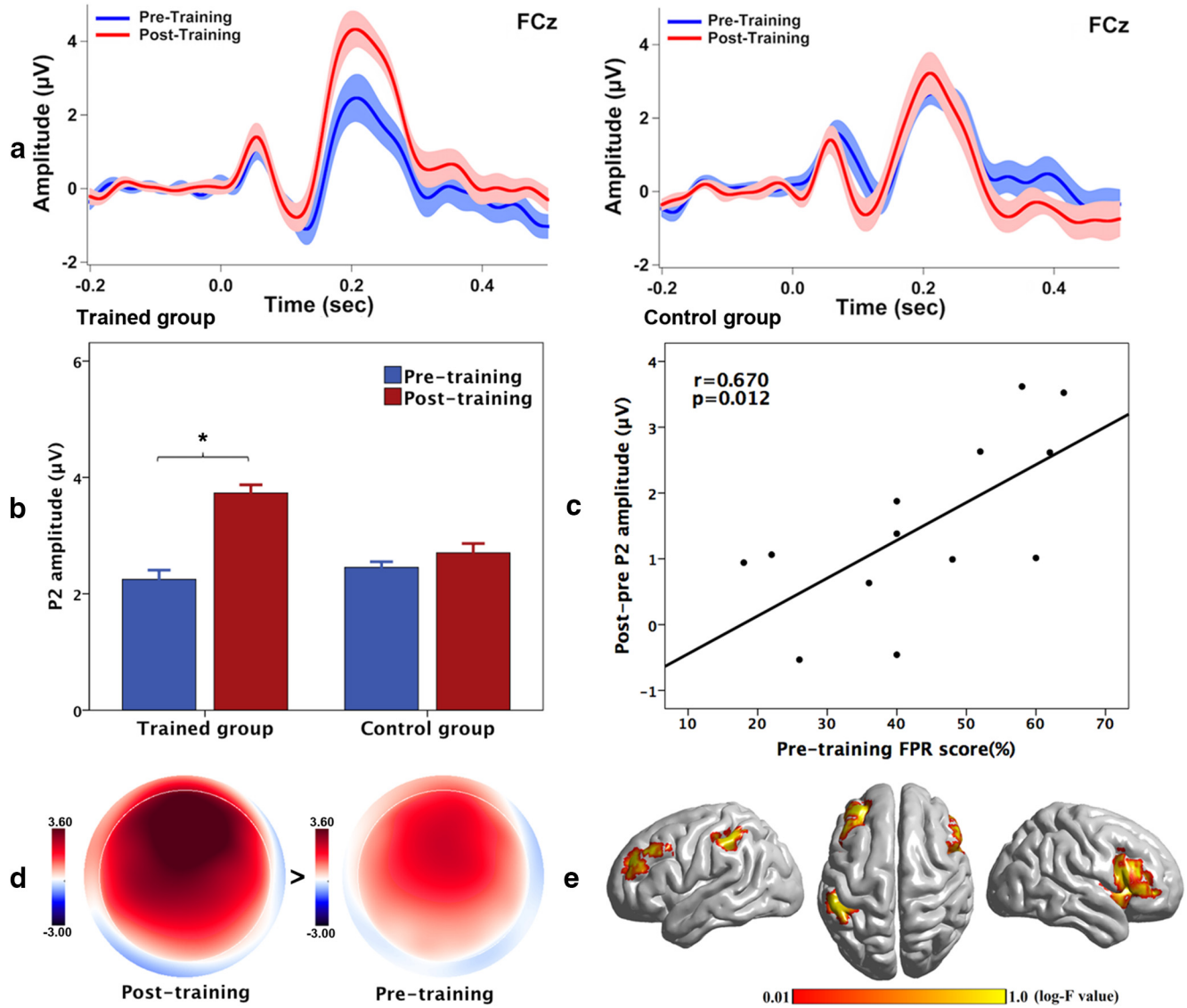

Figure 5. ERP findings of Experiment 2. ***a***, ***b***, Grand-averaged ERP waveforms at FCz and T-bar graphs of the amplitudes of P2 responses (mean and SE) to +200 cents pitch perturbations during the pre- (blue) and post-training (red) phases for the trained and active control groups, respectively. Highlighted areas indicate the SE of the mean ERP responses. The trained group exhibited significantly decreased response magnitudes after training ($p = 0.002$). ***c***, There was a significant positive correlation between the pretraining FPR score and the post-pre P2 amplitudes ($r = 0.670$, $p = 0.012$). ***d***, ***e***, Topographical distributions, and source reconstructions of the amplitudes of P2 responses to +200 cents pitch perturbations, showing enhanced brain activity in the P2 time window for the trained group during the post-training phase relative to the pretraining phase.

caused by larger P2 amplitudes recorded from the frontal electrodes as compared with the central electrodes. P2 amplitudes also significantly differed as a function of phase ($F_{(1,23)} = 9.821$, $p = 0.005$, partial $\eta^2 = 0.299$), and a significant interaction existed between phase and group ($F_{(1,23)} = 4.981$, $p = 0.036$, partial $\eta^2 = 0.178$); however, there was no main effect of group ($F_{(1,23)} = 0.605$, $p = 0.445$, partial $\eta^2 = 0.026$). Follow-up analyses showed that, although the active control group did not show significant training effects ($F_{(1,11)} = 0.359$, $p = 0.561$, partial $\eta^2 = 0.032$), P2 amplitudes for the trained group significantly increased following training ($F_{(1,12)} = 16.393$, $p = 0.002$, partial $\eta^2 = 0.577$; Fig. 5*a*,*b*,*d*). Moreover, there was a significant positive correlation between the FPR scores in the pretraining phase and the post-pre P2 amplitudes ($r = 0.670$, $p = 0.012$; Fig. 5*c*), indicating that auditory WM capacity was predictive of training-induced changes in the neural processing of auditory feedback during vocal pitch regulation. Source reconstruction revealed that the training-induced enhancement of P2 responses resulted from increased brain activity in the left middle frontal gyrus (MFG; BA 9), IPL (BA 40), right IFG (BA 45), and insula (BA 13; Table 3; Fig. 5*e*). In addition, analyses of the P2 latencies showed no main effects of phase ($F_{(1,23)} = 0.354$, $p = 0.558$, partial $\eta^2 = 0.015$), electrode ($F_{(9,207)} = 3.247$, $p = 0.085$, partial $\eta^2 = 0.123$), or group ($F_{(1,23)} = 0.589$, $p = 0.451$, partial $\eta^2 = 0.025$), nor were there significant interactions between these variables ($p > 0.05$).

Table 3. sLORTA *log*-F statistics for maximum activations obtained from comparisons between pre- and post-training sessions in the P2 time window (MNI coordinates)

Condition	BA	Brain region	*Log*-F value	*X*	*Y*	*Z*
Pretraining vs Post-training	9	Left MFG	0.846	−40	35	35
	40	Left IPL	0.713	−50	−35	60
	45	Right IFG	0.751	41	25	5
	13	Right insula	0.769	30	20	15

Discussion

Our results provide two lines of converging evidence to support the hypothesis that WM can exert modulatory effects on auditory-motor control of speech production. The DMS task that engages WM elicited not only significantly increased vocal compensations for pitch perturbations but also enhanced N1 responses in the left MTG and STG, and suppressed P2 responses in the left STG, MTG, IPL, S1, right IFG, and insula. On the other hand, extensive auditory WM training led to suppressed vocal compensations that correlated with improved WM capacity and enhanced P2 responses that were predicted by pretraining WM capacity in the left MFG, IPL, right IFG, and insula. These findings demonstrate, for the first time, that WM is critically involved in the detection and correction of feedback errors during vocal pitch regulation. Rather than being an exclusively bottom-up and automatic process, auditory-motor integration for voice control can be modulated by top-down influences arising from WM.

The enhanced N1 and suppressed P2 responses elicited by the DMS task were comparable with observations in previous WM-related studies. For example, participants who performed sound identity tasks (e.g., human, animal, and music sounds) exhibited enhanced N1 and suppressed P2 responses under conditions with increased levels of WM load (Alain et al., 2009). Also, the observed relationship between pretraining WM capacity and enhanced P2 responses following training was in line with our previous work that reported a significant correlation between improved WM capacity and the degree of P2 enhancement (Li et al., 2015). More importantly, the present study revealed the first behavioral evidence of the influence of WM on auditory-vocal integration, as reflected by significantly increased vocal compensations when WM was engaged by the DMS task, and significantly decreased vocal compensations as auditory WM capacity improved after training.

The role of WM in speech motor control: error detection

An important component of speech motor control is the detection of errors in vocal output through a comparison of incoming and predicted auditory feedback (Guenther, 2006; Houde and Nagarajan, 2011). Given that N1 responses were mostly reduced to participants' own unaltered voice relative to their pitch-shifted and alien voice (Heinks-Maldonado et al., 2005), and that the degree of N1 suppression decreased as the size of feedback perturbation increased (Behroozmand and Larson, 2011), this component is hypothesized to reflect the extent to which incoming auditory feedback matches feedback predictions derived from motor efference copy at an early stage of vocal motor control (Korzyukov et al., 2012a; Chen et al., 2015). In the present study, enhanced N1 responses to pitch perturbations during the DMS task relative to the control task received contributions from the left MTG and STG, regions that have been implicated in the detection of feedback errors during vocal pitch regulation (Behroozmand et al., 2015; Huang et al., 2016a). Particularly, an auditory error map in the posterior STG sends feedback commands to the motor control system in the DIVA model (Guenther et al., 2006), and activity in the STG is predictive of the magnitude of vocal compensation (Chang et al., 2013). Considering the important roles of the STG and MTG in the rehearsal and representation of information in auditory and verbal WM (Crottaz-Herbette et al., 2004; Buchsbaum et al., 2005), increased activity within these two regions during the DMS task may reflect an allocation of more neural resources related to auditory WM to evaluation of pitch feedback errors. Therefore, we propose that an important role of WM in speech motor control is to facilitate the online detection of discrepancies between incoming and predicted auditory feedback.

Interestingly, activation of the prefrontal cortex that has been considered critical in the storage and manipulation of WM information (D'Esposito, 2007), was not observed in the source analysis of the N1 responses. One possible reason is that the N1 response is considered as a sensory response that mainly receives contributions from primary and secondary auditory cortices (Näätänen and Picton, 1987), and that precise representations of auditory WM information can be stored in the auditory cortex (Scott et al., 2014; Huang et al., 2016b). Therefore, activation of left MTG and STG, but deactivation of the prefrontal regions in the WM processing of pitch feedback errors, lends support to the idea that sensory nature of sustained WM information can be stored and represented in the sensory cortices (Sreenivasan et al., 2014).

The role of WM in speech motor control: error correction

Once errors between incoming and predicted feedback are detected, corrective motor commands are generated and executed to adjust vocal motor behaviors. The DMS task elicited suppressed P2 responses in the left STG, MTG, IPL, S1, and right IFG and insula. Moreover, extensive auditory WM training led to enhanced P2 responses in the left MFG, IPL, right IFG, and insula. These brain regions have been identified both in the verbal WM tasks (Paulesu et al., 1993; Buchsbaum et al., 2005) and in the generation of vocal compensation for pitch feedback errors (Zarate et al., 2010; Behroozmand et al., 2015). Particularly, the left IPL and right IFG have been implicated in transforming feedback errors into compensatory speech motor commands (Hickok et al., 2003; Rauschecker and Scott, 2009). Thus, the P2 component is thought to reflect the later activity of higher-level auditory-motor interactions (Chen et al., 2015). Because WM capacity is limited (Luck and Vogel, 2013), the DMS task may recruit more neural resources for the storage and rehearsal of pitch perturbations such that fewer neural sources are allocated to the auditory-motor representations, leading to suppressed P2 responses. Accordingly, enhanced P2 responses in Experiment 2 can be accounted for by an increase in WM-related resources available for allocation to coordinate auditory-motor transformation because of training improved WM capacity. Additionally, participants' pretraining auditory WM capacity was predictive of enhanced P2 responses following training. Together with our previous work that reported a positive relationship between improved WM capacity and P2 responses (Li et al., 2015) and a significant correlation between intrinsic brain activity in the WM-related regions (e.g., IFG, STG) and P2 amplitude (Guo et al., 2016); this finding suggests that the later-stage auditory-motor processing of vocal feedback errors as reflected by the P2 response is closely related to WM capacity.

At the level of vocal output, suppressed vocal compensations were predicted by improved WM capacity. Similarly, Ranasinghe et al. (2017) reported significant correlations between enhanced vocal compensations and executive dysfunction and between reduced compensation durations and memory dysfunction in AD patients. Therefore, WM appears to exert an inhibitory influence on vocal motor behavior. As an important function that inhibits reflex-like or inappropriate behavioral responses (Burle et al., 2004), inhibitory control is closely interrelated with WM (Chmielewski et al., 2015) and neural networks involved in inhibitory control and WM largely overlap (e.g., IPL, right IFG, and anterior insula; Aron et al., 2004; Barber et al., 2013; Chmielewski et al., 2017). Furthermore, inhibitory control depends directly on the amount

of WM resources available to suppress the to-be-ignored processes: decreased activity in brain regions that support WM (e.g., IFG, MFG) results in impaired inhibitory control processes (Barber et al., 2013; Chmielewski et al., 2015). In the present study, enhanced or suppressed vocal compensations were associated with suppressed or enhanced P2 responses in the left IPL, right IFG, and insula, regions that serve as a core network for inhibitory control (Aron et al., 2004; Barber et al., 2013). Thus, another important role of WM in speech motor control is to exert inhibitory control over vocal adjustment to prevent vocal production from being excessively influenced by auditory feedback. Abnormally enhanced vocal compensations observed in AD patients may be a result of impaired inhibitory control over speech motor behavior (Ranasinghe et al., 2017).

Effects of WM or attention?

WM and attention are conventionally viewed as overlapping cognitive functions that rely on the frontoparietal regions (Gazzaley and Nobre, 2012). Thus, one may argue that top-down modulation of auditory-vocal integration also receives contributions from attentional mechanisms. There is increasing evidence, however, suggesting that the distinct neural networks underlie WM and attention in the auditory domain (Rinne et al., 2009; Huang et al., 2013). Moreover, contrary to our findings in the DMS task, attending to pitch perturbations led to intact N1 but enhanced P2 responses (Hu et al., 2015; Liu et al., 2015). Interestingly, enhanced N1 and suppressed P2 responses were found in a divided attention task (Liu et al., 2015), in which WM was required to process multiple independent sensory information (Johnson and Zatorre, 2006). Thus, the observed modulation of neurobehavioral responses to pitch perturbations would be most likely due to top-down influences arising from WM, and confirming this hypothesis will require future studies.

In summary, our results demonstrate that auditory WM can exert modulatory influences on auditory-motor integration for voice control. The functional role of auditory WM is speculated to include two subprocesses: enhance the perception of mismatches between incoming and predicted auditory feedback (error detection); inhibit vocal compensations for feedback errors to stabilize voice control (error correction). Our study provides insights into a top-down mechanism by which auditory WM facilitates the feedback-based control of speech production.

References

Alain C, McDonald KL, Kovacevic N, McIntosh AR (2009) Spatiotemporal analysis of auditory "what" and "where" working memory. Cereb Cortex 19:305–314. CrossRef Medline

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8:170–177. CrossRef Medline

Barber AD, Caffo BS, Pekar JJ, Mostofsky SH (2013) Effects of working memory demand on neural mechanisms of motor response selection and control. J Cogn Neurosci 25:1235–1248. CrossRef Medline

Behroozmand R, Larson CR (2011) Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci 12:54. CrossRef Medline

Behroozmand R, Karvelis L, Liu H, Larson CR (2009) Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clin Neurophysiol 120:1303–1312. CrossRef Medline

Behroozmand R, Shebek R, Hansen DR, Oya H, Robin DA, Howard MA 3rd, Greenlee JD (2015) Sensory-motor networks involved in speech production and motor control: an fMRI study. Neuroimage 109:418–428. CrossRef Medline

Behroozmand R, Oya H, Nourski KV, Kawasaki H, Larson CR, Brugge JF, Howard MA 3rd, Greenlee JD (2016) Neural correlates of vocal production and motor control in human Heschl's gyrus. J Neurosci 36:2302–2315. CrossRef Medline

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

Buchsbaum BR, Olsen RK, Koch P, Berman KF (2005) Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron 48:687–697. CrossRef Medline

Burle B, Vidal F, Tandonnet C, Hasbroucq T (2004) Physiological evidence for response inhibition in choice reaction time tasks. Brain Cogn 56:153–164. CrossRef Medline

Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. J Acoust Soc Am 103:3153–3161. CrossRef Medline

Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc Natl Acad Sci U S A 110:2653–2658. CrossRef Medline

Chen SH, Liu H, Xu Y, Larson CR (2007) Voice F0 responses to pitch-shifted voice feedback during English speech. J Acoust Soc Am 121:1157–1163. CrossRef Medline

Chen Z, Chen X, Liu P, Huang D, Liu H (2012a) Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. BMC Neurosci 13:55. CrossRef Medline

Chen Z, Liu P, Wang EQ, Larson CR, Huang D, Liu H (2012b) ERP correlates of language-specific processing of auditory pitch feedback during self-vocalization. Brain Lang 121:25–34. CrossRef Medline

Chen Z, Wong FC, Jones JA, Li W, Liu P, Chen X, Liu H (2015) Transfer effect of speech-sound learning on auditory-motor processing of perceived vocal pitch errors. Sci Rep 5:13134. CrossRef Medline

Chmielewski WX, Mückschel M, Stock AK, Beste C (2015) The impact of mental workload on inhibitory control subprocesses. Neuroimage 112:96–104. CrossRef Medline

Chmielewski WX, Mückschel M, Ziemssen T, Beste C (2017) The norepinephrine system affects specific neurophysiological subprocesses in the modulation of inhibitory control by working memory demands. Hum Brain Mapp 38:68–81. CrossRef Medline

Crottaz-Herbette S, Anagnoson RT, Menon V (2004) Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. Neuroimage 21:340–351. CrossRef Medline

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21. CrossRef Medline

D'Esposito M (2007) From cognitive to neural models of working memory. Philos Trans R Soc Lond B Biol Sci 362:761–772. CrossRef Medline

Ferree TC, Luu P, Russell GS, Tucker DM (2001) Scalp electrode impedance, infection risk, and EEG data quality. Clin Neurophysiol 112:536–544. CrossRef Medline

Fuchs M, Kastner J, Wagner M, Hawes S, Ebersole JS (2002) A standardized boundary element method volume conductor model. Clin Neurophysiol 113:702–712. CrossRef Medline

Gazzaley A, Nobre AC (2012) Top-down modulation: bridging selective attention and working memory. Trends Cogn Sci 16:129–135. CrossRef Medline

Guenther FH (2006) Cortical interactions underlying the production of speech sounds. J Commun Disord 39:350–365. CrossRef Medline

Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 96:280–301. CrossRef Medline

Guo Z, Huang X, Wang M, Jones JA, Dai Z, Li W, Liu P, Liu H (2016) Regional homogeneity of intrinsic brain activity correlates with auditory-motor processing of vocal pitch errors. Neuroimage 142:565–575. CrossRef Medline

Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK (2000) Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. Exp Brain Res 130:133–141. CrossRef Medline

Hawco CS, Jones JA, Ferretti TR, Keough D (2009) ERP correlates of online monitoring of auditory feedback during vocalization. Psychophysiology 46:1216–1225. CrossRef Medline

Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM (2005) Fine-tuning of auditory cortex during speech production. Psychophysiology 42:180–190. CrossRef Medline

Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci 15:673–682. CrossRef Medline

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69:407–422. CrossRef Medline

Houde JF, Chang EF (2015) The cortical computations underlying feedback control in vocal production. Curr Opin Neurobiol 33:174–181. CrossRef Medline

Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. Science 279:1213–1216. CrossRef Medline

Houde JF, Nagarajan SS (2011) Speech production as state feedback control. Front Hum Neurosci 5:82. CrossRef Medline

Hu H, Liu Y, Guo Z, Li W, Liu P, Chen S, Liu H (2015) Attention modulates cortical processing of pitch feedback errors in voice control. Sci Rep 5:7812. CrossRef Medline

Huang S, Seidman LJ, Rossi S, Ahveninen J (2013) Distinct cortical networks activated by auditory attention and working memory load. Neuroimage 83:1098–1108. CrossRef Medline

Huang X, Chen X, Yan N, Jones JA, Wang EQ, Chen L, Guo Z, Li W, Liu P, Liu H (2016a) The impact of Parkinson's disease on the cortical mechanisms that support auditory-motor integration for voice control. Hum Brain Mapp 37:4248–4261. CrossRef Medline

Huang Y, Matysiak A, Heil P, König R, Brosch M (2016b) Persistent neural activity in auditory cortex is related to auditory working memory in humans and nonhuman primates. eLife 5:e15441. CrossRef Medline

Jacquemot C, Scott SK (2006) What is the relationship between phonological short-term memory and speech processing? Trends Cogn Sci 10:480–486. CrossRef Medline

Johnson JA, Zatorre RJ (2006) Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. Neuroimage 31:1673–1681. CrossRef Medline

Keough D, Hawco C, Jones JA (2013) Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. BMC Neurosci 14:25. CrossRef Medline

Korzyukov O, Karvelis L, Behroozmand R, Larson CR (2012a) ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. Int J Psychophysiol 83:71–78. CrossRef Medline

Korzyukov O, Sattler L, Behroozmand R, Larson CR (2012b) Neuronal mechanisms of voice control are affected by implicit expectancy of externally triggered perturbations in auditory feedback. PLoS One 7:e41216. CrossRef Medline

Li W, Chen Z, Liu P, Zhang B, Huang D, Liu H (2013) Neurophysiological evidence of differential mechanisms involved in producing opposing and following responses to altered auditory feedback. Clin Neurophysiol 124:2161–2171. CrossRef Medline

Li W, Guo Z, Jones JA, Huang X, Chen X, Liu P, Chen S, Liu H (2015) Training of working memory impacts neural processing of vocal pitch regulation. Sci Rep 5:16562. CrossRef Medline

Liu H, Meshman M, Behroozmand R, Larson CR (2011) Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. Clin Neurophysiol 122:951–957. CrossRef Medline

Liu P, Chen Z, Larson CR, Huang D, Liu H (2010) Auditory feedback control of voice fundamental frequency in school children. J Acoust Soc Am 128:1306–1312. CrossRef Medline

Liu Y, Hu H, Jones JA, Guo Z, Li W, Chen X, Liu P, Liu H (2015) Selective and divided attention modulates auditory-vocal integration in the processing of pitch feedback errors. Eur J Neurosci 42:1895–1904. CrossRef Medline

Luck SJ, Vogel EK (2013) Visual working memory capacity: from psychophysics and neurobiology to individual differences. Trends Cogn Sci 17:391–400. CrossRef Medline

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, et al. (2001) A probabilistic atlas and reference system for the human brain: international consortium for brain mapping (ICBM). Philos Trans R Soc Lond B Biol Sci 356:1293–1322. CrossRef Medline

Mulert C, Jäger L, Schmitt R, Bussfeld P, Pogarell O, Möller HJ, Juckel G, Hegerl U (2004) Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. Neuroimage 22:83–94. CrossRef Medline

Munhall KG, MacDonald EN, Byrne SK, Johnsrude I (2009) Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. J Acoust Soc Am 125:384–390. CrossRef Medline

Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24:375–425. CrossRef Medline

Parkinson AL, Flagmeier SG, Manes JL, Larson CR, Rogers B, Robin DA (2012) Understanding the neural mechanisms involved in sensory control of voice production. Neuroimage 61:314–322. CrossRef Medline

Pascual-Marqui RD (2002) Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol 24:5–12. Medline

Paulesu E, Frith CD, Frackowiak RS (1993) The neural correlates of the verbal component of working memory. Nature 362:342–345. CrossRef Medline

Ranasinghe KG, Gill JS, Kothare H, Beagle AJ, Mizuiri D, Honma SM, Gorno-Tempini ML, Miller BL, Vossel KA, Nagarajan SS, Houde JF (2017) Abnormal vocal behavior predicts executive and memory deficits in Alzheimer's disease. Neurobiol Aging 52:71–80. CrossRef Medline

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724. CrossRef Medline

Richardson JTE (2011) Eta squared and partial eta squared as measures of effect size in educational research. Educ Res Rev 6:135–147. CrossRef

Rinne T, Koistinen S, Salonen O, Alho K (2009) Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. J Neurosci 29:13338–13343. CrossRef Medline

Scott BH, Mishkin M, Yin P (2014) Neural correlates of auditory short-term memory in rostral superior temporal cortex. Curr Biol 24:2767–2775. CrossRef Medline

Sheft S, Shafiro V, Cortese K (2014) Relationship of frequency-pattern training to speech perception. In: Auditory plasticity: listening with the brain (Dau T, Santurette S, Dalsgaard JC, Tranebjærg L, Andersen T, and Poulsen T, eds), pp 53–60. Ballerup, Denmark: Danavox Jubilee.

Sreenivasan KK, Curtis CE, D'Esposito M (2014) Revisiting the role of persistent neural activity during working memory. Trends Cogn Sci 18:82–89. CrossRef Medline

Tumber AK, Scheerer NE, Jones JA (2014) Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. PLoS One 9:e109968. CrossRef Medline

Zarate JM, Zatorre RJ (2008) Experience-dependent neural substrates involved in vocal pitch regulation during singing. Neuroimage 40:1871–1887. CrossRef Medline

Zarate JM, Wood S, Zatorre RJ (2010) Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. Neuropsychologia 48:607–618. CrossRef Medline

Zumsteg D, Lozano AM, Wennberg RA (2006) Depth electrode recorded cerebral responses with deep brain stimulation of the anterior thalamus for epilepsy. Clin Neurophysiol 117:1602–1609. CrossRef Medline