Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production

Matthias K. Franken [a,b,1], Frank Eisner [a], Daniel J. Acheson [a,b], James M. McQueen [a], Peter Hagoort [a,b], Jan-Mathijs Schoffelen [a,*]

[a] Donders Institute for Brain, Cognition and Behaviour, Radboud University, P.O. Box 9101, Nijmegen, 6500 HB, The Netherlands
[b] Max Planck Institute for Psycholinguistics, P.O. Box 310, Nijmegen, 6500 AH, The Netherlands

ABSTRACT

Speaking is a complex motor skill which requires near instantaneous integration of sensory and motor-related information. Current theory hypothesizes a complex interplay between motor and auditory processes during speech production, involving the online comparison of the speech output with an internally generated forward model. To examine the neural correlates of this intricate interplay between sensory and motor processes, the current study uses altered auditory feedback (AAF) in combination with magnetoencephalography (MEG). Participants vocalized the vowel /e/ and heard auditory feedback that was temporarily pitch-shifted by only 25 cents, while neural activity was recorded with MEG. As a control condition, participants also heard the recordings of the same auditory feedback that they heard in the first half of the experiment, now without vocalizing. The participants were not aware of any perturbation of the auditory feedback. We found auditory cortical areas responded more strongly to the pitch shifts during vocalization. In addition, auditory feedback perturbation resulted in spectral power increases in the θ and lower β bands, predominantly in sensorimotor areas. These results are in line with current models of speech production, suggesting auditory cortical areas are involved in an active comparison between a forward model's prediction and the actual sensory input. Subsequently, these areas interact with motor areas to generate a motor response. Furthermore, the results suggest that θ and β power increases support auditory-motor interaction, motor error detection and/or sensory prediction processing.

Introduction

Speaking is a remarkably complex motor skill. We speak at a rate of often more than 10 speech sounds per second, each of which require accurate coordination of more than 100 different muscles. We make use of this skill day in day out, throughout our lives, usually without conscious awareness of the complexity of the task. If attention is paid to phonological aspects of speech production, it is mostly focused on wording, while articulation follows effortlessly. In order to perform this motor task almost without errors, a good quality control system is needed. Recent developments in speech motor control have shown that integration of sensorimotor information, including auditory feedback (i.e. the sound of our own voice), is key in this respect. The current study investigates the neural underpinnings of sensorimotor integration during speech production.

The role of auditory feedback in speech production has been investigated by providing speakers with online manipulated feedback (Houde and Jordan, 1998; Burnett et al., 1998; Jones and Munhall, 2000). For example, speakers could be hearing their own speech in real time at a higher pitch or with a lower first formant. It turns out that speakers usually compensate for these manipulations by changing their speech in the opposite direction (that is, by lowering the pitch, or by increasing the frequency in the first formant, which results in a change in vowel quality). This compensatory response occurs even when participants are told to ignore the altered feedback (Keough et al., 2013). This suggests that speakers automatically monitor their auditory feedback during speech production. Cognitive modeling work in this context has drawn from principles in motor control more generally, in order to explain such a fast feedback monitoring mechanism (Wolpert et al., 1995; Wolpert and Ghahramani, 2000). These models hypothesize the use of internally generated forward models (Houde and Nagarajan, 2011; Tourville and Guenther, 2011). Specifically, all articulatory motor

---

programs which are generated in (and will be executed by) the motor system are sent to the auditory system. Each of these efference copies can be used to create a forward model, which models the sensory (auditory) consequences of the articulation. This sensory prediction can then be compared with the observed sensory consequences, and if necessary generate a prediction error that could signal the need for behavioral adaptation.

Using the altered auditory feedback paradigm, several functional magnetic resonance imaging studies have shown that feedback processing is supported by an extended bilateral functional neural network including auditory and motor-related areas (Behroozmand et al., 2015b; Zarate et al., 2010; Zarate and Zatorre, 2005; Zheng et al., 2010; Zheng et al., 2013). Electrophysiological studies using electroencephalography (EEG) to investigate the temporal dynamics of feedback processing have shown that altered feedback leads to a brain response as early as 100 m s after perturbation onset (Behroozmand et al., 2009; Behroozmand et al., 2011; Behroozmand and Larson, 2011; Hawco et al., 2009). The early latency of these findings suggests that auditory processing and motor control already interact at an early processing stage. In addition, in a MEG study, Kort et al., (2014) show responses of a broad bilateral cortical network to an unexpected 100-cent pitch shift in auditory feedback. These authors found enhanced neural activity in response to pitch perturbations in sensorimotor, auditory and premotor cortices.

The current study investigates the neural correlates of pitch perturbation processing and of the subsequent automatic responses to these perturbations. Importantly, we used a small perturbation magnitude (25 cents), to make sure that the participants did not consciously detect the perturbation. This was done to substantiate the claim that speakers' responses to altered auditory feedback are not subject to conscious awareness (Behroozmand et al., 2015a). In most studies, the perturbations used are large enough to trigger conscious processing, and therefore possibly recruit attentional resources. Since it has been established that attention can indeed modulate speakers' responses to unexpected auditory feedback (Hu et al., 2015; Korzyukov et al., 2012; Liu et al., 2015), it is crucial to avoid attentional effects by keeping the perturbation small.

In addition, in this study we performed a detailed analysis of neural oscillatory activity in relation to the feedback perturbations. So far, only a small number of studies on feedback perturbations have looked beyond evoked responses. This may be surprising, as recent dynamic approaches to cognition have linked cortical oscillations to predictive processing (Engel et al., 2001) and sensorimotor integration more generally (Caplan et al., 2003), as well as to speech production specifically (Cruikshank et al., 2012; Gehrig et al., 2012; Jenson et al., 2014). Two recent studies suggested that spectral power increases in the δ (1–4 Hz), θ (4–8 Hz) and γ (65–150 Hz) bands over motor and sensory areas reflect sensorimotor speech processing (Behroozmand et al., 2015a; Kort et al., 2016). The current study looks at responses in the lower frequency range to a much smaller pitch shift (only 25 cents instead of 100 cents).

We also investigated the neural correlates of the different types of response (opposing versus following) to the perturbation. Although the typical response to a feedback pitch perturbation (for instance: an increase) is a compensatory change in the opposing direction (for instance: a decrease), occasionally participants respond by actually following the direction of the perturbation (Behroozmand et al., 2012; Franken et al., 2018; Larson et al., 2007).

Materials and methods

Subjects

Thirty-nine healthy volunteers (age: M = 22, range = 18–34; 27 females) participated after providing written informed consent in accordance with the Declaration of Helsinki and the local ethics board committee (CMO region Arnhem/Nijmegen). All participants had normal hearing, were native speakers of Dutch and had no history of speech and/or language pathology.

Paradigm

An experimental session consisted of two tasks, a speaking and a listening task, always performed in the same order (speaking, then listening), while brain activity was measured using MEG.

In the speaking task, participants performed a tone-matching task (Liu and Larson, 2007; Hawco et al., 2009). This task was chosen to keep participants attentive. A trial started with the presentation of a short tone (duration 700 m s). 200 m s after the tone offset, a visual cue ("EE", in Dutch pronounced as /e/) instructed the participants to start vocalizing /e/, while trying to match the pitch of the tone they just heard. The visual cue disappeared after 3s, cueing the participant to stop vocalizing. During speech production/vocalization, the participant's voice was recorded using a microphone, positioned about 1.5 m from the participant to avoid any artifacts in the MEG signal. The recorded signal was used to provide the participants with online auditory feedback. In half of the trials, participants received normal auditory feedback throughout the trial, i.e. participants' speech was recorded and played back to them unaltered (henceforth control trials). In the other half of the trials (perturbation trials), auditory feedback was normal at first, but, starting between 500 and 1500 m s after speech onset (randomly jittered), the feedback's pitch was increased by 25 cents for a duration of 500 m s, before returning back to normal feedback for the remainder of the trial. The only difference in auditory feedback between control and perturbation trials was this 500 m s pitch shift. The duration of the pitch shift is rather long compared to previous studies (Burnett et al., 1998; Hain et al., 2000), in order to have a broad time window during the shift for time-frequency analyses. The shift duration is not much longer compared to the 400 m s shifts in Kort et al., (2014, 2016). Overall, participants received 99 perturbation trials and 99 control trials, randomly mixed in two blocks of 99 trials each. After the speaking task, participants did the passive listening task, in which the participants were shown the same visual cues as in the production task, but were instructed not to speak. Instead, they listened to recordings of the very same feedback they were given in the speaking task.

Finally, after the experiment, participants filled out a short debriefing questionnaire, which asked whether they noticed any feedback manipulations and if so, what kind of manipulations.

Materials

The tone stimuli were 700 m s pure tones at one of three pitch frequencies. The pitch of the tones was individually tailored to the participants at 4, 8 and 11 semitones above their conversational pitch. This was done by having participants produce the vowel /e/ five times (they were not yet aware the experiment would involve pitch), and the average pitch was considered their conversational pitch.

The auditory feedback shifts were implemented using Audapter software (Cai et al., 2008; Tourville et al., 2013). In brief, the software performs a near-real-time autocorrelation analysis to track the pitch. In order to shift the pitch, the short-time Fourier spectra were stretched and interpolated along the frequency axis. The pitch-shifted sounds were played back to the speaker through audio air tubes with a latency of 10–20 m s.

All voice recordings were made on one channel using a Sennheiser ME64 cardioid microphone, which was set up in the MEG magnetically shielded room and connected through an in-house-built audio mixer to a dedicated soundcard Motu MicroBook II outside the room, which was connected to a Windows computer. Auditory feedback was delivered through the same soundcard which was connected to CTF (VSM/CTF systems, Port Coquitlam, Canada) audio air tubes. Stimulus presentation and sound recording times were controlled by the same Windows computer running Audapter and MathWorks Matlab (MathWorks, Version 8 Release 5, Natick, MA).

MEG acquisition

We used an MEG system (VSM/CTF systems, Port Coquitlam, Canada) with 275 axial gradiometers. Three localization coils, fixed to anatomical landmarks (nasion, left and right preauricular points), were used to determine head position relative to the gradiometers. All data were low-pass filtered by an anti-aliasing filter (300 Hz cut-off), digitized at 1200 Hz and stored for offline analysis. Participants were seated upright, with the head rested against the back of the helmet and touching the top of the helmet. A small cushion was used to fix the head's position so as to minimize free head movement. The participant's head movement and position was monitored in realtime and, if necessary, adjusted between blocks (Stolk et al., 2013). A headband was used to cover the audio air tubes and the participants' ears, minimizing the effect of air-conducted auditory feedback.

MRI acquisition

In order to reconstruct the sources of the sensor-level MEG results, T1-weighted anatomical MRI scans were acquired for 34 out of the 39 subjects. Scans were acquired using Siemens 1.5T Avanto scanner for 24 participants, a Siemens 3T Prisma scanner for 6 participants, and a Siemens 3T Skyra scanner for 4 participants, depending on scanner availability.

Analysis

Behavioral

For every trial of the speaking task, the pitch of participants' vocalization was determined using the autocorrelation method implemented in Praat (Boersma and Weenink, 2013). Subsequently, the pitch contours of all trials were exported to MATLAB (The MathWorks Inc., Natick, MA, 2012) for further processing.

Pitch contours were epoched from 500 m s before perturbation onset to 1000 m s after perturbation onset. For the control trials, in which there was no perturbation onset, random time points were chosen, while making sure the distribution of these time points across trials was equal to the distribution of perturbation onsets within the same subject. The data was de-trended and converted from Hertz to the Cents scale using the following formula:

$$F0\ [cents] = 1200*\log_2\left(\frac{F}{F_{baseline}}\right)$$

Here, F is the original pitch frequency in Hertz, while $F_{baseline}$ is the average pitch frequency in Hertz across a baseline window (-200 m s–0 m s before perturbation onset). Subsequently, trials that contained artifacts were removed from analysis. Artifacts were detected by visual inspection, looking for sharp discontinuities in the pitch contour, or the absence of a pitch contour.

MEG preprocessing

All analyses were performed in MATLAB (The MathWorks Inc., Natick, MA, 2012), using custom scripts and the Fieldtrip toolbox (Oostenveld et al., 2011).

First, data was epoched from 1s before speech onset (or audio onset in the listening task) to 6s after speech onset. Bad channels were removed, and the data was de-meaned and visually inspected for artifacts. Segments containing artifacts were removed. Subsequently, an independent component analysis (ICA) algorithm (Hyvärinen, 1999) was applied to identify eye movement and heartbeat artifacts. The ICA components whose time course showed strong coherence with EOG and ECG channels were removed from the data. Furthermore, the spatial topographies and time courses of all components were visually inspected, and components showing artifacts were removed from the data. On average, about 6 components were removed for each subject (ranging from 3 to 9 components).

Event-related field (ERF) analysis

MEG data was time-locked to perturbation onset, or, for the control trials, to a randomly chosen time point (see behavioral analysis). Every trial was filtered using a zero-phase forward windowed sinc FIR filter with a Hamming window and a 1–40 Hz passband. Subsequently, the data was cut into time windows from -1s to 2s after perturbation onset, de-trended, averaged per condition and per participant and converted into synthetic planar gradients (Bastiaansen and Knosche, 2000).

An additional analysis examined the neural correlates of the distinction between following and opposing responses to the altered auditory feedback. Trials were classified as having either an opposing or a following response using the method described in Franken et al., (2018). In brief, trials were automatically classified using two methods, one based on the slope of the pitch contour after perturbation onset and a second one based on automatic detection of the change in the distribution of the pitch contour over time (using the Castellan change-point test). If the two methods led to a different classification, trials were classified using visual inspection. As the distribution of trials was very uneven between the opposing and following classes, the following procedure was used to enable an unbiased statistical comparison between the neural responses for opposing and following trials. For every participant, the minimum number of trials in a response class was determined (most often this was in the following-response class). Five times, a random subset of that number of trials was selected (without replacement) from the other response class. The event-related field (ERF) response for that response class was calculated by averaging the ERF across the five trial subsets. This way, each ERF was calculated by averaging, within each participant, over the same number of trials in both response classes. However, this procedure could have affected our results by smoothing the data in one condition compared to the other. Therefore, we made sure the results were not affected by performing the same analysis without averaging across five trial subsets (instead, just one trial subset was selected). This led to the same pattern of results.

Time-frequency analysis

For the time-frequency analyses, the data (defined from -1s to 2s relative to perturbation onset) was de-meaned and transformed to the frequency domain using a sliding 500 m s Hanning tapered window, sliding in steps of 50 m s from -500 m s to 1500 m s after perturbation onset. The frequency band of interest ranged from 2 Hz up to 30 Hz (in steps of 2 Hz). Before transformation, the data was zero-padded to 4s.

Statistical inference

To determine whether the difference between the perturbation and the control condition in the speaking task was statistically significant, we performed a non-parametric permutation test with a clustering method to correct for multiple comparisons (Maris and Oostenveld, 2007). This was done for the time window between 0 and 1s after perturbation onset. For each sample (channel-time-frequency point) the difference Perturbation-Control was expressed as a dependent samples t-statistic. Samples for which these t-statistics exceeded an uncorrected α threshold of .05 were clustered based on spatial, temporal, and spectral adjacency. Cluster-level test statistics were calculated by summing the t-statistics of the samples belonging to the same cluster. The largest cluster-level t statistic was used as a test statistic as it was suggested (among others) in Maris & Oostenveld (2007). The advantage of this test statistic is that it is sensitive to both the cluster's (spatial, temporal and spectral) extent, as well as to the effect size at individual samples. Next, a permutation distribution of cluster-level statistics was calculated by randomly exchanging data between the conditions, and calculating the maximal positive and negative cluster-level statistic for every permutation (for a total of 1000 permutations). The observed maximal cluster-level statistic was tested against the permutation distribution.

In order to compare the ERF results of the speaking and the listening

tasks across conditions, the average activity was calculated for every subject in both tasks and both conditions on a 100 m s time window centered at the point of the maximal t-statistic (averaged across channels) for the contrast perturbation-control collapsed across tasks at group level. This was done both for the maximal t-statistic after perturbation onset (222 m s) and after perturbation offset (627 m s). The resulting average activity values were entered in 2-way repeated measures ANOVAs (one for the averages after perturbation onset, and one for perturbation offset), with factors Task (speak vs. listen) and Condition (perturbation vs. control). Post-hoc t-tests were carried out to compare the perturbation and control trials within the listening task.
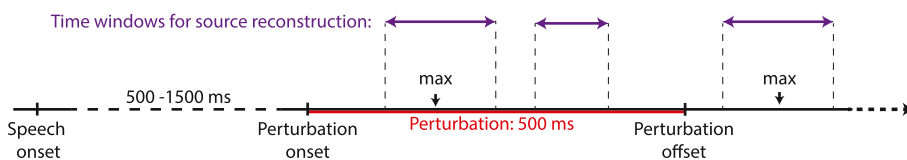
MRI processing

In order to estimate source-level activity, we co-registered the anatomical MRI to the MEG sensors. This was achieved by identifying in the MRI the anatomical locations that were used to place the head localization coils during the MEG measurement (left/right ear canal, and nasion). Subsequently, the aligned image was used to create (1) a volume conduction model based on a single shell model of the inner surface of the skull, and (2) a description of the cortical surface, using Freesurfer 5.1 (Dale et al., 1999). The individual cortical surfaces were surface-registered to a template and downsampled to 4002 nodes per hemisphere, using the Connectome Workbench software (http://www.humanconnectome.org/connectome/connectome-workbench.html).

Beamforming

The sensor-level results were projected onto the individual cortical surfaces using beamforming. Data visualization was performed using the Connectome Workbench of the Human Connectome Project (http://www.humanconnectome.org/connectome/connectome-workbench.html). For the event-related data, we used a time-domain beamformer (LCMV)(Van Veen et al., 1997). The data covariance was calculated across a time window ranging from −150 m s to 800 m s relative to perturbation onset across both (perturbation and control) conditions. Spatial filters were calculated, based on the forward solution, and a regularized inverse of the covariance matrix, averaged across conditions (the regularization parameter was set to 10% of the average sensor signal variance). Next, for each condition separately, these spatial filters were used to estimate the source activity for three time windows of interest: perturbation onset (100–250 m s), perturbation offset (550–700 m s) and intermediate (300–400 m s). These time windows were chosen based on the sensor-level analyses and are illustrated in Fig. 1. The perturbation onset and offset-related windows are 125 m s time windows that reflect the initial part of the main ERF differences between perturbation and control conditions. In addition, the cluster-based permutation test revealed an additional time window (300–400 m s after perturbation onset) where perturbation and control conditions differed.

For the time-frequency results, a frequency-domain beamformer (DICS)(Gross et al., 2001) was used. The data was de-meaned and the cross-spectral density between all pairs of MEG sensors was calculated over a 1.5s time window (−500 m s to 1,000 m s), across conditions, centered on 7 Hz for the θ band (bandwidth 4–10 Hz), and on 17 Hz for

the β band (bandwidth 14–20 Hz), using multitapers. The resulting cross-spectral densities were combined with the forward solution to calculate frequency band-specific spatial filters (regularization parameter was at 10%). Next, condition-specific cross-spectral densities were calculated over the time window 0–500 m s, and combined with the common spatial filters to obtain condition-specific source estimates. The time window of 0–500 m s was chosen to reflect the duration of the perturbation (Fig. 1), as well as based on the fact that the sensor-level time-frequency analyses showed increased theta and beta power in this window (in perturbation trials compared to control trials).

Results

### Behavioral responses

Overall, participants compensated for the pitch increase in the perturbation trials by lowering their pitch (Fig. 2). A cluster-based permutation test revealed that participants' pitch contour in the perturbation trials was different from the control trials (p = 0.002). This difference was mainly driven by a cluster lasting from 144 m s to 765 m s after perturbation onset. Results from the debriefing questionnaire revealed that none of the participants was aware of any pitch perturbations in the auditory feedback. A more in-depth analysis of the behavioral results is described elsewhere (Franken et al., 2018).
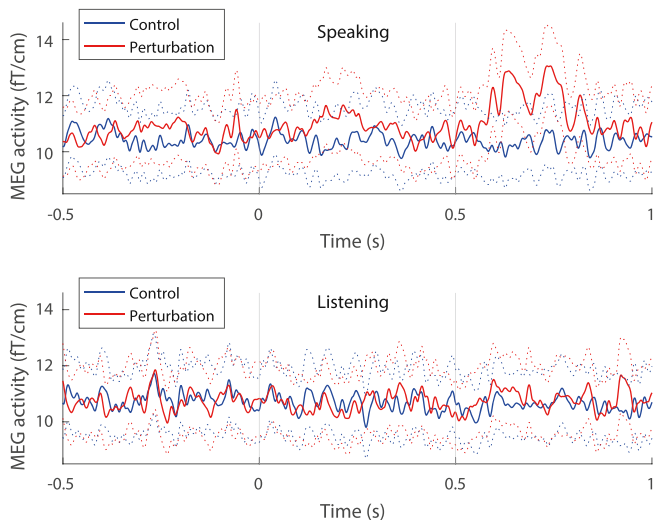
Event-related fields

The main analysis collapsed over both behavioral response types (i.e., following responses and opposing responses). Overall, the event-related fields show a response to both perturbation onset and offset in the speaking task, but not in the listening task (Fig. 3). A cluster-based permutation test on the speaking task data within 1s after perturbation onset revealed a significant difference between perturbation and control trials (p < 0.001). This difference was mainly driven by an increase in activity in the perturbation condition after perturbation onset (from about 85 m s after perturbation onset to about 250 m s) and after perturbation offset (550 m s-850 m s).

The topography plots for the speaking task (Fig. 4) show a mainly right-lateralized pattern in both the onset- and offset-related time
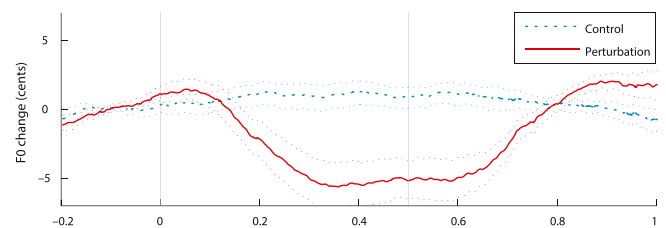
Fig. 2. Average pitch contour in perturbation and control trials. Dotted lines indicate standard error of the mean. For the perturbation trials, the perturbation started at 0s and lasted until 0.5s (marked by vertical grey lines).

Fig. 1. The black horizontal line shows the time course of a trial, where speech onset, perturbation onset (jittered between 500 and 1500 m s after speech onset) and perturbation offset are marked. The perturbation is marked in red. In purple, the three time windows of interest used for the source reconstruction of the ERF results are shown (100–250 m s, 300–400 m s and 550–700 m s after perturbation onset). In addition, the time points of the maximal ERF difference between perturbation and control after onset and offset are indicated. 100 m s time windows centered at these maximum points are used for the averages in Fig. 6

Fig. 3. Event-related field of perturbation (red) and control (blue) trials, averaged across all channels, for the speaking task (top graph) and the listening task (bottom graph). Dotted lines indicate standard error of the mean. For the perturbation trials, the perturbation started at 0s and lasted until 0.5s (marked by vertical grey lines).

windows. A smaller left-lateralized cluster of increased activity in perturbation vs. control trials was found in a later time window (300–400 m s). There was no clear similar cluster after offset that emerged from the cluster analysis, although note that the main cluster after perturbation offset lasted relatively long, until about 850 m s after perturbation onset, that is 350 m s after perturbation offset. Fig. 4 suggests that activity in this last part of the offset-related cluster (800–900 m s) was also left-lateralized. The topography plots for the listening task (Fig. 5) show that there is little difference, if anything,

between the perturbation and control conditions.

A comparison of MEG activity in speaking and listening tasks across conditions (Fig. 6) showed an interaction between task and condition in both the onset-related time window (F (1, 35) = 11.988, $p = 0.001$, $\eta_p^2 = 0.26$) and the offset-related time window (F (1, 35) = 17.464, $p < 0.001$, $\eta_p^2 = 0.33$). Post-hoc t-tests within the listening task showed that neither the difference between perturbation and control conditions in the onset-related time window (t (35) = 1.36, n. s., uncorrected), nor the difference in the offset-related time window (t (35) = −1.56, n. s., uncorrected) led to a significant overall change in MEG activity. This is in contrast to the comparisons between perturbation and control conditions within the speaking task, which showed a significant difference for both the onset (t (35) = −3.10, $p = 0.0038$, Cohen's d = −0.52) and offset time window (t (35) = −6.70, $p < 0.001$, Cohen's d = −1.12).

An LCMV beamformer was used to project the sensor-level activity of the speaking task in three windows of interest (onset: 100–250 m s; offset: 550–700 m s; and a third time window: 300–400 m s) onto the cortical surface. The results are depicted in Fig. 7. Both perturbation onset and perturbation offset-related activity increases were localized to superior temporal and inferior frontal areas, lateralized to the right hemisphere. Activity over the 300–400 m s time window showed increased activity in areas around the central sulcus in the left hemisphere.

Time-frequency responses

A time-frequency analysis of the data time-locked to perturbation onset shows event-related power changes across the low frequency range. A cluster-based permutation test revealed that there was a significant power increase in the perturbation condition, relative to the control condition ($p = 0.041$), which was mainly driven by increased power in the θ (4–8 Hz) and a lower β (12–16 Hz) band between 0 and 500 m s. The topographical distribution of these effects (Fig. 8) suggests involvement of sensorimotor areas.
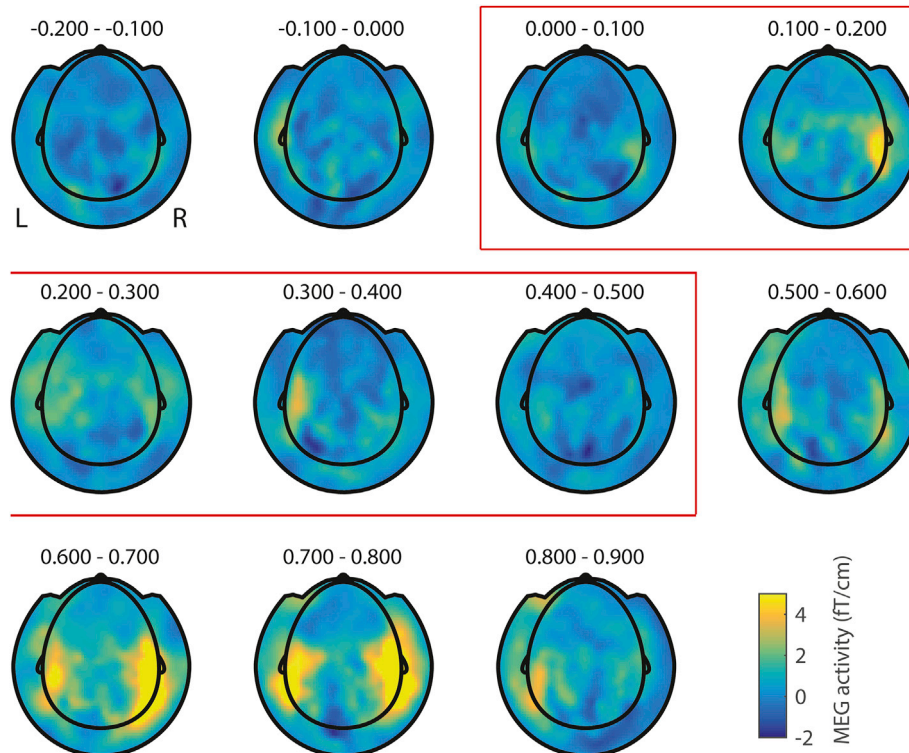
Fig. 4. Topography plots of the difference (perturbation – control) in the speaking task per 100 m s, from −200 m s to −100 m s (top left) to 800 m s-900 m s (bottom right). For the perturbation trials, the perturbation started at 0s and lasted until 0.5s (marked by the red box).
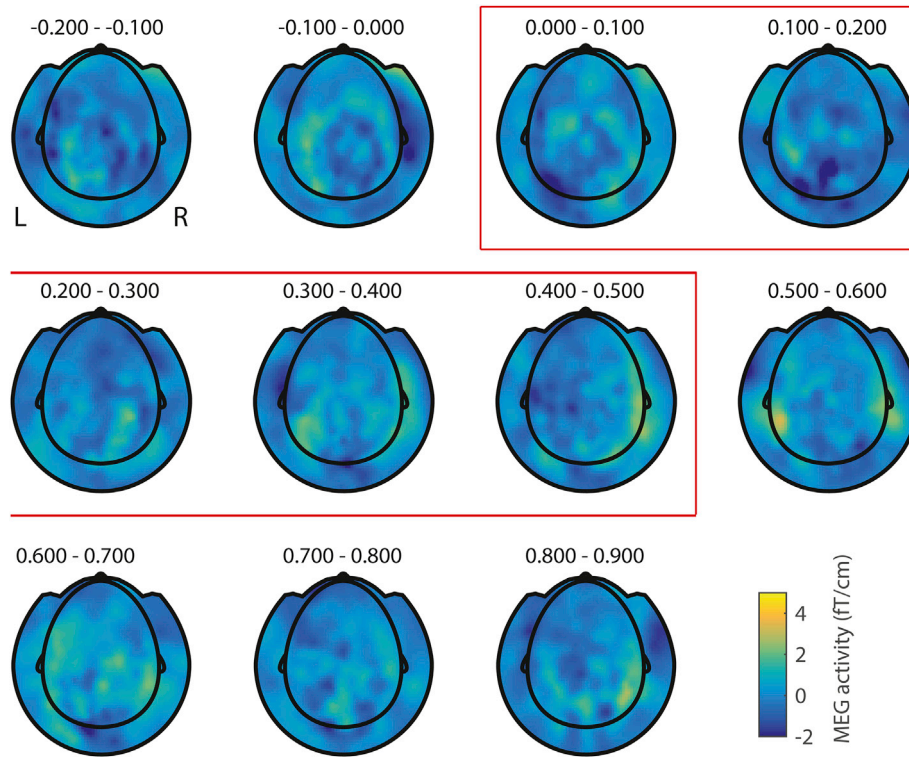
Fig. 5. Topography plots of the difference (perturbation – control) in the listening task per 100 m s, from −200 m s to −100 m s (top left) to 800 m s-900 m s (bottom right). For the perturbation trials, the perturbation started at 0s and lasted until 0.5s (marked by the red box).
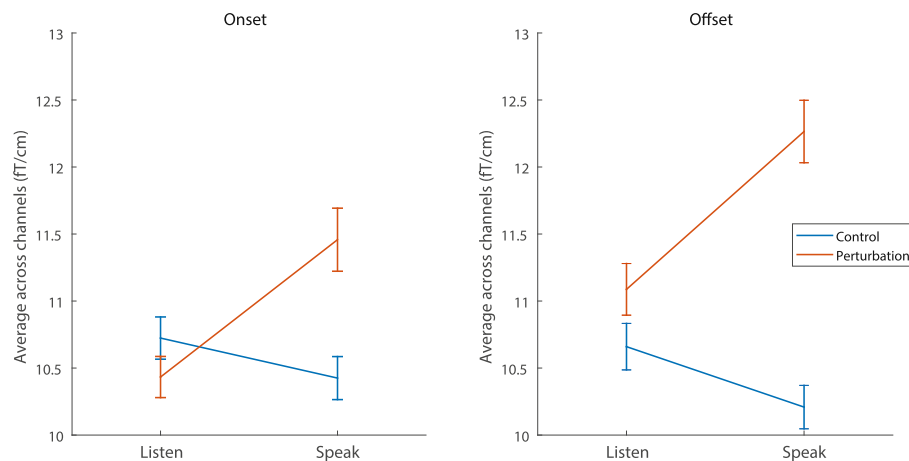

Fig. 6. Average activity across channels in a 100 m s time window after perturbation onset (left, centered at 222 m s) and after perturbation offset (right, centered at 627 m s). These time windows were centered at the point of maximal t statistic (averaged across channels) for the contrast perturbation-control collapsed across tasks at group level. Error bars indicate standard errors. There was a significant interaction between task and condition in both time windows.

For illustrative purposes, Fig. 9 shows the contrast (expressed as *t*-values) of these power changes between perturbation and control trials averaged across the 10 channels that show the strongest effect (marked in Fig. 8). The results of the DICS beamformer (Fig. 10) suggest that θ band activity was associated mostly with areas around inferior primary motor and somatosensory cortical areas (parts of Brodmann areas 1, 2, 3, 4 and 6), whereas the lower β band power increase was projected onto more superior motor areas (parts of Brodmann areas 4 and 6).

Neural correlates of behavioral response type

In a secondary analysis, trials were classified either as following (when the participant's behavioral response followed the direction of the feedback pitch perturbation) or as opposing (when the participant behaviorally opposed the pitch perturbation), as described in more detail elsewhere (Franken et al., 2018). Fig. 11 shows the event-related field responses corresponding to opposing and following trials. A cluster-based permutation test revealed the ERF for following and opposing trials significantly differed from each other ($p = 0.02$). The top left panel in Fig. 11 shows the averaged time course across the 10 channels that contributed most to the largest cluster. From the figure, it can be observed that this difference was mainly driven by a difference in the activity over central channels from 100 to 250 m s after perturbation onset. A second, smaller, cluster that showed up between roughly 550 m s-650 m s is hard to interpret in the current paradigm given its posterior location and given its timing (it is hard to interpret how neural
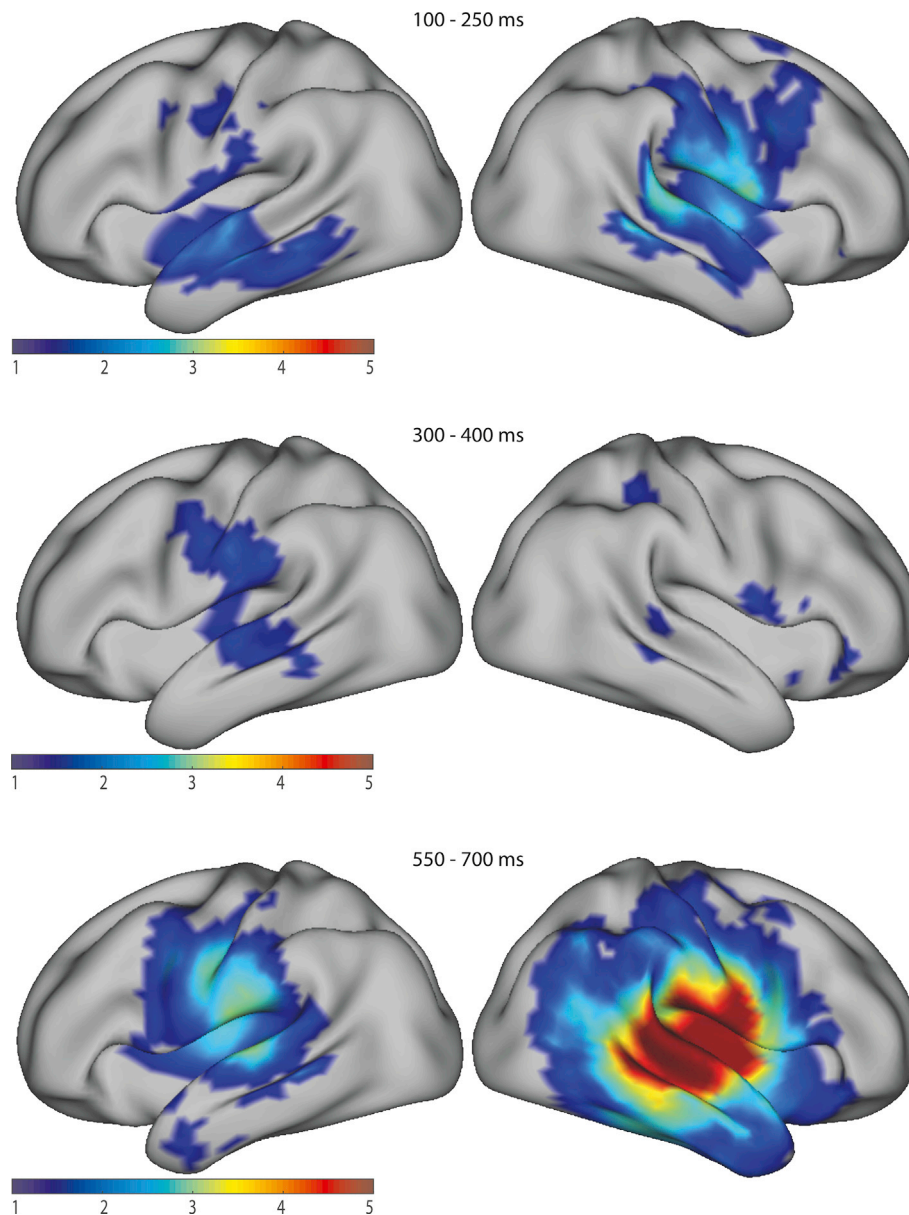
Fig. 7. Source-level event-related fields over the time windows 100–250 m s (top), 300–400 m s (middle) and 550–700 m s (bottom). Right hemispheres on the right and left hemispheres on the left. The colors indicate the contrast (perturbation - control)/baseline and are thresholded at a value of 1.
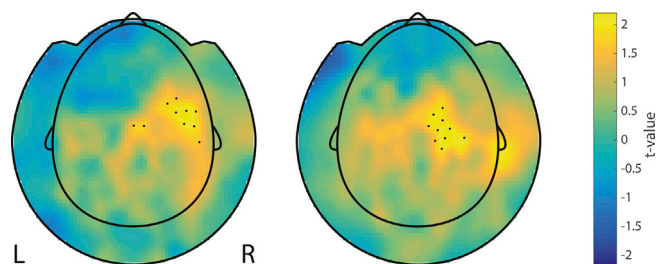
Fig. 8. Topography plots of θ band (4–10 Hz, left) and β band (14–20 Hz, right) power increase in perturbation trials compared to control trials. Colors indicate average t-values over 0–500 m s after perturbation onset. The channels selected for the average spectrograms in Fig. 9 are marked.

activity around 550 m s-650 m s is linked to the earlier behavioral response). The topography of our main cluster (100 m s-250 m s) suggests the activity difference originates from motor-related areas. The results of

a beamforming analysis (shown on the right side of Fig. 11) suggests the motor-related area involved may be the supplementary motor area (SMA). Other areas, mainly the bilateral ventromedial prefrontal cortical areas (vmPFC) and areas in the right middle temporal lobe, show up in the beamforming analysis, though these are not as clear from the results of the sensor-level topography plot.

Discussion

In the current study, we investigated the neural correlates of unexpected shifts in the pitch of auditory feedback during vocalization. While none of the participants were consciously aware of these pitch shifts, they responded by adjusting the pitch of the vocalization, which could have resulted only from unconscious perceptual and motor processing (Hafke, 2008). The neural signals showed a strong time-locked response in auditory cortical areas to both perturbation onset and offset. We also observed spectral power increases in both the θ and the lower β band during the perturbation. These power increases were localized to frontal motor-related cortical areas.
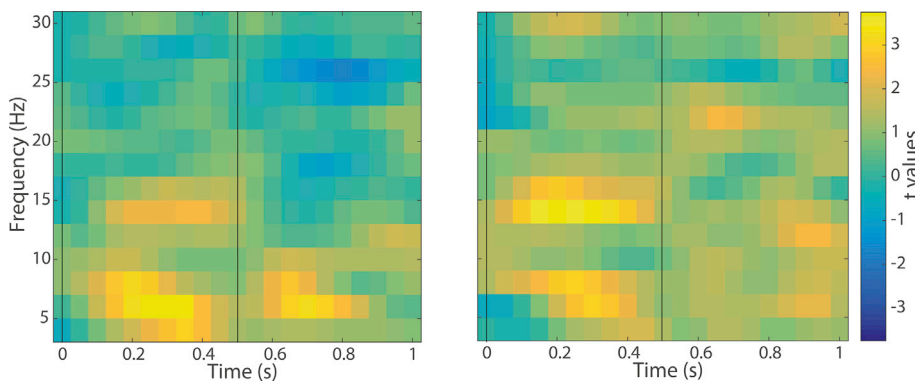
Fig. 9. Average t values indicating power changes in the perturbation condition, relative to the control condition, across the lower frequencies. Data was time-locked to perturbation onset and perturbation on- and offset are marked by vertical grey lines. The left graph shows the power changes averaged across 10 channels (see marked channels in Fig. 8) that were especially sensitive to the θ power difference (4–10 Hz, 0–0.5s), the right graph does the same for the β window (14–20 Hz, 0–0.5s). The color represents the contrast (perturbation – control)/baseline and is thresholded at a value of 0.04.
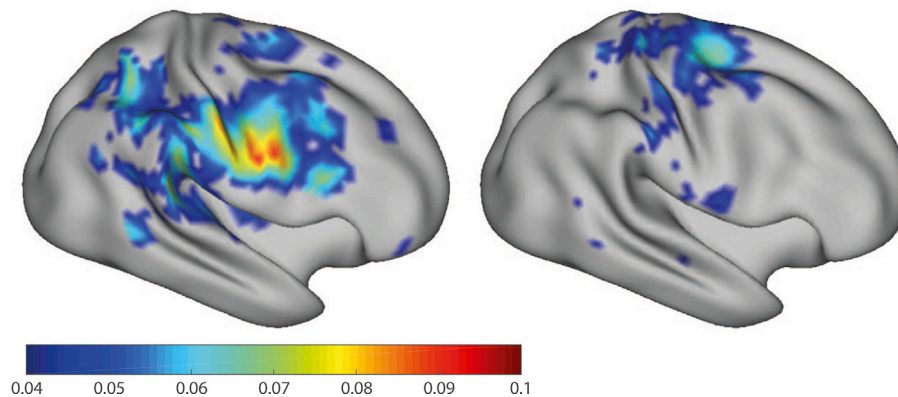
Fig. 10. Source-level projections of θ band (left, 4–10 Hz, 0–500 m s) and β band (right, 14–20 Hz, 0–500 m s) perturbation-related power increases on the right hemisphere.

Behaviorally, the participants responded to the perturbation by compensating for about 20% of the pitch shift on average. This finding is consistent with the vast literature on altered auditory feedback, and supports cognitive models hypothesizing that sensory feedback is continuously monitored to update and maintain adequate motor commands, both within and outside the domain of speech production (Houde and Nagarajan, 2011; Tourville and Guenther, 2011; Wolpert and Ghahramani, 2000).

At the neural level, event-related field analyses showed a response to both perturbation onset and offset, as well as a smaller left-lateralized response at a longer latency after perturbation onset. Source reconstructions suggest that the early latency onset- and offset-related responses originate from bilateral auditory cortical areas, which is consistent with earlier work (Behroozmand and Larson, 2011; Liu et al., 2011; Kort et al., 2014). The effect was stronger in the right hemisphere, which is in line with the well-established view that the right hemisphere is dominant in pitch processing (Johnsrude et al., 2000; Zatorre et al., 1992). However, Kort et al., (2014) did not find such a clear right-hemisphere advantage. Participants did not show similar responses to the perturbation in the listening task. This finding speaks against the interpretation that the response in auditory areas reflects the sensory event in isolation. Instead, it suggests that the onset- and offset-related responses reflect the mismatch between the forward model's prediction and the observed auditory feedback (Behroozmand et al., 2009; Chang et al., 2013; Curio et al., 2000; Franken et al., 2015; Houde et al., 2002). Although Fig. 5 (500–600 m s) suggests a small non-significant response to the perturbation offset in the listening task, the lack of a significant response to a sudden pitch shift is surprising. We suggest this can be explained by the magnitude of the pitch shift (25 cents), which is much smaller than in previous studies (e.g., 100 cents in Kort et al., 2014, 2016). In fact, the standard deviation of the pitch contour in a control

trial of the current study was only 10.8 cents, with a range of 49.8 cents. This suggests that the amplitude of natural fluctuations in pitch may have partly obscured the pitch shift in the perturbation trials.

The ERF results reported here for the speaking task show an effect at both perturbation onset and offset. Perhaps surprisingly, the effect appeared stronger at perturbation offset. Interpreting these responses in the light of an internal forward model, the ERF peaks may reflect the detection of the discrepancy between perturbed perceptual input, and the predicted perceptual signal, as generated by the forward model. After behavioral adaptation to the altered feedback, and an update of the forward model, the offset of the perturbation essentially is just a new (reversed) perturbation. In interpreting the difference between the effect's strength at perturbation onset and offset, it is important to point out the difference in predictability between the perturbation's onset and offset. While the presence and timing of the onset pitch shift is unpredictable, the offset pitch shift always follows 500 m s after an onset pitch shift. Although our participants were not aware of any pitch manipulations, the pitch matching task explicitly draws attention to pitch, and a first, unexpected, pitch shift may unconsciously trigger more attention to the pitch tracking task, and hence to the pitch of the auditory feedback. So even without conscious awareness, the perturbation offset may be more salient compared to perturbation onset. In addition, the offset pitch shift of course occurs *later*. As the pitch contour is more variable close to speech onset, the perturbation onset shift may be less salient to the (unconscious) speech processing machinery, compared to the perturbation offset.

In addition to the involvement of auditory cortical areas, our ERF results also suggest involvement of sensorimotor areas. The smaller ERF peak at 300–400 m s after perturbation onset was localized to left (pre) motor and sensorimotor areas, which suggests that this response may reflect the compensatory motor response. There was no distinct similar
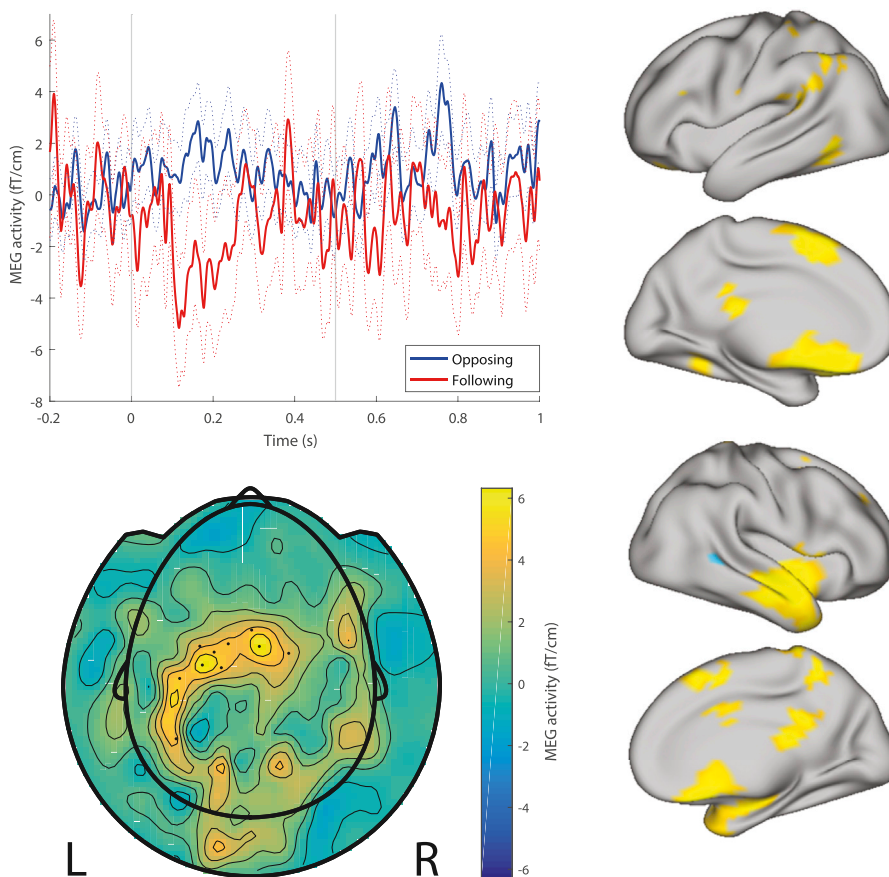
Fig. 11. Top left: event-related field responses to opposing (blue) and following (red) trials, averaged across the channels highlighted in the topography plot (10 channels that contribute most to the cluster). Dotted lines indicate standard error of the mean. Perturbation onset is at 0 m s (onset and offset marked by vertical grey lines). Bottom left: topography plot of the condition difference opposing – following, average over the time window 100–250 m s. Highlighted channels are the channels used for the top left plot. Right: source plots (top to bottom: left lateral view, left medial, right lateral, right medial) of the results of a LCMV beamformer. Colors indicate the difference (opposing – following)/(common_baseline) and are thresholded at values −4 and 4.

left-lateralized cluster after perturbation offset, but the large early latency offset-related cluster may have included this response, especially given the left-lateralized topography in the 800–900 m s time window in Fig. 4. Although most research has focused on a right-lateralized network of brain regions involved in pitch-shifted feedback processing, some fMRI studies have reported activity in similar left-hemisphere areas (Toyomura et al., 2007), which possibly relates to the behavioral response (Behroozmand et al., 2015b). In addition, neuro-anatomically constrained speech production models like DIVA [Directions Into Velocities of Articulators, Tourville and Guenther, 2011] and the state feedback control model (Houde and Nagarajan, 2011) also include these areas, and posit they support articulatory motor programs or auditory-motor interactions.

Overall, the event-related field analysis shows right-lateralized responses in auditory cortical areas to both pitch perturbation onset and offset, as well as a left-lateralized response in motor-related areas around 300 m s after perturbation onset. A similar response could be seen around 300 m s after perturbation offset (800–900 m s in Fig. 4). These results suggest an interconnected sensory-motor network that supports auditory-motor integration, including auditory and motor-related areas in both hemispheres.

In addition to the event-related neural effects, we also studied the time-frequency response to pitch perturbations. We found evidence of increased θ and β band power during and after feedback pitch perturbation. An increase in θ band power has previously been suggested to be a compelling candidate mechanism to provide a temporal window for auditory-motor interactions and ongoing feedback monitoring (Behroozmand et al., 2015a). In the current study, a θ band increase was found with small, not consciously perceived, pitch perturbations, and was reconstructed to involve sources in inferior motor areas and the

posterior temporal areas. This is in line with the hypothesis that auditory and motor-related areas are jointly involved in feedback-based vocal pitch adjustments. These results agree with earlier findings of the involvement of right parietal and temporal areas in integrating sensory information and continuous sensorimotor monitoring after pitch-shifted feedback (Kort et al., 2016). However, looking at high gamma power, Kort et al., (2016) also found early left-lateralized response to pitch shifts, which is in contrast with the low-frequency results in the current study.

Furthermore, a power increase in the lower β band during the perturbation was found. To the best of our knowledge, a β band increase has not been described previously in an altered auditory feedback paradigm. One reason may be that with much shorter pitch perturbations (100–200 m s), many previous studies are less equipped to demonstrate reliable β modulation during the perturbation. A β power increase during pitch-shifted auditory feedback is puzzling, given that β power usually is found to decrease during motor planning and execution (Doyle et al., 2005; Pfurtscheller and Lopes da Silva, 1999). Although it is unclear how the current findings fit with the established role of β desynchronization during movement or β rebound after movement, a crucial difference with classic motor-related β findings is the fact that participants vocalized from well before up to well after the perturbation. In addition, past research suggests that increased β power is associated with lower sensory gating (Cheng et al., 2017). This could indicate that increased β power is related to an increased need for perceptual processing, in line with the peripheral sensory sampling hypothesis (Khanna and Carmena, 2015) stating that β activity may reflect active sensory sampling. In other words, detection of unexpected auditory feedback may lead to more active sensory sampling reflected by increased β power. With respect to auditory perception specifically, previous studies on motor control and auditory processing have suggested the involvement of β power in

auditory-motor interactions (Fujioka et al., 2009; Iversen et al., 2009). Based on these studies, we suggest two possible accounts for the role of β power in the current study. First, in motor control, increased β power has been associated with error monitoring (Koelewijn et al., 2008). Here, we may speculate that increased β power could reflect the detection of erroneous pitch production. The source reconstruction also supports this interpretation, with the dominant source located in superior (pre)motor cortical areas. Second, a recent EEG study suggests that β oscillations play a role in auditory prediction (Chang et al., 2016). Therefore, β oscillations could be related to the prediction of the sensory consequences of a vocalization action (i.e., the forward model), and thus a β increase may reflect the detection of a prediction error. Note, however, that the source of the β increase in the present study was in motor areas, while most studies of auditory β have shown the β increase in auditory cortex or surrounding areas.

Two possible accounts of the role of β power in the current paradigm emerge: motor error detection or auditory prediction errors. Future studies are necessary to disentangle these hypotheses. If increased β power is related to auditory prediction, it should be modulated by predictability of the pitch perturbation, while it should remain stable if it reflects only action error monitoring.

Finally, we performed a secondary analysis, and investigated the neural correlates of the type of behavioral response (opposing or following) to the pitch shift. The results showed an increased ERF response for the opposing responses (or a decrease for the following responses), shortly after perturbation onset. The locus of this effect included the supplementary motor area (SMA), amongst some other areas, such as the right middle temporal lobe. Given that MEG is less sensitive to activity originating from deeper brain areas, especially the activity in deeper areas such as the vmPFC should be interpreted with caution. The activity in the right temporal lobe may be related to sensorimotor processing, although it is located more anterior compared to our main findings.

According to the DIVA model, SMA is involved in an initiation circuit, which ensures that articulations start at the right time and are timed correctly. In the current study, increased SMA activity during opposing responses may possibly signal the initiation of an opposing behavioral response. However, also in the following trials there was a behavioral response, though simply in the other direction. It is unclear why SMA activity distinguished between the two trial types. It is possible that following responses do not require initiation of a new, compensatory action, but instead reflect simple ongoing convergence to an external auditory stimulus, while opposing responses are generated by the initiation of a new articulatory action. An alternative explanation may come from research in motor cognition, on the sense of agency. Previous studies have implicated the SMA, among other brain regions, in agency processing (David et al., 2008; David, 2012). A difference in SMA activity between opposing and following responses may suggest that in following trials, participants may sometimes consider the perturbation to be externally generated, while it is considered self-generated in opposing trials. This is well in line with an earlier explanation of following trials provided by Hain et al., (2000). Further research is needed to clarify the role of the increased SMA activity with respect to following vs. opposing behavioral responses.

In conclusion, the current study explored the neural underpinnings of auditory feedback processing during speech production. We found that even without conscious awareness, speakers compensate for unexpected pitch shifts in auditory feedback. At the neural level, a strong short-latency response was found in auditory cortical areas during vocalization, reflecting a mismatch between the forward model's prediction and the observed sensory feedback. At a longer latency, neural activity associated with preparation or implementation of motor compensation was observed in left pre-motor areas. In addition, a spectral power increase in both θ and β bands occurred as a response to the pitch perturbation. The θ power effect concurs with the literature and suggests the involvement of mechanisms which incorporate auditory feedback in

voice control. We extend this literature by showing that the increased θ power is indeed related to automatic, unconscious, pitch processing, and by localizing it to motor-related cortical areas. To the best of our knowledge, this study is the first that shows an increase in β power in an altered auditory feedback paradigm. Increased β power may reflect motor error-monitoring or auditory prediction mechanisms. Overall, the results reported here are in line with current models of speech production, which posit a need for constant sensorimotor interactions, not unlike other complex motor skills. Even small unexpected errors are quickly detected by the perceptual system and may lead to subsequent behavioral changes through automatic sensorimotor interactions.

References

Bastiaansen, M.C.M., Knosche, T.R., 2000. Tangential derivative mapping of axial MEG applied to event-related desynchronization research. Clin. Neurophysiol. 111, 1300–1305.

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D.A., Larson, C.R., 2015a. Functional role of delta and theta band oscillations for auditory feedback processing during vocal pitch motor control. Front. Neurosci. 9, 109.

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

Behroozmand, R., Korzyukov, O., Sattler, L., Larson, C.R., 2012. Opposing and following vocal responses to pitch-shifted auditory feedback: evidence for different mechanisms of voice pitch control. J. Acoust. Soc. Am. 132, 2468–2477.

Behroozmand, R., Larson, C., 2011. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci. 12, 54.

Behroozmand, R., Liu, H., Larson, C.R., 2011. Time-dependent neural processing of auditory feedback during voice pitch error detection. J. Cognit. Neurosci. 23, 1205–1217.

Behroozmand, R., Shebek, R., Hansen, D.R., Oya, H., Robin, D.A., Howard, M.A., Greenlee, J.D.W., 2015b. Sensory-motor networks involved in speech production and motor control: an fMRI study. Neuroimage 109, 418–428.

Boersma, P., Weenink, D., 2013. Praat: Doing Phonetics by Computer [Computer Program]. http://www.praat.org.

Burnett, T., Freeland, M., Larson, C., Hain, T., 1998. Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. 103, 3153–3161. https://doi.org/10.1121/1.423073.

Cai, S., Boucek, M., Ghosh, S.S., Guenther, F.H., Perkell, J.S., 2008. A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. In: Proceedings of the 8th Intl. Seminar on Speech Production. Strasbourg, France, pp. 65–68.

Caplan, J.B., Madsen, J.R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E.L., Kahana, M.J., 2003. Human theta oscillations related to sensorimotor integration and spatial learning. J. Neurosci. 23, 4726–4736.

Chang, A., Bosnyak, D.J., Trainor, L.J., 2016. Unpredicted pitch modulates beta oscillatory power during rhythmic entrainment to a tone sequence. Front. Psychol. 7, 1–13.

Chang, E.F., Niziolek, C.A., Knight, R.T., Nagarajan, S.S., Houde, J.F., 2013. Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc. Natl. Acad. Sci. U. S. A. 110, 2653–2658.

Cheng, C.-H., Tsai, S.-Y., Liu, C.-Y., Niddam, D.M., 2017. Automatic inhibitory function in the human somatosensory and motor cortices: an MEG-MRS study. Sci. Rep. 7 (4234). http://www.nature.com/articles/s41598-017-04564-1.

Cruikshank, L.C., Singhal, A., Hueppelsheuser, M., Caplan, J.B., 2012. Theta oscillations reflect a putative neural mechanism for human sensorimotor integration. J. Neurophysiol. 107, 65–77.

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

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis I. Segmentation and surface reconstruction. Neuroimage 9, 179–194.

David, N., 2012. New frontiers in the neuroscience of the sense of agency. Front. Hum. Neurosci. 6, 161.

David, N., Newen, A., Vogeley, K., 2008. The "sense of agency" and its underlying cognitive and neural mechanisms. Conscious. Cognit. 17, 523–534.

Doyle, L.M.F., Yarrow, K., Brown, P., 2005. Lateralization of event-related beta desynchronization in the EEG during pre-cued reaction time tasks. Clin. Neurophysiol. 116, 1879–1888. https://www.sciencedirect.com/science/article/pii/S1388245705001458.

Engel, A.K., Fries, P., Singer, W., 2001. Dynamic predictions: oscillations and synchrony in top–down processing. Nat. Rev. Neurosci. 2, 704–716.

Franken, M., Acheson, D., McQueen, J., Hagoort, P., Eisner, F., 2018. Opposing and following responses in sensorimotor speech control: why responses go both ways. Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-018-1494-x. Advance online publication.

Franken, M.K., Hagoort, P., Acheson, D.J., 2015. Modulations of the auditory M100 in an imitation task. Brain Lang. 142, 18–23.

Fujioka, T., Trainor, L.J., Large, E.W., Ross, B., 2009. Beta and gamma rhythms in human auditory cortex during musical beat processing. Ann. N. Y. Acad. Sci. 1169, 89–92.

Gehrig, J., Wibral, M., Arnold, C., Kell, C.A., 2012. Setting up the speech production network: how oscillations contribute to lateralized information routing. Front. Psychol. 3, 169.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. U. S. A. 98, 694–699.

Hafke, H.Z., 2008. Nonconscious control of fundamental voice frequency. J. Acoust. Soc. Am. 123, 273–278.

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

Hawco, C.S., Jones, J.A., Ferretti, T.R., Keough, D., 2009. ERP correlates of online monitoring of auditory feedback during vocalization. Psychophysiology 46, 1216–1225.

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

Houde, J.F., Jordan, M.I., 1998. Sensorimotor adaptation in speech production. Science 279 (80), 1213–1216.

Houde, J.F., Nagarajan, S.S., 2011. Speech production as state feedback control. Front Hum Neurosci 5.

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.

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Network. 10, 626–634.

Iversen, J.R., Repp, B.H., Patel, A.D., 2009. Top-down control of rhythm perception modulates early auditory responses. Ann. N. Y. Acad. Sci. 1169, 58–73.

Jenson, D., Bowers, A.L., Harkrider, A.W., Thornton, D., Cuellar, M., Saltuklaroglu, T., 2014. Temporal dynamics of sensorimotor integration in speech perception and production: independent component analysis of EEG data. Front. Psychol. 5, 656.

Johnsrude, I.S., Penhune, V.B., Zatorre, R.J., 2000. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123, 155–163.

Jones, J.A., Munhall, K.G., 2000. Perceptual calibration of F0 production: evidence from feedback perturbation. J. Acoust. Soc. Am. 108, 1246.

Keough, D., Hawco, C., Jones, J a, 2013. Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. BMC Neurosci. 14, 25.

Khanna, P., Carmena, J.M., 2015. Neural oscillations: beta band activity across motor networks. Curr. Opin. Neurobiol. 32, 60–67. https://www.sciencedirect.com/science/article/pii/S0959438814002360.

Koelewijn, T., van Schie, H.T., Bekkering, H., Oostenveld, R., Jensen, O., 2008. Motor-cortical beta oscillations are modulated by correctness of observed action. Neuroimage 40, 767–775.

Kort, N.S., Cuesta, P., Houde, J.F., Nagarajan, S.S., 2016. Bihemispheric network dynamics coordinating vocal feedback control. Hum. Brain Mapp.

Kort, N.S., Nagarajan, S.S., Houde, J.F., 2014. A bilateral cortical network responds to pitch perturbations in speech feedback. Neuroimage 86, 525–535.

Korzyukov, O., Sattler, L., Behroozmand, R., Larson, C.R., 2012. Neuronal mechanisms of voice control are affected by implicit expectancy of externally triggered perturbations in auditory feedback. PLoS One 7.

Larson, C.R., Sun, J., Hain, T.C., 2007. Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F[sub 0] and amplitude control. J. Acoust. Soc. Am. 121, 2862.

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

Liu, H., Larson, C., 2007. Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. J. Acoust. Soc. Am. 122, 3671–3677.

Liu, Y., Hu, H., Jones, J.A., 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.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Meth. 164, 177–190.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci.

Pfurtscheller, G., Lopes da Silva, F.H., 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. 110, 1842–1857. https://www.sciencedirect.com/science/article/pii/S1388245799001418.

Stolk, A., Todorovic, A., Schoffelen, J.M., Oostenveld, R., 2013. Online and offline tools for head movement compensation in MEG. Neuroimage 68, 39–48.

Tourville, J.A., Cai, S., Guenther, F., 2013. Exploring Auditory-motor Interactions in normal and Disordered Speech, p. 060180.

Tourville, J.A., Guenther, F.H., 2011. The DIVA model: a neural theory of speech acquisition and production. Lang. Cognit. Process. 26, 952–981.

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

Van Veen, B.D., Van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. 44, 867–880.

Wolpert, D., Ghahramani, Z., 2000. Computational principles of movement neuroscience. Nat. Neurosci. 3, 1212–1217.

Wolpert, D., Ghahramani, Z., Jordan, M.I., 1995. An internal model for sensorimotor integration. Science (80-) 269, 1880–1882.

Zarate, J.M., Wood, S., Zatorre, R.J., 2010. Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. Neuropsychologia 48, 607–618.

Zarate, J.M., Zatorre, R.J., 2005. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. Ann. N. Y. Acad. Sci. 1060, 404–408.

Zatorre, R., Evans, A., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science (80-) 256, 846–849.

Zheng, Z.Z., Vicente-Grabovetsky, A., MacDonald, E.N., Munhall, K.G., Cusack, R., Johnsrude, I.S., 2013. Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. J. Neurosci. 33, 4339–4348.

Zheng, Z.Z., Munhall, K.G., Johnsrude, I.S., 2010. Functional overlap between regions involved in speech perception and in monitoring One's own voice during speech production. J. Cognit. Neurosci. 22, 1770–1781.