Perceiving your hand moving: BOLD suppression in sensory cortices and the role of the cerebellum in the detection of feedback delays

B. Ezgi Arikan
Department of Psychology, Justus-Liebig University Giessen, Giessen, Germany

Bianca M. van Kemenade
Department of Psychiatry and Psychotherapy, Philipps University Marburg, Marburg, Germany

Kornelius Podranski
Department of Psychiatry and Psychotherapy, Philipps University Marburg, Marburg, Germany
Core Facility Brain Imaging, Faculty of Medicine, Philipps University Marburg, Marburg, Germany
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Olaf Steinsträter
Department of Psychiatry and Psychotherapy, Philipps University Marburg, Marburg, Germany
Core Facility Brain Imaging, Faculty of Medicine, Philipps University Marburg, Marburg, Germany

Benjamin Straube
Department of Psychiatry and Psychotherapy, Philipps University Marburg, Marburg, Germany

Tilo Kircher
Department of Psychiatry and Psychotherapy, Philipps University Marburg, Marburg, Germany

Sensory consequences of self-generated as opposed to externally generated movements are perceived as less intense and lead to less neural activity in corresponding sensory cortices, presumably due to predictive mechanisms. Self-generated sensory inputs have been mostly studied in a single modality, using abstract feedback, with control conditions not differentiating efferent from reafferent feedback. Here we investigated the neural processing of (a) naturalistic action–feedback associations of (b) self-generated versus externally generated movements, and (c) how an additional (auditory) modality influences neural processing and detection of delays. Participants executed wrist movements using a passive movement device (PMD) as they watched their movements in real time or with variable delays (0–417 ms). The task was to judge whether there was a delay between the movement and its visual feedback. In the externally generated condition, movements were induced by the PMD to disentangle efferent from reafferent feedback. Half of the trials involved auditory beeps coupled to the onset of the visual feedback. We found reduced BOLD activity in visual, auditory, and somatosensory areas during self-generated compared with externally generated movements in unimodal and bimodal conditions. Anterior and posterior cerebellar areas were engaged for trials in which action–feedback delays were detected for self-generated movements. Specifically, the left cerebellar lobule IX was functionally connected with the right superior occipital gyrus. The results indicate efference copy-based predictive mechanisms specific to self-generated movements, leading to BOLD suppression in sensory areas. In addition, our results support the cerebellum’s role in the detection of temporal prediction errors during our actions and their consequences.

Introduction

Our actions shape the way we perceive the world around us (Blakemore, Wolpert, & Frith, 1998; Desantis & Haggard, 2016b; Haggard, Clark, & Kalogeras, 2002; Wolpert, Miall, & Kawato, 1998). For example, it has been well established that sensory stimuli associated with self-generated movements are experienced as less intense and produce less neural activity in corresponding sensory cortices than if they were externally delivered (Blakemore, Rees, & Frith, 1998; Blakemore, Wolpert, & Frith, 1999; Roussel, Hughes, & Waszak, 2014; Shergill et al., 2013). This is known as sensory suppression. Recent theories of motor control explain suppression within the forward model framework, in which an internal model anticipates the sensory consequences of self-generated movements via efferent copies and monitors whether predicted consequences match actual sensory input (Blakemore, Frith, et al., 1999; Flanagan, Vetter, Johansson, & Wolpert, 2003; Wolpert, Ghahramani, & Flanagan, 2001). Efference copies then modulate neural processing in the corresponding sensory cortex, resulting in reduced sensation for these consequences (Pynn & DeSouza, 2013). Such predictive processing is crucial to discriminate events of external origin from those initiated by our own movement, allowing us to attribute agency over sensory inputs associated with our own actions (Blakemore, Wolpert, & Frith, 2002; Brown, Adams, Parees, Edwards, & Friston, 2013; Hoover & Harris, 2012; Shergill, Samson, Bays, Frith, & Wolpert, 2005).

These studies, along with others investigating sensorimotor prediction of voluntary actions and their consequences, have mostly focused on one sensory modality at a time. In general, sensory suppression has been observed for somatosensory, auditory, and visual domains, both on a behavioral (Bays, Flanagan, & Wolpert, 2006; Cardoso-Leite, Mamassian, Schutz-Bosbach, & Waszak, 2010; Roussel et al., 2014; Sato, 2008) as well as at a neural level (Blakemore, Frith, & Wolpert, 2001; Blakemore, Wolpert, et al., 1998; Desantis, Roussel, & Waszak, 2014; Lange, 2011; Leube et al., 2003; Shergill et al., 2013; Waszak, Cardoso-Leite, & Hughes, 2012). However, our actions in the real world often lead to multiple sensory consequences; i.e., when we knock on a door, we would get visual, auditory, and somatosensory information simultaneously. Moreover, each of these modalities might be affected by the presence of the other. Research on cross-modal processing points to the facilitatory effect of a stimulus modality on another modality when they are spatially and/or temporally in synchrony (Arabzadeh, Clifford, & Harris, 2008; Bresciani, Dammeyer, & Ernst, 2008; Diederich & Colonius, 2004; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005). On the other hand, there is evidence that a specific modality can have an inhibitory effect on another modality (Colavita, 1974; Kawashima, O’Sullivan, & Roland, 1995; Meredith, 2002; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Sinnett, Soto-Faraco, & Spence, 2008; Wang et al., 2012). These context-dependent cross-modal interactions might begin at relatively early stages of information processing in regions known as modality-specific (Calvert, Campbell, & Brammer, 2000; Laurienti et al., 2002; Macaluso, 2006; Macaluso, Frith, & Driver, 2000; Shimojo & Shams, 2001). In the case of action-related effects, studies focusing on the role of cross-modal processing of self-generated movement consequences suggest facilitation of behavioral performance with the inclusion of an additional modality (Desantis & Haggard, 2016a; Desantis, Mamassian, Lisi, & Waszak, 2014; Farrer, Valentín, & Hup, 2013; Kawabe, Roseboom, & Nishida, 2013; van Kemenade, Arikan, Kircher, & Straube, 2016). With respect to neural correlates, our group has previously shown that perceiving single auditory or visual (unimodal) and audiovisual (bimodal) consequences of voluntary button presses lead to reduced BOLD activity in somatosensory, visual, and auditory cortices for self-generated movements and their consequences (Straube et al., 2017).

Despite these findings, a number of issues remain to be explored that are crucial in advancing our understanding of why the processing of voluntary actions and their sensory consequences are distinct. First, most studies investigating cross-modal perception and action so far used passive viewing of sensory stimuli as a control condition (Desantis & Haggard, 2016a; Desantis, Waszak, Moutspoulou, & Haggard, 2016; Kawabe et al., 2013; Straube et al., 2017; van Kemenade, Arikan, Kircher, & Straube, 2017). Although this condition might address the influence of self-generated movements on the perception of sensory inputs by factoring out mere sensory processing, the role of efferent feedback cannot be addressed. Externally generated movements, on the other hand, might prove more efficient in examining the role of efference copies specific to self-generated movements. Self-generated sensory stimuli produce both efferent and reafferent information, whereas sensory consequences of externally generated movements provide only reafferent information (Weiskrantz, Elliott, & Darlington, 1971). By contrasting both movements, the influence of efferent information can be identified. Second, actions and their consequences in the real world are more complex than button presses and abstract dots.
as action outcomes. Although button presses as well as computer-generated sensory stimuli are easier to control and manipulate, more studies are needed to demonstrate that similar mechanisms apply to real life (Ingram & Wolpert, 2011). Importantly, by examining naturalistic action–feedback associations, one can also take into account ownership of the body, which contributes to action control (Gentile et al., 2011; Hoover & Harris, 2012; Kilteni & Ehrsson, 2017; Leube et al., 2003; Limanowski, Kirilina, & Blankenburg, 2017).

Our main aim in the present study was twofold: to investigate (a) whether naturalistic action–feedback associations produce suppression in corresponding sensory cortices as opposed to movements without efferent feedback and (b) whether cross-modal processing influences detection of temporal prediction errors for and concurrent neural processing of sensory stimuli associated with self-generated movements. More specifically, we aimed to examine BOLD suppression regarding naturalistic visual feedback of self-generated wrist movements and the impact of an additional (auditory) modality on the detection of temporal discrepancies between self-generated movements and their visual feedback. To this end, participants were asked to perform wrist movements using the handle of a custom-made passive movement device (PMD) (see van Kemenade et al., 2019). Half of the trials involved an externally generated movement condition in which the PMD moved the hand automatically. This condition allowed us to address the role of efferent feedback specific to self-generated movements as opposed to reafferent feedback arising as a result of both movements. The movements were recorded with a camera and displayed to the participants in real time, which allowed us to better represent a naturalistic feedback of the action. Temporal discrepancies were induced by presenting systematic delays between the movement and the visual display of the movement. The task of the participants was to report whether they detected a delay between the actual movement and the visual feedback of the movement (unimodal condition). This task would address detection of temporal prediction errors associated with one’s own action. Half of the trials involved an auditory beep coupled to the onset of the visual feedback (bimodal condition). The task remained the same in these trials. Based on our previous finding of BOLD suppression (Straube et al., 2017), we expected to observe reduced activity in sensory areas for self-generated compared with externally generated movements. Because unimodal and bimodal conditions would be equally likely, we expected to find reduced BOLD activity in auditory regions in both unimodal and bimodal conditions when the movement was self-generated due to sensorimotor prediction. We also hypothesized that the auditory stimulus triggered by self-generated movements would provide additional information about temporal discrepancies between the movement and its visual feedback, aiding in subjective detection of delays. Finally, we explored possible differences in BOLD activity associated with detecting versus not detecting action–feedback delays across movements and modalities in order to assess whether detection of a temporal discrepancy between the movement and its feedback would lead to distinct brain activation when the movement was self-generated.

**Methods**

Data from all participants was taken from a recent study by our group in which we examined contributions of different regions in action–feedback monitoring of naturalistic self- versus externally generated movements (van Kemenade et al., 2019). The experimental design and data acquisition procedures have been reported previously in van Kemenade et al. (2019) and restated below for the reader’s convenience.

**Participants**

Twenty-three right-handed students (11 females, mean age = 25.83 ± SD 3.09) from Philippus University Marburg participated in the behavioral training and fMRI scanning. One participant who took part in the behavioral training was not included in the fMRI study due to an increased bias to report delays even for the 0-ms delay (see Experimental design and procedure for details). Five participants were excluded from the final analyses due to excessive head movement (N = 1) or technical issues (N = 2), and insufficient number of subjective delay detection trials (N = 2), resulting in a final sample of 18 participants (eight females, age = 26.22 ± SD 3.34). Written informed consent was obtained from all participants. Participants reported normal or corrected-to-normal vision and normal hearing as well as no history of psychiatric or neurological disorders and no current use of psychoactive medications. Right-handedness was confirmed by the Edinburgh Handedness Inventory (Oldfield, 1971). The experiment was approved by the local ethics committee and performed in accordance with the Declaration of Helsinki. The participants were paid for their participation.

**Apparatus**

A custom-made magnetic resonance (MR)-compatible PMD was used for the execution of both self-
Externally generated movements. The device was built using nonferromagnetic materials (PVC, titan). A handle was designed for the execution of the movement with a horizontal trajectory. Importantly, the movement range of the handle and, therefore, the to-be-executed movement was restricted so that movement range would remain constant throughout trials and for different movements. Approximate movement range was \( \pm 30 \) and \( \pm 5.5 \) cm (see Figure 1a). Motion to the device could be induced automatically with compressed air (6 bar). Approximate force used when the device was working automatically was 20 N.

An MR-compatible camera (MRC High Speed, MRC Systems GmbH, Heidelberg, Germany) with a 2-ms refresh rate was used to record movements of the hand. The camera was mounted to an adjustable custom-made support. Auditory beeps were presented via MR-compatible headphones (MR-Confon, Opti- mel, Magdeburg, Germany). Responses were given using MR-compatible button pads (Cedrus, Lumina, San Pedro, CA). Visual display of the movement recorded by the camera in real time was presented on a mirror-projected computer screen (60 Hz refresh rate). Five LEDs were attached to the PMD, not visible to the participant, for the purpose of detecting movement onsets. A motion-detection algorithm written specifically for this study was applied to the camera images, detecting the spatial position of one moving LED with regard to the four fixed LEDs in each camera frame. For each trial, movement onset was defined as a minimum of 1° change in the relative position of the moving LED within one frame (requiring an increase of at least 0.5° per frame). Delays were presented from movement onset onward. All equipment was controlled by custom written software running on a personal computer (Dell Optiplex 9020, Intel Core i5-4570, four cores, 3.2 GHz, 4 GB RAM).

**Experimental design and procedure**

An event-related design was used with a delay-detection paradigm as shown in Figure 2. On each trial, the participant was asked to perform a wrist movement (extension and flexion) with the right hand using the handle of the PMD. The movements were recorded by the camera and presented to the participant in real time via mirror-projection, constituting the visual feedback of the movement (see Figure 1b & c). Six delays (0, 83, 167, 250, 330, and 417 ms, determined based on the screen refresh rate) were introduced between the movement and the real-time display of the movement. The participants were asked to judge whether there was a delay between actual movement of their hand and visual feedback of the movement. Half of the trials contained 440-Hz sine wave pure tones (500 ms), which were presented with the same delay used to delay the visual feedback. The task remained the same in these trials.

One week prior to scanning, participants completed a behavioral training in order to familiarize themselves with the task. First, they were trained to perform hand movements both in the self- and externally generated movement conditions. Accordingly, they were instructed to grab the handle of the PMD and move the handle from left to right and then back in approximately 1,500 ms while trying to maintain a constant pace with the help of a metronome. They then practiced externally generated movements by grabbing the handle of the PMD while trying to relax the hand and wrist as much as possible and letting the device move their wrist. In order to prevent participants from using visual cues from the hand, a curtain was placed to hide the right hand from participants’ sight. Participants executed both movements first without and then with the visual display recorded by the camera and presented in real time via a computer screen. The participants were instructed to observe the displayed movement on the screen and to judge whether there was a delay between their actual movement and the visual feedback of the movement or not. They were also told that, in some trials, they would additionally receive...
Participants first completed trials with no delay and a 417-ms delay and received feedback indicating whether they answered correctly. They then completed three runs that were similar in length to the main experimental runs in the scanning sessions. Each run contained a self and an external movement block (miniblocks) in alternating order (144 trials in total). The order of movement was counterbalanced across participants. Because being able to detect the delays was crucial for our paradigm, only participants who detected at least 50% of delays or more at the most extreme delay and who reported detecting a delay less than 50% of the 0-ms delay trials were invited to the fMRI experiment. One participant did not fulfill these criteria and was, therefore, not invited to the fMRI experiment. The training procedure took approximately 1 hr.

In the fMRI experiment, participants were positioned inside the bore in supine position with their right hand placed on the PMD handle, which was located beside them. Each participant received five experimental blocks, each of which contained two miniblocks (self- and externally generated movement conditions, each having 24 trials) in counterbalanced order. At the beginning of each miniblock, the participant was informed about the specific movement they had to perform (self- or externally generated). Each trial started with a cue, “Ready,” that lasted for 1,500 ms. This cue indicated that participants had to prepare themselves for the upcoming movement. Immediately after the disappearance of the cue, the camera was turned on for 4,000 ms. This timing frame was chosen to give participants some freedom in performing a self-generated, intentional movement. The participant was instructed to perform the movement (in the self-generated miniblock) or let the device move their hand (in the externally generated miniblock). The onset of the externally generated trial was jittered (500–1,000–1,500 ms). The participants saw their hand movement on the screen in real time or with one of five delays. After a 500-ms interval, the question “Delay?” appeared on the screen. Participants used their left index and middle fingers to respond “Yes” or “No,” respectively. Button assignment was counterbalanced across participants. The maximum time allowed for the response was 2,000 ms. Following the response, an intertrial interval of 2,000–5,000 ms was presented. Unimodal visual and bimodal audiovisual trials were randomized within blocks. Total number of trials for each participant was 240. Participants were allowed to have short breaks between the blocks. Each movement was monitored online and recorded for the purpose of ensuring compliance with instructions and for post hoc screening of movement parameters. The scanning procedure took approximately 1 hr.

**Analysis**

**Behavioral data analysis**

Statistical analyses (repeated-measures ANOVA and paired samples t tests) were performed using SPSS21 (IBM Corp., released 2012). Behavioral data were grouped based on participant’s responses regarding delay-detection judgments, referred to as the detection of delays. The proportion of delay responses was calculated for each condition by dividing the number of
trials in which the participant detected a delay between the movement and its visual feedback by the total number of trials in that condition.

fMRI Data Acquisition and Analysis

Functional imaging was carried out in a 3 T TIM Trio scanner (Siemens, Erlangen, Germany), using a 12-channel head coil. A gradient echo EPI sequence was used (TR: 1,650 ms, TE: 25 ms, flip angle: 70°). The whole brain was covered in 34 axial slices (gap: 15%, slice thickness: 4 mm), each of which was acquired as a 64×64 matrix (FoV: 192×192 mm², voxel size: 3×3×4.6 mm) in descending order. Three hundred thirty volumes were obtained during each experimental block. Anatomical images were obtained using a T1-weighted MPRAGE sequence (TR: 1,900 ms, TE: 2.26 ms, flip angle: 9°, FoV: 256×256, slice thickness: 1 mm, matrix: 256×256, voxel size: 1×1×1 mm).

In order to account for possible stimulus-correlated motion artifacts in the self-generated movement condition, we performed independent components analysis (ICA) on the raw data using the MELODIC tool (Beckmann & Smith, 2004) in FSL (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012; Smith et al., 2004). The data were prepared for the ICA with the implementation of slice-time correction, realignment using the MCFLIRT procedure, and spatial smoothing with an 8-mm full-width-at-half-maximum isotropic Gaussian kernel. After ICA, components that seem to indicate motion artifacts were removed. We also checked whether head motion differences were similar across self- and externally generated movements by conducting statistical analyses on head motion parameters using a freely available R script to estimate scan-to-scan motion of fMRI data (https://github.com/kesslerr/motionEstimator). The script estimates changes in absolute head motion over time by calculating differences on the x-, y-, and z-axes between successive scans and transforming these values into a single scan-to-scan difference (see Mazaika, Hoef, Glover, & Reiss, 2009; Wilke, 2012). For each run, we initially divided the head motion parameters obtained after realignment based on movement type (head motion parameters during self- vs. externally generated movements). We then calculated maximum scan-to-scan values for each run and movement type. These values were then averaged over all runs for each participant. Finally, we conducted a paired samples t test on the maximum scan-to-scan values averaged across runs for self- and externally generated movements. Our results showed no significant differences in scan-to-scan head motion values between the self-generated and externally-generated movements, t(17) = 1.76, p = 0.97.

The remaining preprocessing steps and analyses were performed using statistical parametric mapping (SPM12, Wellcome Department of Imaging Neuroscience, University College London, UK) implemented in MATLAB 2012b (MathWorks, Natick, MA). Accordingly, each participant’s anatomical image was coregistered to their first functional images, segmented and normalized to the standard Montreal Neurological Institute (MNI) standard space, and resampled to 2×2×2 mm³ voxels.

Prior to statistical analyses, we screened recorded video images from all participants to examine whether the movements were properly and/or actually performed. We excluded those trials in which the participant did not execute the instructed movement and bimodal trials in which the movement detection algorithm did not work (1.4% of all trials).

Statistical analyses were performed based on the following purposes: First, we aimed to identify possible BOLD-suppression effects during self-generated movements and the influence of auditory modality on BOLD activity when the movement was self-generated. As a proof of principle, we initially examined BOLD activity during the execution of both movements as these movements have been shown to engage overlapping as well as different brain regions (Ciccarelli et al., 2005; Miall, Christensen, Cain, & Stanley, 2007; Szameitat, Shen, Conforto, & Sterr, 2012; Weiller et al., 1996). We then examined BOLD-suppression effects in the self-generated condition by contrasting it with the externally generated condition. We also explored whether BOLD-suppression effects were similar across different modalities, i.e., when an auditory modality was presented. Second, we assessed possible differences in BOLD activity across movements and modalities for trials in which a delay was detected versus undetected. Here, we were interested in examining neural responses for detected versus undetected delays for self-generated movements and their sensory consequences and whether such processing would be modulated by the presence of an additional modality. Importantly, delay detection could be confounded by the amount of physical delay; i.e., participants are more likely to detect action–feedback discrepancies when these delays are large. In our statistical analyses, we also attempted to account for this confound by adopting a parametric approach in which both delay detection performance (detected vs. undetected) and the effect of physical delay were accounted for (see below for details).

For the statistical analyses, we implemented a standard summary statistic approach (Friston et al., 1995; Holmes & Friston, 1998; Worsley & Friston, 1995). This method rests on a mass univariate general linear model (GLM) in which time series from each voxel are modeled as a linear combination of experimental variables (effects of interest) plus an error term.
(Friston et al., 1995; Pernet, 2014). For both analyses, time series data from each participant and session were initially modeled using a design matrix that consisted of six regressors: four experimental conditions time-locked to the onset of camera and two additional regressors modeling the cue and response periods, respectively. In addition, we included two parametric modulators for each of the four experimental conditions in the following order: physical delay (delay in milliseconds for that trial) and delay detection (defined as a binary variable: $1 = \text{detected}, 0 = \text{not detected}$). As our focus was on BOLD activity linked with subjective delay detection, we put the regressor for physical delay first, to which SPM attributes higher variance to physical delay by orthogonalizing subjective delay detection with regard to the physical delay. This way, we aimed to capture variance associated specifically with the subjective detection of delay (Mumford, Poline, & Poldrack, 2015). The regressors were convolved with the canonical hemodynamic response function employed by SPM. In addition, six realignment parameters were included as nuisance regressors to account for residual motion artifacts. Time series data were high-pass filtered with a cutoff frequency of 128 s. Temporal autocorrelation was modeled by an autoregressive model of order 1 (AR(1)). After this model was estimated, parameter estimates from regressors of interest were used to compute contrast images corresponding to eight conditions of interest: self-generated unimodal (SelfUni), self-generated bimodal (SelfBi), externally generated unimodal (ExtUni), externally generated bimodal (ExtBi), delay detection (detected vs. undetected) during self-generated unimodal (DetSelfUni), delay detection during self-generated bimodal (DetSelfBi), delay detection during externally generated unimodal (DetExtUni), delay detection during externally generated bimodal (DetExtBi). These contrast images were passed to a second-level random effects flexible-factorial analysis. As a proof of principle, a conjunction of self- and externally generated movements was calculated to identify areas commonly activated during both movements ($(\text{SelfUni} + \text{SelfBi}) \cap (\text{ExtUni} + \text{ExtBi}))$. We also examined activity differences in the self- compared with externally generated movements using the contrast $[(\text{SelfUni} + \text{SelfBi}) - (\text{ExtUni} + \text{ExtBi})]$. BOLD suppression effects were assessed by conducting a $t$ test for externally versus self-generated conditions using the contrast $[(\text{ExtUni} + \text{ExtBi}) - (\text{SelfUni} + \text{SelfBi})]$. Moreover, we examined whether the same regions showed reduced BOLD activity during self-generated movements in both unimodal and bimodal conditions with a conjunction analysis $[(\text{ExtUni} - \text{SelfUni}) \cap (\text{ExtBi} - \text{SelfBi})]$. Differences in BOLD activity linked with the presentation of additional modality were tested by contrasting bimodal with the unimodal condition $[(\text{SelfBi} + \text{ExtBi}) - (\text{SelfUni} + \text{ExtUni})]$. Because we did not have a specific hypothesis regarding the direction of the interaction between movement and modality, an $F$ test was initially conducted to see whether there was any interaction between movement and modality, which would presumably be followed by a $t$ test when necessary. Two $t$ tests were conducted to assess whether BOLD activity linked with delay detection (detected vs. undetected) differed when the movement was self- as opposed to externally generated. These were $[(\text{DetSelfUni} + \text{DetSelfBi}) - (\text{DetExtUni} + \text{DetExtBi})]$ for self-generated movements and $[(\text{DetExtUni} + \text{DetExtBi}) - (\text{DetSelfUni} + \text{DetSelfBi})]$ for externally generated movements. The influence of auditory modality on BOLD activity related to the detection of delays were examined using the contrasts $[(\text{DetSelfUni} + \text{DetExtUni}) - (\text{DetSelfBi} + \text{DetExtBi})]$ for unimodal versus bimodal and $[(\text{DetSelfBi} + \text{DetExtBi}) - (\text{DetSelfUni} + \text{DetExtUni})]$ for bimodal versus unimodal conditions. We also examined whether activity associated with detection of delays was different when the movement was self-generated and in the presence of the auditory modality. The respective contrasts were $[(\text{DetSelfUni} - \text{DetSelfBi}) - (\text{DetExtUni} - \text{DetExtBi})]$ and $[(\text{DetSelfBi} - \text{DetSelfUni}) - (\text{DetExtBi} - \text{DetExtUni})]$.

We additionally assessed psychophysiological interactions (PPIs) between posterior cerebellar regions, namely bilateral cerebellar lobules IX and visual areas in the brain as a function of delay detection and movement type (see Results section for details). We restricted our region of interest to the posterior cerebellum, namely bilateral cerebellar lobules IX, as this region has been found to be highly involved in the visual guidance of movement (Stoodley & Schmahmann, 2010). We initially concatenated each participant’s session data using our original GLM on subjective delay detection, in which the number of sessions was modeled as additional block regressors. Accordingly, for each participant, we used two mask images of the left and right cerebellar lobules IX using the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) in order to look at the time course of each voxel. We then extracted the first eigenvariate from all voxels within a 4-mm radius of the seed region based on the local maxima of the participant within the mask and adjusted by the effects of interest. Extracted eigenvariances were passed to a GLM at the subject level, at which signal time course from left and right cerebellar lobules IX, our effects of interest (movement, modality) along with parametric modulators defining physical delay and delay detection, their psychological factors and PPI regressors were modeled. We also included cue and responses as well as movement parameters as regressors of no interest. “Detected > undetected” parametric modulators (i.e., the parameter estimates of the interaction term) of
respective PPI regressors were then passed to a random effects flexible-factorial GLM at the group level.

To correct for multiple comparisons at the whole-brain level, we used family-wise error (FWE) correction at \( p < 0.05 \) based on Gaussian random field theory (Worsley, Evans, Marrett, & Neelin, 1992; Worsley et al., 1996) implemented in SPM12. In addition, to avoid type II errors, Monte Carlo simulations were run using the estimated smoothness of our functional data (14.6 mm). This resulted in a minimum cluster size of 88 resampled voxels to achieve correction for multiple comparisons at \( p < 0.05 \), assuming an individual voxel type I error of \( p = 0.001 \) (Slotnick, Moo, Segal, & Hart, 2003). This method also allows one to detect moderate effects (Forman et al., 1995; Slotnick et al., 2003). The AAL atlas (Tzourio-Mazoyer et al., 2002) and SPM Anatomy Toolbox (Eickhoff et al., 2005) were used to label significant activations based on peak activation voxels. Visualization of significant activations was performed by SPM12 and MRICron (Rorden & Brett, 2000), and GraphPad Prism version 7.00 for Windows (GraphPad Software, La Jolla, CA). Effect sizes were reported as mean beta estimates using the MarsBar toolbox (Brett, Anton, Valabregue, & Poline, 2006). We calculated the beta values separately for each participant and then obtained an average value for each effect of interest and participant.

**Results**

**Behavioral data**

Figure 3a and b displays proportion of delay responses as a function of movement for unimodal and bimodal conditions, respectively (see also Supplementary Figure S1). A 2 (Movement: Self- vs. Externally generated) \( \times 2 \) (Modality: Unimodal vs. Bimodal) repeated-measures ANOVA was conducted on the proportion of delay responses. There was a main effect of movement, \( F(1, 19) = 6.83, p < 0.05, \eta^2 = 0.21 \). Accordingly, the proportion of delay responses was significantly higher in the externally generated (\( M = 0.52, SEM = 0.03 \)) than in the self-generated (\( M = 0.45, SEM = 0.04 \)) condition (Figure 3c). There was no main effect of modality, \( F(1, 19) = 2.24, p = 0.15 \), nor was there an interaction, \( F(1, 19) = 0.08, p = 0.77 \).

In order to test whether movement duration was similar across self- and externally generated movements, we used recorded trials to determine the onset and offset of each movement (18 participants). A post hoc paired-samples \( t \) test showed that average movement durations across self- and externally generated trials were not significantly different from each other, \( t(17) = 1.23, p = 0.23 \); average movement duration for
self-generated movements (M = 1,373.82, SEM = 81.84), average movement duration for externally generated movements (M = 1,282.78, SEM = 44.22).

**fMRI data**

We first explored common activations during self- and externally generated conditions by using a conjunction of self- and externally generated movements that were contrasted against an implicit baseline. We found that both movements led to activity in visual and somatosensory areas, including bilateral somatosensory areas along with bilateral inferior and middle occipital gyri. In addition, both movements led to activations in premotor and motor areas, inferior frontal and parietal areas, and bilateral thalamus and insula (see Figure 4, and Supplementary Figure S2). The anatomical locations of these clusters are shown in Table 1. Examination of common areas across different movements and modalities revealed similar results.

**BOLD-suppression effects** were assessed by contrasting externally with self-generated conditions. The contrast revealed bilateral activations in somatosensory cortices; middle and superior occipital gyri, including the V5; and middle and superior temporal gyri (MTG and STG, respectively). Other areas involved premotor and motor regions, inferior parietal lobules (IPL),...
precuneus, thalamus, and insula (see Figure 5a and Supplementary Figure S3a). Anatomical locations of these clusters are shown in Table 2.

We also identified common BOLD-suppression effects in sensory areas across unimodal and bimodal conditions. Significant BOLD-suppression effects were found in bilateral primary and secondary somatosensory cortices, middle and inferior occipital gyri, MTG, and STG. Additional activations were observed in supplementary motor area (SMA), bilateral IPL, putamen, angular gyri, supramarginal gyri, and thalamus (see Figure 5b and Supplementary Figure S3b). Contrasting self- with externally generated movements did not lead to significant activations (both at $p_{FWE} < 0.05$ and $p < 0.001$ uncorrected level).

As expected, bimodal compared with unimodal trials led to significant activations in bilateral temporal cortices with two large clusters in bilateral STG (MNI coordinates: $x, y, z = 64, -24, 10, T = 13.76, k_E = 2,429, p_{FWE} < 0.05$; MNI coordinates: $x, y, z = -50, -26, 8, T = 11.53, k_E = 1,588, p_{FWE} < 0.05$). However, when relative strength of these activations was examined separately for each movement, activations in the self-generated bimodal condition were weaker than in the externally generated bimodal condition. Figure 6 illustrates relative BOLD amplitudes in these areas across different movements and modalities (see also Supplementary Figure S4). Using a mask created from the ([$\text{Self}_{Bi} + \text{Ext}_{Bi}$] – [$\text{Self}_{Uni} + \text{Ext}_{Uni}$]) contrast, we further conducted a $t$ test ([Ext$_{Bi}$/Self$_{Bi}$]) to confirm whether activation in auditory areas were significantly higher in the externally than in the self-generated condition. We found significant activations in bilateral STG (MNI coordinates: $x, y, z = 46, -32, 22, T = 8.95, k_E = 1,747, p_{FWE} < 0.05$; MNI coordinates: $x, y, z = -44, -32, 18, T = 7.65, k_E = 986, p_{FWE} < 0.05$). Finally, there were no significant interactions between movement and modality.

Our second analysis regarding delay detection resulted in a main effect. First, delay detection

<table>
<thead>
<tr>
<th>Anatomical label</th>
<th>Hemisphere</th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
<th>$t$ value</th>
<th>No. voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occipital mid R</td>
<td>Right</td>
<td>36</td>
<td>-90</td>
<td>2</td>
<td>16.53</td>
<td>2,835</td>
</tr>
<tr>
<td>Temporal mid R</td>
<td>Right</td>
<td>48</td>
<td>-72</td>
<td>2</td>
<td>13.46</td>
<td></td>
</tr>
<tr>
<td>Calcarine R</td>
<td>Right</td>
<td>20</td>
<td>-100</td>
<td>2</td>
<td>13.18</td>
<td></td>
</tr>
<tr>
<td>Occipital mid L</td>
<td>Left</td>
<td>-36</td>
<td>-88</td>
<td>4</td>
<td>14.99</td>
<td>2,518</td>
</tr>
<tr>
<td>Calcarine L</td>
<td>Left</td>
<td>-14</td>
<td>-94</td>
<td>6</td>
<td>13.65</td>
<td></td>
</tr>
<tr>
<td>Occipital mid L</td>
<td>Left</td>
<td>-30</td>
<td>-96</td>
<td>2</td>
<td>13.63</td>
<td></td>
</tr>
<tr>
<td>Postcentral R</td>
<td>Right</td>
<td>50</td>
<td>-24</td>
<td>48</td>
<td>8.85</td>
<td>692</td>
</tr>
<tr>
<td>Postcentral R</td>
<td>Right</td>
<td>44</td>
<td>-30</td>
<td>54</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>Postcentral R</td>
<td>Right</td>
<td>32</td>
<td>-36</td>
<td>46</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Postcentral L</td>
<td>Left</td>
<td>-44</td>
<td>-32</td>
<td>46</td>
<td>8.49</td>
<td>841</td>
</tr>
<tr>
<td>Parietal inf L</td>
<td>Left</td>
<td>-50</td>
<td>-28</td>
<td>50</td>
<td>8.09</td>
<td></td>
</tr>
<tr>
<td>Postcentral L</td>
<td>Left</td>
<td>-44</td>
<td>-36</td>
<td>54</td>
<td>7.94</td>
<td></td>
</tr>
<tr>
<td>Supp motor area L</td>
<td>Left</td>
<td>0</td>
<td>2</td>
<td>56</td>
<td>7.11</td>
<td>183</td>
</tr>
<tr>
<td>Supp motor area R</td>
<td>Right</td>
<td>4</td>
<td>14</td>
<td>50</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>Temporal sup R</td>
<td>Right</td>
<td>64</td>
<td>-36</td>
<td>22</td>
<td>6.64</td>
<td>79</td>
</tr>
<tr>
<td>Frontal inf oper R</td>
<td>Right</td>
<td>50</td>
<td>8</td>
<td>22</td>
<td>5.77</td>
<td>136</td>
</tr>
<tr>
<td>Precentral R</td>
<td>Right</td>
<td>52</td>
<td>6</td>
<td>34</td>
<td>5.54</td>
<td></td>
</tr>
<tr>
<td>Supramarginal L</td>
<td>Left</td>
<td>-50</td>
<td>-26</td>
<td>20</td>
<td>5.72</td>
<td>110</td>
</tr>
<tr>
<td>Supramarginal L</td>
<td>Left</td>
<td>-60</td>
<td>-22</td>
<td>16</td>
<td>5.46</td>
<td></td>
</tr>
<tr>
<td>Insula R</td>
<td>Right</td>
<td>36</td>
<td>20</td>
<td>6</td>
<td>5.48</td>
<td>93</td>
</tr>
<tr>
<td>Thalamus L</td>
<td>Left</td>
<td>-12</td>
<td>-18</td>
<td>6</td>
<td>5.39</td>
<td>26</td>
</tr>
<tr>
<td>Rolandic oper L</td>
<td>Left</td>
<td>-46</td>
<td>-2</td>
<td>8</td>
<td>5.15</td>
<td>9</td>
</tr>
<tr>
<td>Cerebellum VIII L</td>
<td>Left</td>
<td>-12</td>
<td>-70</td>
<td>-44</td>
<td>5.06</td>
<td>41</td>
</tr>
<tr>
<td>Cerebellum VIII L</td>
<td>Left</td>
<td>-20</td>
<td>-66</td>
<td>-48</td>
<td>4.94</td>
<td></td>
</tr>
<tr>
<td>Insula L</td>
<td>Left</td>
<td>-34</td>
<td>18</td>
<td>4</td>
<td>4.92</td>
<td>8</td>
</tr>
<tr>
<td>Frontal inf oper L</td>
<td>Right</td>
<td>54</td>
<td>14</td>
<td>4</td>
<td>4.84</td>
<td>6</td>
</tr>
<tr>
<td>Rolandic oper R</td>
<td>Right</td>
<td>56</td>
<td>-22</td>
<td>20</td>
<td>4.71</td>
<td>3</td>
</tr>
<tr>
<td>Frontal inf oper R</td>
<td>Right</td>
<td>56</td>
<td>12</td>
<td>10</td>
<td>4.71</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1. Anatomical locations of common peak activations during self- and externally generated movements ($p_{FWE} < 0.05$, height threshold = 4.67).
During self- as opposed to externally generated movements, bilateral regions in the anterior and posterior cerebellar areas including bilateral cerebellar lobules IV-VI (MNI coordinates: $x, y, z = -24, -44, 36, T = 4.31$; MNI coordinates: $x, y, z = 12, -46, -20, T = 4.32$) as well as lobules IX and the cerebellar vermis (MNI coordinates: $x, y, z = -2, -36, -38, T = 4.51, k_E = 1.592, p < 0.05$ cluster corrected; see Figure 7 and Supplementary Figure S5). We found no significant activations when we contrasted externally with self-generated movements. Finally, there were no interaction effects.

In order to examine whether activity linked with detecting versus not detecting a delay in the cerebellum was correlated with activity in visual areas during self-generated movements, we employed a PPI analysis. PPI employs a simple regression method in which task-related (psychological factor) modulations in the time series of a seed region (physiological factor) can be investigated in terms of connectivity changes in other voxels of the brain (Friston et al., 1997). Our hypothesis was that, if the cerebellum were specifically involved in detecting discrepancies between self-generated movements and their visual feedback, increased...
connectivity between this area and the visual cortex would be evident only for self-generated movements. Due to our a priori hypothesis, the connectivity analysis was restricted to visual areas previously found to be involved in perceiving visual feedback of one’s own movement (Leube et al., 2003). Accordingly, after an initial cluster defining threshold of $p < 0.001$, we applied peak-level family-wise error small volume correction within a 10-mm radius region of interest (ROI) centered on the coordinates corresponding to the right and left occipital gyri reported in Leube et al. (2003). As a statistical threshold, we chose $p_{FWE} = 0.025$ because we additionally applied a Bonferroni correction for two tests (a PPI for the left and right cerebellar lobules IX as seed, respectively). The analysis revealed a significant cluster in the right superior occipital gyrus (SOG; MNI coordinates: $x = 18, y = 94, z = 14$, $T = 3.65, k_E = 15$, small volume correction $p_{FWE} = 0.014$; see Figure 8). This suggests that activity in the right superior occipital gyrus correlated significantly with activity in the left cerebellar lobule IX when a delay was detected during self- as opposed to externally generated movements. For the right cerebellar lobule IX, the analysis revealed a region in the left SOG; however, this did not reach significance after correcting for the number of tests performed ($p_{FWE} = 0.038$).

<table>
<thead>
<tr>
<th>Anatomical label</th>
<th>Hemisphere</th>
<th>Coordinates</th>
<th>t value</th>
<th>No. voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supramarginal R</td>
<td>Right</td>
<td>44, -32, 24</td>
<td>8.57</td>
<td>3,408</td>
</tr>
<tr>
<td>Temporal sup R</td>
<td>Right</td>
<td>58, -36, 20</td>
<td>7.83</td>
<td></td>
</tr>
<tr>
<td>Temporal sup L</td>
<td>Left</td>
<td>-46, -30, 22</td>
<td>8.25</td>
<td>3,317</td>
</tr>
<tr>
<td>Rolando oper L</td>
<td>Left</td>
<td>-44, -38, 26</td>
<td>7.78</td>
<td></td>
</tr>
<tr>
<td>Temporal sup L</td>
<td>Left</td>
<td>-56, -44, 18</td>
<td>6.84</td>
<td></td>
</tr>
<tr>
<td>Precentral L</td>
<td>Left</td>
<td>-28, -28, 60</td>
<td>7.26</td>
<td>7,567</td>
</tr>
<tr>
<td>Supp motor area L</td>
<td>Left</td>
<td>-6, 4, 58</td>
<td>7.08</td>
<td></td>
</tr>
<tr>
<td>Supp motor area R</td>
<td>Right</td>
<td>8, 6, 52</td>
<td>6.78</td>
<td></td>
</tr>
<tr>
<td>Insula R</td>
<td>Right</td>
<td>30, 22, 8</td>
<td>6.78</td>
<td>230</td>
</tr>
<tr>
<td>Precentral L</td>
<td>Left</td>
<td>-54, 2, 16</td>
<td>6.09</td>
<td>160</td>
</tr>
<tr>
<td>Insula L</td>
<td>Left</td>
<td>-30, 20, 2</td>
<td>5.98</td>
<td>584</td>
</tr>
<tr>
<td>Putamen L</td>
<td>Left</td>
<td>-24, 6, 6</td>
<td>5.67</td>
<td></td>
</tr>
<tr>
<td>Putamen L</td>
<td>Left</td>
<td>-22, 18, 0</td>
<td>5.62</td>
<td></td>
</tr>
<tr>
<td>Frontal mid R</td>
<td>Right</td>
<td>36, -4, 52</td>
<td>5.65</td>
<td>258</td>
</tr>
<tr>
<td>Precentral R</td>
<td>Right</td>
<td>42, -8, 46</td>
<td>5.11</td>
<td></td>
</tr>
<tr>
<td>Temporal sup L</td>
<td>Left</td>
<td>-48, -14, 6</td>
<td>5.51</td>
<td>94</td>
</tr>
<tr>
<td>Cuneus R</td>
<td>Right</td>
<td>20, -68, 36</td>
<td>5.2</td>
<td>57</td>
</tr>
<tr>
<td>Insula L</td>
<td>Left</td>
<td>-30, 12, -12</td>
<td>5.13</td>
<td>6</td>
</tr>
<tr>
<td>Lingual R</td>
<td>Right</td>
<td>26, -68, -2</td>
<td>5.04</td>
<td>25</td>
</tr>
<tr>
<td>Thalamus L</td>
<td>Left</td>
<td>-6, -16, 0</td>
<td>4.93</td>
<td>18</td>
</tr>
<tr>
<td>Precentral L</td>
<td>Left</td>
<td>-46, -2, 36</td>
<td>4.88</td>
<td>18</td>
</tr>
<tr>
<td>Frontal sup R</td>
<td>Right</td>
<td>20, -8, 64</td>
<td>4.87</td>
<td>4</td>
</tr>
<tr>
<td>Postcentral L</td>
<td>Left</td>
<td>-48, -16, 50</td>
<td>4.86</td>
<td>5</td>
</tr>
<tr>
<td>Precentral R</td>
<td>Right</td>
<td>54, -12, 42</td>
<td>4.84</td>
<td>5</td>
</tr>
<tr>
<td>Frontal mid L</td>
<td>Left</td>
<td>-36, 20, 32</td>
<td>4.84</td>
<td>9</td>
</tr>
<tr>
<td>Cerebellum crus II R</td>
<td>Right</td>
<td>14, -82, -36</td>
<td>4.83</td>
<td>8</td>
</tr>
<tr>
<td>Putamen R</td>
<td>Right</td>
<td>26, -2, 14</td>
<td>4.82</td>
<td>4</td>
</tr>
<tr>
<td>Frontal inf tri R</td>
<td>Right</td>
<td>36, 16, 28</td>
<td>4.8</td>
<td>18</td>
</tr>
<tr>
<td>Postcentral L</td>
<td>Left</td>
<td>-44, -10, 50</td>
<td>4.79</td>
<td>1</td>
</tr>
<tr>
<td>Frontal mid R</td>
<td>Right</td>
<td>30, 44, 14</td>
<td>4.77</td>
<td>2</td>
</tr>
<tr>
<td>Precentral R</td>
<td>Right</td>
<td>20, -18, 66</td>
<td>4.74</td>
<td>1</td>
</tr>
<tr>
<td>Occipital mid L</td>
<td>Left</td>
<td>-32, -72, 28</td>
<td>4.73</td>
<td>1</td>
</tr>
<tr>
<td>Putamen R</td>
<td>Right</td>
<td>24, 12, 4</td>
<td>4.72</td>
<td>4</td>
</tr>
<tr>
<td>Thalamus R</td>
<td>Right</td>
<td>10, -14, 0</td>
<td>4.67</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Anatomical locations of peak activations during externally versus self-generated movements (Ext — Self contrast) ($p_{FWE} < 0.05$, height threshold = 4.67).
Discussion

In the present study, we assessed neural correlates of perceiving self-generated movements and resulting visual feedback and whether and how an additional (auditory) modality would alter corresponding processing in the brain associated with these movements. By contrasting self- with externally generated movements of the wrist, we were able to address efference copy-based predictive processes. Moreover, our experimental setup allowed us to examine these processes with naturalistic action–feedback associations, which are usually more complex than simple button presses. Behaviorally, we found significantly higher proportion of delay responses for externally compared with self-generated movements. fMRI analysis revealed activations in several motor-related and sensory areas for both movements. As hypothesized, we found reduced BOLD activity in somatosensory, visual, and auditory regions of the brain during self- compared with externally generated movements, demonstrating BOLD suppression in sensory cortices. This effect was similar across unimodal and bimodal conditions, implying predictive processing and subsequent suppression for self-generated sensory inputs. Detection of delays was found to be linked with increased recruitment of bilateral anterior and posterior regions of the cerebellum including lobules IV–VI as well as lobules IX when the movement was self-generated as opposed to when it was externally generated. PPI analysis showed enhanced functional connectivity in the left cerebellar lobule IX with right SOG, further supporting cerebellum’s involvement in detection of discrepancies between self-generated movements and accompanying visual feedback. Together, these new findings extend previous findings regarding BOLD suppression in sensory cortices associated with voluntary movements.

Figure 6. Main effect of auditory modality in the (a) right STG (MNI coordinates: x, y, z = 64, −24, 10; pFWE < 0.05) and the (b) left STG (MNI coordinates: x, y, z = −50, −26, 8; pFWE < 0.05) overlaid on a standardized T1-weighted image. Dots and bar graphs show mean beta estimates from individual participants and across conditions (±SEM) for the corresponding peak voxel, respectively. See also Supplementary Figure S4 for data from a single representative participant.

Figure 7. Significant activation cluster in the cerebellum including bilateral cerebellum IV–VI and bilateral cerebellum IX (peak MNI coordinates: x, y, z = 24, −44, −36) associated with detecting a delay during self- compared with externally generated movements (p < 0.05 cluster corrected) overlaid on a standardized T1-weighted image. Dots and bar graphs show mean beta estimates from individual participants and across conditions (±SEM) for the corresponding peak voxel. See also Supplementary Figure S5 for data from a single representative participant.
to the processing of naturalistic action–feedback associations and specifically as a result of efference copy-based processes. Moreover, our results support the relevance of the cerebellum in the detection of delays between self-generated movements and accompanying visual feedback. We now discuss our findings in detail.

**BOLD suppression in sensory cortices during self-generated movements**

Both self- and externally-generated conditions led to activations in areas involved in the planning and control of movements, such as the SMA, M1, thalamus, and insula. These results are in line with existing evidence showing similar activations in pre-motor and motor structures during self- and externally generated movements (Balslev, Nielsen, Lund, Law, & Paulson, 2006; Ciccarelli et al., 2005; Miall et al., 2007; Szameitat et al., 2012; Weiller et al., 1996). As hypothesized, self-generated movements, in contrast with externally generated movements led to reduced BOLD activity in somatosensory and visual regions. The BOLD-suppression effect was specific to higher order visual areas, including V5, which has been implicated in the processing of visual motion (Zeki et al., 1991). This result points to a highly specific down-regulation of visual feedback associated with self-generated movements. Together, our results suggest increased predictive processing of sensory signals associated with self-generated movements compared to when these signals were externally generated.

Strikingly, BOLD-suppression effects were not only evident in somatosensory and visual areas, but also in the auditory cortex in both unimodal and bimodal conditions. In other words, activity in auditory cortices was suppressed during self-generated movements independent of whether an auditory stimulus occurred or not. This is similar to our previous finding, demonstrating suppressed BOLD signal in auditory cortices during the presentation of audiovisual and visual-only stimuli (Straube et al., 2017). These findings can be explained in terms of expectation-based processing. Expected sensory inputs have been shown to result in reduced neural processing, implying a predictive processing for such inputs (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; den Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010; Kok & de Lange, 2015; Kok, Jehee, & de Lange, 2012; Murray, Kersten, Olshausen, Schrater, & Woods, 2002). In our study, participants were likely to anticipate the auditory stimulus, which led to reduced BOLD activity in auditory cortices because unimodal and bimodal trials were intermixed. This effect should be specific for self-generated trials as opposed to externally generated trials as we propose that sensory inputs associated with our own actions are more likely to be expected due to increased control we can exert on these inputs, which led to BOLD suppression in corresponding sensory areas. In other words, sensory consequences of self-generated movements might have led to heightened expectation and subsequent down-regulation for these consequences (Blakemore, Frith, et al., 1999; Blakemore, Wolpert, et al., 1998; Blakemore, Wolpert, & Frith, 2000; Wolpert, 2007; Wolpert & Flanagan, 2001). Our results are in line with these findings and suggest that people might be more ready to form expectations about upcoming sensory inputs when they are in control of the likely consequences.

Suppression effects observed in sensory regions are in line with our previous study (Straube et al., 2017).
along with several other studies demonstrating reduced activity in sensory areas linked with one’s own action (Benazet, Thénault, Whittingstall, & Bernier, 2016; Blakemore, Frith, & Wolpert, 1999; Shergill et al., 2013). Together, these findings support the notion of efference copy-based predictions in processing sensory consequences of voluntary actions (Blakemore et al., 2000; Cui et al., 2014; Wolpert et al., 2001; Wolpert et al., 1998). Externally generated movements also yielded increased activity in the cingulate cortex, MTG, precuneus, and putamen compared with self-generated movements. The MTG, precuneus, and putamen have been implicated in time perception (Matell & Meck, 2004; Stevens, Kiehl, Pearlson, & Calhoun, 2007) and processing of visuomotor incongruencies (Balslev et al., 2006; David et al., 2007; Dreher & Grafman, 2002; Leube et al., 2003; MacDonald & Paus, 2003).

Moreover, externally generated movements led to increased BOLD activity in the IPL, an area found to be relevant in integrating visuomotor signals (Balslev et al., 2006; Farrer et al., 2008; Limanowski et al., 2017; Schnell et al., 2007), distinction between self- and other-generated actions (Blakemore & Frith, 2003; Jackson, 2004; Leube et al., 2003; Shimada, Hiraki, & Oda, 2005; Weiller et al., 1996), and mental imagery and complex actions (Blakemore & Sirigu, 2003; Haggard, 2009; Shergill et al., 2013; Sirigu, Daprati, Pradat-Diehl, Franck, & Jeannerod, 1999). Our results, therefore, provide support for the involvement of these areas in processing visuomotor incongruencies, possibly contributing to self–other distinction.

Cerebellum’s role in the detection of delays associated with self-generated movements and accompanying feedback

fMRI analysis on contrasting trials in which a delay was detected versus not detected revealed two important findings. First, regions of the bilateral anterior and posterior cerebellum, including lobules IV–V and IX, were found to be significantly engaged for trials in which a delay was detected versus not detected during self-generated movements. This is in line with the notion of the cerebellum as a state estimator, modulating predicted sensory consequences of motor commands and providing error signal for the unpredicted consequences of these commands (Blakemore & Frith, 2003; Blakemore et al., 2001; Blakemore, Wolpert, et al., 1999; Leube et al., 2003; Miall, Weir, Wolpert, & Stein, 1993; Miall et al., 2007; van Kemenade et al., 2019; Wolpert et al., 1998). Our finding of higher involvement in the cerebellar areas for subjectively delayed trials during self- compared with externally-generated movements not only supports the cerebellum’s role as generating predictions and detecting errors during sensorimotor processing, but it also provides a subjective component to the involvement of the cerebellum in processing these discrepancies. It has been claimed that awareness of action–feedback discrepancies might not be available to the cerebellum (Blakemore & Sirigu, 2003; Farrer et al., 2008). In this sense, our finding suggests the opposite, that the cerebellum might be involved in the detection of discrepancies between predicted and actual feedback. This was further confirmed by our PPI analysis demonstrating increased functional connectivity between the left cerebellum IX and the right SOG for self-compared with externally generated movements when a delay was detected.

Limitations and future directions

In the current study, we restricted the movement range by the PMD in order to have self- and externally generated movements as similar as possible regarding movement kinematics. We also trained participants prior to the fMRI experiment and monitored their movements throughout the fMRI sessions. Moreover, we tracked movement recordings of 18 participants to check compliance with the instructions and found no significant differences in the average movement duration across movements. Nevertheless, it could be that self-generated movements might be more demanding to execute than externally generated movements, which might, in turn, influence attentional orienting to accompanying sensory stimuli (Hughes, Desantis, & Waszak, 2012).

Second, our main hypothesis concerned BOLD suppression in sensory cortices during self-generated movements as opposed to externally generated movements. We, therefore, focused on differences in sensory regions. However, our main effect also revealed increased BOLD activity in premotor and motor structures. Together with the possibility that self-generated movements might influence processes other than motor control, such as attention, future experiments might benefit from including these movements alone as baseline conditions or scrutinizing different components that define volition, such as what or when to perform an action in terms of their neural correlates (Brass & Haggard, 2008; Hoffstaedter, Grefkes, Zilles, & Eickhoff, 2013).

Another potential limitation of the current study concerns BOLD activity associated with delay detection judgments. We aimed to overcome the covariance of physical delay and delay detection by accounting for the effect of physical delay first and attributing only the remaining variance to subjective delay detection (Mumford et al., 2015). Despite this, it is not entirely possible to disentangle the effect of physical delays.
from that of delay detection judgments (Wearden & Jones, 2007). Therefore, our results with regard to
detection of delays must be interpreted with caution.

We did not find a behavioral advantage in bimodal
conditions, which goes against previous findings on
voluntary actions and multisensory processing (Arab-
zadeh et al., 2008; Desantis & Haggard, 2016b;
Desantis, Mamassian, et al., 2014; van Kemenade et al.,
2016). This could be explained by the fact that the
auditory stimulus in the current study was not
perceived as part of the movement because the visual
feedback was naturalistic, whereas the auditory stimu-
lus consisted of a beep, which was much more
abstract. We chose a simple sine wave instead of a
naturalistic sound recording because a sound recording
would have introduced unwanted additional delays
between action and its feedback due to additional
computational processing. However, this means, at the
same time, that we cannot rule out the possibility that
this might have reduced integration of the sound to the
movement and its visual feedback. Another possible
explanation lies in the different movement and feed-
back types between the different studies. In previous
studies in which multisensory processing enhanced
behavioral performance, the feedback occurred at the
end of the movement, i.e., a button press leading to a
visual dot or a beep. Here, the visual feedback occurred
during movement. In their study, David, Skoruppa,
Gulberti, Schultz, and Engel (2016) found that
participants were more sensitive to outcome- as
opposed to movement-related feedback. It could,
therefore, be the case that multisensory processing aids
particularly behavioral performance when the feedback
is outcome-related and less when it is movement-related
to which participants are generally less sensitive. We
suggest future studies thoroughly investigate the
interplay between different types of movements and
multiple sensory modalities associated with these
movements as well as the role of the naturalness of the
sensory feedback.

Although perception of the sensory consequences
triggered by voluntary movements has been extensively
investigated, studies using naturalistic action–feedback
associations that would better mimic real-world situa-
tions are limited. A study by Benazet et al. (2016)
focused on this aspect by using visual feedback of a
moving limb. In line with our results, they found that
cortical processing of visual reafferent signals are
suppressed, but only when the signals are correctly
predicted. We believe that integrating real-world
situations into experimental paradigms is of crucial
importance, and more studies are needed to investigate
natural movements executed in everyday life. In
addition, recent investigations have found contradic-
tory evidence for the reduced sensation and neural
processing of self-generated movement consequences
(Mifsud et al., 2016; Reznik, Henkin, Levy, & Mukamel,
2015; Reznik, Henkin, Schadel, & Mukamel, 2014; Roussel,
Hughes, & Waszak, 2013; Yon & Press, 2017). Therefore, more studies are needed to clarify
mechanisms involved in perceiving sensory inputs
associated with our own actions, whether and how
predictions are influenced by these actions, and the
impact of cross-modal processing on subsequent
perceptual experience.

Conclusions

In conclusion, we investigated for the first time
action-related BOLD suppression during self-generated
movements leading to unimodal visual versus bimodal
audiovisual feedback using naturalistic action–feed-
back associations. Externally generated movements
allowed us to effectively distinguish the influence of
efferece copy-based predictive processes from reaf-
ferent feedback. We found reduced BOLD activity in
somatosensory, visual, and auditory regions of the
brain during self- compared with externally generated
movements and across different modalities, indicating
heightened suppression of upcoming multisensory
stimuli associated with self-generated movements.
Detection of delays was found to be linked with
increased cerebellar processing specifically for self-
generated movements. Together, these findings extend
previous accounts regarding BOLD suppression to
realistic action–outcome associations and for multiple
sensory inputs associated with these actions. Futher-
more, our results highlight the cerebellum’s role in the
detection of temporal discrepancies between voluntary
actions and accompanying feedback, supporting its
relevance in action–perception coupling.

Keywords: self-generated movements, suppression,
cross-modal, cerebellum

Acknowledgments

This study was funded by the Deutsche
Forschungsgemeinschaft (DFG, German Research
Foundation) project number 222641018 - SFB/TRR
135, Cardinal mechanisms of perception: prediction,
valuation, categorization and through the International
Research Training Group, IRTG 1901, The Brain in
Action-BrainAct. Benjamin Straube is supported by
DFG grants STR 1146/8-1 and STR 1146/9-1. The
authors confirm that there are no known conflicts of
interest associated with this publication. We thank Jens
Sommer for technical support, Christina Lubinus for
assistance with data collection, and Inessa Kraft for
assistance with data analysis. The MRI data acquisition was supported by the Core Facility Brain Imaging, Faculty of Medicine, Philippus University Marburg, Germany. The data are available at https://doi.org/10.5281/zenodo.3479339.

Commercial relationships: none.
Corresponding author: B. Ezgi Arikan.
Address: Department of Psychology, Justus-Liebig University Giessen, Giessen, Germany.
Email: arikan.ezgi@gmail.com.

References


Calvert, G. A., Campbell, R., & Brammer, M. J.


Dreher, J.-C., & Grafman, J. (2002). The roles of the cerebellum and basal ganglia in timing and error prediction. *The European Journal of Neuroscience*, 16(8), 1609–1619.


