Brain Structure and Function (2019) 224:2577–2586
https://doi.org/10.1007/s00429-019-01920-x

SHORT COMMUNICATION

Role of the supplementary motor area in auditory sensory attenuation

Han-Gue Jo[1,2] · Ute Habel[1,2] · Stefan Schmidt[3]

Received: 28 March 2019 / Accepted: 6 July 2019 / Published online: 16 July 2019

Abstract

Self-generated tones elicit smaller brain responses as compared to externally generated tones. This phenomenon known as sensory attenuation has been explained in terms of an internal forward model in which the brain predicts the upcoming events and thereby attenuates the sensory processing. Such prediction processes have been suggested to occur via an efference copy of the motor command that is sent from the motor system to the lower order sensory cortex. However, little is known about how the prediction is implemented in the brain's network organization. Because the supplementary motor area (SMA) is a primary brain structure of the motor system, we attributed the implementation of the prediction to the SMA. To address this question, we examined generative network models for auditory ERPs. ERPs were evoked by either a self-generated or externally generated tone, while subjects were paying attention to their motor action or to the tone. The tone itself was the same throughout all conditions. The network models consisted of three subsets embedding alternative hypotheses of the hierarchical structures: (1) auditory fields of the temporal lobe, (2) adding connections to the SMA, and (3) adding prediction signal to the SMA. The model comparison revealed that all ERP responses were mediated by the network connections across the auditory cortex and the SMA. Importantly, the prediction signal to the SMA was required when the tone was self-generated irrespective of the attention factor, whereas the externally generated tone did not require the prediction. We discussed these results in the context of the predictive coding framework.

Keywords Sensory attenuation · Auditory system · ERP · DCM

Introduction

The effect of a self-initiated action, compared to an externally triggered effect, elicits smaller cortical responses and is perceived as less intense with respect to its subjective experience. This so-called sensory attenuation has been attributed to an internal forward model, such that an "efference copy" of the leading motor command is used to predict the action effect (Blakemore et al. 2000; Wolpert and Ghahramani 2000). The prediction is compared to the actual sensory feedback, and an accurate prediction of the feedback attenuates the cortical processing of the sensory inputs. If a sensory attenuation results from such an internal forward mechanism, there might be a prediction process that is internally generated and alters the sensory processing of the inputs.

One example for cortical evidence of sensory attenuation comes from the auditory domain. Auditory event-related potentials (ERPs) of self-generated tones have been shown to differ from those of externally generated tones (Schafer and Marcus 1973), demonstrating suppressed amplitudes of the ERPs measured over the medial frontal cortex. More recently, suppression of the N1–P2 complex of the evoked potential has been seen as a result of the internal forward mechanism (Martikainen et al. 2005; Bäss et al. 2008; Baess et al. 2011; Saupe et al. 2013; Ford et al. 2014; Poonian et al. 2015; Ghio et al. 2018). In these studies, the suppression has

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00429-019-01920-x) contains supplementary material, which is available to authorized users.

✉ Han-Gue Jo
hjo@ukaachen.de

[1] Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen University, Aachen, Germany

[2] JARA-Institute Brain Structure Function Relationship (INM-10), Research Center Juelich and RWTH Aachen University, Aachen, Germany

[3] Department of Psychosomatic Medicine and Psychotherapy, Medical Faculty, Medical Center, University of Freiburg, Freiburg, Germany

been discussed as a prediction process when the tone feedback was triggered by a voluntary action. The N1–P2 suppression has also been linked to movement planning rather than movement execution, suggesting that the intention-to-act plays a considerable role in sensory attenuation (i.e., voluntary action planning; Timm et al. 2014). In line with this, the supplementary motor area (SMA), a key brain structure for voluntary action planning (Deiber et al. 1999; Lau et al. 2004), has been reported to signal an efference copy to the auditory cortex to modify auditory processing (Reznik et al. 2015). Indeed, growing evidence supports the involvement of the SMA in not only motor-related auditory processing, but also auditory processing in general (for a review, see Lima et al. 2016). Moreover, it has been shown that the experience of a temporal shift of a self-generated tone was in accordance with the early part of the readiness potential (Jo et al. 2014), which is considered as a brain representation of the emergence of the intention-to-act localized in the SMA (for a review, see Shibasaki and Hallett 2006; but also see a recent review by Schmidt et al. 2016). This finding suggests that prediction process generated from the SMA may alter the experience of the tone feedback. Together, it is presumed that during voluntary tone generation such brain prediction could be implemented in the SMA from where it might be signaled to the auditory cortex, which in turn would result in an N1–P2 suppression.

The present study aimed to examine whether the sensory attenuation as revealed in auditory ERPs is attributed to such a prediction process in the brain. We acquired brain responses from a previous intentional binding study (Jo et al. 2014), in which self-generated and externally generated tones elicited auditory ERPs. Intentional binding refers to the observation that voluntary action and its consequence effect are perceived as being closer together in time, which is a critical determinant for the experience of agency (Haggard et al. 2002). This phenomenon has been explained as a result of the prediction process, since intentional binding is sensitive to the accurate prediction of the action–effect (Moore and Haggard 2008; Wolpe et al. 2013) and to the intention-to-act, i.e., voluntary action planning (Jo et al. 2014; Engbert and Wohlschläger 2007; Dogge et al. 2012). In our study (Jo et al. 2014), we administered four task conditions in separate experimental blocks, where subjects were either asked to judge the onset time of their voluntary action or the action–effect tone (Fig. 1). In three of the four conditions, the identical tone was applied, which elicited an N1–P2 complex. The N1–P2 complex differed depending on whether the tone was self-generated or externally generated. In the fourth control condition, no tone was applied. In the present study, we used a Bayesian approach to compare alternative model hypotheses for the generation of the auditory ERP responses. These alternative model hypotheses consisted of hierarchical neural networks across the auditory cortex and SMA. A subset of models with a prediction signal was included, in which a driving input enters the SMA, implementing a prediction process. We identified the optimal models of individual task conditions empirically and compared them across conditions. We hypothesized that models with the prediction signal to the SMA better explain the brain responses to the self-generated tone, whereas externally generated tones do not require the prediction.

Materials and methods

Subjects, task design, and data acquisition

Subjects, task design, data acquisition, and intentional binding effects of this study have been described in detail previously (Jo et al. 2014). Thirty-nine healthy adults (mean ± SD age, 40.5 ± 7.4 years; range of 25–50 years; 25 females) participated in the study, twenty of which have had

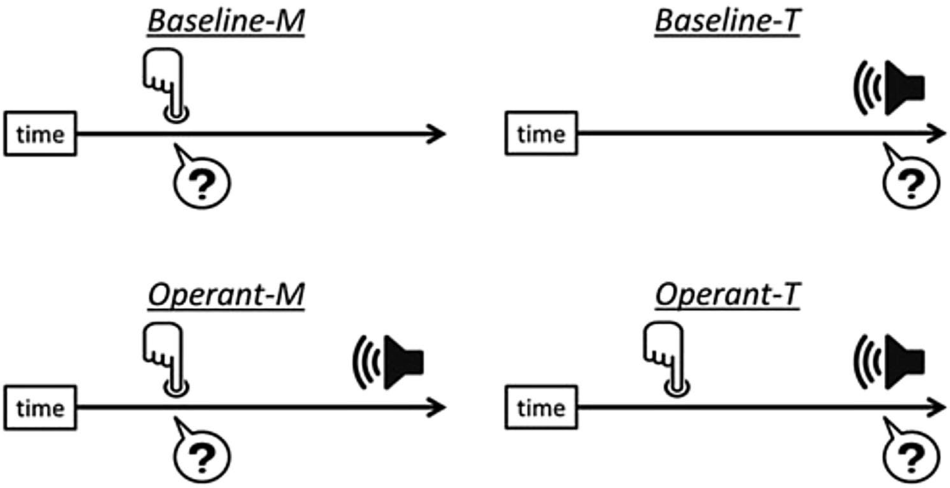

Fig. 1 Intentional binding paradigm. In four different task conditions, subjects reported the onset time of voluntary action (question mark pointing at the button press icon) or tone (speaker icon). In operant conditions (operant-M and operant-T), a voluntary action generates the tone 250 ms after a button press (modified from Jo et al. 2014)

previous experience in mindfulness meditation. No effect of meditation experience was found with respect to the intentional binding effects and ERPs. The task was the classic intentional binding paradigm (Fig. 1), in which subjects participated in a pseudo-random order in four blocks with different experimental conditions (baseline-M, operant-M, baseline-T, and operant-T). For all conditions, an analog clock (visual angle, 3° in diameter) was presented at the center of a monitor screen—a clock method adopted from the Libet task (Libet et al. 1983). A clock hand appeared after a short latency period of 1–2 s and started rotating clockwise with 2550 ms per cycle. Subjects were asked to focus on the center of the clock and to refrain from blinking during clock-hand rotations. In the baseline-M condition, subjects were asked to voluntarily press a button (pressing the left mouse button using their right index finger) at a time point of their own choice. After pressing the button, the clock hand continued rotating for a short interval of 1–2 s and then disappeared. Subjects were then asked to indicate with the mouse pointer the clock-hand position at the moment when they started to move their finger to press the button. The operant-M condition was identical to the baseline-M condition except that a 500 Hz-tone was presented for 100 ms following the button press after a delay of 250 ms. The operant-T condition was identical to the operant-M condition, but subjects were asked to indicate the onset time of the tone instead of the movement onset. In baseline-T, subjects performed no voluntary button press. Instead, the tone occurred at a random time between 2.6 and 7.7 s after the clock hand had started rotating. After the tone, the clock hand continued rotating for a short interval between 1 and 2 s and then disappeared. Subjects were then asked to indicate the clock-hand position at the time of tone onset. In short, in the operant-M and operant-T conditions, the tone was generated by a voluntary action (self-generated tone); while in the baseline-T condition, the tone was externally generated (externally generated tone). The operant-M and operant-T conditions contain the same voluntary action and its consequence effect, but differ in the reporting task. Each condition contained 40 trials and the mean waiting times (the time from the start of a trial to button press or tone onset) were 7.21 s for baseline-M, 6.98 s for operant-M, 7.06 s for operant-T, and 5.06 s for baseline-T.

With respect to the intentional binding effect (Jo et al. 2014), it has been shown that participants reported the occurrence of the tone significantly earlier in the operant-T condition (− 127.31 ms ± 19.4) as compared to the baseline-T condition (− 32.64 ms ± 11.7; paired t test, $p < 0.001$). In contrast, the reported time of the action did not differ significantly between the operant-M (− 42.94 ms ± 19.2) and baseline-M conditions (− 38.54 ms ± 24.6; two-tailed paired t test, $p = 0.74$). In addition, the reported time relative to the action onset differed significantly between the operant-M and the operant-T conditions (two-tailed paired t test, $p < 0.001$), suggesting that participants attended to task-relevant information, i.e., either to their voluntary action or to the tone, respectively.

EEG data were collected using a Quickamp amplifier and 64-channel active electrodes (Brain Products, Germany). A ground electrode was placed on participants' foreheads, and an initial reference was set at P9 according to the 10–20 system. Electrode impedance of all electrodes was kept under 5 kΩ. One channel electrooculography (EOG) was recorded to detect ocular artifacts. To estimate the onset of finger movement, a single-axis accelerometer (1.7 g) was placed on the left mouse button to measure the exact time of movement onset. All electrophysiological data were recorded at a sampling rate of 1000 Hz. Pre-processing of EEG data was performed with the help of EEGLAB (Delorme and Makeig 2004). EEG records were down-sampled to 250 Hz and re-referenced to linked mastoids. A band-pass filter from 0.01 to 45 Hz (zero-phase filter with − 6 dB cutoff) was applied. Continuous EEG data were then segmented into event-locked epochs ranging from 2.5 s before the event (either the onset of the button press or the tone) to 1 s after the event, with baseline corrections for the first 200 ms. Epochs affected by artifacts (± 100 μV) of any electrodes except ocular movement were excluded from further analyses. Remaining ocular artifact components were removed using independent component analysis (ICA). Trials in which the button was pressed during the first rotation of the clock hand were also excluded. On average, 92.7% (SD = 8.6) of all epochs were analyzed. ERPs were the averages of the epochs. They were computed separately for each condition and subject (see the supplementary figure 1 for grand-average ERPs).

Auditory ERP data analysis

We used EEGLAB to extract auditory ERPs. ERP responses were first referenced to common average. To isolate sound-evoked brain activity from movement-related activity present during the self-generated tone conditions (operant-M and operant-T), we subtracted the ERPs of the movement-only condition (baseline-M) from the ERPs of the operant-M and operant-T conditions separately for each subject (Timm et al. 2013; Ford et al. 2014). These motor-corrected responses, which are time-locked to the tone onset (auditory ERPs), were then compared to the responses in the tone-only condition (baseline-T). The responses were baseline-corrected using a time window of 100 ms prior to tone onset. The time course of auditory ERPs at electrode location FCz was entered as a dependent variable into a repeated-measure ANOVA with condition (operant-M, operant-T, and baseline-T) as a within-subject variable. It was applied for each time point from 0 to 400 ms (250 Hz) after the tone; the significance level was set at 0.001 to reduce the risk of false-positive error. Post hoc paired sample t tests were applied to determine the differences between the two conditions.

Source space analysis

Cortical sources of the auditory ERP responses were reconstructed using the source reconstruction toolbox implemented in SPM12 (ver. 7219). A forward-head model was computed with the boundary element method (BEM), using a standard T1 MRI template co-registered with EEG electrode coordinates. The cortical spatial resolution was set to the size of 8196 vertices and multiple sparse prior algorithm (Friston et al. 2008) was chosen to compute the source inversion. The time courses of five sources of interest (see below) were then extracted for each task condition between 0 and 400 ms after the tone onset. Here too, a repeated-measure ANOVA with condition (operant-M, operant-T, and baseline-T) as a within-subject variable was applied for each source and time point, and the significance level was set at 0.001.

Network modeling

We used dynamic causal modeling (DCM; David et al. 2006), a hypothesis-driven Bayesian approach, to estimate the generative model that best explains the observed auditory ERP. Based on a biophysically constrained neural-mass model of a cortical column (Jansen and Rit 1995; David and Friston 2003), DCM infers the generative mechanism behind the observed evoked potentials. The sources of auditory fields of the temporal lobe were approximated to previous neuroimaging studies (Opitz et al. 2001; Rademacher et al. 2002). The source coordinates were converted to the Montreal Neurological Institute (MNI) space using the algorithm described in http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach; bilateral primary auditory cortices (A1; [−42, −22, 7], [46, −14, 8]) and bilateral superior temporal gyri (STG; [−61, −32, 8], [59, −25, 8]). In addition, based on our hypothesis that the SMA plays a central role in the process of perceiving the action effect (Jo et al. 2014; Moore et al. 2010; Reznik et al. 2015), we added a source within the SMA (MNI coordinate: [2, 4, 54]). This source coordinate was taken from a previously published study (Lau et al. 2004) which examined the source of the intention-to-act using the same Libet task as the present study. Each source was modeled as a single equivalent current dipole (ECD) based on the forward-head model using the standard algorithm with default settings implemented in SPM12. The source locations of the dipoles were fixed, but allowed for variation in direction.

A model space (Fig. 2) was constructed to test alternative hypotheses of the mechanism underlying the auditory ERPs. An external sensory input representing the tone (sensory input) was entered bilaterally to A1 (model 1), which was forward connected to bilateral STG with or without backward connections (model 2 and 3, respectively). Bilateral STGs were connected forward to the SMA with or without backward connections as well as the backward connections between the STG and ipsilateral A1 (model 4–7). Furthermore, in addition to the conventional external sensory input, the SMA was chosen as a cortical input station for processing the prediction (prediction input). This internally generated prediction had the same functional response as the bottom-up external sensory input. The prediction input thus represented an internal expectation of the sensory input, enabling the brain to predict the action effects. From model 8 to 11, where the prediction input was entered in the SMA, the SMA was backward connected to the bilateral STG with or without forward connections as well as the backward connections between STG and ipsilateral A1. These resulted

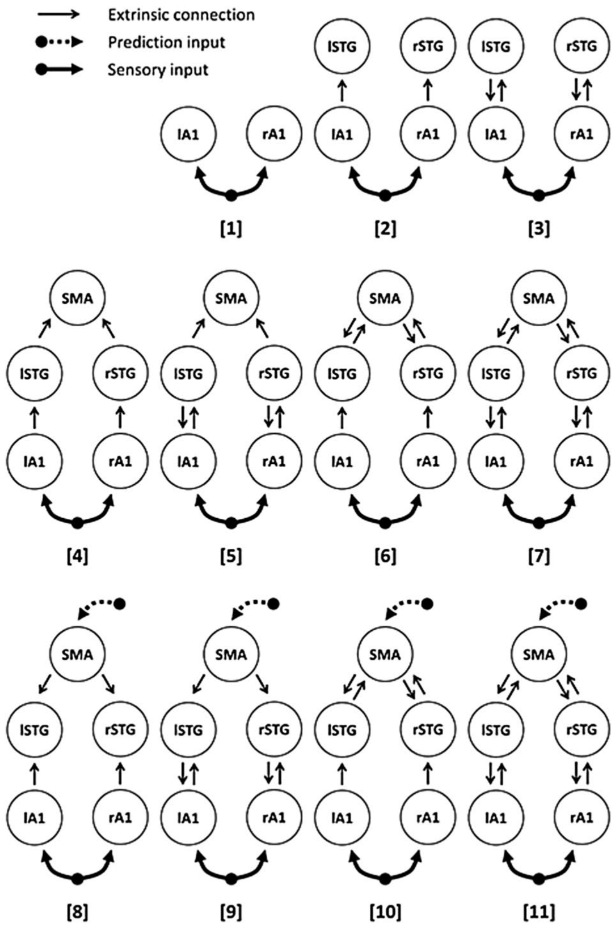

Fig. 2 DCM (dynamic causal modeling) model space with 11 different structures. Model 1 consists of sensory input into bilateral A1, followed by extrinsic connections to STG (model 2–3), connections to SMA (model 4–7), and the presence of prediction input into the SMA (model 8–11). *lA1* left primary auditory cortex, *rA1* right primary auditory cortex, *lSTG* left superior temporal gyrus, *rSTG* right superior temporal gyrus, *SMA* supplementary motor area

in 11 plausible models to be evaluated. Here, we hypothesized that a predictive process of the self-generated tone is initiated, which induces the SMA to implement high-level prediction effects on the auditory cortex, which then alters the N1–P2 complex. The time window of 0–400 ms upon tone onset, which encompasses the N1–P2 complex, was estimated to fit the models.

Model estimation and selection

The model evidence, an index of model fitness that accounts for both the accuracy and complexity of the model, was computed for the proposed 11 models (David et al. 2006). Conventionally, strong evidence in favor of one model requires three or more differences in a log scale of model evidence (ΔF). Posterior probability was calculated to demonstrate the probability of a model within the 11 models. For group-level inference, we used fixed-effects (FFX) analysis, assuming that the basic physiological mechanism of sensory attenuation is unlikely to vary across the subjects (Stephan et al. 2010). We also compared subsets of model space using family-based inference, which can quantify the probability that the presence versus the absence of a particular model structure improves model performance (Penny et al. 2010). We split the model space into three families with respect to the SMA effect: models without SMA (model 1–3; family 1), models with SMA (model 4–7; family 2), and models with SMA and prediction signal to the SMA (model 8–11; family 3).

Results

Auditory ERP

Figure 3 shows grand-averaged auditory ERP responses for the three conditions in which the same tone was presented. Across the three conditions, scalp topographical maps indicated the maximum amplitude of N1 and P2 components at electrode location FCz. The baseline-T condition showed the highest negative N1 and positive P2 amplitudes compared to the other two conditions, while the operant-M and operant-T conditions showed a similar pattern of ERPs. To illustrate the task effects over the time course of the ERPs, we applied a repeated-measure ANOVA for each time point with the condition (operant-M, operant-T, and baseline-T) as a within-subject variable. The analysis revealed a significant main effect over time for N1 and P2 components (between 68 and 124 ms and between 224 and 332 ms, $p < 0.001$; top panel in Fig. 3). Post hoc tests confirmed that this effect was

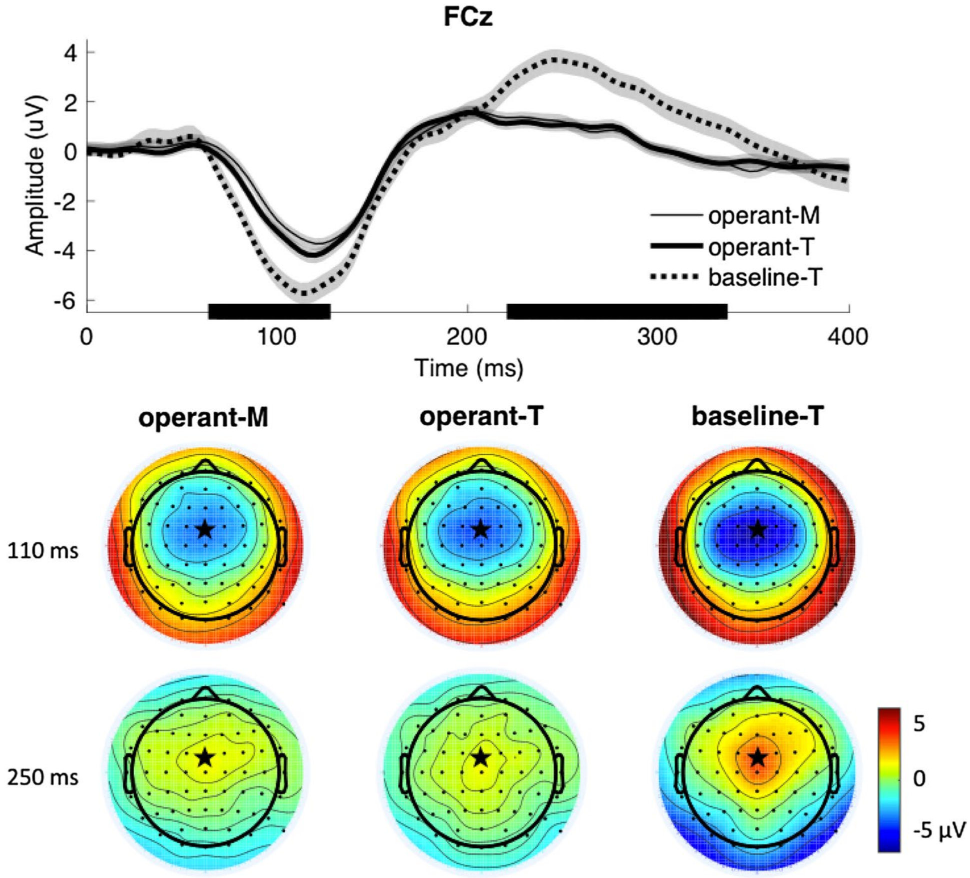

Fig. 3 Auditory event-related potentials (ERPs) in response to a tone. Top: the time courses of the grand-averaged ERPs for each of the three task conditions (operant-M, operant-T, and baseline-T) at mid-frontal electrode (FCz). Gray shades around the lines are standard error of the mean. The black bars on the *x*-axis indicate temporal ranges of significant difference between three task conditions (main effect $p < 0.001$). Bottom: scalp topographies of the N1 and P2 amplitudes at the time point of 110 and 250 ms, respectively. The asterisk marks the FCz electrode

due to the difference between the baseline-T condition and the other two conditions (i.e., baseline-T vs. operant-M and baseline-T vs. operant-T), rather than the difference between the operant-M and operant-T conditions (supplementary figure 2). Together, these results demonstrate that the N1–P2 complex was suppressed when the tone was self-generated as compared to an externally generated one.

Source analysis

We examined auditory responses at source level focusing on the regions of interest that were defined as informed priors of DCM analyses. As above, a repeated-measure ANOVA was applied for each time point and source. All selected sources demonstrated a significant main effect of condition ($p < 0.001$), spanning several temporal ranges (Fig. 4). This significant effect started around 68 ms after tone onset, covering the range of the N1 component. Each source activity also showed a significant main effect in the time range of the P2 component. This activity started at approx. 172 ms in the bilateral A1 and was present until 380 ms in the right STG. Although the left STG showed a similar task effect, it displayed the opposite polarity of a waveform as in the other source regions. Post hoc tests confirmed these effects when the baseline-T condition was compared with the operant-M and operant-T conditions, while there was no significant difference in the two conditions in which a tone was self-generated, i.e., between the operant-M and operant-T conditions.

Network structures

The auditory ERP responses for each condition and subject were fitted to the models in Fig. 2. Using Bayesian model selection (BMS) at the group-level FFX analysis (Stephan et al. 2010), we identified the winning model that best explains the observed ERP data. The model evidence and

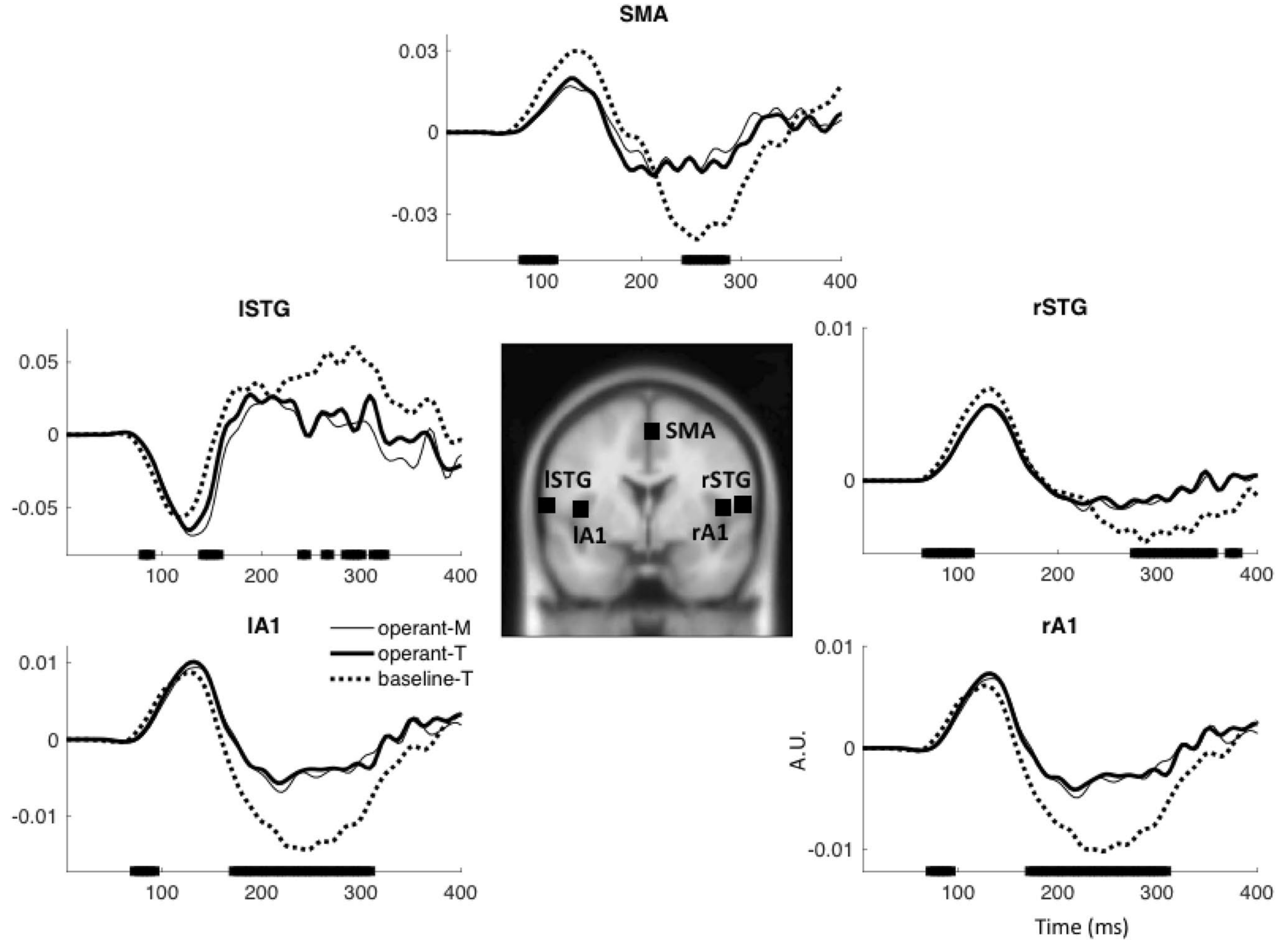

Fig. 4 Auditory cortical activity in response to a tone. Grand-averaged source responses are shown in five regions of interest, and their anatomical locations are illustrated in the center in a coronal plane. For each location, the black bars on the *x*-axis indicate temporal ranges of significant difference between three task conditions (main effect $p < 0.001$). *lA1* left primary auditory cortex, *rA1* right primary auditory cortex, *lSTG* left superior temporal gyrus, *rSTG* right superior temporal gyrus, *SMA* supplementary motor area

BMS results for the auditory ERPs are shown in Fig. 5 (see also the supplementary figure 3 for the model fit). The winning models of individual conditions were fully connected structures, including the presence of the SMA and bidirectional connections between A1 and the STG and between the STG and SMA. For the operant-T and operant-M conditions (self-generated tone), model 11, which included the SMA and the prediction input into the SMA, had the highest log-evidence. The difference between the winning model and the second-place model (ΔF) was 376.10 and 36.17 for the operant-M and operant-T conditions, respectively. In contrast, for the baseline-T condition (externally generated tone), model 7 was the winning model with $\Delta F = 700.20$, which included the SMA, but without prediction input. All winning models exceeded a posterior probability of 0.99 (middle panels in Fig. 5). We further compared subsets (families) of the model space, grouping models with respect to the SMA effect. Model family 3 (models 8–11), containing

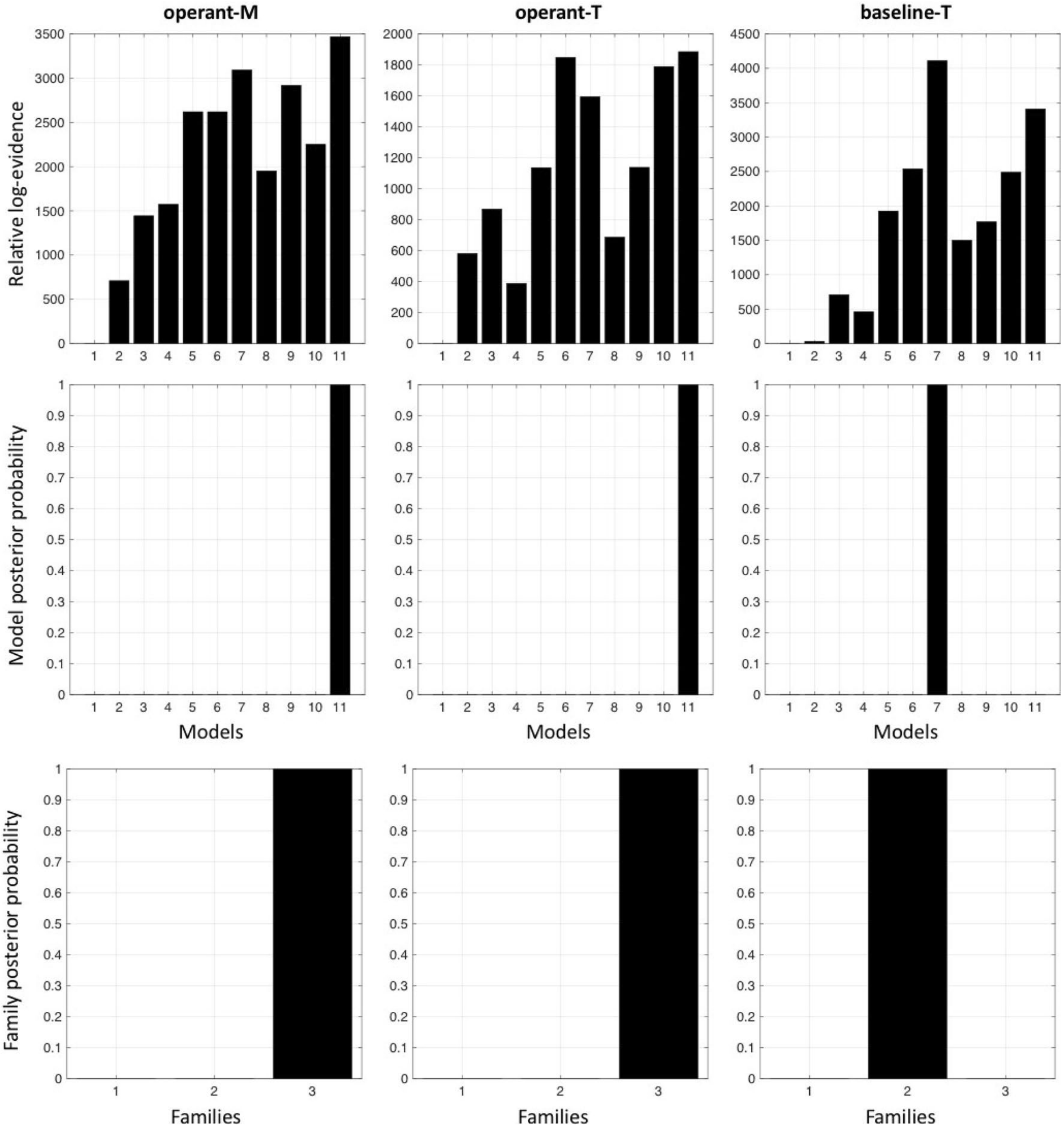

Fig. 5 Bayesian model selection results. Top and middle panels plot the relative log-evidence over the 11 models in Fig. 2 and their model posterior probability, respectively. Model 11 was the winning model of the operant-M and operant-T conditions (self-generated tone), whereas model 7 was the winning model of the baseline-T condition (externally generated tone). Bottom panels plot the family posterior probability. Eleven models were split into three families; models 1–3 without SMA (family 1), models 4–7 with SMA (family 2), and models 8–11 with SMA and prediction input into the SMA (family 3). Family 3 was the winning family of the operant-M and operant-T conditions, whereas family 2 was the winning family of the baseline-T condition

only models that include the SMA with the prediction input into the SMA, best explained auditory ERP responses in the operant-M and operant-T conditions (bottom panels in Fig. 5). For the baseline-T condition, the winning family 2 (models 4–7) also includes the SMA, but, without prediction input, best explained the observed ERP. All winning families exceeded a posterior probability of 0.99. In sum, across individual models and family-based inferences, auditory ERP responses in all conditions required the presence of the SMA. Importantly, both self-generated tone conditions, i.e., operant-T and operant-M, required a prediction input into the SMA, whereas this was not the case for the baseline-T condition.

Discussion

The intentional binding paradigm distinguished brain responses to auditory stimuli that were either the direct consequence of the subject's behavior (self-generated) or were presented without any relation to the subject's behavior (externally generated). It demonstrated suppressed brain responses to self-generated tones in comparison to the externally generated tones. Alternative network models for the generation of the auditory ERPs were examined by the use of DCM. Our model estimation accounted for the presence of a brain prediction signal to the SMA for self-generated tones (operant-M and operant-T), but not for the baseline-T condition where the tone was externally generated.

The SMA was found to play a crucial role in the development of the intention-to-act that is internally determined during voluntary action planning (e.g., Fried et al. 1991; Lau et al. 2004). Action planning is thought to involve an internal forward model that predicts the sensory consequence of the action and—as a result of this process—the sensory input is perceived as less intense (Wolpert and Ghahramani 2000). More recently, the intention-to-act rather than the movement execution itself has been shown to be essential for this attenuation (Timm et al. 2014). Specifically, within the auditory domain, growing evidence proposes the transfer of information between the SMA and auditory cortex (for a review, see Lima et al. 2016). The SMA is recruited during the processing of auditory information that may allow for a controlled generation of a sensory prediction. Based on this reasoning, we hypothesized that an internally generated prediction is implemented in the SMA and passed to the auditory cortex to modify the processing of the voluntary action effect. The observed source activity in the present study suggests the involvement of the SMA in the sensory attenuation, i.e., the difference in brain activity between self-generated and externally generated tones, and thus supported our hypothesis.

Over the decades, there has been remarkable headway in understanding auditory perception under hierarchical generative models, both theoretically and empirically. This progress rests mainly on a predictive coding scheme, which implies that the brain is constantly trying to minimize the discrepancy between external sensory inputs and internally generated predictions based on prior expectations (Friston 2005). In this view, the brain is a hierarchical system in which the prediction is made in the higher order cortical areas and updated via continuous interaction with the actual sensory inputs in the lower order cortical areas. This predictive coding view has been supported by empirical evidence resulting from the auditory mismatch negativity (MMN; Sams et al. 1985). A change in sensory input, i.e., the presentation of a deviant tone in a stream of repeated standard tones, exhibits a stronger negative ERP (MMN) than the standard tone. This MMN is considered a failure to predict the sensory input and consequently to suppress the discrepancy, the so-called prediction error (Garrido et al. 2009). This predictive coding view seems to hold in sensory attenuation, as well (Randeniya et al. 2018). The self-generated tone was predicted and thereby the prediction error was minimized through a prediction, whereas the externally generated tone was less predictable and thus the prediction error retained. According to this view, the change in N1–P2 amplitude from self-generated to externally generated tones would be a measurable prediction error in sensory attenuation, such as the MMN.

Using DCM, we were further able to test hierarchical generative models underlying the N1–P2 complex in which the prediction was modeled as a driving input into the SMA. The most likely network structure for the suppressed N1–P2 amplitude during a self-generated tone was the fully connected network with the presence of the prediction input, which would enable the network to predict auditory activity and consequently to suppress the activity. The top-down and bottom-up connections of the network suggest that the prediction process seems to be coordinated via reciprocal interaction between the SMA and auditory cortex. In contrast, the winning-model structure resisted the presence of the prediction input when the tone was externally generated and the network would retain a prediction error, as revealed in a larger N1–P2 amplitude. In line with this notion, internal prediction input in the higher order cortical area has been shown to be a key feature in explaining the prediction process of upcoming events. A network structure that includes a higher order prediction input provided a better model fit for the ERP response to the acoustic stimuli that violated a temporal prediction (Phillips et al. 2015), and the prediction was to specifically modulate downward connections, instead of forward, from higher to lower order cortical areas (Chennu et al. 2016). Notably, despite differences in the reporting task used in the present study, the winning-model structure was consistent across the operant-M and operant-T conditions. Hence, this observed stability across conditions cannot be attributed to the difference in attention to action (movement) vs. action effect (tone). Rather, our results appear to be robust with respect to variations in the self-generated tone conditions.

The supplementary motor area (SMA) is a classical term defined as a single cortical area located in the medial part of the Brodmann's area 6 (Penfield and Welch 1951). Recent anatomical and functional accounts suggest that the SMA comprises two distinct areas; the rostral area pre-SMA and the caudal area SMA proper (Nachev et al. 2008; Ruan et al. 2018). Although both regions are known to be involved in sensorimotor integration for movement generation, the pre-SMA has a stronger association with a higher order aspect of movements such as in procedural learning, cognitive control, and intentional movements, whereas the SMA proper is more concerned with movement execution (for a review, see Nachev et al. 2008). Further support of this notion is provided by functional connectivity patterns that dichotomize these two cortical areas. While the SMA proper was strongly co-activated with motor-related areas such as precentral gyrus and caudal dorsal premotor cortex, the pre-SMA was more co-activated with inferior frontal and posterior parietal cortices (Ruan et al. 2018). As the SMA source of the present models lies in the pre-SMA, our results support the notion that pre-SMA attributes to the integration of voluntary movement planning, rather than movement itself, and sensory processing of its consequent effect (Jo et al. 2014). Further studies including the pre-SMA and SMA proper into the model may prove useful in extending the findings of the present study.

Several limitations should be considered when interpreting the present findings. We aimed to examine whether the auditory ERP attenuation was attributed to a prediction process implemented in the SMA. Consequently, we used a model space confined to the SMA and auditory cortex. Therefore, the present model space provides a limited repertoire that the models with the prediction signal to the SMA better explain the brain responses to the self-generated tone, but not to the externally generated tone. One could potentially model left and right SMAs separately, but we modeled as a single source, in part because SMA activation during self-initiated movement was well observed as one cluster and because a priori coordinate of the potential left and right SMA could not be specified. Furthermore, we took the source coordinate from a previous study (Lau et al. 2004), in which they used the same Libet task as the present study and found a single SMA cluster. On the other hand, previous studies modeled the auditory interhemispheric pathway (lateral connection) and found evidence for better model fits for auditory paradigms (Boly et al. 2011; Phillips et al. 2015). However, to understand the prediction input to the SMA, this was not necessary. We believe that future studies could build on the present findings to establish the complete brain network underpinning the functional organization of auditory sensory attenuation. Another limitation was a possible confound of motor-related activity during the operant conditions. Although, in the context of the general linear model, the subtraction method of button alone condition from the button-tone condition has often been used to isolate the auditory ERP (e.g., Timm et al. 2013; Ford et al. 2014), we cannot conclusively rule out that ERP activity of button press might affect the N1–P2 complex. Alternatively, one could include the primary motor cortex as a cortical source into the model to examine the possible effects of the remaining motor-related activity. Further study with a careful modification of the task paradigm, for instance, including an operant condition with an unexpected tone, would help to dissociate the prediction process from the motor component and provide additional validation of the present finding.

To attenuate self-generated sensations, the brain needs a predictive signal that conveys the motor commands to the sensory cortex and thereby impacts the processing of the sensory input. The present findings suggest that brain prediction feeding into the SMA is a key network structure in explaining the suppressed brain responses to a self-generated tone. This prediction process seems to be coordinated via the reciprocal interaction between the SMA and auditory cortex.

Acknowledgements This work was supported by the International Research Training Group (IRTG 2150) of the German Research Foundation (DFG).

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of Ethics Committee of the University Medical Center Freiburg, Germany, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Conflict of interest The authors declare that they have no conflict of interest.

References

Baess P, Horváth J, Jacobsen T, Schröger E (2011) Selective suppression of self-initiated sounds in an auditory stream: an ERP study. Psychophysiology 48(9):1276–1283

Bäss P, Jacobsen T, Schröger E (2008) Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. Int J Psychophysiol 70(2):137–143

Blakemore SJ, Wolpert D, Frith C (2000) Why can't you tickle yourself? NeuroReport 11(11):11–16

Boly M, Garrido MI, Gosseries O, Bruno MA, Boveroux P, Schnakers C, Massimini M, Litvak V, Laureys S, Friston K (2011) Preserved feedforward but impaired top-down processes in the vegetative state. Science 332(6031):858–862

Chennu S, Noreika V, Gueorguiev D, Shtyrov Y, Bekinschtein TA, Henson R (2016) Silent expectations: dynamic causal modeling of cortical prediction and attention to sounds that weren't. J Neurosci 36(32):8305–8316

David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. Neuroimage 20(3):1743–1755

David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ (2006) Dynamic causal modeling of evoked responses in EEG and MEG. Neuroimage 30:1255–1272

Deiber MP, Honda M, Ibañez V, Sadato N, Hallett M (1999) Mesial motor areas in self-initiated versus externally triggered movements examined with fMRI: effect of movement type and rate. J Neurophysiol 81(6):3065–3077

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

Dogge M, Schaap M, Custers R, Wegner DM, Aarts H (2012) When moving without volition: implied self-causation enhances binding strength between involuntary actions and effects. Conscious Cogn 21(1):501–506

Engbert K, Wohlschläger A (2007) Intentions and expectations in temporal binding. Conscious Cogn 16(2):255–264

Ford JM, Palzes VA, Roach BJ, Mathalon DH (2014) Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. Schizophr Bull 40(4):804–812

Fried I, Katz A, McCarthy G, Sass KJ, Williamson P, Spencer SS, Spencer DD (1991) Functional organization of human supplementary motor cortex studied by electrical stimulation. J Neurosci 11(11):3656–3666

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360(1456):815–836

Friston K, Harrison L, Daunizeau J, Kiebel S, Phillips C, Trujillo-Barreto N, Henson R, Flandin G, Mattout J (2008) Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39(3):1104–1120

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol 120(3):453–463

Ghio M, Scharmach K, Bellebaum C (2018) ERP correlates of processing the auditory consequences of own versus observed actions. Psychophysiology 55(6):e13048

Haggard P, Clark S, Kalogeras J (2002) Voluntary action and conscious awareness. Nat Neurosci 5:382–385

Jansen BH, Rit VG (1995) Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern 73(4):357–366

Jo HG, Wittmann M, Hinterberger T, Schmidt S (2014) The readiness potential reflects intentional binding. Front Hum Neurosci 8:421

Lau HC, Rogers RD, Haggard P, Passingham RE (2004) Attention to intention. Science 303(5661):1208–1210

Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. Brain 106:623–642

Lima CF, Krishnan S, Scott SK (2016) Roles of supplementary motor areas in auditory processing and auditory Imagery. Trends Neurosci 39(8):527–542

Martikainen MH, Kaneko K, Hari R (2005) Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb Cortex 15(3):299–302

Moore J, Haggard P (2008) Awareness of action: inference and prediction. Conscious Cogn 17(1):136–144

Moore JW, Ruge D, Wenke D, Rothwell J, Haggard P (2010) Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the sense of agency. Proc Biol Sci 277(1693):2503–2509

Nachev P, Kennard C, Husain M (2008) Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci 9(11):856–869

Opitz B, Rinne T, Mecklinger A, von Cramon DY, Schröger E (2001) Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. Neuroimage 1:167–174

Penfield W, Welch K (1951) The supplementary motor area of the cerebral cortex; a clinical and experimental study. AMA Arch Neurol Psychiatry 66(3):289–317

Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP (2010) Comparing families of dynamic causal models. PLoS Comput Biol 6(3):e1000709

Phillips HN, Blenkmann A, Hughes LE, Bekinschtein TA, Rowe JB (2015) Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. J Neurosci 35(25):9255–9264

Poonian SK, McFadyen J, Ogden J, Cunnington R (2015) Implicit agency in observed actions: evidence for N1 suppression of tones caused by self-made and observed actions. J Cogn Neurosci 27(4):752–764

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2002) Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13(4):669–683

Randeniya R, Oestreich LKL, Garrido MI (2018) Sensory prediction errors in the continuum of psychosis. Schizophr Res 191:109–122

Reznik D, Ossmy O, Mukamel R (2015) Enhanced auditory evoked activity to self-generated sounds is mediated by primary and supplementary motor cortices. J Neurosci 35(5):2173–2180

Ruan J, Bludau S, Palomero-Gallagher N, Caspers S, Mohlberg H, Eickhoff SB, Seitz RJ, Amunts K (2018) Cytoarchitecture, probability maps, and functions of the human supplementary and pre-supplementary motor areas. Brain Struct Funct 223(9):4169–4186

Sams M, Paavilainen P, Alho K, Näätänen R (1985) Auditory frequency discrimination and event-related potentials. Electroencephalogr Clin Neurophysiol 6:437–448

Saupe K, Widmann A, Trujillo-Barreto NJ, Schröger E (2013) Sensorial suppression of self-generated sounds and its dependence on attention. Int J Psychophysiol 90:300–310

Schafer EW, Marcus MM (1973) Self-stimulation alters human sensory brain responses. Science 181(4095):175–177

Schmidt S, Jo HG, Wittmann M, Hinterberger T (2016) 'Catching the waves'—slow cortical potentials as moderator of voluntary action. Neurosci Biobehav Rev 68:639–650

Shibasaki H, Hallett M (2006) What is the Bereitschaftspotential? Clin Neurophysiol 117(11):2341–2356

Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ (2010) Ten simple rules for dynamic causal modeling. Neuroimage 49(4):3099–3109

Timm J, SanMiguel I, Saupe K, Schröger E (2013) The N1-suppression effect for self-initiated sounds is independent of attention. BMC Neurosci 14:2

Timm J, SanMiguel I, Keil J, Schröger E, Schönwiesner M (2014) Motor intention determines sensory attenuation of brain responses to self-initiated sounds. J Cogn Neurosci 26(7):1481–1489

Wolpe N, Haggard P, Siebner HR, Rowe JB (2013) Cue integration and the perception of action in intentional binding. Exp Brain Res 229(3):467–474

Wolpert DM, Ghahramani Z (2000) Computational principles of movement neuroscience. Nat Neurosci 3:1212–1217