Violation of identity-specific action-effect prediction increases pupil size and attenuates auditory event-related potentials at P2 latencies when action-effects are behaviorally relevant

Elisabeth Lindner, Andrea Desantis, Felicia Pei-Hsin Cheng, Alexander Gail

▶ **To cite this version:**

Elisabeth Lindner, Andrea Desantis, Felicia Pei-Hsin Cheng, Alexander Gail. Violation of identity-specific action-effect prediction increases pupil size and attenuates auditory event-related potentials at P2 latencies when action-effects are behaviorally relevant. NeuroImage, 2024, 297, 10.1016/j.neuroimage.2024.120717 . hal-04937339

HAL Id: hal-04937339
https://hal.science/hal-04937339v1

Submitted on 10 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

NeuroImage 297 (2024) 120717

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Violation of identity-specific action-effect prediction increases pupil size and attenuates auditory event-related potentials at P2 latencies when action-effects are behaviorally relevant

Elisabeth Lindner [a,b,*], Andrea Desantis [c,d,e], Felicia Pei-Hsin Cheng [a], Alexander Gail [a,b]

[a] Cognitive Neuroscience Laboratory, German Primate Center – Leibniz Institute for Primate Research, Kellnerweg 4, 37077, Germany
[b] Faculty of Biology and Psychology, Georg-August University, Wilhelm-Weber-Str. 2, 37073 Göttingen, Germany
[c] The French Aerospace Lab ONERA, Département Traitement de l'Information et Systèmes, BA 701 13661 Salon Cedex AIR, 13661 Salon-de-Provence, France
[d] Institut de Neurosciences de la Timone (UMR 7289), CNRS, Aix-Marseille Université, Faculté de Médecine, 27, boulevard Jean Moulin, F-13005 Marseille, France
[e] INCC - Integrative Neuroscience & Cognition Center UMR 8002, CNRS, Université de Paris, 45 Rue des Saint-Pères, F-75006 Paris, France

ARTICLE INFO

Keywords:
EEG
N1
P2
Sensory attenuation
Action-effect prediction
Pupil
Cancellation

ABSTRACT

Self-initiated sensory action effects are widely assumed to lead to less intense perception and reduced neural responses compared to externally triggered stimuli (sensory attenuation). However, it is unclear if sensory attenuation occurs in all cases of action-effect prediction. Specifically, when predicted action-effects are relevant to determine follow-up actions attenuation could be detrimental. We quantified auditory event-related potentials (ERP) in electroencephalography (EEG) when human participants created two-sound sequences by pressing two keys on a keyboard associated with different pitch, giving rise to identity-specific action-effect prediction after the first keypress. The first sound corresponded to (congruent) or violated (incongruent) the predicted pitch and was either relevant for the selection of the second keypress to correctly complete the sequence (Relevance) or irrelevant (Control Movement), or there was only one keypress and sound (Baseline). We found a diminished P2-timed ERP component in incongruent compared to congruent trials when the sound was relevant for the subsequent action. This effect of action-effect prediction was due to an ERP reduction for incongruent relevant sounds compared to incongruent irrelevant sounds at P2 latencies and correlated negatively with modulations of pupil dilation. Contrary to our expectation, we did not observe an N1 modulation by congruency in any condition. Attenuation of the N1 component seems absent for predicted identity-specific auditory action effects, while P2-timed ERPs as well as pupil size are sensitive to predictability, at least when action effects are relevant for the selection of the next action. Incongruent relevant stimuli thereby take a special place and seem to be subject to attentional modulations and error processing.

1. Introduction

When playing a musical instrument, specific actions such as pressing a piano key are connected to distinct sounds. If the practiced pianist is playing a melody and suddenly misses a key, the unexpected sound becomes rather salient compared to the intended correct sound. This phenomenon is thought to arise from our ability to form predictions on the sensory effects of our own actions (Blakemore et al., 1998). Here we ask if and how modulations of the sensory processing of action effects due to predictability are affected by the behavioral relevance of the action effect.

The underlying concept of internal forward models within the framework of a comparator model suggests that when we plan and execute a voluntary action (like pressing a key), a copy of the motor command, known as the efference copy, is used to form a prediction of the sensory consequences of our action in terms of both, time and identity of the sound resulting from the keypress (Crapse and Sommer, 2008; Miall and Wolpert, 1996; Sperry, 1950; von Holst and Mittelstaedt, 1950; Wolpert, 1997). This prediction is then compared to the incoming sensory signals (the actual sound) and the predictable component is subtracted from the incoming sensory input, leading to the perceptual reduction of predicted action effects (Miall and Wolpert, 1996).

Sensory attenuation has not only been observed when predicting the

* Corresponding author at: Cognitive Neuroscience Laboratory, German Primate Center – Leibniz Institute for Primate Research, Göttingen, Germany.
E-mail address: elindner@dpz.eu (E. Lindner).

https://doi.org/10.1016/j.neuroimage.2024.120717
Received 14 February 2024; Received in revised form 20 June 2024; Accepted 1 July 2024
Available online 5 July 2024
1053-8119/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

identity of sensory action effects (Hughes et al., 2013a; Kühn et al., 2011), but especially when comparing self-initiated action effects to externally generated stimuli, highlighting the distinct role of self-generated movement and prediction (Asimakidou et al., 2022; Bays et al., 2006; Kilteni and Ehrsson, 2022; Kilteni et al., 2019; Sato, 2009). Such mechanism to support discrimination between internally and externally generated actions is suggested to be an important basis for the formation of a coherent sense of agency, i.e., the attribution of having caused one's own motor acts and their effects in the external world (Haggard and Tsakiris, 2009). A link between sensory attenuation and the sense of agency has been proposed (Baess et al., 2011, 2008; Ford and Mathalon, 2012; Sato and Yasuda, 2005; Weiss et al., 2011).

Sensory attenuation has not only been demonstrated perceptually, but also at the neural level. Using electroencephalography (EEG) sensory attenuation in the auditory modality has been observed specifically in the N1 and P2 component of the event-related potential (ERP; Baess et al., 2011; Hughes et al., 2013a; Mifsud, Oestreich, et al., 2016; SanMiguel, Todd, et al., 2013; Timm, SanMiguel, Saupe, and Schröger, 2013). The N1 component is associated with automatic processing, the encoding of simple physical features, and an orienting towards a sensory event (Hofmann-Shen et al., 2020; Korka et al., 2022; Näätänen and Picton, 1987; SanMiguel et al., 2013), so that a reduced N1 amplitude can be seen as indicator for attenuated early sensory processing.

Attenuation in N1 in the auditory modality has primarily been demonstrated for self-initiated compared to externally produced sounds (Asimakidou et al., 2022; Bays et al., 2006; Kilteni and Ehrsson, 2022; Kilteni et al., 2019; Sato, 2009). Assuming that motor prediction is responsible for sensory attenuation in this case, attenuation should also occur when comparing self-initiated predicted with self-initiated unpredicted sounds (Hughes et al., 2013a), i.e., when the sound occurrence per se fulfills the prediction, but its quality ("identity") does not, e.g., due to an incongruent pitch. An attenuation in N1 based on identity-specific motor prediction has indeed been demonstrated (Hughes et al., 2013a; Kühn et al., 2011), while more recent studies did not find identity-specific attenuation (Chung et al., 2022; Korka et al., 2019; Le Bars et al., 2019). Instead, modulations in the P2 or P2-timed components related to identity-specific action-effect prediction have been reported (Chung et al., 2022; Hsu et al., 2015; Jones et al., 2013). The P2 component or P2-timed components, i.e., changes in ERP at P2 latencies independent of a confirmed topographic P2 characteristic, have amongst others been suggested to reflect a comparison between sensory inputs and internal predictions (Chung et al., 2022; Garrido et al., 2009; Hughes and Waszak, 2014; Pinheiro et al., 2019), where both enhancements and attenuations have been observed in response to violated predictions.

Previous studies on sensory attenuation have focused on situations where participants perform an action that generated a sensory outcome unrelated to subsequent actions (Kiepe et al., 2021). However, in real-life scenarios, like the piano example, actions and their outcomes are often embedded in chains of movements and the specific effects can become relevant for the subsequent action. In case of behavioral relevance, attenuation of an action-effect could be counterproductive. The purpose of this study is to investigate whether sensory attenuation occurs in all cases of action-effect prediction, including cases when action effects are relevant for the choice of a subsequent action. We hypothesize that if the action effect is relevant, sensory attenuation may diminish or be reversed, with predicted action effects potentially being enhanced compared to unpredicted action effects (Kok et al., 2012; Saupe et al., 2013; Schröger, Marzecová, and SanMiguel, 2015; Timm et al., 2013). We expect that with behavioral relevance of the action effect for future actions, mechanisms of attention, response cancellation or motor-goal updating may interact with mechanisms of prediction and sensory attenuation. In that case, the forward model may not be sufficient to explain the mechanism between action-effect prediction and sensory processing. Previous studies have yielded mixed findings regarding the relationship between attention and sensory attenuation (Kok et al., 2012; Saupe et al., 2013; Schröger, Marzecová, and SanMiguel, 2015; Timm et al., 2013), therefore it remains unclear how prediction-based attenuation might be modified by behavioral relevance. For instance, some authors suggest that sensory predictions and voluntary selective attention affect the auditory N1 via distinct processes in an additive way (Timm et al., 2013; Saupe et al., 2013). Meanwhile, Kok and colleagues (2012) propose that attention leads to a reversal of sensory attenuation by increasing the precision of the prediction and thereby increasing the weighting of sensory data of predicted sounds, which results in an increased response to attended, predicted stimuli.

We further assumed that action-effect prediction and potentially associated sensory attenuation should be reflected in pupil dilation responses, as prediction error for prediction-incongruent sounds may lead to higher arousal and consequently larger pupil size. Arousal has been identified as a key driver of pupil dilation response (Wang et al., 2018). An attenuation effect of prediction may therefore be visible in pupil size. Indeed, deviant auditory stimuli have been reported to be associated with an elevated pupil dilation response (Bianco et al., 2020; Kamp and Donchin, 2014; Liao et al., 2016; Wetzel et al., 2016). Given that pupil size should serve as an indicator of prediction error, we expect it to (anti-)correlate with modulations in those ERP components that depend on action-effect prediction.

Based on previous research regarding prediction, prediction error, attention and motor inhibition, we also exploratively analyzed the P3 as a later component (Bokura et al., 2001; Escera et al., 1998; Ford et al., 2010; Jones et al., 2013; Patel and Azzam, 2005; Polich, 2007; Randall and Smith, 2011; Schröger, Marzecová, and SanMiguel, 2015; Smith et al., 2007).

To test the effects of action-effect prediction on different ERP components, we developed an auditory paradigm using identity-specific action-effect prediction. We ensured an equal overall attribution of attention to prediction-congruent and -incongruent sounds, in contrast to designs comparing actively created to passively perceived sounds (Schröger, Marzecová, and Sanmiguel, 2015). Participants were asked to create sounds by pressing specific keys on a keyboard. Sounds could either be prediction-congruent or prediction-incongruent. In different conditions participants were invited to create two-sound sequences by pressing the respective keys. The first sound of the sequence could either be relevant or irrelevant in order to plan and initiate the follow-up sound. Sensory attenuation was assessed using EEG, by extracting the event-related potentials to the sounds. We expected an attenuation in the N1 and P2-timed event-related components for predicted compared to unpredicted sounds, as previously reported by others (Baess et al., 2011; Hughes et al., 2013a; Mifsud, Oestreich, et al., 2016; SanMiguel, Todd, et al., 2013; Timm et al., 2013). We hypothesized, that this attenuation diminishes, disappears or even reverses into an enhancement for task-relevant predicted action-effects.

2. Materials and methods

2.1. Participants

Thirty volunteers participated in our study with an allowance of 8.50 Euro per hour. The sample size was based on similar studies stating identity-specific sensory attenuation in N1 (Hughes et al., 2013a; Kühn et al., 2011). Data of five participants was excluded from the analysis, due to technical issues regarding the eye tracking (three participants) or EEG recording (two participants). The average age of the remaining 25 participants (12 males, 2 left-handed) was 27.32 (range 19 - 40, SD = 6.20) years. Prior to the experiment, all participants provided written informed consent and confirmed to have normal or corrected vision and hearing and no history of neurological or psychiatric illness. They confirmed not to have taken drugs or medication prior to the experiment or to have an alcohol dependency. Participants with experience in piano playing were excluded from the study during recruitment to avoid bias

in the association between actions and sounds (Rusconi et al., 2005, 2006). The methods and purpose of the study was approved by the local ethics committee of the Georg-Elias-Mueller-Institute for Psychology, University of Goettingen and adhered to the principles of the Declaration of Helsinki.

2.2. Materials and stimuli

Data collection was performed in a dark, sound-attenuated, and electromagnetically shielded chamber (Desone Modulare Akustik, Berlin, Germany). The task was controlled with a custom-written Matlab program (version R2015b) using the Psychophysics toolbox (version 3.0.14; Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) on a 22.5-inch monitor (ViewPixx/EEG Inc., resolution=1440 × 980 pixels, refresh rate = 120 Hz). Participants were seated with a distance of 91 cm from the screen with their heads resting on a chinrest. Auditory stimuli were displayed using on-ear headphones (Sennheiser HD100) and consisted of pure tones at 1000 Hz (low-pitched sound) and at 2000 Hz (high--pitched sound) with 50 ms duration (including a ramp-up and ramp-down of 5 ms each) at 73 db SPL. Sound intensity in db SPL was calibrated prior to experimental testing using an artificial ear (artificial ear: Brüel & Kjær, Copenhagen – Type 4152; precision sound level meter: Brüel & Kjær, Copenhagen – Type 2203, octave filter set: Brüel & Kjær, Copenhagen – Type 1613). Visual stimuli were depicted on a mid-grey background and included a fixation point (empty white circle with a diameter of 0.72° visual angle (VA)), a go-signal (empty white square, 0.72° VA side length), a second go-cue which included a second smaller empty white square inside the previous square (0.5° VA side length) and a fixation which indicated the possibility to blink (empty green circle, 0.72° VA).

An EyeLink 1000 eye tracker system (SR Research, Ontario, Canada) in a desktop mount configuration was used to record the movements of the right eye at a sample rate of 1000 Hz.

Electrophysiological data were recorded at 1000 Hz from 64 active electrodes placed on an elastic cap, amplified by an actiCHanp amplifier (BrainVision Recorder 1.23.0001 and actiCap, Brain Products GmbH, Gilching, Germany). Arrangement of the electrodes corresponded to the international 10–20 system. Two electrodes were placed on the mastoids (M1 and M2, respectively). Online reference was the right mastoid (M2). Two other electrodes (TP9 and TP10) were used to record horizontal eye movements (left and right HEOG, respectively). All data used in this paper are available at https://doi.org/10.25625/MV3NIM.

2.3. Procedure

The experiment consisted of two sessions on two different days with three experimental parts in total. Two parts were performed on day one and the third on the other day. Each part tested a different condition: The Baseline condition, the Relevance condition and the Control Movement condition (details see below). The order of the conditions was counterbalanced across participants. Each condition was divided into blocks that comprised a learning phase of 24 trials and a test phase of 12 trials per block. Each block started with the learning phase, followed by the test phase. Participants were informed about the existence of two phases without revealing the statistical structure of congruent and incongruent sounds before the experiment. The first block of each condition comprised 100 learning trials instead of 24 to consolidate the learning of the specific action-sound association and 12 test trials. Each condition ended upon completing the entire block, at which point 360 correct trials in the test phases were achieved. Trials in which the participant pressed the key too early, too late (first keypress), created an incorrect sequence, blinked or deviated their gaze from the middle of the screen after the first sound, were aborted and repeated at the end of the experimental condition within a new block. Prior to each part of the experiment, participants completed a training, after reading the specific instructions, to ensure understanding of the task.

2.3.1. The learning phase

In the learning phase participants learned an association between two specific actions and their specific auditory consequences. In each trial participants were asked to fixate on a fixation circle, which was presented in the middle of the screen (Fig. 1A). They were asked to repeatedly press a left or a right key on the computer keyboard (letters ‘Q’ and ‘P’, respectively) with their left and right index finger, randomly and about equally often. In mapping 1, a left button press was associated with a high-pitched sound and a right keypress with a low-pitched sound. In mapping 2, a left keypress generated a low-pitched sound and a right keypress a high-pitched sound. These mappings were counterbalanced across participants. Catch trials were included to ensure that participants were attending to the task. In catch trials, a keypress resulted in the sound associated with the other keypress. In that case, for mapping 2 for instance, a left keypress resulted in a high-pitched sound. Participants were asked to count the appearance of these catch trials and report the number at the end of each learning phase. Feedback on the correct number was depicted on the screen. The number of catch trials in each learning phase varied randomly between 0 %, 5 %, 10 %, or 15 % of all trials (trial number rounded for given block). Before reporting on the number of catch trials, participants received feedback on the screen on the proportion of left and right keypresses that they produced and were asked to try to press equally often with the left and right index finger.

2.3.2. The test phase

Following each learning phase, a corresponding test phase was initiated, as specified on the screen. Additionally, participants were presented with information regarding the block number and the number of correct test trials completed thus far, relative to the total number of required correct trials for the specific experimental condition. Participants were asked to start the phases by pressing the space bar. A black empty fixation circle was displayed for 500 ms and then gradually disappeared while a white empty square started gradually appearing around the circle (Fig. 1B). The disappearance of the circle and appearance of the square took one second. As soon as the white square was fully white and the circle completely disappeared the participant had a 500 ms window to perform a left (“Q”) or right (“P”) keypress. If the participant pressed the key too early or too late, this was depicted on the screen and the trial was aborted. 400 ms after a successful keypress one of two sounds was played for 50 ms. The 400 ms interval was selected so that action-related and effect-related neural transient activity would not interfere (Hughes et al., 2013a). Previous research stating sensory attenuation implemented this interval between action and stimulus (Hughes et al., 2013a; Hughes and Waszak, 2011). In a ‘predicted’ case, the sound appeared which was associated with this specific keypress as learned in the learning phase. In the ‘unpredicted’ case, the sound associated with the other keypress was played. In this latter case, the prediction of the sound was violated as the action-effect was incongruent with the learned association. The predicted or the unpredicted sound appeared with an even chance in these short test phases. The empty white square remained at the screen for another 1.5 s, before the trial continued in a way that depended on the task conditions (see below).

Although predicted and unpredicted sounds were presented with equal probability in the test phase, unpredicted sounds after the first keypress overall were only presented with a probability of 22 % over the course of the experiment and in each block. Interleaving of learning between short test phases ensured that participants refresh and maintain expectations about which sound follows which keypress.

Before the start of the experiment, participants were instructed to keep fixating the middle of the screen as indicated by the circle and the square and to blink and move as little as possible. The inter-trial time could be used to blink or close the eyes for a brief moment. Further, participants were asked to initiate the sounds roughly equally often and without pattern by pressing left and right keys.

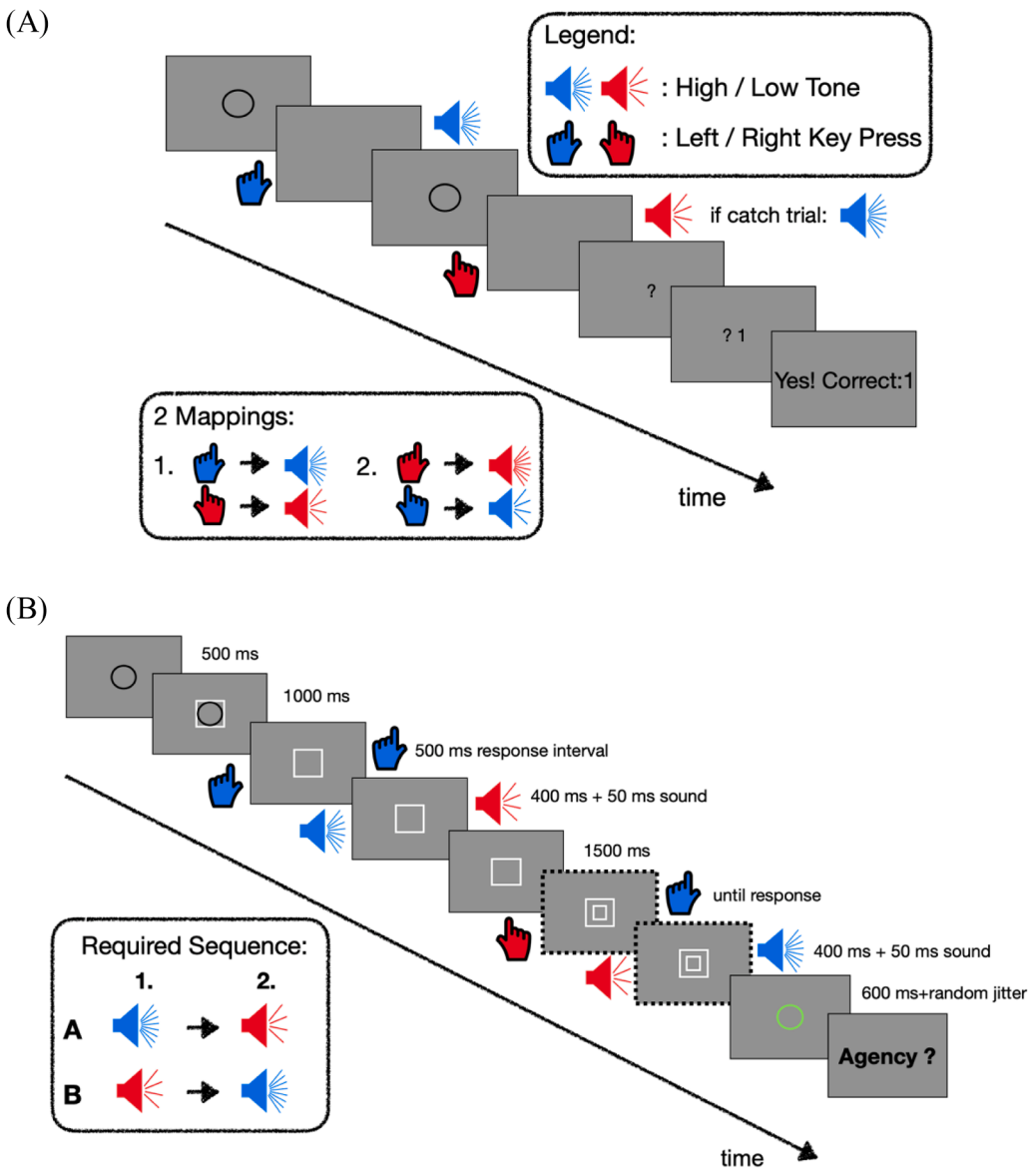

Fig. 1. (A): Learning Phase. In the learning phase participants learn an association between two keypresses and two sounds. Participants are asked to press specific keys with their right or left index finger and the respective sound appears. At the end of each block, participants report the numbers of catch trials. (B): Test Phase: The test phase of the Relevance condition is depicted with a congruent trial on the bottom and incongruent on the top, for the example of mapping 1. After a gradually appearing go cue, participants press the left or right key. Based on the learned contingency, the incongruent or congruent sound appears (here mapping 1). When a second square is displayed, participants were asked to produce a second sound. The second keypress and sound should align, so that tone sequence A or B is fulfilled. At the end of each block participants rate their sense of agency. The depicted trial outline without the dotted frames refers to a Baseline trial. Note: Illustration of the visual stimuli not to scale for illustration purposes.

In the Baseline condition, the empty square was followed by a green circle for 600 ms plus a random jitter between 20 and 60 ms to avoid phase locking of brain activity (Luck, 2014). After each block, participants received written feedback on the screen on the percentage of left and right keypresses in this block and were asked to initiate each sequence equally often with the left or right index finger. By pressing the space bar, participants were then invited to answer the question 'How strongly do you feel that you caused the tones' on a scale from 1 to 7 from "Not at all" to "Completely". The question and the scale, including the definition of its endpoints, were visible on the screen. Participants could type the respective number on the keyboard and continue with the next trial by pressing 'Enter'. These block-wise ratings on the sense of agency were included to test for a potential influence of condition on the sense of agency.

To ensure that participants attended to the task, 20 % of the test trials were presented as catch trials wherein the square in the middle of the screen turned fully white after the sound was played. While in the learning phase participants were asked to count violations of key-sound association, here they were requested to indicate which sound they had just heard by pressing the respective key as associated in the learning phase. In case of mapping 1, if a low sound was played the left key would be the correct response and in mapping 2 the right key. If the catch trial was correct, the fully white square turned green and red if the response was incorrect. The trial stopped afterwards. Correct indication of the sounds in catch trials was taken as evidence for attentive task performance.

The Relevance condition aimed at testing our main hypothesis, whether relevance of the action-effect for a follow up action would result in a reduction or disappearance of sensory attenuation of the predicted action-effect. In order to make the sound relevant, participants were

asked to create sequences of sounds, either low sound → high sound or high sound → low sound. Instead of the trial ending after the keypress and the appearance of the sound as in the Baseline condition, 1.5 s after the first sound of each trial a second smaller white empty square appeared within the first empty white square. This second square served as a cue to initiate the second keypress of the trial to create a successful sequence of sounds. After onset of this second go cue participants could freely time their keypress. The second sound corresponded to the key-sound association learned in the learning phase. In case of mapping 1, if the first sound was a predicted low-pitched sound, corresponding to a left keypress, the second sound had to be a high-pitched sound caused by a right keypress. If the first sound was an unpredicted high-pitched sound, the second sound needed to be a low-pitched sound, corresponding to a left keypress. This means, in the unpredicted case, participants were asked to press the same key twice. If the sequence was not completed successfully, an error message appeared on the screen, no sound appeared and the trial was aborted. If the trial was successful, it ended as in the Baseline condition with a green circle, signaling the possibility of eye blinks.

The Control Movement condition was implemented to control for differences between Relevance and Baseline condition in response to the first sound. These differences could arise from the additional key press and subsequent sound in the Relevance condition. In contrast to the Relevance condition, in the Control Movement condition participants were not required to create specific sound sequences; the identity of the sound resulting from the initial keypress held no behavioral relevance. Participants could freely choose, which key to press in order to cause a second sound. However, the timing of the required physical action and produced sound remained consistent between Relevance and Control conditions.

4. Behavioral, EEG and pupil analysis

4.1. Behavioral analysis

4.1.1. Distribution of keypresses

To assess whether key-sound contingencies were equally well acquired for left and right keypresses, the distribution of keypresses in the learning phases were compared using Wilcoxon Signed Rank Tests for each condition. The same analysis was performed on the distribution of keypresses in the test phases. Additionally, we conducted Wilcoxon Signed Rank Tests to compare the number of left and right keypresses for both the initiation of a sequence and the second keypress in the test phases for the Relevance and the Control Movement conditions. None of the comparisons yielded significance. These results suggest that key-sound contingencies likely were equally acquired for left and right keypresses.

To further compare whether the distribution of left and right keypresses differed between conditions we fitted two Generalized Linear Mixed Models with binomial error distribution and logit link function, one for the learning and one for the testing phases. The response in both models was a two-columns matrix with the number of keypresses to the left and right, respectively (Baayen, 2008). Hence, the models practically evaluated the proportion of left versus right keypresses. The models included a fixed effect of Condition, a random intercepts effect of Keypress side, and random slopes for Condition within participant. In both models, we did not find an effect of Condition (likelihood ratio test, learning phases: LRT = 0.06, df = 2, $p = 0.97$; test phases: LRT = 2.28, df = 2, $p = 0.32$). Hence, we dropped the fixed effect of Condition from both models and refitted the models. The intercept in the resulting models then informs about the probability of a right keypress. In the model we fitted to the learning data we obtained a very slight albeit significant preference for right keypresses (estimate ± SE = 0.02 ± 0.01, $z = 2.23$, $p = 0.03$). This equates to an estimated proportion of 50.5% right keypresses. For the data obtained in the test phase, we did not observe a significant side preference (0.03 ± 0.03, $z = 1.14$, $p = 0.26$).

We further assessed the frequency of incorrect sequences created by participants in the Relevance condition. On average, 3.61% ± 2.88% of trials containing a second keypress (non-catch trials) resulted in incorrect sequences (absolute average: 14 ± 11.74).

4.1.2. Catch trials

To ensure that participants acquired the association between the specific keypresses and the respective sounds and that they attended to the identity of the sounds, we report the number of detected catch trials in the learning phase and the number of correct catch trials in the test phases for each condition. Participants detected on average 90% of the catch trials in the learning phase (Baseline: 90.84 % ± 7 %; Relevance: 90.04 % ± 8.3 %; Control Movement: 90.04 % ± 8.3 %). In the test phase participants attained around 94 % correct catch trials (Baseline: 95.61 % ± 4.86 %; Relevance: 92.91 ± 8.90 %; Control Movement condition: 92.58 ± 6.55 %). These results indicate that participants were paying attention to the identity of the sounds in all conditions and acquired their association with their specific keys.

4.1.3. Agency ratings

To estimate the extent to which agency ratings differed between conditions, we fitted an ordinal (i.e., cumulative logit link) mixed model (Agresti, 2002). We included condition (i.e., Baseline, Relevance, Control Movement) as fixed factor and Session order as a control variable. Participant ID was again added as a random intercept effect and random slopes of Condition and Session order within Participant were included. We compared the full model with a null model, which comprised the same structure as the full model, but without the fixed effect Condition. The model was fitted in R using the function clmm of the package ordinal (version 2019.12–10; Rune Haubo Bojesen Christensen, 2019). No collinearity was observed (maximum VIF: 1.67). The model revealed a good stability, based on the previously described model stability test. By dichotomizing the agency ratings, fitting logistic models with the rated agency levels, and examining the model estimates, we verified whether the assumption of proportional odds was fulfilled. The results suggest that this assumption was not fully fulfilled, as the estimates differed between the conditions. This suggests that results of our model have to be taken with caution. The full-null model comparison did not yield a significant finding, indicating that the factor Condition did not have a significant effect on agency ($\chi^2 = 0.37$, df = 2, $p = 0.83$; see Supplementary Material S2 for statistical tables of all models). Participants indicated an agency rating with a mean of 3.76 ± 1.82, 3.74 ± 1.74 and 3.61 ± 1.73 for the test phases of Baseline, Relevance and Control Movement condition, respectively.

4.2. EEG analyses

We referenced our data to the average left and right mastoids (M1 and M2 respectively) and resampled our data to 500 Hz. A causal Butterworth filter was applied with a low pass filter of 80 Hz and a high pass filter of 0.01 Hz. Epochs were aligned and segmented around the first (or only) sound of each trial ranging from −200 ms before sound onset to 1500 ms after sound onset, excluding catch and error trials. Baseline level activity was then subtracted (average within the −200 ms to 0 ms time window prior to sound onset). Trials with eye blinks and deviations from fixation (outside of a radius of 5° visual angle around fixation), as signified by the eye tracker and behavioral error trials, were removed. Epochs that contained amplitudes greater than 100 μV or less than −100 μV were marked as potential artefacts and removed after confirmation through visual inspection. In sum, there were 228 (115 congruent, 113 incongruent), 338 (186 congruent, 152 incongruent), 305 (155 congruent, 150 incongruent) trials removed in this manual artefact rejection in the Baseline, Relevance and Control Movement condition respectively. This accounts to 0.03 %, 0.04 %, 0.03 % of total trials, respectively. The same “clean” trials were used for all analyses (EEG, pupil size, behavioral). An independent component analysis was

performed in order to identify and remove remaining eye movement components. The preprocessing was performed individually for the conditions Baseline, Relevance and Control Movement.

4.2.1. N1

To investigate whether the relevance of the action-effect (i.e., a sound) for a follow-up action interferes with the sensory attenuation of the predicted versus unpredicted sound, we first focused our analysis on the auditory N1 component. The auditory N1 component is associated to an early cortical reaction towards auditory stimulation (Kaiser and Schütz-Bosbach, 2018). Based on previous findings, we extracted the N1 component from frontal-central electrodes (Dogge et al., 2019; Hughes et al., 2013a; Korka et al., 2019). We analyzed the average N1 from F1, Fz, F2, FC1, FCz, FC2, C1, Cz and C2. We identified the N1 component by extracting a population peak in a time window of 80–120 ms. For this, the signal was averaged across our nine electrodes, all participants, and the conditions congruent and incongruent. Average N1 amplitudes around the peak ± 20 ms were computed for every participant to extract individual N1 values independently for the condition Baseline, Relevance and Control Movement (peak identical in all three conditions: 108 ms, range 88–128 ms).

To test our hypothesis whether sensory attenuation in N1 is reduced by relevance of the action-effect, we fitted linear mixed models including Congruency (i.e., Congruent, Incongruent) and Condition (i.e., Baseline, Relevance, Control Movement) and their interaction as fixed factors and Session order as a factorial control variable in order to control for potential effects of fatigue (see Supplementary Material S2 for all formulas of models used in this article). We included Participant as random intercept. In order to avoid overconfidence of the model with regard to the precision of fixed effects estimates and to keep type I error rate low, we further added random slopes of Congruency and Condition within Participant (Barr et al., 2013; Schielzeth and Forstmeier, 2009). Originally, we also included correlations among random intercepts and slopes (Barr et al., 2013). As these values contained values close to one, which indicated that they are not identifiable, these correlations were removed. This led to only a minor deterioration in model fit (log-likelihoods, model with correlation parameters included: −227.60; model without correlation parameters: −229.75). To overall test the fixed effects of Congruency and Condition and their interaction, a full-null model comparison based on a likelihood ratio test was conducted (Dobson, 2002; Forstmeier and Schielzeth, 2011) in order to avoid multiple testing. The null model structure was identical to the full-model, besides lacking the fixed effects of Congruency and Condition and their interaction. The effect of individual fixed effects was tested using the Satterthwaite approximation (Luke, 2017), with the function lmer of the R package lmerTest (Kuznetsova et al., 2017) and a model fitted with restricted maximum-likelihood. As the interaction Congruency and Condition did not yield a significant effect on N1, we also fitted a reduced model without this interaction (see Results section).

After we fitted the full model, assumptions of normally distributed and homogeneous residuals were inspected visually with QQ-plots (Field, 2005) of residuals and of residuals plotted against fitted values (Quinn and Keough, 2002). There was no indication of a violation of these assumptions. To investigate collinearity, we calculated variance inflation factors from a standard linear model entailing all predictor and control variables without interaction terms. There was no observation of collinearity (maximum VIF: 1.02). Variance inflation factors (VIF; (Field, 2013) were implemented in R using the vif function of the car package (version 3.0–3; Fox and Weisberg, 2019). Model stability was assessed on the level of the estimated coefficients and standard deviations by excluding the levels of the random effect one by one (Nieuwenhuis et al., 2012), using a user-defined function. The model yielded acceptable stability. Using the function bootMer of the package lme4, we bootstrapped model estimates ($N = 1000$ bootstraps). Models were fitted in R version 4.2.0 (R Core Team, 2022) using the function lmer of the lme4 package version 1.1.29 and the lmerTest package (Kuznetsova et al., 2017). The sample for the models contained 150 N1 values, comprising 25 participants and six conditions (3 Conditions, 2 Congruencies). Post-hoc tests were applied using the emmeans package (Version: 1.7.4.1, Lenth, 2022).

4.2.2. P2

The same structure of analysis without a reduced model was used for analyzing the P2 component of the ERP with the peak latency of 184 ms (range: 154 ms – 214 ms) for the Baseline condition, 184 ms (range: 154 ms – 214 ms) for the Relevance condition, and 182 ms (range: 152 ms – 212 ms) for the Control Movement condition. Note that for convenience we refer to the measured positive ERP component at typical P2 latencies at our pre-defined electrode locations as “P2” in our figures and statistical models. When discussing possible functional implications of our findings, we will refer to ERP components at P2 latencies as “P2-timed” if we cannot specify whether they belong to a process with the characteristic fronto-central and right parieto-occipital topography seen in Fig. 2C. The same electrodes as for N1 were included into the analysis (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2). Variance inflation factors reached a maximum of 1.02.

4.2.3. P3

Drawing from prior research regarding prediction, prediction error and attention, we also analyzed the P3 as a later component (Escera et al., 1998; Ford et al., 2010; Jones et al., 2013; Polich, 2007; Schröger et al., 2015). Based on the literature (Ford et al., 2010; Jones et al., 2013; Polich, 2007), the P3 component was extracted using average signals from three regions of interest (“Clusters”) each, Frontal (F1, Fz, F2, FC1, FCz, FC2), Central (C1, Cz, C2) and Parietal (CP1, CPz, CP2, P1, Pz, P2). The P3 peaks were identified by visual inspection, separately for the three conditions Baseline, Relevance and Control Movement, based on the grand average over all regions of interest. Because of the missing prominent peak in congruent trials, we did not identify the P3 based on a population peak identification as for N1 and P2. However, we chose a wide P3 window to keep our analysis robust against smaller shifts in P3 timing. The final P3 comprised the peak interval of 340 – 400 ms after the sound for the Baseline condition and 320 – 400 ms after the sound for the Relevance and Control Movement condition. The interaction of Congruency, Condition and Cluster and the control variable Session order were included into a mixed model as fixed effects as well as random slopes within Participant. Correlations among random intercepts and slopes were not included. As the interaction between Congruency, Condition and Cluster did not yield significance, a reduced model was calculated without the three-way interaction. Variance inflation factors did not exceed 1.03.

4.3. Pupil size analysis

We applied a baseline correction (−200 to 0 ms before sound onset) and chose an analysis window of 600 to 1500 ms after the first sound of each trial, based on visual inspection.

Pupil size analysis was performed in the same way as the N1 analysis without a reduced model and also including Session order as random slope within Participant. Correlations between random intercepts and slopes were excluded. Variance inflation factors remained below 1.03. Residual homogeneity deviated from the assumed pattern in that the residuals were on average smaller than for intermediate fitted values. Therefore, results on pupil size should be treated with caution. The model was moderately stable for most fitted effects. Wide confidence intervals for several effects of this model suggested considerable uncertainty in the precise estimates.

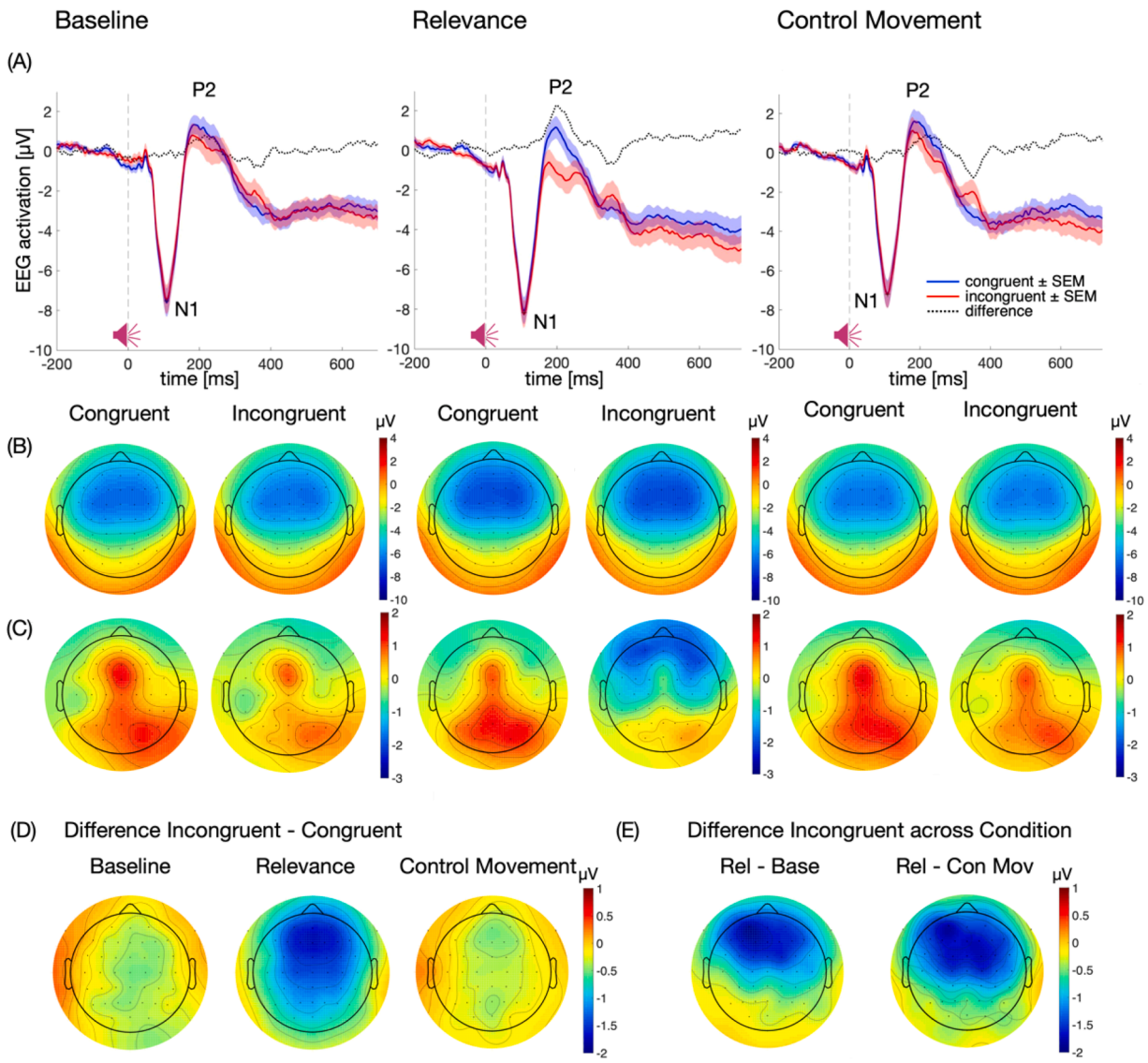

Fig. 2. (A): ERP grand-average waveforms ($N = 25$) with standard errors from frontocentral electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz and C2) in Baseline, Relevance and Control Movement condition. (B): Topographic maps in the N1 time window (78 – 118) for congruent and incongruent trials for Baseline, Relevance and Control Movement condition. (C): Topographic maps in the P2 time window for congruent and incongruent trials for Baseline (154 – 214 ms), Relevance (154 – 214 ms) and Control Movement (152 – 212 ms) condition, $N = 25$. Note the characteristic positivity at fronto-central and right parietooccipital electrodes in all but one condition. (D): Topographic maps in the P2 time window on the difference between incongruent and congruent trials per condition. (E): Topographic maps in the P2 time window on the difference between incongruent trials of conditions.

5. Results

5.1. EEG results

5.1.1. N1 is modulated by relevance, but not congruency

Firstly, we assessed how congruency and relevance influenced the auditory N1 response in fronto-central electrodes (see Figures 2 and 4). Overall, Congruency and Condition and their interaction did not show a significant effect on the N1 amplitude (full-null model comparison: $\chi^2 = 10.27$, df = 5, $p = 0.07$). Using restricted maximum likelihood with the Satterthwaite's method, the interaction clearly yielded no significance ($F = 0.87$, NumDf = 2, DenDf = 43.65, $p = 0.42$), which indicates that relevance (Condition) did not interact with prediction (Congruency) in N1. The reduced model without the interaction of Congruency and Condition showed significant effects of Condition ($F = 4.42$, NumDf = 2, DenDf = 27.56, $p = 0.02$) and Session order ($F = 4.74$, NumDf = 2, DenDf = 43.89, $p = 0.01$), but not of Congruency, on the N1. Post-hoc tests revealed a significant decrease (increase of the absolute amplitude) of N1 for the Relevance condition compared to the Baseline ($\beta = -0.53$, df = 25, $p = 0.03$) and the Control Movement ($\beta = -0.78$, df = 32.5, $p = 0.01$) conditions.

Fig. 3

These results show that no effect of prediction was reflected in N1, but that relevance of the action-effect increased the amplitude of the N1 ERP component.

5.1.2. P2 amplitude is modulated by an interaction of congruency and task condition

Congruency and Condition and their combination also led to a significant modulation of the P2 value, as depicted by the full-null model comparison ($\chi^2 = 47.97$, df = 13, $p < 0.001$). The ANOVA of the model reflected this result with a significant interaction ($F = 4.74$, NumDf = 2, DenDf = 47.36, $p < 0.001$). Session order did not affect P2. Post-hoc testing revealed a significant decrease of P2 in incongruent trials for the Relevance condition compared to the Baseline condition ($\beta = -1.46$, df = 39.5, $p < 0.001$) and to the Control Movement condition ($\beta =$

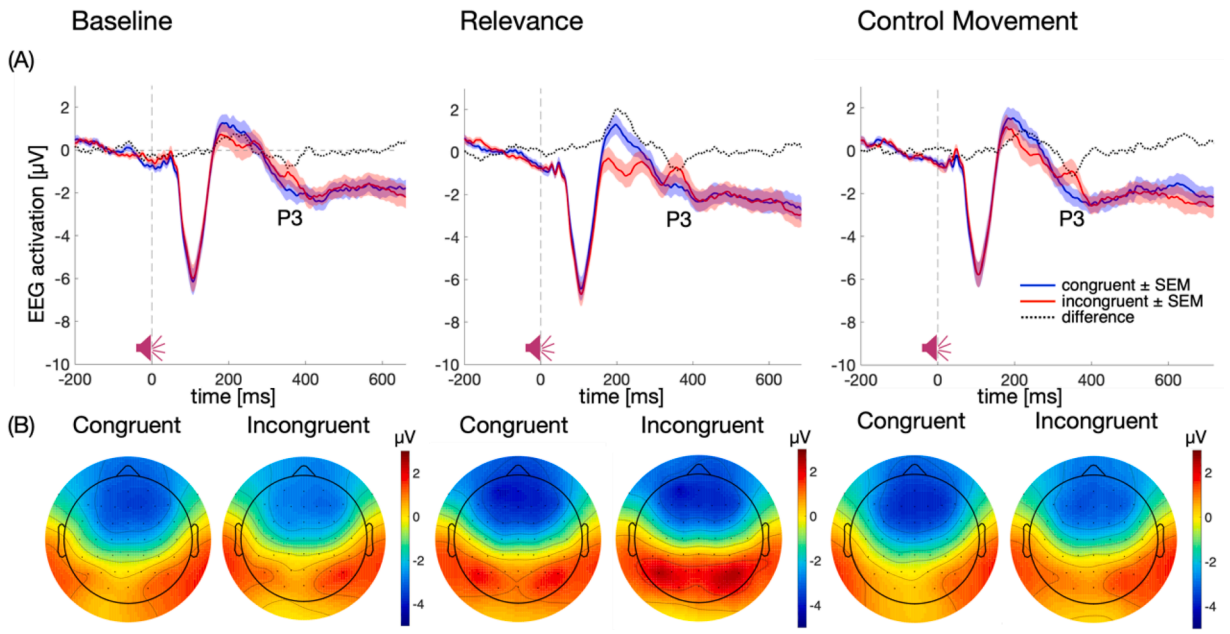

Fig. 3. (A): ERP grand-average waveforms ($N = 25$) with standard errors from electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2) in Baseline, Relevance and Control Movement condition. (B): Topographic maps in the P3 time window for congruent and incongruent trials for Baseline (340 - 400 ms), Relevance (320 - 400 ms) and Control Movement (320 – 400 ms) condition, $N = 25$.

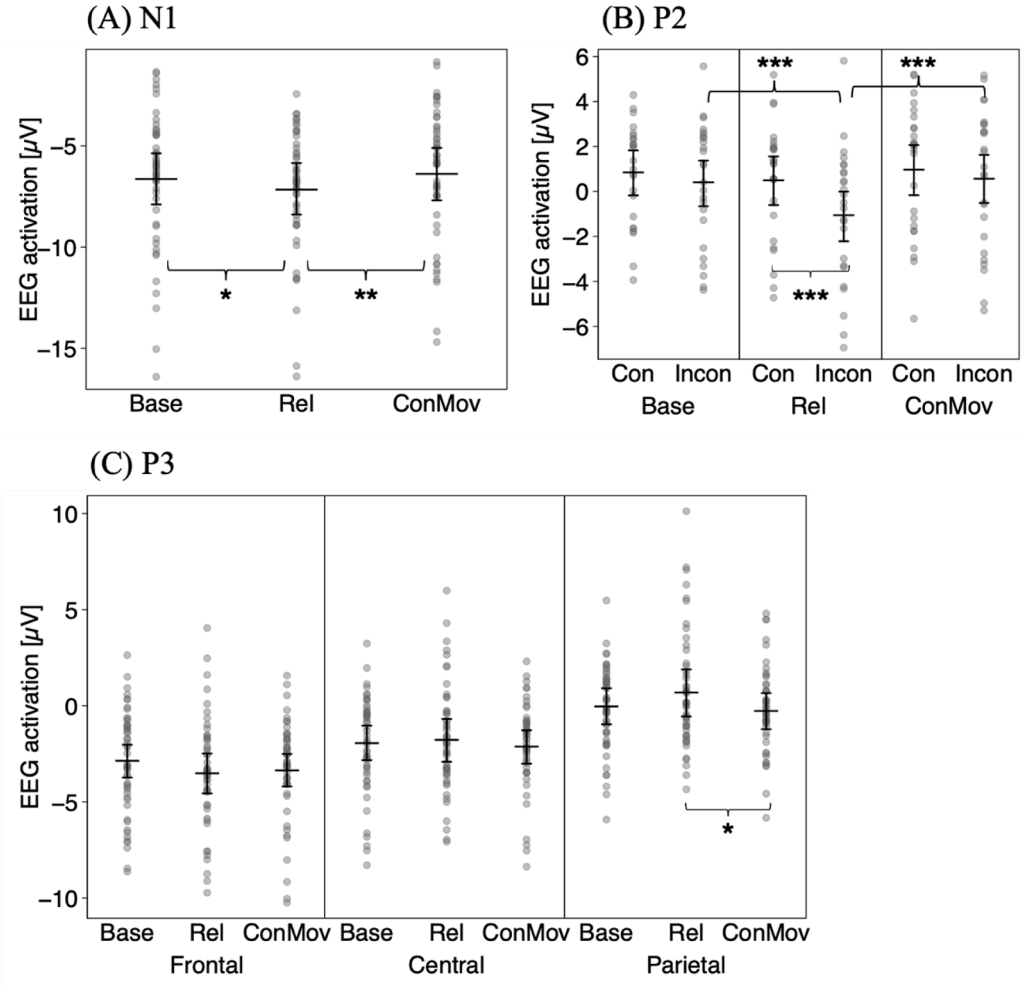

Fig. 4. Medians and confidence intervals of fitted values for (A) Condition on N1, (B) the interaction Condition x Congruency on P2 and (C) Condition x Cluster on P3. Grey dots refer to datapoints. *** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$.

−1.62, df = 46.2, $p < 0.001$). There was a significant difference between congruent and incongruent trials for the Relevance condition (β = 1.55, df = 65.8, $p < 0.001$). At the same time, a trend to significance was observed for the difference between congruent and incongruent trials for Baseline (β = 0.44, df = 65.8, $p = 0.06$) and for Control Movement (β = 0.41, df = 65.8, $p = 0.08$). Summarizing, prediction-congruency and relevance exerted an interactive effect on the P2 ERP component, with relevance decreasing the P2 amplitude for incongruent sounds (see Figs. 2 and 4).

5.1.3. P3 is modulated by task condition

Congruency, Condition and Cluster and their interaction led to an overall significant modulation on the P3 (full-null model comparison: $\chi^2 = 100.22$, df = 17, $p < 0.001$). However, restricted maximum likelihood with the Satterthwaite's method showed no significance of the interaction between Congruency, Condition and Cluster ($F = 1.86$, NumDf = 4, DenDf = 81.58, $p = 0.13$). The reduced model without the interaction of Congruency, Condition and Cluster demonstrated a significant interaction between Condition and Cluster ($F = 9.01$, NumDf = 4, DenDf = 30.53, $p < 0.001$) and a significant effect of Session order ($F = 3.99$, NumDf = 2, DenDf = 25.79, $p = 0.03$) on the P3. Post-hoc tests revealed a significant increase of parietal P3 for Relevance compared to Baseline P3 (β = 0.89, df = 26.4, $p = 0.045$) and a close to significant trend with an increase of P3 from Control movement to Relevance condition (β = 0.73, df = 26.4, $p = 0.095$; see Figures 2 and 4). There was no significant modulation of P3 between Conditions at the central or frontal Clusters. Since the effect of Cluster was not in the focus of this study, the exact significances of the pairwise comparison between Clusters are not reported, but all comparisons of Clusters reached a p-value below 0.001.

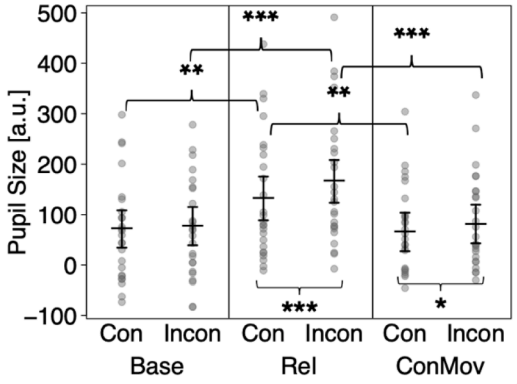

Fig. 5. Medians and confidence intervals of fitted values for pupil size for the interaction Congruency x Condition. Grey dots refer to datapoints. *** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$.

Relevance increased the P3 independent of prediction-congruency of the initiated sound.

5.2. Pupil size results

5.2.1. Pupil dilation response is modulated by an interaction of congruency and condition

Congruency and Condition and their interaction showed an overall significant impact on the pupil dilation response (full-null model comparison: $\chi^2 = 41.02$, df = 5, $p < 0.001$). The ANOVA supported this result with a significant interaction ($F = 7.97$, NumDf = 2, DenDf = 46.90, $p = 0.001$). Post-hoc tests indicated a significant increase of pupil dilation for the Relevance condition compared to the Baseline (β = 67.6, df = 26.8, $p = 0.002$) and to the Control Movement condition (β = 58.9, df = 27.3, $p = 0.006$) in congruent trials. For incongruent trials, relevance led to an increase in pupil size compared to Baseline (β = −96.9, df = 26.8, $p < 0.001$) and to Control Movement (β = 78.3, df = 27.3, $p < 0.001$). Pupil dilation response was significantly greater for incongruent compared to congruent trials in the Relevance condition (β = −34.35, df = 6.11, $p < 0.001$) and Control Movement condition (β = −14.96, df = 6.70, $p = 0.03$; see Figs. 5 and 6). These results indicate that pupil size was sensitive to congruency and interacted with the relevance of the action-effect.

5.3. Correlations

Separate Pearson's correlations were computed for the difference between congruent and incongruent trials in N1 and P2 to the same difference for pupil dilation response. N1 and pupil size did not correlate significantly ($r = 0.1$, $p = 0.37$), whereas P2 and pupil size anti-correlated significantly ($r = -0.25$, $p = 0.03$).

6. Discussion

We investigated whether sensory attenuation is present in all cases of action-effect prediction and therefore has to be seen as an inevitable correlate of controlled action. Specifically, we asked whether sensory attenuation of expected auditory action-effects remains when the action-effect gains behavioral relevance for planning and execution of a follow-up action. Participants were asked to create sounds with different pitches by pressing respective keys, either as a single sound (Baseline condition) or in the context of a specific (Relevance condition) or unspecific (Control Movement condition) sound sequence. The sounds played could be either prediction-congruent or prediction-incongruent based on a learned key-sound contingency.

Contrary to our expectation, we did not observe sensory attenuation (reduction of evoked neural response amplitude to congruent compared to incongruent sounds), neither for relevant nor for irrelevant auditory

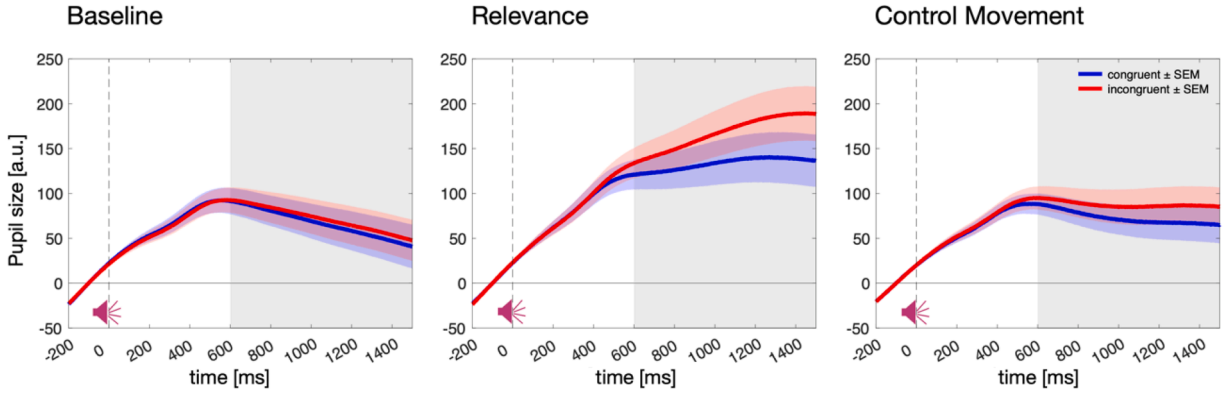

Fig. 6. Pupil response with standard error averaged over 25 participants, for congruent (blue) and incongruent (red) trials for Baseline, Relevance and Control Movement condition. The grey area marks the time period that was included into statistical analysis.

action-effects in the N1 or P2-timed component of the ERP. We did, however, observe an attenuation effect in pupillary response for relevant and irrelevant congruent sounds, whenever a follow-up movement after the sound was performed. The attenuation effect of pupil response was strongest for relevant sounds. We also observed an attenuated P2-timed component for relevant incongruent action-effects compared to irrelevant incongruent and to all congruent action-effects.

Our findings suggest, first, that sensory attenuation is not consistently present in all cases of action-effect prediction and, second, that identity-specific action-effect prediction may not be represented in the N1 component of the ERP, contrary to prior suggestions in the literature (Hughes et al., 2013a). At the same time, we could show that identity-specific action-effect prediction was present in our task, as reflected in congruency effects in pupillary responses and a P2-timed ERP component. These results imply that the idea of a forward model prediction, which cancels out early sensory input, may not apply to all cases of action-effect prediction.

P2 responses have been suggested to represent a comparison between sensory inputs and internal predictions (Chung et al., 2022; Garrido et al., 2009). Sensory attenuation in P2 for auditory action-effects has been reported especially for self- vs externally initiated sounds, where self-initiation helps to increase prediction of the sound (Mifsud, Beesley, et al., 2016; SanMiguel et al., 2013; Timm et al., 2014). Recent literature focusing on identity-specific prediction, i.e., where features of the sound matter (like the pitch in the present experiment), instead reported an enhancement for predicted compared to unpredicted sounds (Chung et al., 2022; Hsu et al., 2015; Jones et al., 2013; Le Bars et al., 2019), similar to our findings. This identity-prediction related P2 enhancement aligns with findings on mismatch negativity in oddball paradigms, in which deviant (unpredictable) sounds showed a reduced P2 amplitude compared to predictable standards sounds (Korka et al., 2019; Näätänen, Paavilainen, Rinne, and Alho, 2007). Consequently, the P2 may represent processes related to the identity of the action-effect, which is supported by the finding that the P2 component could reflect the processing of the specific features of auditory stimuli (Shahin et al., 2005). This idea would go in line with the suggestion that while the N1 reflects lower-level forward prediction that prepares the auditory cortex, the P2 reflects a rather cognitive response (Knolle et al., 2019). Hence, the fact that we find P2-timed but not N1-associated congruency effects could be a result of our task design, in which stimulus identity rather than pure presence or timing defined predictability was behaviorally relevant. The behavioral relevance of the incongruent stimulus for selecting future actions thereby enhances the effect of congruency by reducing the response to an incongruent sound.

Because action-effects gained behavioral relevance for subsequent motor behavior in our task, it could be that mechanisms related to the planning and execution of the following action explain our results at least partly. Especially the relevance of the incongruent stimulus enhanced the congruency effect in our data. In contrast to relevant congruent and to all irrelevant sounds, the relevant incongruent sound indicated to participants that a change of initial planning was necessary. In relevant incongruent trials, participants had to cancel their initially planned consecutive action that would have been needed to finish the original sequence. Instead, they had to switch their motor plan to complete the alternative sequence by repeating the action (press the same key) as was used to trigger the first action-effect. This means that the reduction of the ERP at P2-latencies in relevant incongruent trials could in principle reflect response cancellation and updating of an action plan.

Yet, we do not believe that motor inhibition explains our P2-timed finding. In contrast to our data, successful motor inhibition previously has been associated with an enhanced, not an attenuated P2. Senderecka et al. (2012) demonstrated that amplitudes of P2 components to auditory stop-signals were elevated for successful compared to unsuccessful stop-signal trials. Also, rather than the P2, the N2 and P3 components have been proposed to be involved in motor inhibition (Groom and Cragg, 2015; Ramautar et al., 2006; Smith et al., 2008, 2013).

Our data does not support the hypothesis that attention leads to a reversal of sensory attenuation by heightening the weighting of sensory evidence (Kok et al., 2012). According to the theory by Kok and colleagues (2012), prediction manifests as attenuation of predicted effects, while attention increases the precision of the prediction. Therefore, the sensory data of predicted sounds would be weighted more strongly, which results in an increased response to attended, predicted stimuli (Kok et al., 2012). This hypothesis is not supported by our data. The congruency effect in the relevant condition in our data was driven by a decrease of the P2-timed amplitude for incongruent sounds for relevant compared to irrelevant tones, not by an enhancement for congruent sounds.

The results at P2-latencies could further indicate an enhanced representation and processing of predicted sensory effects in the brain, as postulated by the sharpening account (de Lange et al., 2018). According to this idea, neurons encoding unexpected features (here: unexpected pitch of the sounds) are suppressed leading to more selective ("sharper") neural population response with a reduced overall amplitude. Computationally, so the idea, inputs inconsistent with high-level feedback signals are subject to stronger inhibition compared to consistent input, thereby allowing this top-down feedback to selectively enhance the representation of expected sensory signals (de Lange et al., 2018). Comparable physiological processes have been attributed to underlie selective attention (Martinez-Trujillo and Treue, 2004). The precise impact of sharpening on specific ERP components is not explicitly defined within the theoretical framework. However, based on the account's assumptions, one may infer, that an overall signal, as ERPs, may represent unpredicted events as even more reduced compared to their predicted counterparts. This assumption is based on the theoretical concept of the sharpening account, that in a context in which prediction takes place, units encoding unexpected features are suppressed. Consequently, we might observe attenuation for unpredicted compared to predicted sounds in the N1 ERP component as an indicator of sharpening in early auditory processing (Näätänen and Picton, 1987; Picton, 2013) or the P2, possibly reflecting sharpening towards more complex stimulus characteristics (Shahin et al., 2005). In terms of ERP amplitudes, our findings at P2-latency seem to align with the expectations of the sharpening account for the Relevance condition. Relevance may amplify sharpening, potentially mediated by selective attention to the predicted sound, which needs to be recognized in order to complete the action sequence as planned. Yet, an additional discriminant analysis did not indicate better discriminability of sound identity (pitch) in our P2 measure for congruent and / or for relevant sounds, hence do not support the idea of sharpening as explanation for the observed congruency effect in the Relevance condition (Supplementary Material S1).

Our observed modulation at P2 latency may not represent an isolated P2 component. Instead, it could be influenced by overlapping signals such as mismatch negativity (MMN) or an early N2 response. The topographical differences between congruent and incongruent trials, as well as between incongruent relevant and incongruent irrelevant conditions (Baseline and Control Movement, respectively), reveal a fronto-central modulation at P2 latency (Fig. 2D-E). This localization goes in line with findings on auditory P2 in fronto-central electrodes (e.g. Chung et al., 2022; Harrison et al., 2022; Timm et al., 2016; Van Elk et al., 2014; Woods et al., 1993). However, we cannot exclude the possibility that components as MMN or an early N2 response may overlap with our P2 time window. The subcomponent N2c, which matches our fronto-central scalp distribution, has been associated with subprocesses of stimulus classification. The auditory MMN, peaking around 100–250 ms after an auditory deviation with fronto-central distribution, has been linked to an indexing of different stages of mismatch detection and is involved in subcomponents of the N2 (Folstein and Van Petten, 2008; Garrido et al., 2009; Näätänen et al., 2007). Both, the MMN and the N2 have been linked to processes related to deviance processing and error

detection (Folstein and Van Petten, 2008; Garrido et al., 2009; Korka et al., 2021, 2022).

Furthermore, the frontal N2 component has been connected to response inhibition (Bokura et al., 2001; Pfefferbaum et al., 1985; Randall and Smith, 2011) and could therefore be involved in the modulation of the ERP signal to incongruent relevant sounds, where inhibition of the second keypress could play a role. However, since the difference between congruent and incongruent sounds yielded a trend to significance in the P2 time frame also for non-relevant sounds and because we further did not observe typical inhibition-related signals in P3 (Groom and Cragg, 2015; Ramautar et al., 2006; Smith et al., 2008, 2013) we assume that inhibition may not be the primary driver of the observed modulation.

The increase of pupil size in the Control Movement and Relevance condition, i.e., conditions which comprised a follow-up action, and the additional significant difference between congruent and incongruent pupil dilation response may indicate enhanced levels of attention, task demand or cognitive effort compared to Baseline (Alamia et al., 2019; Rondeel et al., 2015; van der Wel and van Steenbergen, 2018; Wetzel et al., 2016). Recent literature investigating action-effect prediction found enhanced pupil size for sounds aligned with motor responses compared to sounds presented without a motor response (Paraskevoudi and SanMiguel, 2021) and also for visual stimuli related to self-generation (Lubinus et al., 2022). We could demonstrate that prediction effects on pupil responses are also present for identity-specific action-effects, at least when the effect is followed by a second action.

The significant correlation between the P2-timed congruency effect and the pupil dilation response in our data suggest that both signals might reflect similar processes. Larger pupil dilation responses in incongruent trials are reminiscent of results from oddball paradigms in which deviant sounds created a larger pupil size response compared to standard auditory stimuli. Alamia et al. (2019) found a significant correlation between pupillary response for the difference between attended but unconscious rare and frequent events and visual mismatch negativity (indexed as difference between rare and frequent events) in fronto-central electrodes between 158 and 246 ms. Previously it was claimed that MMN at least partly originates in ACC (Hughes et al., 2013b; Hyman et al., 2017), which encode surprise- and prediction-error related signals (Alexander and Brown, 2019; Carter et al., 1998; Ide et al., 2013; Jahn et al., 2014) and to respond to task features that trigger pupil response (Ebitz and Platt, 2015; O'Reilly et al., 2013). Therefore, the authors argue that these results suggest a causal relationship between anterior cingulate cortex (ACC) and pupil responses or a common origin to the two physiological responses. In our study, pupil sizes were largest for incongruent trials especially in the Relevance condition. This may further reflect processes associated with error detection, and response inhibition and switching, Larson and Clayson, 2011), i.e., processes which are also discussed in the context of ACC function (Gehring et al., 2018; Holroyd and Coles, 2002). In conclusion, pupil response and our P2-timed ERP effects seem to respond to similar mechanism, potentially related to error processing, attention and arousal, and are both sensitive to violations of identity-specific action-effect prediction, especially when relevant for the selection of future actions.

Contrary to our expectations, the auditory N1 did not show an attenuation effect for prediction-congruent compared to -incongruent sounds. One may hypothesize that identity-specific stimulus features are not represented in N1, but only occur at later stages of processing, in the P2 time window, as discussed above. N1 attenuation has previously been demonstrated especially in tasks comparing prediction in self- versus externally generated effects (Asimakidou et al., 2022; Bays et al., 2006; Kilteni and Ehrsson, 2022; Kilteni et al., 2019; Sato, 2009). At the same time, a lack of an N1-associated congruency effect, as seen in our data, aligns with recent literature investigating identity-specific auditory action-effect prediction (Chung et al., 2022; Korka et al., 2019; Le Bars et al., 2019). Chung et al. (2022) suggested that since both the predicted and mis-predicted sounds are self-generated and the occurrence of the sound as such can be predicted via the action, sensory attenuation in N1 might not be visible, as both types of sounds would be suppressed equally. Baess et al. (2008) pointed out that for self- and externally generated stimuli N1 attenuation remained as long as the participants identified themselves as the agent of the respective sounds, even when the frequency of the sound could not be accurately predicted. However, when both the frequency and the onset of the sound were predictable the attenuation effect in N1 was most pronounced, suggesting that both predictable timing and predictable identity of action effects should contribute to attenuation.

In our study, the presentation of the sound reliably occurred 400 ms after the keypress during the test phases. Therefore, temporal predictability might have dominated N1 attenuation effects and have affected both the congruent and incongruent sounds similarly. Yet, previous research reporting identity-specific sensory attenuation in N1 employed a 400 ms delay between action and sound, as in our study (Hughes et al., 2013a). In cases where no attenuation in N1 was observed previously, the sound was either presented with no delay (Chung et al., 2022; Korka et al., 2019) or after 400 ms (Le Bars et al., 2019). Baess et al. (2008) showed that prediction of the correct sound frequency led to a stronger attenuation effect, i.e., an identity-based congruency effect enhanced the attenuation in N1 independent of temporal action-effect prediction. Generally, auditory N1 attenuation has further been observed for action-effects that included an unpredictable sound onset between 500 and 1000 ms (Baess et al., 2008). Taken together, the temporal distance between action and sound does not explain why we did not see N1 attenuation in identity-specific action-effect prediction.

The division into learning and test phases, with equal presentation frequency of congruent and incongruent sounds in the test phases, may have further weakened an effect of motor prediction on N1. Participants in principle could have differentiated between the phases and consequently lost their prediction of sounds. However, very prominent congruency effects are evident in pupil size and the P2-timed component and participants successfully identified the keys corresponding to the specific sounds in the catch trials. This indicates that predictions of action effects were maintained in the test phases. Also, previous studies have successfully implemented a congruency effect using similar learning and test phases (Chung et al., 2022; Jones et al., 2013; Kühn et al., 2011). Furthermore, several studies using a higher frequency of congruent sounds in the test phase could not observe an attenuation in N1 as well (Chung et al., 2022; Le Bars et al., 2019). These observations together suggest, that the separation into distinguishable training and test phases with different congruency ratio does not explain the lack of an N1 effect in our data. The fact that not even a trend to significance of a N1 congruency effect is visible therefore suggests that N1 attenuation does not correspond to identity-specific action-effect prediction.

Some studies also suggest that previously observed suppression in N1 in self-vs-other paradigms might merely be due to unspecific mechanisms, e.g., related to the action per se rather than the action-effect prediction (Horváth et al., 2012; Paraskevoudi and SanMiguel, 2022; SanMiguel et al., 2013). Further, it was shown that attenuation in N1 can be sensitive to global regularities rather than to the predictability of a sound (Korka et al., 2019). An N1 enhancement by relevance (independent of congruency) is visible in our data in the Relevance compared to the other conditions, likely attributable to heightened attention (Harrison et al., 2021; Hillyard et al., 1973; Näätänen et al., 1978).

We further did not observe a prediction effect in P3. This means, we could not replicate an effect of match/mismatch detection (Verleger et al., 2017). As congruent and incongruent P3 did not significantly differ, specifically for relevant sounds, a potential effect of motor inhibition, as could be assumed for incongruent relevant trials, was not visible in P3 (Bokura et al., 2001; Patel and Azzam, 2005; Randall and Smith, 2011; Smith et al., 2007). However, relevance enhanced the P3 in parietal regions, which may indicate attentional or recognition processes (Hillyard et al., 1973; Näätänen et al., 1978; Polich, 2007) or

mechanisms related to decision making (Twomey et al., 2015).

In conclusion, when a predicted action-effect gains relevance for a follow-up action, markers of identity-specific action-effect prediction are enhanced. These markers however differ from prediction effects shown in paradigms comparing passive vs active trials, with missing N1 attenuation and an attenuation in P2-timed ERPs for incongruent sounds in contrast to an attenuation of congruent sounds. Pupil dilation response underlines this enhancement of the prediction effect with relevance, as seen in P2-timed ERPs.

CRediT authorship contribution statement

Elisabeth Lindner: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Andrea Desantis:** Writing – review & editing, Writing – original draft, Resources, Methodology, Conceptualization. **Felicia Pei-Hsin Cheng:** Validation, Software, Formal analysis. **Alexander Gail:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

None.

Data availability

Data is available at https://doi.org/10.25625/MV3NIM.

Acknowledgments

We thank Roger Mundry for his support with statistical analysis and Marcus Jeschke for his support with the artificial ear.

Funding

This work was supported by the German Research Foundation (DFG) Collaborative Research Consortium CRC-1528 “Cognition of Interaction”, and by Agence Nationale de la Recherche, research funding ANR JCJC, project number ANR-18-CE10–0001.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2024.120717.

References

Agresti, A., 2002. Categorical Data Analysis. Categorical Data Analysis. Wiley. https://doi.org/10.1002/0471249688.

Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A., Zenon, A., 2019. Pupil-linked arousal responds to unconscious surprisal. J. Neurosci. 39 (27), 5369–5376. https://doi.org/10.1523/JNEUROSCI.3010-18.2019.

Alexander, W.H., Brown, J.W., 2019. The role of the anterior cingulate cortex in prediction error and signaling surprise. Top. Cogn. Sci. 11 (1), 119–135. https://doi.org/10.1111/tops.12307.

Asimakidou, E., Job, X., & Kilteni, K. (2022). The positive dimension of schizotypy is associated with a reduced attenuation and precision of self-generated touch. *Schizophrenia*, 8(1). https://doi.org/10.1038/s41537-022-00264-6.

Baayen, R.H., 2008. Analyzing linguistic data: a practical introduction to statistics using R. Analyzing Linguist. Data 1–353. https://doi.org/10.1017/CBO9780511801686.

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. https://doi.org/10.1111/j.1469-8986.2011.01196.x.

Baess, 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. https://doi.org/10.1016/J.IJPSYCHO.2008.06.005.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68 (3), 255–278. https://doi.org/10.1016/J.JML.2012.11.001.

Bays, P.M., Flanagan, J.R., Wolpert, D.M., 2006. Attenuation of self-generated tactile sensations is predictive, not postdictive. PLoS Biol. 4 (2), e28. https://doi.org/10.1371/journal.pbio.0040028.

Blakemore, S.J., Wolpert, D.M., Frith, C.D., 1998. Central cancellation of self-produced tickle sensation. Nat. Neurosci. 1 (7), 635–640. https://doi.org/10.1038/2870.

Bokura, H., Yamaguchi, S., Kobayashi, S., 2001. Electrophysiological correlates for response inhibition in a Go/NoGo task. Clin. Neurophysiol. 112 (12), 2224–2232. https://doi.org/10.1016/S1388-2457(01)00691-5.

Brainard, D.H. (1997). The psychophysics toolbox. In *Spatial Vision* (Vol. 10, Issue 4, pp. 433–436). VSP. https://doi.org/10.1163/156856897X00357.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science (1979) 280 (5364), 747–749. https://doi.org/10.1126/science.280.5364.747.

Christensen, R.H.B. (2019). *Ordinal - regression models for ordinal data* (R package version 2019.12-10. CRAN.R-project.org/package=ordinal). https://cran.r-project.org/web/packages/ordinal/citation.html.

Crapse, T.B., Sommer, M.A., 2008. Corollary discharge across the animal kingdom. Nat. Rev. Neurosci. 9 (8), 587–600. https://doi.org/10.1038/nrn2457.

de Lange, F.P., Heilbron, M., Kok, P., 2018. How do expectations shape perception? Trends Cogn. Sci. (Regul. Ed.) 22 (9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002.

Dobson, A.J., 2002. An Introduction to Generalized Linear Models, 2nd ed. Chapman & Hall/CRC Boca Raton.

Dogge, M., Custers, R., Aarts, H., 2019. Moving forward: on the limits of motor-based forward models. Trends Cogn. Sci. (Regul. Ed.) 23 (9), 743–753. https://doi.org/10.1016/j.tics.2019.06.008.

Ebitz, R.B., Platt, M.L., 2015. Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. Neuron 85 (3), 628–640. https://doi.org/10.1016/j.neuron.2014.12.053.

Escera, C., Alho, K., Winkler, I., Näätänen, R., 1998. Neural mechanisms of involuntary attention to acoustic novelty and change. J. Cogn. Neurosci. 10 (5), 590–604. https://doi.org/10.1162/089892998562997.

Field, A., 2005. Discovering Statistics Using SPSS, 2nd ed. Sage Publications, Inc.

Field, A., 2013. Discovering Statistics Using IBM SPSS Statistics, 4th ed. Sage Publications.

Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology. 45 (1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x.

Ford, J.M., Mathalon, D.H., 2012. Anticipating the future: automatic prediction failures in schizophrenia. Int. J. Psychophysiol. 83 (2), 232–239. https://doi.org/10.1016/j.ijpsycho.2011.09.004.

Ford, J.M., Roach, B.J., Miller, R.M., Duncan, C.C., Hoffman, R.E., Mathalon, D.H., 2010. When it’s time for a change: failures to track context in schizophrenia. Int. J. Psychophysiol. 78 (1), 3–13. https://doi.org/10.1016/J.IJPSYCHO.2010.05.005.

Forstmeier, W., Schielzeth, H., 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner’s curse. Behav. Ecol. Sociobiol. (Print) 65 (1), 47–55. https://doi.org/10.1007/S00265-010-1038-5/FIGURES/3.

Fox, J., Weisberg, S., 2019. *An R Companion to Applied Regression* (Third Edit). Sage publications.

Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: a review of underlying mechanisms. Clin. Neurophysiol. 120 (3), 453–463. https://doi.org/10.1016/J.CLINPH.2008.11.029.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 2018. The error-related negativity. Perspect. Psychol. Sci. 13 (2), 200–204. https://doi.org/10.1177/1745691617715310.

Groom, M.J., Cragg, L., 2015. Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. Brain Cogn. 97, 1–9. https://doi.org/10.1016/J.BANDC.2015.04.004.

Haggard, P., Tsakiris, M., 2009. The experience of agency: feelings, judgments, and responsibility. Curr. Dir. Psychol. Sci. 18 (4), 242–246. https://doi.org/10.1111/j.1467-8721.2009.01644.x.

Harrison, A.W., Christensen, B.K., Whitford, T.J., 2022. Action-effect prediction in sensory attenuation and error monitoring: distinguishing stimulus-driven and volitional movement. PsyArXiv.

Harrison, A.W., Mannion, D.J., Jack, B.N., Griffiths, O., Hughes, G., Whitford, T.J., 2021. Sensory attenuation is modulated by the contrasting effects of predictability and control. Neuroimage 237, 118103. https://doi.org/10.1016/J.NEUROIMAGE.2021.118103.

Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. Science (1979) 182 (4108), 177–180. https://doi.org/10.1126/SCIENCE.182.4108.177.

Holroyd, C.B., Coles, M.G.H., 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109 (4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679.

Horváth, J., Maess, B., Baess, P., Tóth, A., 2012. Action–sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. J. Cogn. Neurosci. 24 (9), 1919–1931. https://doi.org/10.1162/jocn_a_00215.

Hsu, Y.F., Hämäläinen, J.A., Moutsopoulou, K., Waszak, F., 2015. Auditory event-related potentials over medial frontal electrodes express both negative and positive prediction errors. Biol. Psychol. 106, 61–67. https://doi.org/10.1016/j.biopsycho.2015.02.001.

Hughes, G., Desantis, A., Waszak, F., 2013a. Attenuation of auditory N1 results from identity-specific action-effect prediction. Eur. J. Neurosci. 37 (7), 1152–1158. https://doi.org/10.1111/ejn.12120.

Hughes, G., Desantis, A., Waszak, F., 2013b. Mechanisms of intentional binding and sensory attenuation: the role of temporal prediction, temporal control, identity prediction, and motor prediction. Psychol. Bull. 139 (1), 133–151. https://doi.org/10.1037/a0028566.

Hughes, G., Waszak, F., 2011. ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. Neuroimage 56 (3), 1632–1640. https://doi.org/10.1016/J.NEUROIMAGE.2011.02.057.

Hughes, G., Waszak, F., 2014. Predicting faces and houses: category-specific visual action-effect prediction modulates late stages of sensory processing. Neuropsychologia 61 (1), 11–18. https://doi.org/10.1016/j.neuropsychologia.2014.06.002.

Hyman, J.M., Holroyd, C.B., Seamans, J.K., 2017. A Novel Neural Prediction Error Found in Anterior Cingulate Cortex Ensembles. Neuron 95 (2), 447–456. https://doi.org/10.1016/j.neuron.2017.06.021 e3.

Ide, J.S., Shenoy, P., Yu, A.J., Li, C.R., 2013. Bayesian prediction and evaluation in the anterior cingulate cortex. J. Neurosci. 33 (5), 2047. https://doi.org/10.1523/JNEUROSCI.2201-12.2013, 2039 LP –.

Jahn, A., Nee, D.E., Alexander, W.H., Brown, J.W., 2014. Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. Neuroimage 95, 80–89. https://doi.org/10.1016/j.neuroimage.2014.03.050.

Jones, A., Hughes, G., Waszak, F., 2013. The interaction between attention and motor prediction. An ERP study. NeuroImage 83, 533–541. https://doi.org/10.1016/J.NEUROIMAGE.2013.07.004.

Kaiser, J., Schütz-Bosbach, S., 2018. Sensory attenuation of self-produced signals does not rely on self-specific motor predictions. Eur. J. Neurosci. 47 (11), 1303–1310. https://doi.org/10.1111/ejn.13931.

Kamp, S.-M., & Donchin, E. (2014). *ERP and pupil responses to deviance in an oddball paradigm*. https://doi.org/10.1111/psyp.12378.

Kilteni, K., Ehrsson, H.H., 2022. Predictive attenuation of touch and tactile gating are distinct perceptual phenomena. iScience 25 (4), 104077. https://doi.org/10.1016/j.isci.2022.104077.

Kilteni, K., Houborg, C., Ehrsson, H.H., 2019. Rapid learning and unlearning of sensory delays in self-touch. Elife 1–17.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., 2007. What's new in psychtoolbox-3. Perception. 36 (14), 1–16. https://doi.org/10.1068/v070821/JQUERY.MIN.JS.

Kok, P., Rahnev, D., Jehee, J.F.M., Lau, H.C., de Lange, F.P., 2012. Attention reverses the effect of prediction in silencing sensory signals. Cerebral Cortex 22 (9), 2197–2206. https://doi.org/10.1093/cercor/bhr310.

Korka, B., Schröger, E., Widmann, A., 2019. Action intention-based and stimulus regularity-based predictions: same or different? J. Cogn. Neurosci. 31 (12), 1917–1932. https://doi.org/10.1162/JOCN_A_01456.

Korka, B., Schröger, E., Widmann, A., 2021. The encoding of stochastic regularities is facilitated by action-effect predictions. Scientif. Rep. 2021 11:1 11 (1), 1–14. https://doi.org/10.1038/s41598-021-86095-4.

Korka, B., Widmann, A., Waszak, F., Darriba, Á., Schröger, E., 2022. The auditory brain in action: intention determines predictive processing in the auditory system—A review of current paradigms and findings. Psychon. Bull. Rev. 29 (2), 321–342. https://doi.org/10.3758/s13423-021-01992-z.

Kühn, S., Nenchev, I., Haggard, P., Brass, M., Gallinat, J., Voss, M., 2011. Whodunnit? electrophysiological correlates of agency judgements. PLoS. One 6 (12), e28657. https://doi.org/10.1371/journal.pone.0028657.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–26. https://doi.org/10.18637/JSS.V082.I13.

Larson, M.J., Clayson, P.E., 2011. The relationship between cognitive performance and electrophysiological indices of performance monitoring. Cognit. Affect. Behav. Neurosci. 11 (2), 159–171. https://doi.org/10.3758/s13415-010-0018-6.

Le Bars, S., Darriba, Á., Waszak, F, 2019. Event-related brain potentials to self-triggered tones: impact of action type and impulsivity traits. Neuropsychologia 125, 14–22. https://doi.org/10.1016/j.neuropsychologia.2019.01.012.

Lenth, R.V. (2022). *Estimated Marginal Means, aka Least-Squares Means* (1.7.4-1). CRAN.

Liao, H.-I., Yoneya, M., Kidani, S., Kashino, M., Furukawa, S., 2016. Human pupillary dilation response to deviant auditory stimuli: effects of stimulus properties and voluntary attention. Front. Neurosci. 10. https://doi.org/10.3389/fnins.2016.00043 (FEB).

Lubinus, C., Einhäuser, W., Schiller, F., Kircher, T., Straube, B., van Kemenade, B.M., 2022. Action-based predictions affect visual perception, neural processing, and pupil size, regardless of temporal predictability. Neuroimage 263, 119601. https://doi.org/10.1016/j.neuroimage.2022.119601 (September).

Luck, S.J., 2014. An Introduction to the ERP technique, Second Edition. MIT Press. MIT Press.

Luke, S.G., 2017. Evaluating significance in linear mixed-effects models in R. Behav. Res. Methods 49 (4), 1494–1502. https://doi.org/10.3758/S13428-016-0809-Y.

Martinez-Trujillo, J.C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biol.*, 14(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028.

Miall, R.C., Wolpert, D.M., 1996. Forward models for physiological motor control. Neural Netw. 9 (8), 1265–1279. https://doi.org/10.1016/S0893-6080(96)00035-4.

Näätänen, R., Gaillard, A.W.K., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. 42 (4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118 (12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.

Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology. 24 (4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x.

Nieuwenhuis, R., te Grotenhuis, M., Pelzer, B., 2012. Influence.ME: tools for detecting influential data in mixed effects models. R Journal 4 (2), 38–47. https://doi.org/10.32614/RJ-2012-011.

O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc. Natl. Acad. Sci. U.S.A. 110 (38), E3660–E3669. https://doi.org/10.1073/pnas.1305373110.

Paraskevoudi, N., SanMiguel, I., 2021. Self-generation and sound intensity interactively modulate perceptual bias, but not perceptual sensitivity. Sci. Rep. 11 (1), 17103. https://doi.org/10.1038/s41598-021-96346-z.

Paraskevoudi, N., SanMiguel, I., 2022. Sensory suppression and increased neuromodulation during actions disrupt memory encoding of unpredictable self-initiated stimuli. Psychophysiology. 0034, 1–52. https://doi.org/10.1111/psyp.14156.

Patel, S.H., Azzam, P.N., 2005. Characterization of N200 and P300: selected studies of the Event-Related Potential. Int. J. Med. Sci. 2 (4), 147–154. https://doi.org/10.7150/ijms.2.147.

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10 (Issue 4), 437–442. https://doi.org/10.1163/156856897X00366.

Pfefferbaum, A., Ford, J.M., Weller, B.J., Kopell, B.S., 1985. ERPs to response production and inhibition. Electroencephalogr. Clin. Neurophysiol. 60 (5), 423–434. https://doi.org/10.1016/0013-4694(85)91017-x.

Picton, T., 2013. Hearing in time: evoked potential studies of temporal processing. Ear Hear. 34 (4), 385–401. https://doi.org/10.1097/AUD.0b013e31827ada02.

Pinheiro, A.P., Schwartze, M., Gutierrez, F., Kotz, S.A., 2019. When temporal prediction errs: ERP responses to delayed action-feedback onset. Neuropsychologia 134, 107200. https://doi.org/10.1016/j.neuropsychologia.2019.107200.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Quinn, G.P., Keough, M.J., 2002. Experimental Design and Data Analysis For Biologists. Cambridge University Press. https://doi.org/10.1017/CBO9780511806384.

R. Core Team. (2022). *R: a Language and environment for statistical computing* (4.2.0). R Foundation for Statistical Computing. https://www.r-project.org/.

Ramautar, J.R., Kok, A., Ridderinkhof, K.R., 2006. Effects of stop-signal modality on the N2/P3 complex elicited in the stop-signal paradigm. Biol. Psychol. 72 (1), 96–109. https://doi.org/10.1016/J.BIOPSYCHO.2005.08.001.

Randall, W.M., Smith, J.L., 2011. Conflict and inhibition in the cued-Go/NoGo task. Clin. Neurophysiol. 122 (12), 2400–2407. https://doi.org/10.1016/j.clinph.2011.05.012.

Rondeel, E.W.M., Steenbergen, H., Holland, R.W., van Knippenberg, A., 2015. A closer look at cognitive control: differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. Front. Hum. Neurosci. 9, 1–13. https://doi.org/10.3389/fnhum.2015.00494 (September).

Rusconi, E., Kwan, B., Giordano, B.L., Umiltà, C., Butterworth, B., 2006. Spatial representation of pitch height: the SMARC effect. Cognition 99 (2), 113–129. https://doi.org/10.1016/J.COGNITION.2005.01.004.

Rusconi, E., Kwan, B., Giordano, B., Umiltà, C., Butterworth, B., 2005. The mental space of pitch height. Ann. N.Y. Acad. Sci 1060, 195–197. https://doi.org/10.1196/annals.1360.056.

SanMiguel, I., Todd, J., Schröger, E., 2013. Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. Psychophysiology. 50 (4), 334–343. https://doi.org/10.1111/psyp.12024.

Sato, A., 2009. Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. Cognition 110 (1), 74–83. https://doi.org/10.1016/j.cognition.2008.10.011.

Sato, A., Yasuda, A., 2005. Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. Cognition 94 (3), 241–255. https://doi.org/10.1016/J.COGNITION.2004.04.003.

Saupe, K., Widmann, A., Trujillo-Barreto, N.J., Schröger, E., 2013. Sensorial suppression of self-generated sounds and its dependence on attention. Int. J. Psychophysiology 90 (3), 300–310. https://doi.org/10.1016/J.IJPSYCHO.2013.09.006.

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behav. Ecol.*, 20(2), 416–420. https://doi.org/10.1093/BEHECO/ARN145.

Schröger, E., Marzecová, A., Sanmiguel, I., 2015. Attention and prediction in human audition: a lesson from cognitive psychophysiology. Eur. J. Neurosci. 41 (5), 641–664. https://doi.org/10.1111/EJN.12816.

Senderecka, M., Grabowska, A., Szewczyk, J., Gerc, K., Chmylak, R., 2012. Response inhibition of children with ADHD in the stop-signal task: an event-related potential study. Int. J. Psychophysiol. 85 (1), 93–105. https://doi.org/10.1016/j.ijpsycho.2011.05.007.

Shahin, A., Roberts, L.E., Pantev, C., Trainor, L.J., Ross, B., 2005. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. Neuroreport 16 (16), 1781–1785. https://doi.org/10.1097/01.wnr.0000185017.29316.63.

Smith, J.L., Jamadar, S., Provost, A.L., Michie, P.T., 2013. Motor and non-motor inhibition in the Go/NoGo task: an ERP and fMRI study. Int. J. Psychophysiol. 87 (3), 244–253. https://doi.org/10.1016/J.IJPSYCHO.2012.07.185.

Smith, J.L., Johnstone, S.J., Barry, R.J., 2007. Response priming in the Go/NoGo task: the N2 reflects neither inhibition nor conflict. Clin. Neurophysiol. 118 (2), 343–355. https://doi.org/10.1016/j.clinph.2006.09.027.

Smith, J.L., Johnstone, S.J., Barry, R.J., 2008. Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. Clin. Neurophysiol. 119 (3), 704–714. https://doi.org/10.1016/j.clinph.2007.11.042.

Sperry, R.W., 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. J. Comp. Physiol. Psychol. 43 (6), 482–489. https://doi.org/10.1037/h0055479.

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. https://doi.org/10.1162/jocn_a_00552.

Timm, J., SanMiguel, I., Saupe, K., Schröger, E., 2013. The N1-suppression effect for self-initiated sounds is independent of attention. BMC. Neurosci. 14 (1), 2. https://doi.org/10.1186/1471-2202-14-2.

Timm, J., Schönwiesner, M., Schröger, E., SanMiguel, I., 2016. Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. Cortex 80, 5–20. https://doi.org/10.1016/J.CORTEX.2016.03.018.

Twomey, D.M., Murphy, P.R., Kelly, S.P., O'Connell, R.G, 2015. The classic P300 encodes a build-to-threshold decision variable. Eur. J. Neurosci. 42 (1), 1636–1643. https://doi.org/10.1111/ejn.12936.

van der Wel, P., van Steenbergen, H., 2018. Pupil dilation as an index of effort in cognitive control tasks: a review. Psychonomic Bull. Rev. 25 (6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y.

Van Elk, M., Salomon, R., Kannape, O., Blanke, O., 2014. Suppression of the N1 auditory evoked potential for sounds generated by the upper and lower limbs. Biol. Psychol. 102 (1), 108–117. https://doi.org/10.1016/j.biopsycho.2014.06.007.

Verleger, R., Cäsar, S., Siller, B., Śmigasiewicz, K., 2017. On why targets evoke P3 components in prediction tasks: drawing an analogy between prediction and matching tasks. Front. Hum. Neurosci. 11 https://doi.org/10.3389/fnhum.2017.00497.

von Holst, E., Mittelstaedt, H., 1950. Das Reafferenzprinzip. Naturwissenschaften. 37 (20), 464–476. https://doi.org/10.1007/BF00622503.

Wang, C.A., Baird, T., Huang, J., Coutinho, J.D., Brien, D.C., Munoz, D.P., 2018. Arousal effects on pupil size, heart rate, and skin conductance in an emotional face task. Front. Neurol. 9, 1–13. https://doi.org/10.3389/fneur.2018.01029 (December).

Weiss, C., Herwig, A., Schütz-Bosbach, S., 2011. The self in action effects: selective attenuation of self-generated sounds. Cognition 121 (2), 207–218. https://doi.org/10.1016/j.cognition.2011.06.011.

Wetzel, N., Buttelmann, D., Schieler, A., Widmann, A., 2016. Infant and adult pupil dilation in response to unexpected sounds. Dev. Psychobiol. 58 (3), 382–392. https://doi.org/10.1002/dev.21377.

Wolpert, D.M., 1997. Computational approaches to motor control. Trends Cogn. Sci. (Regul. Ed.) 1 (6), 209–216. https://doi.org/10.1016/S1364-6613(97)01070-X.

Woods, D.L., Knight, R.T., Scabini, D., 1993. Anatomical substrates of auditory selective attention: behavioral and electrophysiological effects of posterior association cortex lesions. Cognit. Brain Res. 1 (4), 227–240. https://doi.org/10.1016/0926-6410(93)90007-R.