Motor and Context Information in the Attenuation of Auditory ERP Amplitudes

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Alexander Seidel
aus Dormagen

Düsseldorf, Juni 2024

aus dem Institut für Experimentelle Psychologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Christian Bellebaum

2. Prof. Dr. Tobias Kalenscher

Tag der mündlichen Prüfung: 22.08.2024

Contents

1 GENERAL INTRODUCTION 1

1.1 Predicting Sensory Action Consequences 1

1.2 Non-Motor Based Predictions 9

1.3 Research Objectives 15

2 OVERVIEW OF STUDIES 18

2.1 Study 1 18

2.2 Study 2 23

2.3 Study 3 28

3 GENERAL DISCUSSION 35

3.1 N1 and P2 attenuation as markers of agency 36

3.2 Motor Based Forward Models in the N1 attenuation 41

3.3 Predictive Coding 47

3.4 Conclusion 52

4 BIBLIOGRAPHY 54

5 APPENDIX A: AFFIDAVIT 67

6 APPENDIX B: ORIGINAL RESEARCH ARTICLES 68

List of Abbreviations

EEG	electroencephalography
ERP	event related potential
TMS	transcranial magnetic stimulation
P(O)	probability of the outcome

Acknowledgements

It was the best of times, it was the worst of times. And while everyone has some bad tales to tell about their PhD, I feel very privileged to have lived through the best of times because of the best of people. First of the all, the god of patience, source of reason and enduring support, my supervisor Christian Bellebaum. There was always a solution, and you were always willing to find it, thank you.

My bachelor thesis supervisor Laura Bechtold turned into an essential pillar of the life at the HHU, maybe because she provided dogs, maybe because she laughed louder than anyone else. She also introduced me into the group that had a lasting change on my life, the HHU Running Heads, who would drive me to new heights and show me that I can always keep improving: Herr Prof. Jutta Peterburg, the head of the Running Heads, facilitator of connections, Adam the colorful, wizard of the coast, Dr. Benne Bänäne, who would always like to know more and Snadra Tralála, night cheese connoisseur and wildcard 5ever. Thank you for so many new experiences in these last years. A special thanks must go to the last member of the Running Heads, the office neighbor that raced me to the finish line (and of course won, as always), Tine Turner. From discussing scientific problems and building our own BelleLab Cinematic Universe to random rants about every topic under the (Sander-) sun, there was always some way to keep boredom at bay. Many thanks also to Marta Ghio for many sanity checks, scientific super support and only the best shared meals, as well as Constanze Weber for all the help in turning my first experimental idea into reality and a really complicated paper. Und endlich, danke Mama und Papa, ohne eure Unterstützung stände hier gar nichts.

Summary

The attenuation of neuronal responses to sounds that are produced by our own actions compared to sounds that we perceive in the environment has been suspected to result from sensory predictions that are matched to the actual sensory input we receive. The classic explanation is that of cerebellar forward models, which use information about planned motor activity to generated predictions, and attenuate sensory processing for correctly predicted stimuli. However, self- and externally generated sounds also differ in their temporally predictability, and temporally predictable externally generated sounds indeed display an attenuation of the N1 ERP component in EEG studies that is similar to that for self-generated sounds. The studies in this dissertation have added evidence for separate systems of sensory predictions based on temporal predictability, and motor-information. Study 2 showed that the N1 attenuation for observed motor acts developed slower over time when viewed from a first- instead of a third-person perspective, possibly as an effect of familiarity with third-person action observation. In Study 3, we reported an N1 attenuation for pro-, but not anti-saccade generated sounds, which are likely suffering from a disturbance of the forward model system. Crucially, the effects in both studies were found despite equal temporal predictability between motor related conditions, suggesting an independent influence of motor information based predictive processing. In study 1 we confirmed that assumptions about the context sounds are produced in impact not the N1, but the P2 component, which might reflect an aspect of the sense of agency. These studies also demonstrated the value of the analysis of ERP amplitude time courses as an important tool to uncover differing temporal patterns that would be hidden in aggregated datasets.

1 GENERAL INTRODUCTION

1.1. PREDICTING SENSORY ACTION CONSEQUENCES

It is unlikely that perception is simply a passive process of capturing sensory stimuli. Instead, some understand it as a process of active inference, in which the brain generates sensory predictions that are compared with the actual sensory input (Clark, 2013; Pickering & Clark, 2014). One line of research focuses specifically on predictions of the sensory consequences of our actions, inspired by the phenomenon of sensory attenuation, i.e. the reduction in perceived stimulus intensity for self- compared to externally generated movements. It was typically illustrated by the inability to tickle oneself, even when the self-generated haptic stimulation is identical to the external stimulation that causes a tickling sensation (Blakemore et al., 1999). However, the phenomenon is also found as a general attenuation of perceptual intensity for self- compared to externally-generated stimuli, not only in the haptic domain, but also for auditory and visual stimuli (Blakemore et al., 2000; Cardoso-Leite et al., 2010; Sato, 2008; Weiss et al., 2011).

An additional phenomenon is the attenuation of neuronal responses, which was shown accompanying the attenuation of perceptual intensity when self-generated haptic stimuli were compared to identical external stimuli (Blakemore et al., 1998). More recently, this phenomenon was mostly investigated as attenuations of the amplitudes of event-related potentials (ERPs) that have been found for self- compared to externally-generated auditory stimuli in electroencephalography (EEG) studies (Baess et al., 2011; Baess et al., 2008; Ghio et al., 2018; Hughes et al., 2013a; Timm et al., 2016). First reported by Schafer and Marcus (1973), self-generated sounds exhibit reductions of the amplitudes of the N1 and P2 components compared to externally generated sounds. These components and possibly a later P3 were examined extensively to investigate different processes behind the attenuation effects (Baess et al., 2008; Klaffehn et al., 2019; Knolle et al., 2013a; Kühn et al., 2011; Timm et al., 2016), and this dissertation focuses on the results for attenuations of ERP amplitudes in the auditory domain.

Most studies comparing self- and externally generated sounds employ a variation of the *contingent paradigm* is (for an overview, see Horváth, 2015): In an action-sound condition, sounds are produced by some kind of (typically) self-initiated movement. In an action-only condition, the same movements are performed without eliciting any sounds. This condition is used in an offline motor-correction procedure, where the ERPs of action-only button presses are subtracted from ERPs elicited in the action-sound condition to remove movement-related neuronal activity. In a sound-only condition, sounds are presented without any cues or actions. The inter-stimulus-interval has been recognized early as a factor determining the magnitude of the N1, which increases as a function of the inter-stimulus-interval up to around 10 seconds (Budd et al., 1998; Davis et al., 1966; Nelson & Lassman, 1968). As such, to keep inter-stimulus-intervals identical between the action-sound and sound-only conditions, some studies replayed the sounds produced by participants in a preceding action-sound condition in the following sound-only condition, especially when responses are self-timed (Baess et al., 2011; Baess et al., 2008; Ford et al., 2014; Ghio et al., 2018).

1.1.1 The internal forward model

Early on, sensory and neuronal attenuations were attributed to a system of *internal forward models* that directly rely on motor plans to generate predictions of sensory consequences following their execution. These models suppose that copies of motor commands or motor intentions are created during motor planning, and that these so-called *efference copies* are relayed to the cerebellum, where the forward models compute the sensory consequences for the planned actions (Blakemore et al., 2001; Ishikawa et al., 2016; Miall & Wolpert, 1996; Wolpert et al., 1998). The processing of sensory input fitting the predictions made by these models might then be attenuated, as this input does not carry novel information. A likely candidate for the source providing such efference copies is the supplementary motor area (Haggard & Whitford, 2004), which could provide input to forward models even before action execution (Reznik et al., 2018).

This idea of matching the predicted and actual sensory consequences of an action assumes that the sensory predictions include specific stimulus features that can be compared to the actual sensory input. Therefore, changing the identity of the stimulus should lead to a measurable misprediction when compared to stimuli matching the prediction. Baess et al. (2008) first examined ERP attenuations when the identity of a self-produced stimulus deviates from what was previously learned. They showed that a deviation in sound frequency led to a reduced attenuation of the N1, and this effect was confirmed by a similar study (Knolle et al., 2013b). Hughes et al. (2013a) found that even when both standard and deviant sounds were regularly produced in a task by either the right or left hand, incongruency of sound type and hand side led to a reduced attenuation.

However, even if sensory predictions play a role in the attenuation of the N1, the question remains whether they are based on information arising from the motor system, or post-hoc haptic, proprioceptive or environmental cues. The most direct evidence for this was reported by Timm et al. (2014a), who compared the processing of sensory consequences of voluntary with those of involuntary hand movements. For the latter, no a-priori motor plan should be created, as they were induced via transcranial magnetic stimulation (TMS). Only voluntary actions led to an attenuation of the N1-P2 complex, indicating that motor intention is involved in creating sensory predictions for self-generated actions. Using electrical stimulation of the median nerve of the inner forearm, Jack et al. (2021) also elicited involuntary sound producing finger movements and replicated this result for the N1. In addition, they showed that the N1 attenuation was also missing in a semi-voluntary condition (participants pressed a button to elicit the electrical stimulation on the other hand). For haptic stimuli, an intact attenuation of perceptual intensity was found even when a movement had not yet occurred, either by delivering the stimulus right before the movement (Voss et al., 2008), or at the time movement would have occurred, if it was not delayed by TMS (Voss et al., 2006). This was interpreted in that motor intention suffices to generate sensory predictions that lead to sensory attenuation, and thus seen as further evidence for an involvement of motor planning.

The forward model theory also supposes a cerebellar involvement in the generation of predictions leading to sensory attenuation (Blakemore et al., 2001). For this, evidence was provided by Knolle et al. (2012), who reported a missing N1 attenuation for cerebellar lesion patients. Knolle et al. (2013a) later replicated this finding, and additionally showed an intact P2 attenuation, indicating that the process leading to the P2 attenuation is distinct, and does not rely on cerebellar motor-specific predictions.

However, the role of forward models in sensory attenuation has been questioned, since self- and externally-generated sounds differ in more than their preceding motor activity. Producing button presses in a constant self-timed rhythm, as it is usually done in the contingent paradigm, might shift attention from the produced sound to the button press, and introduce an attentional difference to the conditions in which external sounds are presented. Though attention is known to increase the

N1 (Hillyard et al., 1973), the attenuation of self-generated sounds was found despite differences in the allocation of attention (Saupe et al., 2013; Timm et al., 2013).

Dogge et al. (2019) also argued that the forward model systems identified in animal studies usually involve action consequences closely related to the body, such as vocalizations (Schneider & Mooney, 2018), or retinal displacement (Sperry, 1950), which are trained from birth or possibly even hardcoded and not aquired. However, the action-effect associations in human studies are typically environment-related, such as visual or auditory stimuli, and likely have to be acquired during the course of the experiment. According to Dogge et al. (2019), the slow acquisition process for environment-related action-effects found in animal studies (Rummell et al., 2016) makes it questionable whether the appearance of neuronal attenuation effects in human studies relies on forward models. It must be noted though, that many studies use button presses as actions, which are very well known from the handling of modern devices like light switches or microwaves, and often accompanied by sounds as feedback. This might lead to a retuning of known action-effect associations, instead of an entirely new acquisition during experimental sessions.

One study examined a novel environment-related action-effect association, that cannot be trained in every-day life: Mifsud et al. (2016) contrasted the ERP attenuations for sounds generated by either the classic button presses, or saccadic eye movements (saccades), when compared to externally generated sounds. They found an N1 and P2 attenuation for button-press-elicited sounds, but a significantly less attenuated N1, and no P2 attenuation for saccade-elicited sounds. This could be interpreted as evidence for a slower acquisition process of the relevant action-effect association in the case of saccade-generated sounds, that is directly expressed in the magnitude of the attenuation of neuronal responses. For a similarly uncommon action-effect association, van Elk et al. (2014) reported a missing P2 attenuation as well, in this case for sounds elicited by button presses with the foot, that was present for button-press elicited sounds by hand-movements. This might indicate that the predictions underlying the N1 and P2 attenuations might rely on separate systems that are retuned at different rates to capture a new action-effect association.

1.1.2 Attributing Agency

The proposed system of efference copies (Von Holst, 1954), and forward models employing them (Wolpert, 1997), has long been assumed to enable the organism to broadly classify stimuli as self- or externally produced. The perceptual and neurophysiological differences between stimuli of those two categories have then been re-attributed to serving such a classification mechanism (Blakemore et al., 1999; Hughes et al., 2013a).

A related contextual factor when comparing self- and externally generated stimuli is the perceived agency over the elicited stimuli, i.e. the perception of being the originator of a stimulus. The first study examining the influence of agency (Kühn et al., 2011) manipulated it by changing the context of sound production, as well as the sound identity to induce an erroneous belief that sounds were not self-generated. Participants were led to believe that each elicited sound could be caused by either themselves or a concealed experimenter, while the sounds deviated in frequency, as well as their latency in a part of the trials. The resulting variance in the perception of agency was captured on trial level by an explicit rating. The N1, but not the later positive component P3, was sensitive to alterations of the sound identity (similar to Baess et al. [2008]) in that an N1 attenuation was only found for sounds with the same frequency as learned in the preceding training compared to sounds with a deviating frequency. On the other hand, the P3, but not the N1, was attenuated for sounds judged as self- rather than externally-produced, suggesting that this later positive component is either influenced by top-down agency judgements, or is itself reflecting a precursor to reaching a decision about these judgements.

In another study, Timm et al. (2014b) established that the perception of agency can be reduced by first training an action-effect association between button presses and sounds that includes a specific delay, and removing the delay in later trials whereby sounds are perceived by the participants to occur prior to the button press. Employing this phenomenon, Timm et al. (2016) found a similar pattern of results as Kühn et al. (2011) and reported that the N1 was not influenced by the illusory agency loss in the non-delayed sound condition (or the resulting variance in mean agency judgements) but instead simply attenuated for self- vs. externally generated sounds in general. For a later component, in this case the P2, this attenuation was smaller for sounds presented in the reduced agency condition (without the trained delay). In addition, the agency judgement difference of participants between the delayed and undelayed sound condition was directly correlated with the P2 attenuation difference between conditions.

Interestingly, the concept of agency has been extended by Synofzik et al. (2008) into a two-step process that starts with a rather automatic feeling of agency, which is then followed by a more conscious judgement of agency. This distinction might be reflected in the differing results for the N1 and the later positive components P2/P3. N1 attenuations seem to be closely linked to motor-intention or activity, independent of contextual changes (Knolle et al., 2013a; Kühn et al., 2011; Saupe et al., 2013; Timm et al., 2014a; Timm et al., 2016), but also sensitive to changes in the stimulus identity (Baess et al., 2008). This would fit in with the N1 reflecting the formation of a fast and automatic feeling of agency, which breaks down when ambiguity of the base features of a stimulus is introduced. The P2/P3 on the other hand has been directly related to explicit ratings of agency in studies aiming to induce agency ambiguity (Kühn et al., 2011; Timm et al., 2016), and possibly reflect the conscious judgement of agency that relies on further information than just basic action-effect associations. Contrary to the purely predictive forward model idea, the formation of the sense of agency is furthermore seen as an interplay of predictive as well as postdictive cues (Synofzik et al., 2013), that lead to bottom up sensorimotor or top down contextual influences. To examine whether the N1 and P2/P3 reflect different aspects of the sense of agency, which might be differently impacted by bottom up or top down influences, it is important to consider how paradigms experimentally induce agency variance.

In that regard, both agency manipulation studies (Kühn et al., 2011; Timm et al., 2016) changed basic stimulus features to reduce the accuracy of sensorimotor predictions. Kühn et al. (2011) introduced agency ambiguity by varying stimulus identity, which was reflected in the N1 attenuation. Timm et al. (2016) used an established agency loss illusion introduced by the omission of a latency during sound production that participants had previously adapted to, which was not reflected in the N1 attenuation. In both cases the P2/P3, but not the N1, was sensitive not only to the ambiguity itself, but directly to the subjective ratings of agency.

In addition to the bottom-up ambiguity introduced by varying stimulus features, both studies also introduced a possible top-down ambiguity by creating a context in which some of the sounds might not be self-generated. Kühn et al. (2011) directly informed participants that some sounds might be generated by the experimenter, but both studies implied that sounds might not be self-generated by letting participants rate each (relevant) tone as self-generated or not. That is why it cannot clearly be determined whether the process reflected in the P2/P3 attenuation is resulting from further evaluation of the bottom-up stimulus properties, or top-down control caused by contextual information.

One perceptual phenomenon that is suspected to stem from an impairment to attribute agency to oneself is the experience of auditory hallucinations in schizophrenic patients. Several studies have shown that the N1 attenuation for these patients is reduced for self-generated sounds, both by button press (Ford et al., 2014), and vocalization (Ford et al., 2007; Ford et al., 2001; Heinks-Maldonado et al., 2007). If the N1 is as strongly related to early, stimulus driven prediction mechanisms as previous studies have indicated, and less influenced by top-down control, the diminished N1 attenuation could provide evidence for an involvement of bottom-up information in the formation of agency judgements.

1.2. NON-MOTOR BASED PREDICTIONS

Despite the presented evidence for motor-signal based prediction mechanisms being involvement in the attenuation of the N1, the most common objection to the forward model theory is that these attenuations could be explained at least in part by the difference in temporal predictability between self-and externally generated sounds. This stems from a number of studies that show neurophysiological attenuations for stimuli made predictable by non-motor cues from the environment (Kaiser & Schütz-Bosbach, 2018; Klaffehn et al., 2019; Lange, 2009). It was first examined by Lange (2009) who tested the influence of stimulus identity and latency on the N1 for cued external sounds, similar to Baess et al. (2008), but without a comparison to self-generated sounds. The results showed that the N1 was decreased when timing and/or pitch of externally cued sounds was predictable compared to unpredictable, providing evidence for some form of motor-unspecific predictability mechanism involved in the N1 attenuation.

In a second study, Lange (2011) compared sounds produced by self-timed button presses and sounds following a visual cue. To further manipulate the temporal predictability, sounds were produced in a condition with a constant delay and a condition with a random delay after the response/cue. They found the N1 for self-generated compared to visually cued sounds to be reduced, but the N1 did not differ between constant vs. random delay. Even though the missing uncued external sound condition of the standard contingency paradigm makes it impossible to determine whether the sounds were reduced or enhanced compared to (neutral) external sounds, the results indicated that motor-specific prediction mechanisms affect the N1 attenuation.

Sowman et al. (2012) added an uncued external sound control condition, and also changed how the sound-generating action was performed: Button presses producing sounds were not self-timed anymore, but visually prompted. For this, the same visual cue as in the cued external sound condition was used. The delay between cues and sounds in the cued sound condition were matched for each participant with the response-dependent delay in the prior self-generation condition. With this method, self-produced sounds and cued sounds were similarly temporally predictable, and mostly differed with respect to the self-generation aspect. The results showed that while there was no significant N1 difference between both sounds, only the N1 for self-generated, but not cued sounds, was significantly reduced compared to the uncued control sounds. This indicates some form of predictive process for visually cued sounds, but does not provide evidence against predictions derived specifically from motor information like it is assumed for the forward-model system.

Kaiser and Schütz-Bosbach (2018) used a countdown from three to one, followed by an X on the screen to visually cue either an external sound, or the participant's response eliciting a sound. The N1 for cued external sounds was reduced compared to their uncued counterparts, while N1 amplitudes for sounds following cued responses did not differ from amplitudes for those that were uncued. Overall, this resulted in enhanced N1 amplitudes for cued self-generated compared to cued external sounds, which was taken as evidence that sensory attenuation does in fact not depend on self-generation. But it has to be noted that the usual motor-correction procedure was performed for the cued and uncued self-produced sounds via the subtraction of additional motor-only conditions, but no such correction was performed for the visual stimuli in the cued external sound condition, questioning the reliability of a direct comparison between the self-production and external sound condition.

In a newer design, Klaffehn et al. (2019) took a different approach to cueing to achieve temporal predictability for cued and self-produced sounds. Instead of a sequence of sudden stimuli (the countdown numbers) as in Kaiser and Schütz-Bosbach (2018) they employed a loading bar that filled up gradually over an interval of 750 ms, after which a sound was played. Button presses in the self-production condition started the loading bar which was followed by a sound, and were thus not prompted by a cued action as in the design by Kaiser and Schütz-Bosbach (2018). This way, temporal predictability based on the visual stimulation was now identical between self-generated and external stimuli, as cue-sound intervals were identical, leading to in improved comparability between those conditions. In the previous design by Kaiser and Schütz-Bosbach (2018), intervals between the appearance of the visual cue and the sound varied between conditions, because they included the response of the participant for self- generated sounds. The results showed an N1 attenuation for self- vs. externally-generated sounds both when the loading bar linearly filled to cue sounds, and when the loading bar increased and decreased randomly, not cueing sounds. This indicates that self-generation does indeed have an effect beyond mere temporal predictability. The P2, on the other hand, was only attenuated for self- vs. externally-generated sounds when they were not cued by the loading bar, suggesting that cued external sounds are similarly attenuated as cued self-generated sounds and thus the two do not differ.

Taken together, these studies suggest that N1 is sensitive for predictive mechanisms based on some form of external cueing, but also seems to reflect the processing of specifically motor-based predictions. Otherwise, conditions with identical temporal predictability that only differ by the participants' action should not have shown the described differences. The predictions reflected in the P2 on the other hand again seem to integrate other, in this case visual, information to form predictions instead of just the motor-based information which seems sufficient to attenuate the N1.

1.2.1 Predicting Consequence of Observed Actions

There is one type of external cues that has been suspected to receive specialized predictive processing: the actions of other humans. The idea stems from the discovery of *mirror neurons*, that fire both when actions are executed, as well as when similar actions are performed by another agent (Fadiga et al., 1995; Rizzolatti & Arbib, 1998; Rizzolatti & Sinigaglia, 2016). Mirror neurons have been linked to internal forward models, and are suspected to provide a motoric representation of the observed action, which might be used as an input for forward models to predict the sensory consequences (Miall, 2003; Wolpert et al., 2003). Considering the already existing predictive potential of external cues discussed above, this would indicate a specialized processing of observed-motor information, just as there seems to be for motor information from performed actions. First evidence for this was provided by Poonian et al. (2015), who measured the N1 during an intentional binding task, that requires participants to estimate the temporal interval between two events. In two experiments, they first compared N1 amplitudes for sounds that were preceded by either a button press, or an observed button press, to a control sound, and found both equally attenuated. In the second experiment, the N1 for sounds preceded by an observed button press was attenuated compared to sounds preceded by either an unrelated observed action, or no action, providing evidence for a specialized processing of human movement as a predictor of sensory consequences. However, atypical cue-sound intervals ranging from 500 – 1500 ms, the omission of a motor-only condition and the embedding of all stimuli in different stages of the additional intentional binding task makes it hard to compare these results to results from the usual contingent paradigm.

Two studies adapted observational versions of the contingent paradigm and measured N1 and P2 amplitudes for sounds produced by self-initiated and observed button presses, including an observational motor-only condition used for motor-correction. Ghio et al. (2018) displayed a three-picture animation on a computer screen that showed a button merely being touched by a finger, and subsequently partially and fully pressed in the observational conditions. When participants produced sounds themselves, they did so at the same desk seen in the animation and wearing the same lab coat as the observed actor. In Ghio et al. (2021), on the other hand, two participants sat side by side, and in alternating blocks either pressed a button themselves to produce a sound, or observed the other person doing it. While in the first study, on-screen observation led to an N1 and P2 attenuation, just like in the condition in which the sounds were self-produced, the second study reported no N1 attenuation when sounds were observed to be produced by another participant, only a P2 attenuation.

The differing N1 results could be an effect of the time available to visually process the observed button press. In Ghio et al. (2018), the partially pressed button picture, starting the animation, was shown around 175 ms before sound onset. Meanwhile, in Ghio et al. (2021), the sound was presented around 50 ms after the button press. The shorter interval between perceptible motion and sound might not have sufficed to process visual input, activate mirror neurons and generate sensory predictions until the time frame of the N1. In contrast, environmental external cues might be processed even faster, as they do not need require additional motor-related processing, which could be the reason that N1 attenuation for external cues were found more reliably. Studies that reported N1 attenuations based on external cues also used far larger intervals between the onset of a predictive stimulus and the following sound, up to 1500 ms (Kaiser & Schütz-Bosbach, 2018; Klaffehn et al., 2019; Lange, 2011).

Moving away from artificial sound production via button-presses, van Laarhoven et al. (2021) instead used video and audio of two hands performing a clap, with around 500 ms between motion and sound onset. In addition, and similar to Lange (2009) and Baess et al. (2008), the sound could vary both in temporal precision and stimulus identity (i.e., changed to diverse environmental sounds). While the N1 and P2 for all sounds following the clap motion were reduced relative to uncued sounds, unexpected temporal or identity characteristics diminished the N1 attenuation, confirming the results of Lange (2009) for external cues. The P2 attenuation was only diminished for changes in sound latency, but not sound identity.

Ultimately, even though it is interesting to see the results for sounds generated by observed action, none of these studies offered evidence for involvement of motor-specific processing beyond what has been shown for neutral external cues (Harrison et al., 2021; Kaiser & Schütz-Bosbach, 2018; Klaffehn et al., 2019; Lange, 2009). Drawing such a conclusion would require either additional externally cued control conditions or the manipulation of possible motor-specific influences, as shown by Poonian et al. (2015) in a different task design.

1.3. RESEARCH OBJECTIVES

The studies in this dissertation aimed to further examine the involvement of motor-specific information on the attenuation of the N1-P2 complex in the processing of action-generated sounds. The first study (Seidel et al., 2021) employed an illusion of control (Allan, 1980) to induce a low or high feeling of control over sound production via a button press, which also alters the perceived agency (Moore et al., 2009). In our design, the manipulation of control was performed in a separate task, prior to the conventional tasks of the contingent paradigm used to measure N1 and P2 attenuations. Using only the aftereffects of this manipulation enabled us to conduct two identical sets of the contingent paradigm tasks per participant, and only change the context they were performed in (low/high control). Previous studies that employed agency manipulations (Kühn et al., 2011; Timm et al., 2016) altered stimulus properties (tone frequency, latency) during the task to suggest a reduction of agency. As such alterations are also known to change the supposedly motor-specific N1 attenuation seen for self-produced sounds (Baess et al., 2008), our design was the first in which these bottom-up influences could be fully excluded. A more pronounced P2 attenuation in the high control condition would furthermore indicate that this attenuation is at least partially caused by top-down control.

The second and third studies comprised in this dissertation both tried to directly alter the motor-related information supplied to the forward model system. In the second study (Seidel et al., 2023), participants watched sound-generating button presses being performed on a computer screen either from the first- or third-person perspective. The mirror neuron system, which might supply forward models with motor-related information during action observation (Wolpert et al., 2003), has been shown to be more active for the first-person perspective (Angelini et al., 2018; Caggiano et al., 2015). The viewpoint of the participants should therefore also impact the forward model-related N1 attenuation for sounds produced during observation in this study. A more pronounced N1 attenuation during first-person observation could indicate that observed actions receive specialized processing beyond external cues.

In the third study (Seidel & Bellebaum, 2023), we aimed to diminish N1 attenuations for self-produced sounds by disturbing the efference copy information received by forward models. This was made possible by using eye movements as sound-producing actions, similar to Mifsud et al. (2016). Anti-saccades, i.e., eye movements in the opposite direction of the presented target stimulus, are known to be accompanied by the fast and automatic planning of saccades towards the target (Coe & Munoz, 2017). This motor plan must be suppressed for anti-saccades to be performed correctly, but we speculated that an efference copy for such a motor plan might still be relayed to forward models and compete with the efference copy of the executed saccade, reducing the precision of the resulting sensory prediction. Similar to sounds which deviate from their sensory prediction (Baess et al., 2008), sounds produced by anti-saccades should then show a diminished N1 attenuation, as their (non-deviating) properties should deviate from the disturbed sensory prediction.

2 OVERVIEW OF STUDIES

2.1. STUDY 1

2.1.1 Research Question and Hypotheses

For self-generated stimuli, perceptual intensity and neuronal responses are typically reduced compared to externally generated ones (Baess et al., 2008; Hughes et al., 2013a; Sato, 2008). It is thought that this sensory attenuation serves to distinguish stimuli caused by our own actions from those generated by external causes. The classic interpretation for both phenomena is that the cerebellum employs an internal forward model that predicts the sensory consequences of voluntary movements (Blakemore et al., 2001; Pickering & Clark, 2014; Wolpert & Flanagan, 2001). These predictions are thought to be based on efference copies of motor commands, enabling the early comparison of predicted and received sensory input, with matches resulting in altered neuronal processing and perception.

In the case of auditory stimuli, EEG studies have revealed attenuations of the auditory ERP components N1 and P2/P3 (Baess et al., 2011; Horváth, 2015; Hughes et al., 2013a; Schafer & Marcus, 1973), which seem to be functionally dissociated. Cerebellar lesion patients showed no attenuation of the N1, but of the P2, suggesting that only the N1 is dependent on the proposed cerebellar forward models. The P2/P3 attenuation on the other hand has been shown to be sensitive to influences of agency, i.e. the sensation of being the author of a stimulus. Two studies reported no difference in the N1 attenuation when agency was manipulated, but a diminished attenuation of the P2 (Timm et al., 2016) and P3 (Kühn et al., 2011) when agency was reduced.

Changes in the perception of agency have also been reported with a varying perceived contingency of actions and (sensory) outcomes (Moore et al., 2009). The contingency is the difference of the probability of an outcome occurring after an event, and its probability of occurrence without this event (Allan, 1980). If outcomes in both cases are equally likely, the contingency is zero, and the relationship is referred to as non-contingent. The illusion of control is a phenomenon of overestimated contingencies (Langer, 1975) and thus control, in non-contingent action-outcome relationships, which has been shown to be enhanced when the number of outcomes increases while the relationship remains non-contingent (Thompson et al., 2007).

In study 1, we used a two-button choice task to induce a low or high illusion of control, which was followed by conditions similar to those of the contingent paradigm while EEG was recorded. Participants repeated the procedure once with a high and once with a low illusion of control version of the preceding task. Keeping the conditions of the contingent paradigm identical while only changing the illusion of control task allowed us to measure only after-effects of the illusion of control induction on the N1 and P2. We expected that as in previous studies on agency (Kühn et al., 2011; Timm et al., 2016), this change in perceived control should not influence the N1, but the P2. More specifically, lower perceived control should lead to a diminished P2 attenuation compared to higher perceived control.

2.1.2 Methods

We tested forty participants, with a mean age of 25.4 years (*SD* = 3.5). In a within-subjects design, each participant was tested in two experimental blocks which were counterbalanced in their order. Both blocks started with an illusion of control induction task in which the participants were instructed to try to produce a desired compared to an undesired sound by pressing one of two buttons, even though participants exercised no real control. Instead, the probability of the desired sound appearing was identical for both buttons, making the design non-contingent. In the low and high illusion of control block, this probability for the (desired) auditory outcome (P[O]) was low and high respectively. The perceived level of control was captured with an explicit rating after each repetition of the illusion of control task.

The remaining conditions in each block, during which we recorded EEG data, were identical between low and high illusion of control blocks, and consisted of three tasks similar to the contingent paradigm (Horváth, 2015). In the auditory-motor condition, the desired sound was produced by a button press action (ACT-sounds) with a chance of 50% (or no sound appeared), while in the auditory-only condition, the desired sound was externally generated (EXT-sounds) and appeared as a playback off all sounds of the auditory-only condition. The motor-only condition was used to motor-correct the EEG activity in the auditory-motor condition, and consisted only of button presses that produced no sounds (motor-only).

After a standard EEG preprocessing procedure, we averaged the segments for the three tasks of the contingent paradigm for each illusion of control block separately, and performed a motor correction, by subtracting the average segments of the motor-only task from those of the ACT-sound tasks per block. We then localized N1 and P2 peaks in the overall grand averages for the (corrected) ACT- and EXT-sound task separately, but across both blocks (high and low P(O)). These peaks were used to construct time windows in which average N1 and P2 amplitudes were determined for each participant from their averaged ERP segments. As a behavioral measure of the strength of the induced illusion of control, we calculated the rating difference between the two ratings given after each repetition of the illusion of control task for each participant (high illusion of control task - low illusion of control task).

The EEG Data were analyzed with a linear mixed model including the factor Sound Type (ACT, EXT), outcome probability P(O) (low, high) of the desired sounds appearing in the prior two-button choice task, and the rating difference (RD) of the explicit level of control ratings between the illusion of control tasks. We included the maximal random effect structure that still converged for Participants (1 + Sound Type + P[O]) and Electrodes (Intercept).

2.1.3 Results and Discussion

The perceived control was higher in the high illusion of control and lower in the low illusion of control task, as intended, independent of the order of the two blocks. As hypothesized, we found that the N1 was not influenced by the illusion of control manipulation, and showed the typical reduction of ACT-sounds compared to EXT-sounds (Baess et al., 2011; Hughes et al., 2013a). The P2 on the other hand, showed the expected amplitude increasse in the low vs high illusion of control block for ACT-sounds, but not EXT-sounds. An interaction with the difference in control ratings (RD) between the two blocks revealed that this pattern was only exhibited by participants with high rating differences. This was in line with previous studies on agency influences on the N1 and P2 amplitudes, which also showed no contextual effect for the N1, only the P2/P3 (Kühn et al., 2011; Timm et al., 2016). Based on the conceptual distinction of the sensation of agency in an automatic feeling, and a conscious judgement of agency by Synofzik et al. (2008), we interpreted the N1 and P2/P3 as neural correlates of these two processes, respectively. Contrary to previous studies, we were able to show the contextual effect on the P2 without also inducing an agency ambiguity over each produced sound by varying sounds features. This indicates that the effect on the P2 does not results from a trial level agency attribution failure, to which P2/P3 amplitudes in the previous studies have been related (Kühn et al., 2011; Timm et al., 2016). Even though we carefully avoided stimulus timing or identity differences between illusion of control conditions, the first task in each block still differed in the total number of desired sounds. We thus cannot completely exclude that our results are in part a product of expectancy. Our finding that the P2 difference between illusion of control conditions was absent for participants that did not perceive a difference in control between them, however, makes that explanation unlikely.

2.1.4 Conclusion

We manipulated the perceived level of control in tasks of the contingent paradigm to examine the effects on the typical attenuation of the N1 and P2 for self- compared to externally-generated sounds. Similar to previous studies on agency (Kühn et al., 2011; Timm et al., 2016), we did not find an influence of the contextual influence of perceived control on the attenuation of the N1, but on that of the P2. Additionally, this attenuation only emerged for participants that reported a large perceived difference between the experimental conditions. This first study in this dissertation thus adds evidence to the idea that the later processing (P2) of self-produced sounds might be affected by contextual influences like the assumed agency and control, while only simpler, motoric information is integrated in the early processing (N1).

2.2. STUDY 2

2.2.1 Research Question and Hypotheses

The sensory attenuation phenomenon of reduced perceptual intensity and neurophysiological responses has also been investigated for sounds generated by observed actions (Ghio et al., 2021; Ghio et al., 2018). This stems from the assumption that mirror neurons, which discharge when performing an action as well as when observing it (Rizzolatti & Sinigaglia, 2016), might supply information to cerebellar forward models and engage similar predictive processes as self-performed actions (Miall, 2003). Two studies measured the N1 and P2 for sounds elicited by self-performed and observed actions: Ghio et al. (2018) showed a short animation of a button press from a first-person view on a screen, while in Ghio et al. (2021), two participants sat next to each other, producing and observing button presses alternately. While the P2 for observed-action-sounds was reduced in both studies compared to external sounds, the N1 was only reduced in the first. The short interval between button presses and sounds (~50 ms) in (Ghio et al., 2021) might have been insufficient to allow for the formation of sensory predictions until the early processing of the sound as reflected in the N1, unlike the animation to sound onset interval of around 200 ms in Ghio et al. (2018).

The second study in this dissertation addressed another difference between the two previous studies, the observer viewpoint. Non-human primate and human studies have shown that the mirror neuron system is likely more active during first-person than third-person observation (Angelini et al., 2018; Caggiano et al., 2015). If the N1 attenuation for sounds elicited by observed motor actions are a result of mirror neuron based predictive mechanisms, this could also explain why there was no N1 attenuation during third-person observation of another person (Ghio et al., 2021). To further investigate the influence of the observer viewpoint, we presented participants in this study with on-screen button press animations from a first-person and third-person perspective. Additionally, we included the time course of the experiment in the analysis to capture the learning process of the action-effect association between observed button presses and sounds.

We expected to find an N1 attenuation for sounds elicited by observed actions compared to external sounds regardless of the observer viewpoint, as we employed an animation to sound onset interval of 270 ms. But considering the larger activity of mirror neurons during first-person observation, we expected a processing advantage for this viewpoint, leading to a stronger N1 attenuation for the first-person perspective. For the P2 we also expected an attenuation as in both previous studies (Ghio et al., 2021; Ghio et al., 2018), but speculated that the first-person perspective might induce an ambiguity in agency attribution, resulting in a diminished P2 attenuation, since this visual input is most common in self-production, not observation.

2.2.2 Methods

We tested twenty-seven participants, with a mean age of 23.7 years (*SD* = 4.4), who watched the hand of a gender-matched person on a screen perform sound-producing button presses from a first- or third-person perspective. We recorded videos of a male and female actor in a white lab coat perform these presses in the same setting that participants later sat in and extracted ten pictures per condition for replay as a video animation at 30 frames per second. Additionally, we captured one frame per actor and viewpoint that showed them resting their hand on the button box, to be used in the condition in which external sounds without action were played.

Participants went through four experimental runs, during which we recorded EEG data, each containing three conditions in blocks: One with sound producing button presses (ACT-sounds, animation onset to sound onset: 300 ms), one showing only button presses without sounds (motor-only), and one in which external sounds were played while the still image of the resting hand was shown (EXT-sounds). The runs alternated in the observer viewpoint between first- and third-person. During the two conditions showing button presses, we also captured gaze data via eye tracking.

Before EEG data preprocessing, we first determined trials in all blocks containing button presses in which participants did not look at the moving finger to exclude these trials from further analysis. For this, we determined in each trial how many gaze position data points in the 300 ms time window after animation onset diverged more than 200 px from the position of the fingertip. Trials in which more than 25% of gaze data points fell outside this limit were removed. After a standard EEG preprocessing procedure, we performed a motor-correction procedure by averaging segments of the motor only condition and subtracting this average ERP from each single segment of the ACT-sound condition, separately for each viewpoint condition. We then created an overall grand average ERP over the (corrected) ACT- and EXT-sound segments and both viewpoints, to locate the overall peaks of the N1 and P2. Visual inspection also revealed a negative peak around 300 ms that differed between conditions, and has been reported as an N2 in connection with self-modulated auditory stimuli (Sugimoto et al., 2021). We thus also located the overall peak of the N2, and furthermore analyzed this component in the same manner as the N1 and P2.

Subsequently, segments for each condition/run/participant/electrode combination were separately averaged and individual N1, P2 and N2 peaks for each combination were identified, in time windows around the peaks from the overall grand average. In a last step, we calculated mean amplitudes for each single trial, in 40 ms time windows around the corresponding individual peaks for the N1, P2 and N2. In each condition/run/participant/electrode combination of the three amplitude data sets, we separately analyzed trials and removed those with amplitudes further than 2.5 SD from the mean. The dataset for each component was analyzed with a linear mixed model including the factor Sound Type (ACT, EXT), Viewpoint (first, third) and, to model the temporal course of the experiment, the factors Run (first, second) and Trialnumber (1-50). We included the maximal random effect structure that still converged for Participants (1 + Sound Type * Viewpoint + Run) and Electrodes (Intercept).

2.2.3 Results and Discussion

The results showed the N1 was reduced for sounds elicited by observed actions compared to external sounds irrespective of observer perspective. But a three-way interaction (Sound Type by Viewpoint by Trialnumber) revealed that while for third-person observation the N1 attenuation remained throughout the experiment, during first-person observation the N1 attenuation only emerged later in each run. For the P2 on the other hand, we found a Sound Type by Run interaction that revealed a P2 attenuation in the first, but not the second run, independent of Viewpoint. An additional analysis of the N2 showed a three-way interaction (Sound Type by Viewpoint by Trialnumber) as well. For first-person, there was an attenuation for sounds generated by observed actions compared to external sounds that increased over time, but for third-person, the attenuation was never significant.

In the typical interpretation of the N1 attenuation as function of internal forward models, the results of a slowly emerging N1 attenuation during first-person, compared to a constant attenuation during third-person observation, suggests a faster acquisition of action-effect associations during third-person observation. This is contrary to our expectation, as we assumed that the preference of the mirror neuron system for first-person observation (Angelini et al., 2018) would lead to a more pronounced N1 attenuation during first-person observation. It is instead possible that the more natural viewpoint of the third-person perspective in which we typically observe others is much better trained from real life action observation, and thus had a much shorter acquisition time for the action-effect association as the one during first-person observation. It is also possible that during action observation in first-person, the forward model system is trained to integrate not only visual, but also motoric and proprioceptive signals, which were missing here, leading to imprecise predictions, that had to be retuned over time. Ultimately, since sounds in both viewpoint conditions were equally predictable, it seems likely that these actions received processing beyond simply visual cues, as the results for both viewpoints should otherwise have been identical. For the P2 we did not find any difference between viewpoints, indicating that the perceived agency in both viewpoints might not have differed. We instead found an N2 attenuation that only appeared during first-person observation, that might be related to a stronger feeling of (self-) agency for this viewpoint.

2.2.4 Conclusion

We examined the effect of the observer perspective on the N1 and P2 attenuation for sounds generated by observed actions compared to external sounds. For this, participants observed sound generating button presses on screen in first- and third-person perspective and listened to externally generated sounds. While we found a general N1 attenuation, as in a previous study (Ghio et al., 2018), interactions revealed a viewpoint dependent temporal dynamic. The attenuation of the N1 in the third-person perspective was stable over time, but for first-person observation we only found a slowly emerging attenuation over time. Since third-person observation is the more familiar viewpoint in everyday life, this previous experience might have led to a much faster acquisition of the action-effect association. For the P2 we replicated the attenuation for sounds elicited by observed actions compared to external sounds found in previous studies, and found no difference between viewpoints.

2.3. STUDY 3

2.3.1 Research Question and Hypotheses

The necessity of motor information for the generation of sensory predictions reflected in the N1 attenuation is still in question: Not only has the N1 attenuation been shown without the involvement of motor activity (Harrison et al., 2021; Kaiser & Schütz-Bosbach, 2018; Klaffehn et al., 2019), but Dogge et al. (2019) have also argued that the formation of action-effect associations needed for accurate predictions would take longer than the typical duration of the experimental paradigms used to study them. However, the button-press action typically used in ERP attenuation studies (Baess et al., 2011; Hughes et al., 2013a; Knolle et al., 2013a) is very familiar from the use of modern devices, which often have buttons that either mechanically generate noise or play sounds as feedback, such as keyboards, microwaves or even light switches. It is thus possible that forward-models are already well trained for such an association, and only need minimal retuning for the specific stimuli of the experimental setting.

Mifsud et al. (2016) have demonstrated that it is possible to learn an entirely novel action-effect association during one experimental setting. They trained participants to generate sounds by saccadic eye movements, which does not occur in every-day life, and found the N1 attenuation for self- compared to externally generated sounds, albeit less pronounced than for button-press generated sounds. The use of saccades for sound production is also interesting because this paradigm can be extended to antisaccades, which are eye movements in the opposite direction of a target stimulus. While motoric execution for pro- and antisaccades is identical, it is assumed that prior to antisaccades, a reflexive saccade toward the target has to be suppressed (Coe & Munoz, 2017). If an efference copy of the motor plan for this suppressed saccade is created, it might conflict with the efference copy of the executed saccade, and lead to imprecise predictions of the sensory consequences, which should be reflected in a diminished N1 attenuation.

In the third study included in this dissertation, we aimed to compare N1 and P2 amplitude attenuations for pro- and antisaccade-generated sounds and capture the acquisition of the novel saccade-sound action-effect contingency. For this, we employed the same statistical model as in the previous study (Seidel et al., 2023) and included the temporal dynamic of ERP amplitudes over the course of each experimental run. As it is necessary to prompt antisaccades with a suddenly appearing target, we also prompted pro-saccades in the same way. Furthermore, to ensure similar visual stimulation in all conditions, the external sounds in this study were also cued by the visual stimuli used as targets in the saccade-generation conditions

We expected an N1 attenuation for prosaccade-generated sounds vs. cued external sounds, replicating the results of Mifsud et al. (2016). We furthermore expected a diminished N1 attenuation for antisaccade-generated sounds caused by conflicting efference copy information. The acquisition process of the novel saccade-sound action-effect contingency should be reflected in the N1 amplitude as a decrease over the course of the experiment. For the P2 we did not expect an attenuation of saccade-generated compared to cued external sounds, in accordance with Mifsud et al. (2016). Considering that the P2 seems independent from cerebellar forward models (Knolle et al., 2013a) we also did not expect a difference between pro- and antisaccade-generated sounds caused by conflicting efference copy information.

2.3.2 Methods

38 participants, with a mean age of 24.8 years (*SD* = 4.5) took part in this study. In a within-subjects design, participants went through four experimental blocks, during which we recorded EEG data. Each block contained three conditions similar to the conditions in the contingent paradigm in a fixed order: an action-sound condition in which sounds were produced by performing visually prompted pro- and antisaccades, an identical action-only condition in which no sounds were produced, and a cue-sound condition in which the same cues and sounds as in the prior action-sound condition were replayed. As this paradigm employed cued instead of uncued sounds, we included a fourth condition in each block that was used to control for visual stimulation, equivalent to the motor-only condition in the contingent paradigm. In this cue-only condition, we presented the same cues as in the cue-sound condition, but without any sounds.

The blocks alternated between those in which prosaccades were performed in the two action-conditions, and those which required antisaccades. In action-conditions, participants waited until a fixation dot in the middle of the screen vanished and was followed by a target dot to the left or right. Depending on the saccade block condition, they performed a prosaccade to the target dot, or an antisaccade in the opposite direction. If the saccade was performed in the correct direction depending on task, a sound was played 170ms after the gaze position on the screen had moved beyoned the halfway point to the intended saccadic landing spot (i.e. on or opposite to the target point). In the two cue-conditions (with or without external sounds), the visual stimulation was nearly identical. However, the initial fixation dot did not appear in the middle of the screen, but in the same position as the following target dot, which is why no saccades were performed in those conditions. Fixation and target dot positions, as well as the sound timings were recorded in the action-sound condition and used for positions and timings in the following cue-conditions.

After EEG preprocessing, we performed a motor-correction of the action-sound condition and an equivalent visual-correction of the cue-sound condition as in study 2. All further analyses were conducted on these corrected action-sound and corrected cue-sound conditions. Next, a grand average was calculated, including all action-sound and cue-sound conditions to extract the overall N1 and P2 peak timings. To extract individual N1 and P2 peaks, we first averaged the segments for each condition/block/participant/electrode combination separately. We then identified individual N1 and P2 peaks around those from the overall grand average. In 40 ms time windows around these individual peak latencies, we calculated the mean amplitude for the N1 and P2 in each single corresponding trial, and then removed outliers which differed by more than 2.5 SD from the mean for each condition/block/participant/electrode combination separately. The ERP amplitudes were analyzed with a linear mixed model including the factor Sound Type (cue-sound, act-sound), Saccade Type (prosaccades, antisaccades), the factor Run to encode the repeated experimental blocks (first, second) and the Trialnumber (1-40) in each block. We included the maximal random effect structure that still converged for Participants (1 + Sound Type * Saccade Type * Run) and Electrodes (Intercept).

2.3.3 Results and Discussion

Resolving the significant Sound Type by Trialnumber interaction we found for the N1 revealed an overall decrease of the N1 amplitude over time for action-sounds, but not for cued externally-generated sounds. An additional four-way interaction, including all factors (Sound Type by Saccade Type by Trialnumber by Run), revealed that only for prosaccades (in the second run) the attenuation of the N1 over time for action-sounds led to significantly reduced amplitudes compared to cued sounds. For antisaccade-generated action sound amplitudes the pattern did not end with a significant attenuation.

For the P2 we found generally enhanced amplitudes for action-sounds compared to cue-sounds, but again a four-way interaction of all factors. Resolving this interaction showed for antisaccades that while amplitudes for cue-sounds did not change over time, action-sound amplitudes increased in the first run over time, and then decreased in the second run. Action-sound amplitudes were enhanced compared to cue-sounds at the end of the first run, and during the entire second run. For prosaccades, we only found a Sound Type by Trialnumber interaction independent of Run, and simple effects analyses revealed significantly increased P2 amplitudes for action-compared to cue-sounds at the start of runs, but no significant difference at the end.

The N1 attenuation for prosaccade-generated sounds compared to cued-sounds, that was revealed by the four-way interaction, is in line with the previously found N1 attenuation in Mifsud et al. (2016), and our expectations. We additionally wanted to examine whether the sensory predictions this attenuation possibly reflects are based on motor information, or just visual cues. We speculated that for the antisaccade-generated sounds, sensory predictions could be disturbed, as motor information in the form of efference copies might conflict between the executed antisaccade, and the reflexive prosaccade that first has to be suppressed. The missing N1 attenuation for antisaccade-generated sounds could be indicative of such a distortion, resulting in imprecise sensory prediction, as mismatches between predicted and actual sensory outcomes of actions have been shown to reduce the N1 (Baess et al., 2008).

That the attenuation of the N1 for prosaccade-generated sounds was only found at the end of the experiment might be explained by a slow acquisition process of a novel action-effect contingency between saccades and sounds as suggested by Dogge et al. (2019). For button press-generated sounds, as used in most studies (Baess et al., 2011; Ghio et al., 2021; Hughes et al., 2013a), the forward models that are assumed to provide the predictions that underly the N1 attenuation could potentially rely on button press-sound contingencies familiar from every-day life.

We did not expect to find significant differences in the P2 amplitude between action- and cue sounds, similar to Mifsud et al. (2016). The overall significant P2 enhancement for action-sounds we found could be explained as an effect of the comparison with cue-sounds, which are known to already be reduced (Sowman et al., 2012) compared to the uncued sounds used by Mifsud et al. (2016). Considering previous reports on diminished P2/P3 attenuations under agency ambiguity (Kühn et al., 2011; Timm et al., 2016), our result of an enhancement mostly for anti-, but not prosaccade-generated sounds might reflect a diminished sense of agency during sound generation by antisaccades.

2.3.4 Conclusion

Participants in this study generated sounds by performing pro- and antisaccades instead of the typical button presses and listened to visually cued sounds. Our results of a missing N1 attenuation for anti-saccades might be an effect of disturbed sensory predictions, caused by conflicting efference copies for antisaccades, and the reflexive prosaccades that have to be suppressed before their execution. The slow temporal dynamic of the N1 attenuation could indicate that forward models take longer to establish the novel saccade-sound association than the familiar button press-sound association. We also found a P2 attenuation that gradually developed over time, but only for prosaccades. For antisaccades, we instead found a largely enhanced P2 amplitude compared to cued externally generated sounds, which could be interpreted as loss of agency.

3 GENERAL DISCUSSION

The three studies comprised in this dissertation examined the influence of context and motor information on the processes underlying the neuronal response to sounds in the environment. This did not only include a comparison of self- and externally-generated sounds, as in classical studies (Baess et al., 2008; Hughes et al., 2013a), but also sounds cued by visual stimuli, as well as sounds generated by the actions of an observed actor.

In the first study, we employed an illusion of control to change the context in which sounds were generated, inducing a belief in participants that they had more or less control over the generation of sounds when pressing buttons. While there was no impact on the attenuation of the N1 for self-generated sounds, the P2 was sensitive to the illusion. In addition, the pattern in the P2 also reflected the self-reported ratings of perceived control.

In the second study, we examined the impact of the observer perspective while participants observed sound producing button presses of another person on a screen. Differing temporal dynamics of the N1 while observing from a first- and third-person perspective indicated that observed actions are not processed like other visual cues, as the identical temporal predictability for both viewpoints should otherwise have led to a similar response of the N1. This adds evidence to the idea that observed actions engage the mirror neuron system, providing a representation of the observed motor activity to predictive processes.

The third study moved away from standard button presses to elicit sounds, and compared the processing of pro- and antisaccade-generated sounds to the processing of visually cued external sounds. We found that N1 amplitudes for both pro- and anti-saccade generated sounds were reduced over time, but a significant attenuation compared to cued sounds only emerged for prosaccade-generated sounds. This indicates a predictive processing disadvantage for anti- compared to prosaccade generated sounds that is independent from temporal predictability. As anti-saccades are thought to require a suppression of an automatically generated saccade towards a target, we suspected a disturbance of predictive processing for antisaccade-generated sounds caused by efference copies for the executed and suppressed saccade.

In the following I will discuss the two major topics that these three studies added evidence for, whether the N1 and P2 reflect different aspects of the sense of agency, and the impact of motor information in the generation of the sensory predictions that are thought to underly the N1 attenuation. At the end I will also comment on the theory of predictive coding, and its implication for the field of attenuated neuronal responses for self-generated stimuli.

3.1. N1 AND P2 ATTENUATION AS MARKERS OF AGENCY

In the first study we manipulated the perceived level of control over sound production in the contingent paradigm, similar to previous agency studies (Kühn et al., 2011; Timm et al., 2016). Attributions of agency have been suspected to be the goal of the classification into self- or externally generated stimuli by the forward model system (Picard & Friston, 2014; Wolpert & Flanagan, 2001). Two previous studies showed that the N1 and later positive components (P2/P3) were differently affected by manipulations of agency: Kühn et al. (2011) changed the latency and frequency of sounds, while suggesting that each sound might be produced by the participant, or the experimenter. Timm et al. (2016) used an illusion to reduce the perceived agency that relied on variance of sound latency as well. In both studies, a larger attenuation of later positive components (P2/P3) was found in conditions with high agency, and a reduced P2/P3 attenuation if the perceived agency was reduced. These studies also gathered explicit ratings of agency and found that the P2/P3 reflected the explicit ratings. The N1, on the other hand, was only reduced for self-produced compared to externally-generated sounds, and not affected by the perceived agency.

Our study did not directly manipulate or measure agency, but the perceived control over sound production, which was induced by a control illusion task that likely also affected agency (Thompson et al., 2007). Crucially, we induced the high or low illusion of control in a separate task before the main task. Fixing the chance of sound production in the main task at 50% allowed us to employ only aftereffects of the control illusion task. This means that during the EEG recording, we did not have to change sound parameters, or suggest that sounds were produced by another agent, as in the previous studies (Kühn et al., 2011; Timm et al., 2016). The design in our study thus ruled out potential confounds that may have skewed results in these previous findings.

Still, our results were in line with the pattern of results in these previous studies, in that the later positive P2 amplitude was decreased under a higher illusion of control, and furthermore showed that the pattern only appeared in participants who actually distinguished between the low and high control conditions with their subjective ratings. Crucially, this makes it unlikely that the P2 attenuation pattern could have resulted from the difference in the number of presented desired sounds in the illusion of control task of each block, or the ratio of desired to undesired sounds leading to differing action-effect associations. Compared to the two previous studies (Kühn et al., 2011; Timm et al., 2016), we could further exclude that variations in stimulus features, such as frequency and latency, impacted the P2 attenuation. The pattern of the N1 confirmed the typical attenuation for self- compared to externally generated sounds which was also found in these previous studies.

Overall, study 1 provided further evidence that the P2 attenuation is based on predictive processes that integrate context information, in contrast to the N1 attenuation, which is largely sensitive to the availability of motoric information, and basic stimulus features (Baess et al., 2008; Jack et al., 2021; Knolle et al., 2013a; Kühn et al., 2011; Timm et al., 2014a; Timm et al., 2016). This dissociation also mirrors the division of the sense of agency by Synofzik et al. (2008) into an automatic feeling of agency, and a conscious judgement of agency, which could be reflected in the N1 and P2/P3 respectively. The judgement of agency is furthermore thought to be an integration of predictive and postdictive information (Synofzik et al., 2013).

Unlike the paradigms used by Timm et al. (2016) and Kühn et al. (2011), the design in study 1 allowed us to show a clear influence of context-related predictive information on the P2 attenuation. The previous studies created uncertainty about the author of each single sound, via explicit instruction at the beginning, and by changing sound identity or latency in each trial. The P2/P3 was then related to after-trial agency ratings which should reflect the judgement of agency, making it impossible to disentangle the influence of pre- and postdictive information.

In our task on the other hand, we never suggested that successfully produced sounds were not authored by the participants, neither by explicit instruction, nor by varying stimulus features. By not introducing uncertainty about agency and varying sound features, the postdictive decision process in the judgement of agency should not have been different between trials, leaving only predictive information as a source of the decreased P2 amplitudes for high compared to low control conditions. In line with this notion, we were able to relate this P2 difference to the ratings of control made before the start of the main sound production task, and thus directly to prior context information.

There is some clinical evidence that changes in conscious agency do rely on early processing as reflected in the N1. Schizophrenic symptoms, like hallucinations, have been described as failures to attribute agency (Swiney & Sousa, 2014). This has been related to impaired efference copies and reflected in a reduced N1 attenuation for button press-produced sounds (Ford et al., 2014), as well as for vocalizations (Ford et al., 2007; Heinks-Maldonado et al., 2007). Even though the N1 attenuation has not been shown to be directly correlated with symptom severity (Ford, 2018), the general impairment of the N1 attenuation in schizophrenic patients might indicate that low-level predictive processing is taken into account for conscious agency judgements.

The attribution of agency has also been suspected to underly the P2 attenuation during the observation of actions of other people (Ghio et al., 2021; Ghio et al., 2018). But in this case, the relevant agency judgement is likely made about whether the observed person caused the stimulus, not oneself. In study 2, we also found a P2 attenuation for sounds generated by an observed action compared to external sounds. We did not find any difference in the P2 attenuation for sounds produced by actions observed from a first- vs. third-person perspective, suggesting no agency difference between viewpoints. This is in line with two previous study that each only used a single viewpoint and reported comparable P2 attenuations (Ghio et al., 2021; Ghio et al., 2018).

In study 2, we also reported an exploratory analysis of the N2 component that notably displayed a similar pattern to the N1. It revealed a gradually increasing attenuation for observed-action-generated compared to external sounds, but only during first-person observation. Unlike in the analysis of the N1, there was no N2 attenuation during third-person observation. An amplitude increase of this component for sounds with diminished predictability has been reported a few times (Knolle et al., 2013b; Pinheiro et al., 2019; Sugimoto et al., 2021), but an examination in an agency-focused paradigm is still pending. If later ERP components are related to more conscious aspects of agency, the N2 might reflect an agency attribution to oneself.

In study 3 we also offered an interpretation of the results for the P2 as an effect of agency. P2 amplitudes for prosaccade-generated sounds were only significantly enhanced compared to cued sounds at the beginning of the experimental blocks, but not at the end. For antisaccade-generated sounds however, P2 amplitudes in the first block even increased, and were then enhanced compared to cued sounds during the entire second block. Considering the possible influence of impaired efference copies on agency judgements in schizophrenic patients discussed above, the efference copy conflict we proposed for antisaccade-execution might have directly impacted the perceived agency, resulting in this pattern of an increased P2.

To summarize the findings on the N1 and P2 attenuation as markers of agency, we provided further evidence for the idea that early and late processing stages for self-generated sounds reflect different aspects of the sense of agency in study 1. We additionally showed that the P2 amplitude, which is assumed to be related to the more conscious judgements of agency, reflected the explicit judgements about the context sounds were produced in. In contrast to previous studies that reported similar results (Kühn et al., 2011; Timm et al., 2016), we did not induce postdictive ambiguity about stimulus authorship, and thereby showed that this reduction of the P2 attenuation does not depend on a failure to attribute agency in a postdictive decision process. In study 2, we reported another late component, the N2, which displayed an attenuation for observed action-generated sounds compared to external sounds specifically during first-person observation, possibly reflecting an agency attribution to oneself. An increased P2 for self- compared to cued external sounds was also found in study 3 when sounds were generated by antisaccades, possibly indicating diminished agency.

3.2. MOTOR BASED FORWARD MODELS IN THE N1 ATTENUATION

Studies 2 and 3 were focused on further examining the role of motor information and temporal predictability in the attenuation of the N1. N1 amplitudes for visually cued stimuli have been shown to be similarly reduced (Harrison et al., 2021; Kaiser & Schütz-Bosbach, 2018; Lange, 2009) as those for self-produced stimuli (Baess et al., 2008; Ghio et al., 2018; Hughes et al., 2013a). This indicates that temporal predictability might be sufficient to explain the N1 attenuation effect, questioning the classic forward model based on motoric information. However, Klaffehn et al. (2019) had shown that when temporal predictability was identical for self-produced and cued sounds, the N1 for self-produced sounds was still reduced in comparison. Temporal predictability was achieved in this study by cueing all sounds with an animated loading bar, including self-produced sounds, which were elicited by self-timed button presses that were followed by the loading bar.

In addition to temporal predictability, it also was proposed that temporal control over the appearance of a sound could play a role in the attenuation of the N1 (Hughes et al., 2013b). To examine the effect of temporal control, Harrison et al. (2021) used a novel design that included cued as well as uncued conditions for self-and externally-generated sounds. At the same time, the cueing was based on a continuous visual stream, that enabled identical temporal predictability for self- and externally-generated sounds.

Visual cueing was accomplished by presenting four rows, each containing multiple white vertical lines that moved horizontally at the same speed. Additionally, a red fixation line was positioned in the middle of the screen, spanning all four rows. In conditions that included temporal predictability, all rows contained the same lines, moving synchronously, making them appear as lines spanning all four rows. Each time such a line reached the fixation line in the middle, an external sound was played, or the participant was instructed to press a button to generate a sound.

In the uncued conditions, the lines in the four rows were distributed randomly but still moved at the same speed. This created a visual stream consisting of four lanes of moving vertical lines, none of which reached the fixation line at the same time. Uncued button presses were self-timed as this stream was presented, and external sounds onsets were not timed to any line reaching the fixation line. The advantage of this design was that the visual stream in cued and uncued conditions was kept very similar, and in addition to the typical motor-only condition, a visual-only condition was included to correct the ERPs in the external sound condition.

They showed that uncued self-produced sounds and cued external sounds were equally attenuated compared to the typical uncued external sounds. The fact that cued self-produced sounds were attenuated even further was interpreted as a sign that the higher temporal control for uncued self-initiated sounds (the default for self-production in other studies) is reflected in the N1 as an enhancement. By this logic, the classic N1 attenuation of uncued external vs uncued self-generated sounds would still include the effect of motoric forward model predictions, as the attenuation by temporal predictability and the enhancing effect of temporal control in self-produced sounds would effectively cancel each other out. However, the even lower N1 amplitude for cued self-generated sounds can also simply be interpreted as an additive attenuation effect of forward models and temporal predictability on the N1. In this view, independent predictive systems, working on different input information, each attenuate the N1 comparably, while a combination of both reduces the N1 even further.

In study 2, we also compared cued to uncued external sounds, but with observed actions serving as cues. This was thought to activate additional predictive processing, as the mirror neuron system might provide motor representations for the observed actions, that are relayed to forward models to enable motor-based sensory predictions. The temporal dynamic we found for the N1 indicated an involvement of such motoric processing, as the N1 attenuation for sounds generated by observed actions compared to external sounds differed based on the observer viewpoint, which should not impact temporal predictability. For the third-person perspective condition, the N1 attenuation was found throughout the experiment. During first-person observation however, the N1 attenuation was not found at the beginning of the experimental blocks, and only gradually built up over time.

One explanation for this could be that during first-person observation, the forward model is well trained to integrate visual, proprioceptive and motor signals that usually occur together, when we see ourselves performing an action. As the corresponding motor and proprioceptive signals were missing, the retuning of the model could have taken more time than for the third- person perspective, in which the combination of multisensory signals was as expected when observing someone else.

Another explanation for the gradual buildup of the N1 attenuation during first-person observation could be a slower acquisition of the action-effect association between the observed button press and the sound, which is necessary to generate accurate sensory predictions through forward models. Without an accurate prediction, comparing the imprecise prediction to the sensory input would create a slight mismatch, and should lead to a reduced N1 attenuation. This is similar to studies that presented sounds with slightly changed frequencies, in which a comparison between an accurate prediction and an (inaccurate/deviating) sensory input leads to a mismatch reflected in the N1 attenuation (Baess et al., 2008; Hughes et al., 2013a). A slower acquisition during first-person observation might indicate that the learning of action-effect associations through visual stimuli is specialized for a third-person perspective, as this is what we encounter most frequently in everyday action observation.

Dogge et al. (2019) argued that in contrast to body-related effects, environmental effects of actions likely have a longer acquisition time. Our results are in line with this idea, considering that we only found the difference between the viewpoints in the temporal dynamics of the N1 amplitude, and would otherwise have missed this effect. However, Dogge et al. (2019) used this argument to exclude forward models as an explanation for the typical N1 attenuation found for self-generated compared to external sounds, because learning the necessary environmental action-effect association should take longer than one experimental session. It is important to note that most studies investigating the N1 attenuation use button press generated sounds (Baess et al., 2008; Ghio et al., 2018; Harrison et al., 2021; Hughes et al., 2013a), which as an action-effect association is very likely well trained through the handling of modern devices such as keyboards or microwaves. Retuning an action-effect to a specific sensory outcome might simply be faster than establishing an entirely new association. The different temporal dynamics of the N1 in study 2 would be in favor of this argument. While the association of observed actions in third-person and resulting stimuli might be well trained, and allow for a fast retuning to a new stimulus, first-person observation of other people is not occurring often in everyday life.

However, the acquisition process of novel action-effects that have never been trained can be a very useful tool for researchers examining the speed of such acquisitions. While button-press generated auditory and visual stimuli serve this purpose well in animal studies, this action-effect association is likely too well trained for human participants. Even foot-generated sounds might be too familiar (van Elk et al., 2014), as we experience many different types of footsteps on different materials. Saccades however, as they have been used by Mifsud et al. (2016), and our third study, are associated with visual consequences, but never with any type of sound. The results based on this unfamiliar action-effect contingency in study 3 mostly showed a similar N1 pattern for action-generated sounds as the possibly untrained first-person perspective in study 2: N1 amplitudes for action-generated sounds decreased over time, while those for (cued) externally generated sounds remained stable. Thus, both conditions might reveal the underlying time course of establishing a new action-effect association, which is usually hidden in studies that do not include such trial-level time related factors.

It is interesting that the temporal dynamic for visually cued sounds does not indicate the learning process of a new cue-effect association. This could either mean that visual to auditory effects in the environment are well trained, and the respective model can be retuned easily, or that the predictive systems relying on visual vs motor signals generally have differing learning rates. Indeed, just as Dogge et al. (2019) suggested that cause-effect relationships between body parts are likely learned faster than effects of the body on the environment, associations between environmental cues and environmental effects might be learned faster.

For prosaccade-generated sounds, the N1 amplitude was eventually attenuated compared to cued externally generated sounds, indicating a successful tuning of forward models for this unfamiliar action-effect contingency. Since cue to sound latencies were set to the same timings that were produced by the participants in the preceding self-production condition, the N1 attenuation indicates an effect beyond the temporal predictability of the cued sounds. For antisaccade-generated sounds however, the decreasing N1 amplitudes over time did not end with a significant attenuation compared to cued sounds. This could mean that the experiment simply was not long enough to capture the entire acquisition process, and/or that the opposite prosaccade that is assumed to be suppressed before each antisaccade hindered this process. The alternative would be that the conflicting motor information led to a permanently imprecise sensory prediction, resulting in an overall predictability that never exceeds the effect of temporal predictability.

Our result for prosaccades matches the one of Harrison et al. (2021), who also showed that cued self-generated sounds elicit a smaller N1 amplitude than visually cued sounds. Regarding the question of whether this is caused by separate effects of temporal predictability and temporal control, or an additive effect of motoric and general temporal prediction, our overall results point to the first. In the antisaccade condition, and the second run of the prosaccade condition, the N1 amplitude was elevated at the start of each block compared to the cued sounds. If the predictive effect of the visual cue prompting the saccades was additive with the motoric forward model predictions, and the latter were yet missing at the beginning of blocks, the N1 amplitudes for saccade-generated sounds should have been on the same level as those for cued sounds.

Overall, studies 2 and 3 added not just further evidence for the involvement of motor specific predictions in the process attenuating the N1, but also revealed the temporal dynamics of N1 amplitudes for action-generated, cued or uncued externally-generated sounds. For externally generated sounds, cued and uncued, the N1 amplitude seems to change less over time than for sounds produced by own or observed actions, especially those without a familiar action-effect association. This not only provides novel evidence for the idea that the classic N1 attenuation for self-produced compared to uncued external sounds at least in part reflects motor-based forward model predictions, but also indicates that temporal predictability might stem from a distinct predictive system.

3.3. PREDICTIVE CODING

While the predictive forward models as the source of sensory predictions and their possible processing to form agency attributions are often the focus of literature examining the neuronal responses to self-produced sounds, there is another theory concerning the processing of predictions in the cortex: predictive coding (Friston, 2005, 2010; Heilbron & Chait, 2018; Rao & Ballard, 1999; Schröger et al., 2015; Walsh et al., 2020). According to this idea, predictions descend in a hierarchical system of processing levels, and are used to generate prediction errors at every level, which are carried upward, perpetually refining predictions and ultimately minimizing the prediction error (Friston, 2010). To accomplish this, it is assumed that there are two different neuron populations at every level, that are located in different cortical layers: pyramidal cells in deeper layers encoding predictions, and pyramidal cells in superficial layers encoding the prediction error. While there is valid criticism of the theory of predictive coding and its proposed wide scope in explaining brain function (Heilbron & Chait, 2018; Hodson et al., 2023; Kogo & Trengove, 2015; Walsh et al., 2020), it offers an interesting view on phenomena of auditory perception discussed in this dissertation.

The proposed continuous relay of prediction error signals in superficial cortical layers through pyramidal cells that are oriented perpendicularly to the cortex surface would suggest that neuron populations that encode prediction errors are physically very well reflected in event related EEG signals (Feldman & Friston, 2010; Friston, 2010). This could mean a fundamental change of the interpretation of ERP components like the auditory N1, from a signal reflecting the magnitude of auditory processing which can be attenuated by a secondary predictive process, to a signal of the thus created prediction error signal itself. Indeed, for an uncued external sound, as often used as a control condition, there would be no difference between the neuronal response as a signal of maximal sensory processing and maximal prediction error signaling indicating that there was no prediction for this stimulus. And accordingly, as sounds are increasingly expected or predicted, the error signal would diminish, or in the classical view, the sensory processing would be attenuated, both resulting in a reduced N1 amplitude.

While there appears to be no discernible difference in these basic patterns of the N1, predictive coding makes further predictions that are different from those of feedforward models. One of the biggest differences concerns the question what happens if a predicted stimulus is omitted. Predictive coding would predict a large error signal, which could result in a measurable N1, as the difference between the expected auditory stimulus and the received silent sensory input is similar to the difference in case of an entirely unexpected auditory stimulus. Meanwhile, forward models would not predict the appearance of an N1, because without the appearance of a sound, no sensory processing should be reflected in the N1 at that moment in time. Furthermore, forward models only modulate processing when the prediction error is small, so in the case of an omitted sound, and thus a large prediction error, no modulation should occur. Studies have shown deflections in the EEG signal that resemble those for actual sounds when sounds are expected but omitted, as assumed by predictive coding, both for omitted sounds otherwise elicited by actions (SanMiguel et al., 2013a; SanMiguel et al., 2013b), as well as omitted sounds in regularities (Hsu et al., 2015; Janata, 2001; Wacongne et al., 2011).

In the case of a stimulus omission, the prediction error is a mirror of the prediction for that stimulus, indicating that a highly predicted stimulus should elicit a larger response than a stimulus for which no specific prediction has been formed. This pattern was reported by SanMiguel et al. (2013b), in an N1 omission response that was only elicited by otherwise self-produced sounds that appeared with a high probability, and not at chance level. Another study showed an omission N1 when a button otherwise produced the same sound, but not when it produced a random sound in every trial (SanMiguel et al., 2013a), indicating a stimulus identity specificity for the prediction, just as in studies on the N1 attenuation (Baess et al., 2008; Hughes et al., 2013a).

There has also been evidence that EEG signals for predicted externally produced sounds, for which no agency can be attributed, as for self- (or other-) produced sounds, display a processing hierarchy between earlier and later ERP components. Wacongne et al. (2011) presented series of five sounds, of which the last sound could deviate from the first four, that were identical. In different blocks, the deviant could appear with a higher or lower probability, resulting in a local rule (the last sounds should be identical to the four preceding sounds), and a global rule (the last sounds should be the one that appears more often as the last sound in this block). They found that deviations from the local rule were reflected in a larger early negative response, in this case the mismatch negativity (Fitzgerald & Todd, 2020), while deviations from the global rule led to an increased later positive response in the P3. This pattern is similar to what we found in study 1, where a later positive response in the P2 was sensitive to purely contextual information. That this pattern was also found without the involvement of motor actions would suggest that our results in study 1 might be explained by a general higher level contextual predictive process instead of an attribution of agency.

A key feature of predictive coding is the weighting of prediction errors according to their precision, and by increasing or decreasing their amplitude, changing their impact on the retuning of predictions (Friston, 2010). This weighting has not only been related to the predictive precision, but also to attention (Feldman & Friston, 2010), and the mirror neuron system (Friston et al., 2011; Kilner et al., 2007). The mirror neuron system was proposed to encode action intention, regardless of actor, and it was assumed that prediction errors are weighted down for observed actions and magnified for self-performed actions. While speculative, this offers another view on the slow buildup of the N1 attenuation during first-person observation in study 2. As the first-person perspective is closer to the visual input received during self-performed actions, it might promote a magnification of the prediction error, resulting in an increased N1 amplitude at the beginning of blocks, until the prediction error is minimized. The sounds produced in the third-person perspective condition, on the other hand, might have received a downward modulation of their prediction error signal.

Similarly, the increased N1 amplitude for self-produced sounds at the start of (most) blocks in study 3 might stem from a magnified prediction error, that is employed as an increased signal to retune predictive models, and facilitate the fast acquisition of the new action-effect association. With self-produced as well as external sounds in study 3 being cued, we wondered why the predictive attenuation effect of the visual cue and the motoric forward model was not additive. This would have resulted in N1 amplitudes for cued self-produced sounds starting at least at the level of cued sounds and decreasing over time from there, but they instead started significantly enhanced. Additive predictions errors, on the other hand, could explain this enhancement, especially if the action-related prediction error is magnified.

All in all, the theory of predictive coding offers an interesting view on previous results, but how well it fits the considerable evidence for cerebellar forward models (Blakemore et al., 2001; Dogge et al., 2019; Horváth, 2015; Jack et al., 2021; Knolle et al., 2012; Knolle et al., 2013a; Timm et al., 2014a) in the attenuation of neuronal responses needs to be further examined. It does, however, make notable predictions that could be important to future studies employing the contingent paradigm.

The hierarchical structure of predictions in predictive processing indicates that at a high level, explicit and conscious knowledge about upcoming stimuli should inform higher order predictions, that lead to measurable, reduced prediction errors if predictions are met. This could be tested in paradigms where the identity of each upcoming (self-) produced stimulus is either announced or left unknown. In the contingent paradigm, this is already practiced, as participants are typically explicitly instructed about the activities and stimuli in upcoming blocks, which could have an unknown effect on later processing.

A bigger problem for tasks in the contingent paradigm is that the expectation of stimuli might still be expressed in EEG signals even when no sounds are presented, as shown in omission studies (Janata, 2001; SanMiguel et al., 2013a; SanMiguel et al., 2013b). This could be relevant not only to the typical motor-only condition, which is used to motor-correct the EEG signal of self-produced sounds, but also the increasingly used and equivalent visual-only condition (Harrison et al., 2021; Seidel & Bellebaum, 2023; Seidel et al., 2023; van Laarhoven et al., 2021). The time course of these conditions should be further investigated, similar to the temporal dynamics reported in studies 2 and 3. This might reveal that early trials at the start of conditions still reflect expectations for the previously trained action-sound contingency are thus not fit to the be used in correction procedures. The problem could also be closely related to the order of conditions in the overall experimental procedure, leading to unwanted carry-over effects, not only into the conditions used for EEG signal correction, but also between experimental conditions. We ourselves have demonstrated this in study 1, as we used a carry-over effect to change the perceived level of control in a follow-up task.

3.4. CONCLUSION

The studies in this dissertation provided novel evidence concerning the involvement of motor and context information in the attenuation of the N1 and P2 amplitudes for action-generated sounds when compared to externally-generated sounds. In study 1 we replicated evidence revealing that the processing reflected in the N1 is not sensitive to contextual information, and only differed based on motor- and implicit temporal predictability. We instead found a contextual influence on P2 amplitudes that might reflect agency judgements. The paradigms in studies 2 and 3 were used to disturb motor information for action-related sounds and we found this to be reflected in N1 amplitudes, showing that the typical N1 attenuation is unlikely to be merely an effect of temporal predictability. We also demonstrated that linear mixed models are a useful tool to investigate the time course of ERP amplitudes. This was especially relevant for study 3 in which we captured the acquisition of a truly novel action-effect association as reflected in N1 amplitudes for action-generated sounds. For classic action-effect associations like those employing button presses and designs involving changing or varying action-effect contingencies, comparing the temporal dynamics through trial-level statistics could be key to further unravel underlying mechanisms.

4 BIBLIOGRAPHY

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147-149. https://doi.org/10.3758/BF03334492

Angelini, M., Fabbri-Destro, M., Lopomo, N. F., Gobbo, M., Rizzolatti, G., & Avanzini, P. (2018). Perspective-dependent reactivity of sensorimotor mu rhythm in alpha and beta ranges during action observation: an EEG study. *Scientific reports*, *8*(1), 1-11. https://doi.org/10.1038/s41598-018-30912-w

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. *International Journal of Psychophysiology*, *70*(2), 137-143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of cognitive neuroscience*, *11*(5), 551-559. https://doi.org/10.1162/089892999563607

Blakemore, S.-J., Smith, J., Steel, R., Johnstone, E., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychological medicine*, *30*(5), 1131-1139. https://doi.org/10.1017/s0033291799002676

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, *1*(7), 635. https://doi.org/10.1038/2870

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, *12*(9), 1879-1884. https://doi.org/10.1097/00001756-200107030-00023

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology*, *31*(1), 51-68. https://doi.org/10.1016/S0167-8760(98)00040-3

Caggiano, V., Giese, M., Thier, P., & Casile, A. (2015). Encoding of point of view during action observation in the local field potentials of macaque area F5. *European Journal of Neuroscience*, *41*(4), 466-476. https://doi.org/10.1111/ejn.12793

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). A new look at sensory attenuation: Action-effect anticipation affects sensitivity, not response bias. *Psychological science*, *21*(12), 1740-1745. https://doi.org/10.1177/0956797610389187

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204. https://doi.org/10.1017/S0140525X12000477

Coe, B. C., & Munoz, D. P. (2017). Mechanisms of saccade suppression revealed in the anti-saccade task. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160192. https://doi.org/10.1098/rstb.2016.0192

Davis, H., Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: recovery process. *Electroencephalography and clinical neurophysiology*, *21*(2), 105-113. https://doi.org/10.1016/0013-4694(66)90118-0

Dogge, M., Custers, R., & Aarts, H. (2019). Moving forward: On the limits of motor-based forward models. *Trends Cogn Sci*. https://doi.org/10.1016/j.tics.2019.06.008

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology*, *73*(6), 2608-2611. https://doi.org/10.1152/jn.1995.73.6.2608

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front Hum Neurosci*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Fitzgerald, K., & Todd, J. (2020). Making sense of mismatch negativity. *Frontiers in psychiatry*, *11*, 538673. https://doi.org/10.3389/fpsyt.2020.00468

Ford, J. M. (2018). The difficulty in finding relationships between ERPs and clinical symptoms of schizophrenia. *Clinical EEG and neuroscience*, *49*(1), 6-7. https://doi.org/10.1177/1550059417737416

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology*, *44*(4), 522-529. https://doi.org/10.1111/j.1469-8986.2007.00533.x

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry*, *158*(12), 2069-2071. https://doi.org/10.1176/appi.ajp.158.12.2069

Ford, J. M., Palzes, V. A., Roach, B. J., & Mathalon, D. H. (2014). Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia bulletin*, *40*(4), 804-812. https://doi.org/10.1093/schbul/sbt072

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138. https://doi.org/10.1038/nrn2787

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological cybernetics*, *104*, 137-160. https://doi.org/10.1007/s00422-011-0424-z

Ghio, M., Egan, S., & Bellebaum, C. (2021). Similarities and Differences between Performers and Observers in Processing Auditory Action Consequences: Evidence from Simultaneous EEG Acquisition. *Journal of cognitive neuroscience*, *33*(4), 1-12. https://doi.org/10.1162/jocn_a_01671

Ghio, M., Scharmach, K., & Bellebaum, C. (2018). ERP correlates of processing the auditory consequences of own versus observed actions. *Psychophysiology*, *55*(6), e13048. https://doi.org/10.1111/psyp.13048

Haggard, P., & Whitford, B. (2004). Supplementary motor area provides an efferent signal for sensory suppression. *Cognitive Brain Research*, *19*(1), 52-58. https://doi.org/10.1016/j.cogbrainres.2003.10.018

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

Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience*, *389*, 54-73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of general psychiatry*, *64*(3), 286-296. https://doi.org/10.1001/archpsyc.64.3.286

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

Hodson, R., Mehta, M., & Smith, R. (2023). The empirical status of predictive coding and active inference. *Neuroscience & Biobehavioral Reviews*, 105473. https://doi.org/10.1016/j.neubiorev.2023.105473

Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain research*, *1626*, 54-65. https://doi.org/10.1016/j.brainres.2015.03.038

Hsu, Y.-F., Le Bars, S., Hämäläinen, J. A., & Waszak, F. (2015). Distinctive representation of mispredicted and unpredicted prediction errors in human electroencephalography. *Journal of Neuroscience*, *35*(43), 14653-14660. https://doi.org/10.1523/JNEUROSCI.2204-15.2015

Hughes, G., Desantis, A., & Waszak, F. (2013a). Attenuation of auditory N 1 results from identity-specific action-effect prediction. *European Journal of Neuroscience*, *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. *Psychological Bulletin*, *139*(1), 133. https://doi.org/10.1037/a0028566

Ishikawa, T., Tomatsu, S., Izawa, J., & Kakei, S. (2016). The cerebro-cerebellum: Could it be loci of forward models? *Neuroscience research*, *104*, 72-79. https://doi.org/10.1016/j.neures.2015.12.003

Jack, B. N., Chilver, M. R., Vickery, R. M., Birznieks, I., Krstanoska-Blazeska, K., Whitford, T. J., & Griffiths, O. (2021). Movement planning determines sensory suppression: An event-related potential study. *Journal of cognitive neuroscience*, *33*(12), 2427-2439. https://doi.org/10.1162/jocn_a_01747

Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain topography*, *13*, 169-193. https://doi.org/10.1023/a:1007803102254

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

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166. https://doi.org/10.1007/s10339-007-0170-2

Klaffehn, A. L., Baess, P., Kunde, W., & Pfister, R. (2019). Sensory attenuation prevails when controlling for temporal predictability of self-and externally generated tones. *Neuropsychologia*, *132*, 107145. https://doi.org/10.1016/j.neuropsychologia.2019.107145

Knolle, F., Schroger, E., Baess, P., & Kotz, S. A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of cognitive neuroscience*, *24*(3), 698-706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schroger, E., & Kotz, S. A. (2013a). Cerebellar contribution to the prediction of self-initiated sounds. *cortex*, *49*(9), 2449-2461. https://doi.org/10.1016/j.cortex.2012.12.012

Knolle, F., Schröger, E., & Kotz, S. A. (2013b). Prediction errors in self-and externally-generated deviants. *Biological Psychology*, *92*(2), 410-416. https://doi.org/10.1016/j.biopsycho.2012.11.017

Kogo, N., & Trengove, C. (2015). Is predictive coding theory articulated enough to be testable? *Frontiers in computational neuroscience*(9), 111. https://doi.org/10.3389/fncom.2015.00111

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

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and cognition*, *69*(1), 127-137. https://doi.org/10.1016/j.bandc.2008.06.004

Lange, K. (2011). The reduced N1 to self-generated tones: an effect of temporal predictability? *Psychophysiology*, *48*(8), 1088-1095. https://doi.org/10.1111/j.1469-8986.2010.01174.x

Langer, E. J. (1975). The illusion of control. *Journal of personality and social psychology*, *32*(2), 311. https://doi.org/10.1037/0022-3514.32.2.311

Miall, R. C. (2003). Connecting mirror neurons and forward models. *Neuroreport*, *14*(17), 2135-2137. https://doi.org/10.1097/00001756-200312020-00001

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

Mifsud, N. G., Beesley, T., Watson, T. L., & Whitford, T. J. (2016). Attenuation of auditory evoked potentials for hand and eye-initiated sounds. *Biological Psychology*, *120*, 61-68. https://doi.org/10.1016/j.biopsycho.2016.08.011

Moore, J. W., Lagnado, D., Deal, D. C., & Haggard, P. (2009). Feelings of control: contingency determines experience of action. *Cognition*, *110*(2), 279-283. https://doi.org/10.1016/j.cognition.2008.11.006

Nelson, D. A., & Lassman, F. M. (1968). Effects of intersignal interval on the human auditory evoked response. *The Journal of the Acoustical Society of America*, *44*(6), 1529-1532. https://doi.org/10.1121/1.1911292

Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, *83*(12), 1112-1118. https://doi.org/10.1212/WNL.0000000000000798

Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends Cogn Sci*, *18*(9), 451-456. https://doi.org/10.1016/j.tics.2014.05.006

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

Poonian, S. K., McFadyen, J., Ogden, J., & Cunnington, R. (2015). Implicit agency in observed actions: evidence for N1 suppression of tones caused by self-made and observed actions. *Journal of cognitive neuroscience*, *27*(4), 752-764. https://doi.org/10.1162/jocn_a_00745

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87. https://doi.org/10.1038/4580

Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia*, *119*, 302-307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in neurosciences*, *21*(5), 188-194. https://doi.org/10.1016/s0166-2236(98)01260-0

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nature Reviews Neuroscience*, *17*(12), 757-765. https://doi.org/10.1038/nrn.2016.135

Rummell, B. P., Klee, J. L., & Sigurdsson, T. (2016). Attenuation of responses to self-generated sounds in auditory cortical neurons. *Journal of Neuroscience*, *36*(47), 12010-12026. https://doi.org/10.1523/JNEUROSCI.1564-16.2016

SanMiguel, I., Saupe, K., & Schröger, E. (2013a). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted" what" but not" when". *Front Hum Neurosci*, *7*, 407. https://doi.org/10.3389/fnhum.2013.00407

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013b). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience*, *33*(20), 8633-8639. https://doi.org/10.1523/JNEUROSCI.5821-12.2013

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and cognition*, *17*(4), 1219-1227. https://doi.org/10.1016/j.concog.2008.01.003

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

Schafer, E. W., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science*, *181*(4095), 175-177. https://doi.org/10.1126/science.181.4095.175

Schneider, D. M., & Mooney, R. (2018). How movement modulates hearing. *Annual review of neuroscience*, *41*, 553-572. https://doi.org/10.1146/annurev-neuro-072116-031215

Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641-664. https://doi.org/10.1111/ejn.12816

Seidel, A., & Bellebaum, C. (2023). Conflicting motor plans and sensory attenuation: Evidence from event-related potentials for sounds generated by pro- and antisaccades. *Manuscript submitted for publication*.

Seidel, A., Ghio, M., Studer, B., & Bellebaum, C. (2021). Illusion of control affects ERP amplitude reductions for auditory outcomes of self-generated actions. *Psychophysiology*, *58*(5), e13792. https://doi.org/10.1111/psyp.13792

Seidel, A., Weber, C., Ghio, M., & Bellebaum, C. (2023). My view on your actions: Dynamic changes in viewpoint-dependent auditory ERP attenuation during action observation. *Cognitive, Affective, & Behavioral Neuroscience*, 1-17. https://doi.org/10.3758/s13415-023-01083-7

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental brain research*, *222*(1-2), 149-157. https://doi.org/10.1007/s00221-012-3204-7

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology*, *43*(6), 482. https://doi.org/10.1037/h0055479

Sugimoto, F., Kimura, M., & Takeda, Y. (2021). Attenuation of auditory N2 for self-modulated tones during continuous actions. *Biological Psychology*, *166*, 108201. https://doi.org/10.1016/j.biopsycho.2021.108201

Swiney, L., & Sousa, P. (2014). A new comparator account of auditory verbal hallucinations: how motor prediction can plausibly contribute to the sense of agency for inner speech. *Front Hum Neurosci*, *8*, 675. https://doi.org/10.3389/fnhum.2014.00675

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and cognition*, *17*(1), 219-239. https://doi.org/10.1016/j.concog.2007.03.010

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in psychology*, *4*, 127. https://doi.org/10.3389/fpsyg.2013.00127

Thompson, S. C., Nierman, A., Schlehofer, M. M., Carter, E., Bovin, M. J., Wurzman, L., Tauber, P., Trifskin, S., Marks, P., & Sumner, J. (2007). How do we judge personal control? Unconfounding contingency and reinforcement in control judgments. *Basic and Applied Social Psychology*, *29*(1), 75-84. https://doi.org/10.1080/01973530701331189

Timm, J., SanMiguel, I., Keil, J., Schroger, E., & Schonwiesner, M. (2014a). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of cognitive neuroscience*, *26*(7), 1481-1489. https://doi.org/10.1162/jocn_a_00552

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

Timm, J., Schönwiesner, M., SanMiguel, I., & Schröger, E. (2014b). Sensation of agency and perception of temporal order. *Consciousness and cognition*, *23*, 42-52. https://doi.org/10.1016/j.concog.2013.11.002

Timm, J., Schonwiesner, M., Schroger, 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

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. *Biological Psychology*, *102*, 108-117. https://doi.org/10.1016/j.biopsycho.2014.06.007

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2021). Suppression of the auditory N1 by visual anticipatory motion is modulated by temporal and identity predictability. *Psychophysiology*, *58*(3), e13749. https://doi.org/10.1111/psyp.13749

Von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour*. https://doi.org/10.1016/S0950-5601(54)80044-X

Voss, M., Ingram, J. N., Haggard, P., & Wolpert, D. M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. *Nature neuroscience*, *9*(1), 26-27. https://doi.org/10.1038/nn1592

Voss, M., Ingram, J. N., Wolpert, D. M., & Haggard, P. (2008). Mere expectation to move causes attenuation of sensory signals. *PloS one*, *3*(8), e2866. https://doi.org/10.1371/journal.pone.0002866

Wacongne, C., Labyt, E., Van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*(51), 20754-20759. https://doi.org/10.1073/pnas.1117807108

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the new York Academy of Sciences*, *1464*(1), 242-268. https://doi.org/10.1111/nyas.14321

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

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

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 593-602. https://doi.org/10.1098/rstb.2002.1238

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current biology*, *11*(18), R729-732. https://doi.org/10.1016/s0960-9822(01)00432-8

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn Sci*, *2*(9), 338-347. https://doi.org/10.1016/s1364-6613(98)01221-2

5 APPENDIX A: AFFIDAVIT

Eidesstattliche Erklärung gemäß § 5 der Promotionsordnung vom 15.6.2018 der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf:

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in der vorliegenden oder ähnlichen Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 18.06.2024

Alexander Seidel

6 APPENDIX B: ORIGINAL RESEARCH ARTICLES

Original article of study 1

Seidel, A., Ghio, M., Studer, B., & Bellebaum, C. (2021). Illusion of control affects ERP amplitude reductions for auditory outcomes of self-generated actions. *Psychophysiology*, 58(5), e13792. https://doi.org/10.1111/psyp.13792

I was the main author of this article along with the second first author Dr. Marta Ghio. I curated the data, planned, and performed the data analysis, and interpreted the results.

Original article of study 2

Seidel, A., Weber, C., Ghio, M., & Bellebaum, C. (2023). My view on your actions: Dynamic changes in viewpoint-dependent auditory ERP attenuation during action observation. *Cognitive, Affective, & Behavioral Neuroscience*, 23(4), 1175-1191. https://doi.org/10.3758/s13415-023-01083-7

I was the main author of this article along with the second first author Constanze Weber. I contributed to the conceptualization and methodology of the paradigm. I planned and supervised the creation of the stimulus material and programmed the paradigm. I planned and supervised the data acquisition. I developed the research question, planned and performed the data analysis, and interpreted the results.

Original article of study 3

Seidel, A., & Bellebaum, C. (2024). Conflicting motor plans and sensory attenuation: Evidence from event-related potentials for sounds generated by pro- and antisaccades. *Manuscript submitted for publication in Psychophysiology*.

I was the main author of this article. I developed the conceptualization and methodology of the paradigm and programmed the paradigm. I planned and supervised the data acquisition. I developed the research question, planned, and performed the data analysis, and interpreted the results.

Received: 30 March 2020 | Revised: 28 January 2021 | Accepted: 29 January 2021

DOI: 10.1111/psyp.13792

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Illusion of control affects ERP amplitude reductions for auditory outcomes of self-generated actions

Alexander Seidel[1] | **Marta Ghio**[2] | **Bettina Studer**[3,4] | **Christian Bellebaum**[1]

[1]Institute of Experimental Psychology, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

[2]CIMeC - Center for Mind/Brain Sciences, University of Trento, Trento, Italy

[3]Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

[4]Department of Neurology, Mauritius Hospital Meerbusch, Meerbusch, Germany

Correspondence
Alexander Seidel, Institute of Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany.
Email: Alexander.Seidel@hhu.de

Abstract

The reduction of neural responses to self-generated stimuli compared to external stimuli is thought to result from the matching of motor-based sensory predictions and sensory reafferences and to serve the identification of changes in the environment as caused by oneself. The amplitude of the auditory event-related potential (ERP) component N1 seems to closely reflect this matching process, while the later positive component (P2/ P3a) has been associated with judgments of agency, which are also sensitive to contextual top-down information. In this study, we examined the effect of perceived control over sound production on the processing of self-generated and external stimuli, as reflected in these components. We used a new version of a classic two-button choice task to induce different degrees of the illusion of control (IoC) and recorded ERPs for the processing of self-generated and external sounds in a subsequent task. N1 amplitudes were reduced for self-generated compared to external sounds, but not significantly affected by IoC. P2/3a amplitudes were affected by IoC: We found reduced P2/3a amplitudes after a high compared to a low IoC induction training, but only for self-generated, not for external sounds. These findings suggest that prior contextual belief information induced by an IoC affects later processing as reflected in the P2/P3a, possibly for the formation of agency judgments, while early processing reflecting motor-based predictions is not affected.

KEYWORDS

auditory ERP, EEG, illusion of control, sense of agency, sensory prediction

1 | INTRODUCTION

When processing sensory inputs, it is essential to distinguish those caused by our own actions (e.g., touching oneself) from those with an external cause (e.g., being touched by another agent). Self- and externally generated stimuli appear to be treated differently by our perceptual system. Sensory attenuation, that is a decreased perceptual intensity for sensory stimuli caused by our own actions compared to physically identical but externally generated stimuli, has been a common finding in different sensory modalities (Cardoso-Leite et al., 2010; Sato, 2008; Shergill et al., 2005). A related finding is the reduction of neuronal responses associated with processing self- versus externally generated sensory stimuli

---

Alexander Seidel and Marta Ghio are co-first authors.

(Baess et al., 2011; Horváth, 2015; Hughes & Waszak, 2011), although a recent study suggested that these two phenomena might be based on different underlying processes (Palmer et al., 2016).

According to the classic interpretation, both phenomena can be explained by assuming internal forward models (Pickering & Clark, 2014; Reznik et al., 2014; Wolpert & Flanagan, 2001) that predict the sensory consequences of voluntary movements. This type of prediction is thought to be performed by the cerebellum and based on efference copies of commands sent by the supplementary motor cortex (Blakemore et al., 2001; Haggard & Whitford, 2004; Popa & Ebner, 2018; Reznik et al., 2015), which are available even before the initiation of a movement (Crapse & Sommer, 2008; Reznik et al., 2018; Vercillo et al., 2018; von Holst & Mittelstaedt, 1950). Sensory input occurring after movements could thus be compared immediately to the predicted input, with matches possibly resulting in altered subsequent neural processing and ultimately adjusted perception. In line with this notion, it has been shown that these phenomena depend on motor intention, not motor execution. Voss et al. (2006) reported sensory attenuation for somatosensory stimuli before the actual movement when transcranial magnetic stimulation (TMS) delayed its execution, while neuronal responses for auditory stimuli caused by involuntary movements that were generated with TMS over the primary motor cortex were not reduced (Timm et al., 2014).

In the auditory domain, which has been studied extensively in the last decades (Horváth, 2015), electroencephalography (EEG) recordings have consistently revealed that the processing of self- versus externally generated auditory stimuli is associated with amplitude reductions of the event-related potential (ERP) component N1 and (although findings are less consistent) of the P2/P3a (Baess et al., 2011; Ghio et al., 2018; Horváth, 2015; Horváth et al., 2012; Knolle et al., 2012; Lange, 2009, 2011; Schafer & Marcus, 1973; Timm et al., 2013, 2016). An increasing number of studies seem to suggest a functional dissociation between the early N1 and the later positive component(s). For example, cerebellar lesion patients showed no reduction of the N1 amplitude for self-generated sounds, but a reduction of P2 amplitudes similar to controls (Knolle et al., 2013). A possible interpretation is that only the early sensory processing as reflected by the N1 is modulated by cerebellar forward model predictions based on motor information. A similar dissociation can be found when comparing the processing of self-generated to cue external sounds. Even though self-generated and cued sounds are similarly predictable, the N1 for cued sounds is not attenuated compared to non-cued external sounds (Lange, 2011; Sowman et al., 2012), indicating that the process underlying the attenuation requires motor information. A P2 attenuation, moreover, was observed for cued external sounds (Sowman et al., 2012), again suggesting that a different, possibly higher order prediction mechanism is used at this stage, not critically depending on motor information.

Previous studies associated the (mis)matching of predicted and perceived sensory input with an internal interpretation of sensory input as self-generated (if it matches the prediction) or as externally generated (if there is a mismatch), suggesting a contribution to the subjective experience of agency, that is, of being responsible for the experienced sensory stimulation (Blakemore et al., 2000, 2002; Synofzik et al., 2013). For example, schizophrenic patients with altered feelings of agency, reflected in the typical symptoms of auditory hallucinations and passivity experiences, showed reduced sensory attenuation for self-produced sounds and forces (Blakemore et al., 2000, 2002; Shergill et al., 2005). Furthermore, diminished N1 amplitude reductions for self-relative to externally generated auditory stimuli was found in schizophrenic patients (Ford et al., 2007, 2013; Heinks-Maldonado et al., 2007), suggesting that impaired prediction mechanisms may play a role for deficits in the distinction between self- and externally generated stimuli and thus for the experience of agency. However, Ford et al. (2013) did not find a significant correlation between the deficient N1 reduction and schizophrenia symptoms, questioning the relationship between the N1 modulation and agency.

Some studies on healthy participants experimentally manipulated the sense of agency to examine whether and which ERP components are modulated by agency in the processing of self- or externally generated stimuli. Kühn et al. (2011) induced uncertainty about the authorship of self-produced sounds by varying their delay and pitch, while suggesting that some sounds may be generated by the experimenter. Agency ratings for each sound were collected, and no difference was found for the auditory N1 amplitude between trials with high and low agency ratings. The P3a, a component associated with unexpected stimuli (Herrmann & Knight, 2001), was, however, significantly reduced for sounds judged as self-generated. While amplitudes of the P2 were not analyzed, visual inspection suggests that it was not affected by agency. In a related study, Timm et al. (2016) successfully manipulated agency by presenting delayed or non-delayed tones when delayed tones were expected, which resulted in a high and low agency condition, respectively. In line with the findings of no agency effect on the N1 (Ford et al., 2013; Kühn et al., 2011), these authors found a comparable reduction of N1 amplitudes for self- versus externally generated sounds in both agency conditions. However, for the P2 an agency effect emerged: amplitude reductions for self-generated tones were less pronounced in the low agency condition. Overall, the emerging pattern in the findings suggests that distinct prediction mechanisms are reflected in the early and late ERP components for the processing of self-generated stimuli. The N1 seems to reflect simple predictions directly linked to motor actions and appears unaffected by context-dependent variations in agency, whereas the later positive components (P2 or P3a) appear to reflect prediction mechanisms sensitive to top-down influences such as context-dependent modulations of subjective agency.

It has been shown that agency over the production of a sound is also modulated by the predictability of its occurrence, as reducing the probability of sounds being played following button presses lowered subjective agency for those sounds that did occur, as revealed by an implicit behavioral measure (Moore & Haggard, 2008; Moore et al., 2009). Furthermore, agency is sensitive to the contingency of action and outcome (Moore et al., 2009). Contingency is commonly calculated as the difference between the probability of an outcome given a potential cause, for example, a movement, and the probability of the outcome in the absence of the potential cause (Allan, 1980; Jenkins & Ward, 1965). If the outcome (e.g., a tone) occurs with the same probability after an action (e.g., button press) and without it, the difference is zero and the relation is thus non-contingent. In this context, it is important to note that contingency in non-contingent action-outcome paradigms is generally overestimated, an effect that is known as the illusion of control (IoC) (Langer, 1975). Research has shown that increasing the probability of the outcome in non-contingent paradigms increases the level of perceived control over the outcome, thereby enhancing the IoC and thus the perceived contingency (Blanco et al., 2013; Jenkins & Ward, 1965; Matute et al., 2015; Studer et al., 2020; Thompson et al., 2007).

In the present study, we combined research on the IoC and on the processing of self-generated stimuli. More specifically, we examined whether differences in the perceived personal control affect subsequent processing of self-generated sounds. To manipulate perceived control, we used a new version of a classic two-button choice task (Jenkins & Ward, 1965), in which participants chose a button in each trial in order to produce a desired auditory outcome stimulus and the probability of the desired auditory outcome was manipulated to induce a stronger or weaker IoC. We recently showed that induced levels of illusory control affected subsequent behavioral persistence in two different motivationally challenging situations (Studer et al., 2020). This suggests that IoC effects can extend beyond the conditions in which IOC was induced. In the current study, we aimed to test if induced IoC over generating a specific sound could also affect auditory processing of physically identical self-generated sounds in a subsequent structurally different task, namely in the self-generation paradigm. In analogy to findings on agency manipulations (Kühn et al., 2011; Timm et al., 2016), we expected no effect of the IoC on the N1 amplitude reduction for self- compared to externally generated sounds. Instead, we expected an effect of the IoC on later processing in the P2 time window. In particular, amplitude reductions for self-generated tones were expected to be larger for higher levels of IoC, since the P2 seems to be sensitive to top-down influences like expectancy and agency (Kühn et al., 2011; Sowman et al., 2012; Timm et al., 2016).

2 | METHOD

2.1 | Participants

Forty participants took part in the experiment (33 women, $M_{\text{Age}} = 25.4$ years, $SD_{\text{Age}} = 3.5$ years) and received either course credit or monetary compensation. Normal or corrected-to-normal vision and normal hearing (according to self-report) were requirements for participation. The experiment was approved by the Ethics Committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf, Germany, and written informed consent was given by all participants.

2.2 | Procedure

To examine the effect of IoC on the processing of self- versus externally generated sounds, we set up a within-subject design that included two experimental blocks, each consisting of two tasks. In both blocks, participants were first exposed to a two-button choice task, designed to induce either a high or low level of perceived control over the production of desired sound (as opposed to an undesired sound), over which they actually exercised no control (from now on referred to as the IoC task). We aimed to achieve this by varying the base probability P(O) for the desired auditory outcome regardless of the button choice (see below). With this manipulation we wanted to modulate the processing of self- (Act-sounds) versus externally generated sounds (Ext-sounds) in a subsequent, so-called self-generation paradigm (Horváth, 2015), which was conducted as a second task in each experimental block. Indeed, our main interest in this study was to examine the effect of the IoC, as induced by the IoC task, on the neuronal processing of Act-sounds versus Ext-sounds in the self-generation paradigm, assessed by means of EEG. To distinguish this effect from a possibly carried over identity association between action and sound (Hughes et al., 2013), button presses were performed on different buttons in the two tasks. As we employed a within-subject design, the self-generation paradigm was performed twice, once after the IoC task with high P(O) (from now on the sequence of these two tasks is referred to as high P(O) block), and once after the IoC task with low P(O) (referred to as low P(O) block). As we wanted to test if the P(O) in the IoC task could affect neuronal processing of sounds in the subsequent task, the self-generation paradigm in the high P(O) block will be referred to as high P(O) condition, whereas the self-generation paradigm in the low P(O) block will be referred to as low P(O) condition, although the self-generation paradigm itself was identical in both blocks. Whether participants started with the low or high P(O) block was counterbalanced across participants (see Figure 1 for a schematic representation).

Stimulus presentation and response recording were controlled by Presentation software (Version 17.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) on a Windows 10 PC. Sound was delivered through an onboard soundcard (Realtek ALC887-VD) in DirectX Software mode and Sennheiser HD 202 headphones. Button presses were registered with a Cedrus RB-740 response pad (www.cedrus.com) featuring seven response buttons oriented in a straight horizontal line. The leftmost and rightmost buttons were used in the IoC induction task, while the middle button was pressed during the self-generation paradigm. This procedure was applied in order to avoid carryover effects of specific action-outcome associations. Ratings in the IoC task and Act-sound condition (see below) were given with a regular keyboard.

2.2.1 | IoC induction

Our IoC task was a variant of the classic two-button choice task by Jenkins and Ward (1965). In this task, participants were asked to try to elicit a desired auditory outcome by means of button presses. Participants chose between two buttons on every trial, but the outcome (i.e., desired vs. undesired sound) in each trial did not depend on the action that was performed (Allan, 1980; Matute et al., 2015). Instead, the P(O) for the desired sound was fixed by predetermining the outcome of every trial before the start of the task, and the P(O) varied between conditions in order to elicit a high or low IoC. Specifically, in the IoC task with high P(O), the desired sound was presented in 70% of the trials (total number of trials = 100), while for the remaining 30% of the trials an undesired sound was presented, irrespective of what button was pressed in the respective trial. In turn, in the IoC task with low P(O), the desired sound was presented in 30% of the trials (total number of trials = 100) and the undesired sound in 70% of the trials, again regardless of which button was pressed.

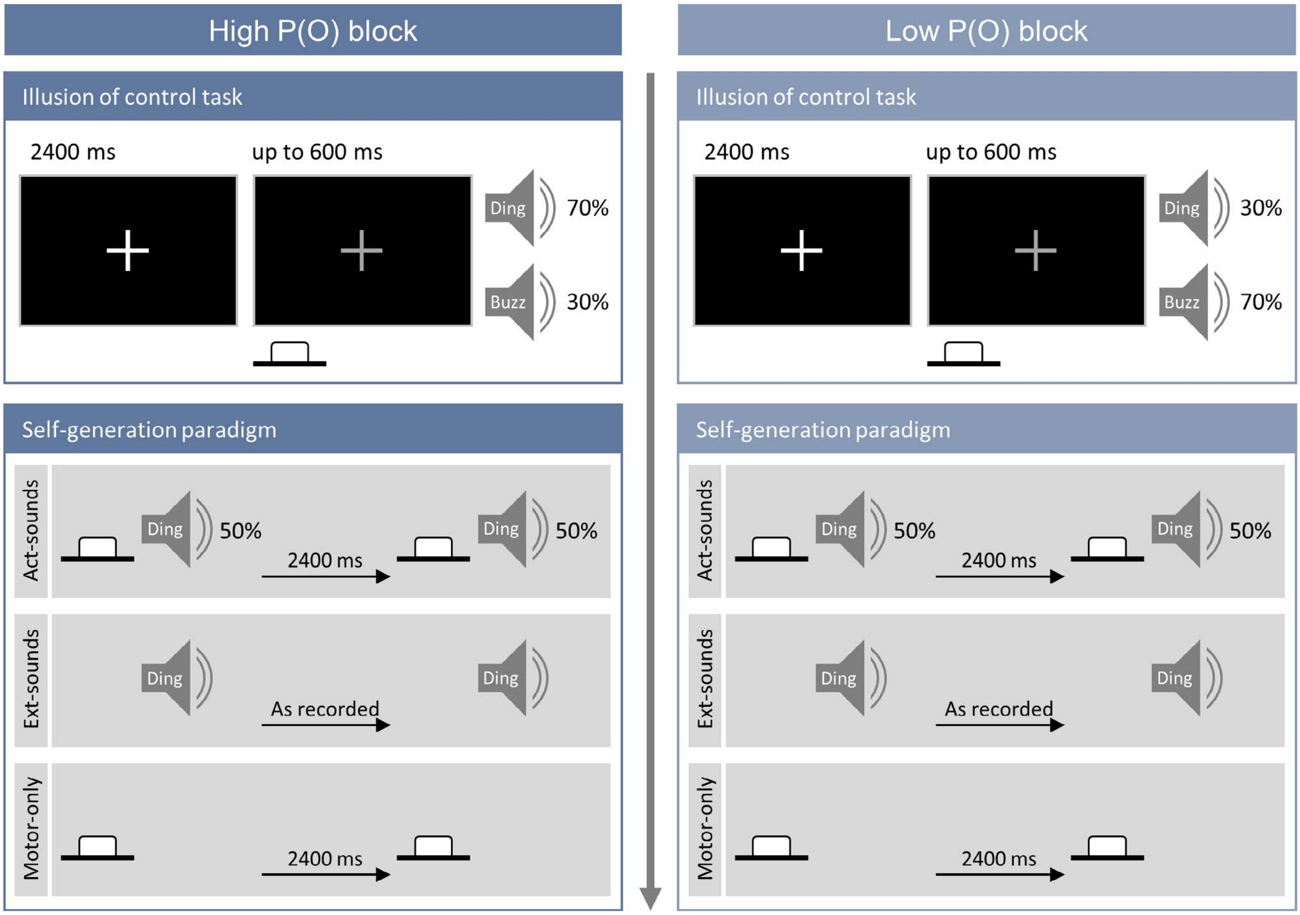

FIGURE 1 Experimental sequence for the high and low probability block (High P(O) block on the left and Low P(O) block on the right). Task order is indicated by the arrow, ending with the identical self-generation paradigm in both blocks

Each trial started with a white fixation cross on a black background. After 2,400 ms the fixation cross color switched to grey for 600 ms. Participants were asked to press one of the two available buttons (i.e., the leftmost or rightmost button on the response pad) as soon as the color of the fixation cross changed from white to grey in order to produce a desired sound that was introduced in the instruction as a positive sound ("ding.wav," distributed with Windows XP, 100 ms duration). The rhythm of one button press every 2,400 ms was introduced in order to train participants to this rhythm for the subsequent self-generation paradigm, which required regular self-paced button presses (see Ghio et al., 2018; Knolle et al., 2013). Each button press was followed either by the desired sound (70% and 30% of the trials in the IoC task with high and low P(O), respectively) or by an undesired sound (introduced in the instructions as a negative sound, namely a synthetic buzzer sound, 100 ms duration), irrespective of what button was pressed in the trial. The sounds were presented 50 ms after button press onset (see Ghio et al., 2018). As a further motivation to try to elicit the desired sound, each occurrence of the sound during the IoC task was associated with a monetary reward of 0.20 €, whereas each undesired sound was associated with a monetary loss of 0.05 €. Button presses occurring during the white fixation cross and thus outside of the required rhythm were penalized with a loss of 0.20 €. On average, every 15 trials (with a random variance of ±5 trials) participants were asked "*How much control did you have over the generation of the positive tone?*" and prompted to rate their level of control on a visual analog scale ranging from 0 (*NO control - the appearance of the positive sound had nothing to do with your button press*) to 100 (*COMPLETE control – the appearance of the positive sound was entirely determined by your button press*) presented on the screen. A medium level of control was described as "*MEDIUM control - your button press had an influence on the appearance of the positive sound. You did however not fully control it*."

2.2.2 | Self-generation paradigm

The self-generation paradigm comprised three experimental conditions in a fixed order, which were presented in separate sub-blocks and did not differ between the high and low P(O) condition. Throughout all conditions, a white fixation cross was displayed.

Act-sound condition

Subjects were instructed to press the middle button on the response pad (thus, a button different from those used in the IoC task) with their right index finger in the same rhythm that was learned in the IoC task (i.e., every 2,400 ms ca.). Different from the standard self-generation paradigm in which each button press generates a sound (Horváth, 2015), button presses in our paradigm ($n = 200$) were followed by a sound only in 50% of the trials ($n = 100$ for the number of sounds). This variation was adopted in order to create uncertainty concerning the association between button presses and sounds and to prevent the preceding IoC induction from decaying during the beginning of the task. Importantly, the sound used in the self-generation paradigm was identical to the desired sound in the IoC task, while the undesired sound was never presented. On average, every 20 trials (with a random variance of ±5 trials) participants were prompted to rate their level of control over the production of the sound, applying the same scale used in the IoC task (see above). After each rating, they also received feedback concerning the length of their button press interval in the previous 20 (±5) trials. If 25% of these intervals deviated more than 600 ms from the required duration of 2,400 ms in one direction, participants were asked to react faster or slower, respectively. Otherwise, they were encouraged to keep their current rhythm.

Ext-sound condition

Subjects were presented with the playback of all the 100 sounds generated in the previous Act-sound condition and instructed to listen to them carefully without performing any action.

Motor-only condition

To control for the motor demand present in the Act-sound condition (see below for details), participants were asked to press the button in the same rhythm applied in the Act-sound condition (i.e., every 2,400 ms ca.). Crucially, no sounds were presented.

2.3 | EEG Data acquisition and preprocessing

EEG data were continuously recorded at 1,000 Hz with BrainVision Recorder software (1.20.0506, Brain Products, GmbH, Germany). Twenty-eight Ag/AgCl passive ring electrodes connected to a BrainAmp amplifier were positioned on the scalp via an elastic cap (EasyCap). According to the international 10–20 System, electrodes were positioned at F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P7, P3, Pz, P4, P8, PO7, PO3, POz, PO4, and PO8. The recorded signal was referenced to linked mastoids, and the ground electrode was placed at AFz. Electrooculogram data were recorded at F9 and F10 for horizontal eye movements. To register vertical eye movements, one electrode was positioned below the right eye, aligning with a second electrode at Fp2. Impedances were kept below 5 kΩ.

Raw EEG data recorded during the self-generation paradigm in the high and low P(O) condition were analyzed with

Brain Vision Analyzer 2.2 (Brain Products) and MATLAB (R2018a, The MathWorks, Inc., Natick, MA). After a global direct current de-trend, a Butterworth zero-phase filter (low cutoff: 0.3 Hz, 12 dB/oct; high cutoff: 30 Hz, 12 dB/oct) and a notch filter (50 Hz) were applied. An independent component analysis (steps = 512, infomax restricted biased) was employed to identify components reflecting blinks, horizontal and vertical eye movements and remove them before a subsequent inverted ICA.

ERP segments were time-locked to the onsets of the sounds in the Act-sound and the Ext-sound condition and had a duration of 600 ms (including a 150 ms pre-sound period). As the interval between button press onset and tone onset was 50 ms (see above) the segment lasted from 100 ms before to 500 ms after the button press onset in the Act-sound condition. For the Motor-only condition, corresponding segments were created from −100 to 500 ms relative to button press onset.

Successive button presses and sounds appearing in an interval smaller than 1 s were excluded from further analyses. Epochs were baseline corrected by the mean amplitude of the first 100 ms of the interval, that is, from 150 ms to 50 ms before sound onset in the Act- and Ext-sound conditions and from 100 ms to 0 ms before button press onset in the Motor-only condition. Segments containing artifacts were detected and rejected using the automatic algorithm provided by BrainVision Analyzer (maximal allowed voltage step = 50 μV/ms, maximal allowed difference of values within 100-ms intervals = 100 μV, maximal/minimal allowed amplitude = ±100 μV, lowest activity of 0.5 μV within 100 ms intervals). Remaining segments were averaged separately for the Act-sound, Ext-sound, and Motor-only conditions and in the high and low P(O) conditions. In order to control for motor activity in the Act-sound condition (see Horváth, 2015), we then subtracted the average ERP of the Motor-only condition from the average Act-sound ERP. From now on the Act-sound condition will refer to the ERPs yielded by the ERP subtraction procedure. As differing action frequencies can cause processing differences both related to the actions and the elicited sounds, the average time interval between button presses was compared in a 2x2 repeated measure ANOVA with the factors Condition (Act-sound, Motor-only) and P(O) (low, high). Although the main effect of Condition was significant, $F(1, 28) = 5.11$, $p = .026$, $\eta_p^2 = .044$, with longer time intervals in the Motor-only ($M = 3{,}088$ ms, $SD = 929$) than in the Act-sound condition ($M = 2{,}731$ ms, $SD = 546$), we still consider the ERPs in the Motor-only condition as an appropriate control for the motor-related ERPs in the Act-sound condition, as in absolute terms the difference was small and the intervals were long enough for the motor ERPs to return to baseline. Importantly, the interaction, $F(1, 28) = 0.35$, $p = .556$, $\eta_p^2 = .003$, and the main effect for P(O), $F(1, 28) = 0.45$, $p = .503$, $\eta_p^2 = .004$, were not significant, indicating that action- and sound-related ERPs could not be affected by P(O)-dependent differences in the timing of actions and/or sounds. Concerning the inter-response-intervals (IRIs) after button presses that did ($M = 2{,}688$ ms, $SD = 569$ ms) and did not elicit sounds ($M = 2{,}666$ ms, $SD = 538$ ms), in the Act-sound condition no significant difference was found, as revealed by a paired sample t test, $t(28) = 0.294$, $p = .771$, $d = .06$. We did, however, see large interindividual differences in the (dis)similarity of these intervals (see section 2.4).

Finally, the mean amplitudes for the two ERP components of interest, N1 and P2, were extracted for the Act-sound and Ext-sound conditions, separately for the high and low P(O) conditions. To determine time windows and electrode positions for the analysis of these components, we created two grand averages, one for the Act- and one for Ext-sounds, across P(O) conditions. We avoided an overall grand average across sound types, as the P2 showed a notable latency difference between Act- and Ext-sounds (see supplementary materials). Based on the topographical maps of the two grand average N1 peaks, we determined FCz and Cz as appropriate electrode positions for our analysis, focusing on the midline electrode sites as previously suggested (Knolle et al., 2013; Timm et al., 2016). For the later positive component, mappings indicated highest activity at Cz and CPz for Act-sounds and at FCz and Cz for the Ext-sounds. We, therefore, included FCz, Cz, and CPz in our P2 analysis. N1 peak latencies were extracted from the two grand averages and then averaged for both electrodes. As the latencies for the two sound type conditions were very similar (Act-sounds: 84 ms; Ext-sounds: 92 ms) the mean signal in one N1 time window from 68 ms to 108 ms was considered to appropriately reflect N1 amplitude for both conditions. For the P2, the latencies in the two conditions clearly differed (Act-sounds: 255 ms; Ext-sounds: 180 ms). Therefore, we defined separate time windows for the two sound type conditions, from 235 ms to 275 ms for Act-sounds and from 160 ms to 200 ms for Ext-sounds. Mean amplitudes were measured in these time windows for all conditions and the respective electrodes.

2.4 | Data analysis

From the sample of 40 participants, we excluded one participant for missing data at the Cz electrode and one for excessive artifacts and a consequent loss of more than 50% of the ERP segments for the analysis. Three participants were determined as outliers due to the length of their IRIs in the Act-sound condition or Motor-only condition and one due to an enlarged difference between IRIs following button presses that did or did not elicit sounds, as these four participants each showed a deviation of more than 2.5 SDs from the sample mean. Furthermore, we excluded five participants for rating their level of control in the high P(O) condition lower than in the low P(O) condition (namely their rating difference was <0, see below for an explanation about how this indicates that the intended IoC induction was unsuccessful). The data of the remaining 29 participants (22 women, 26 right-handed, $M_{\text{Age}} = 25.2$ years, $SD_{\text{Age}} = 3.2$ years) were then entered into the statistical analyses for EEG and behavioral data. All analyses were conducted in R (version 3.6.3) using RStudio (version 1.3.959).

2.4.1 | Behavioral data

To determine whether the IoC task with the high P(O) induced a stronger IoC than the IoC task with low P(O), average control ratings in the two IoC task versions were calculated and compared by means of paired *t* tests. Furthermore, we calculated the difference in the control ratings in the IoC task with high and low P(O) for each participant, which will be referred to as the rating difference (RD) in the following. Positive values indicate higher perceived control in the IoC task with high P(O), as intended, whereas negative values indicate the opposite pattern. To exclude order effects, RD values were compared via Welch *t* test between participants who started with the high ($n = 17$) or low ($n = 12$) P(O) condition. All analyses were performed with the default R stats package, Cohen's *d* values for *t* tests were calculated with the R package lsr (version 0.5). An α level of .05 was considered as statistically significant. For completeness, identical analyses on the control ratings in the self-generation paradigm are reported in the supplementary materials.

2.4.2 | EEG Data

Of central importance for the present study was the relationship between ERP measures of auditory processing and the subjective control ratings in the IoC task with high and low P(O), since they reflect the success of the intended IoC induction. The large interindividual variability for the difference measure (RD), from 0.6% to 49% (see Figure 2 for a histogram), suggested that the high and low P(O) IoC task induced different levels of perceived control in each participant. Consequently, the effect on the processing of the Act- and Ext-sounds likely also varied between participants. We thus aimed to include the continuous variable RD in the statistical model. To achieve this, we analyzed the data by means of a linear mixed-effects analyses (LME), in which both categorical and continuous independent variables can be included (Baayen et al., 2008).

Several separate LME analyses were performed using the lme4 package (version 1.1 23). The N1 analysis was conducted on the mean amplitudes in the N1 time window (see above). P2 was analyzed using the mean amplitudes in the Act-sound P2 peak time window for Act-sounds and in the Ext-sound P2 peak time window for Ext-sounds. See supplementary materials for an alternative analysis that directly compares N1 and P2 mean amplitudes in one model. To explore if the ERPs of the Act-sound condition already differed between P(O) conditions in the earlier Ext-sound P2 peak time window, as suggested by visual inspection, an additional analysis included the mean amplitudes in this time window for all conditions.

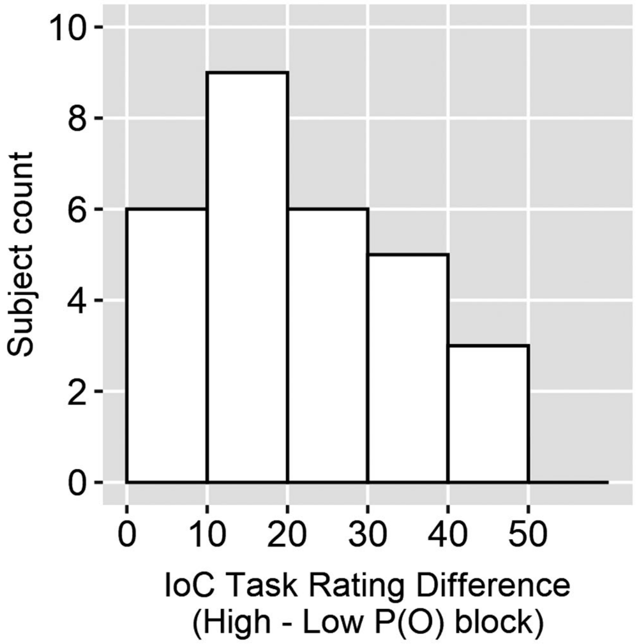

FIGURE 2 Histogram of participants in the final sample ($n = 29$) according to their differences between the level of control ratings in the IoC task of the High and Low P(O) block. From the sample of 40 participants, five participants were excluded for IoC task rating differences below zero (see section 2.4), in addition to six participants excluded as outliers or due to technical reasons

For each analysis, we created a model comprising the fixed-effect predictors Sound type, P(O) and RD. RD was mean-centered and entered into the model as a continuous fixed-effect predictor. The categorical predictors Sound Type (0.5 = Act-sounds, −0.5 = Ext-sounds) and P(O) (0.5 = high, −0.5 = low) were simple coded. Interactions between all three predictors were modeled, and random intercepts and slopes (for Sound type and P[O]) by participants were modeled as random effects. As we included the data of several electrodes in each model (see above), we modeled a random intercept for the electrode as a random effect nested in the random effect Participant (see supplementary materials for the R model syntax). We used the restricted maximum likelihood approach for model estimation and assessed significance with the R package lmerTest 3.0-1 (Kuznetsova et al., 2017) and its built-in Satterthwaite approximation for the degrees of freedom. This is in line with suggestions by

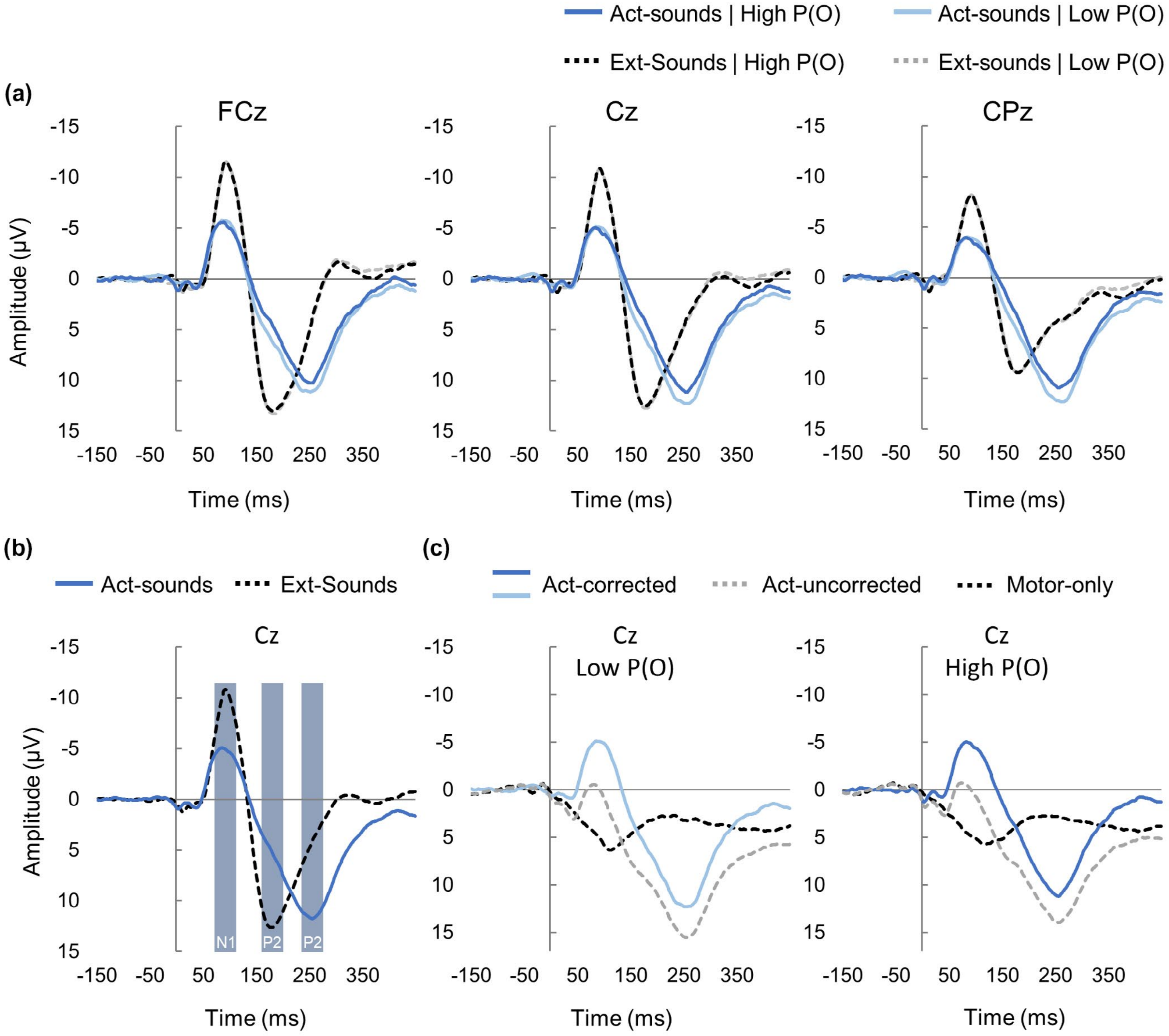

FIGURE 3 (a) Grand average ERPs for all analyzed electrode sites in the Act- and Ext-sound conditions and in the low and high P(O) conditions, with the y-axis intersecting at the start of the sound event. (b) Grand average ERPs at Cz for the Act-sound and Ext-sound condition across P(O) conditions, used to determine the shown time windows (68–108 ms, 160–200 ms, 235–275 ms) for mean amplitude extraction (see section 2.3). (c) Grand average ERPs at Cz showing the uncorrected and corrected Act-sound condition, as well as the Motor-only condition, separately for the low and high P(O) condition

Luke (2017), reporting acceptable Type I error rates, largely independent of sample size, compared to the common likelihood ratio tests. After building our statistical models for the EEG data analysis, we investigated whether any participants emerged as an influential data point by applying the R package influence.ME (Nieuwenhuis et al., 2012). As an exclusion criterion, we defined that participants had to emerge as an influential data point in both the N1 and P2 models in order to be excluded, which was not the case for any of our participants. After performing the LME analyses, however, we additionally investigated the impact of each detected influential data point in one model on the results of that model, by the separate exclusion of each outlier and re-analysis of the corresponding model. We found neither loss of significance for the reported effects, nor a significant new effect in any of the single-participant-exclusion models.

Simple effects analyses for the resolution of significant interactions were conducted by dummy coding the categorical factors' reference condition to 0 and shifting the center of the continuous factor RD by one standard deviation up or down in two separate analyses (Aiken et al., 1991; Liu et al., 2017). We will refer to the simple effects at these recentered values as being "low" and "high" values of RD. Marginal means (for plotting) were calculated with the R package ggeffect (version 0.16.0). An α level of .05 was considered statistically significant. Significant effects of the predictor RD are only reported in interactions with P(O).

3 | RESULTS

3.1 | Behavioral data

Subjective control ratings in the IoC task were significantly higher in the high P(O) block (*M* = 60.3%, *SD* = 13.1%) than in the low P(O) block (*M* = 39.1%, *SD* = 16.3%, $t(28) = 8.50$, $p < .001$, $d = 1.58$). Figure 2 shows a histogram of RD values. No significant differences were found when comparing RD values in the IOC task between participants who started with the high P(O) condition and participants who started with the low P(O) condition, $t(24.92) = -1.46$, $p = .157$, $d = .55$.

3.2 | EEG data

Figure 3a shows grand average ERPs for Cz, FCz, and CPz in the Act- and Ext-sound conditions and in the low and high P(O) conditions. Figure 3b shows overall grand averages for the Act- and Ext-sound condition across P(O) conditions at Cz, which was used for determining the different P2 time windows for these conditions (see Methods section). In Figure 3c, the subtraction procedure for the correction of the Act- sound ERPs is illustrated. Figure 4 shows topographical maps of N1 and P2 amplitudes in all conditions and Figure 5 shows N1 and P2 mean amplitudes in all conditions as a function of the RD values. Table 1 contains descriptive statistics for all analyzed components.

3.2.1 | N1 component

Model fit statistics for the N1 amplitude revealed a significant main effect of Sound type, $F(1, 27) = 25.75$, $p < .001$, and parameter estimates suggested significantly smaller amplitudes for Act-sounds compared to Ext-sounds ($b = 4.25$, $p < .001$, see the left scatterplot in Figure 5). No further significant main or interaction effects were found (all *p*s > .114).

3.2.2 | P2 component

First, we performed the analysis on the P2 mean amplitude by separate time windows for the Act-sound and Ext-sound condition as the P2 amplitude showed a noticeable difference in their latencies between the two conditions (for an analysis of the latencies see supplementary materials). The analysis of the amplitudes revealed no significant main effect of Sound type, $F(1, 27) = 0.01$, $p = .934$, but a significant Sound type by P(O) condition interaction, $F(1, 201) = 9.02$, $p = .003$. A follow-up simple effects analysis showed that for Act-sounds amplitudes in the high P(O) condition were significantly lower than in the low P(O) condition, $F(1, 35.12) = 5.22$, $p = .028$, $b = -1.32$, but no effect of P(O) emerged for Ext-sounds, $F(1, 35.12) = 0.03$, $p = .869$. Furthermore, we found a significant Sound type by P(O) condition by RD interaction, $F(1, 201) = 6.91$, $p = .009$ (see the middle scatterplot in Figure 5 and for further visualization of this interaction, Figure 6). Resolving the three-way interaction with a simple slope analysis showed that the interaction between Sound type and P(O) condition was significant for high RD values, $F(1, 201) = 15.86$, $p < .001$, but not for low RD values, $F(1, 201) = 0.07$, $p = .793$ (see Figure 6). In a follow-up simple effects analysis to resolve this interaction, a trend

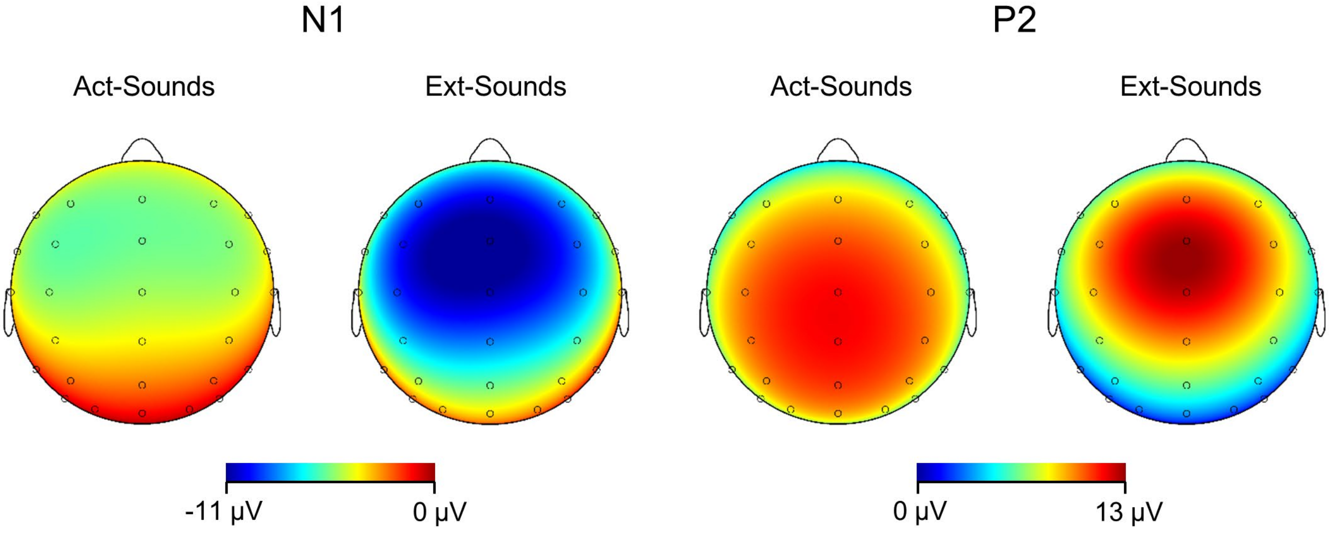

FIGURE 4 Topographical maps showing scalp potentials for the N1 and P2 at grand-average peak latencies separately for the Act-sound condition and Ext-sound condition (pooled across P(O) conditions)

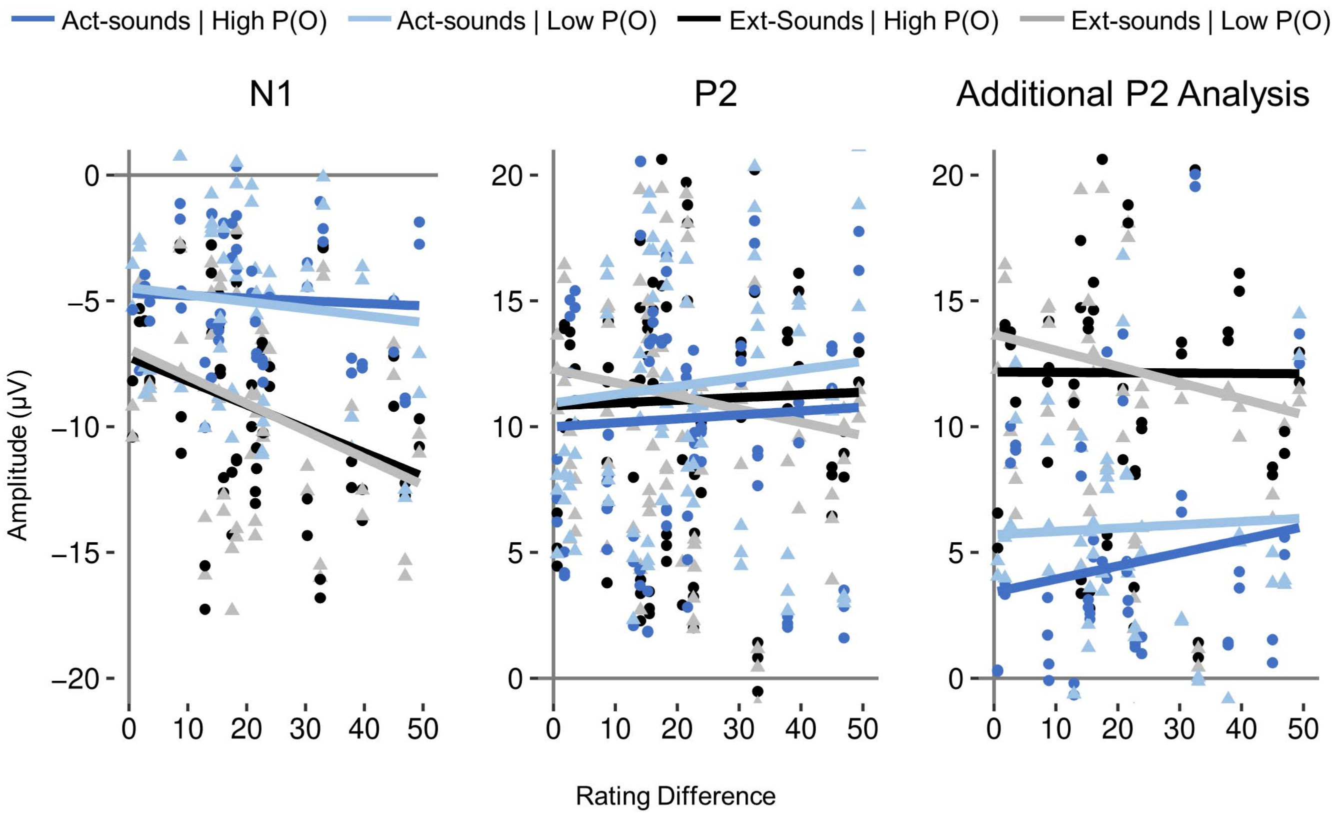

FIGURE 5 Scatter plots for the extracted N1 and P2 amplitudes as a function of participants' rating difference values. Data points and regression lines are shown separately for Act- and Ext-sounds in the low and high P(O) conditions. The plots are restricted to the y-axis segment from 0 µV to −20 µV for the N1 data, and from 0 µV to 20 µV for the P2 data to facilitate visibility of the regression lines, which are still based on the entire data

TABLE 1 Descriptive statistics of data used in the analyses

		Act-sounds		Ext-sounds		Act-sounds		Ext-sounds	
		High P(O)		High P(O)		Low P(O)		Low P(O)	
	Electrode	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
N1 analysis	FCz	−5.18	2.54	−9.55	4.01	−5.37	4.33	−9.49	4.25
	Cz	−4.63	2.66	−8.95	3.77	−4.76	4.32	−8.94	4.07
	CPz	−3.60	2.61	−6.83	3.04	−3.74	3.64	−6.89	3.43
P2 analysis	FCz	9.79	5.61	12.40	6.42	10.87	5.80	12.59	6.69
	Cz	10.69	5.72	11.89	5.99	12.06	6.09	12.03	6.43
	CPz	10.51	5.23	8.90	4.87	12.01	5.82	8.85	4.95
Additional P2 analysis	FCz	4.58	5.30	12.40	6.42	6.01	5.48	12.59	6.69
	Cz	4.47	5.22	11.89	5.99	5.95	5.32	12.03	6.43
	CPz	4.06	4.55	8.90	4.87	5.40	4.78	8.85	4.95

Note: Estimated means and standard deviations for the data used in the N1, P2, and additional P2 analysis (using only data from the Ext-sound P2 peak time window) separately for each electrode and the levels of the factors Sound type and P(O).

emerged for participants with higher values of RD concerning the effect of P(O) for Act-sounds, $F(1, 35.12) = 3.60$, $p = .066$, while the effect was not significant for Ext-sounds, $F(1, 35.12) = 0.83$, $p = .369$. According to parameter estimates, amplitudes for Act-sounds were reduced in the high compared to the low P(O) condition for participants with high RD values ($b = -1.55$, $p = .066$). The other main effects and interactions did not reach significance (all $ps > .201$).

Second, we report an additional analysis of the mean amplitudes for both conditions in the Ext-sound P2 peak time

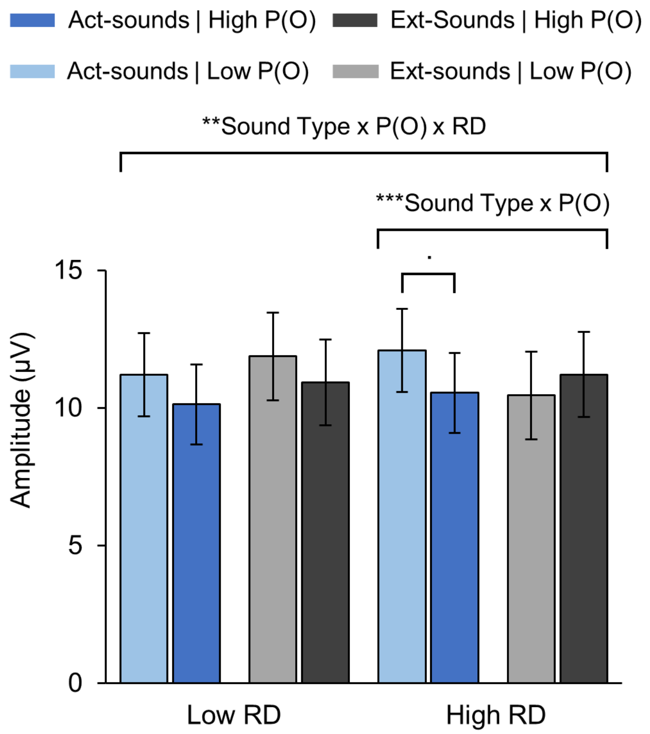

FIGURE 6 Bar plot for the marginal means of the P2 model with separate bars for Act- and Ext-sounds in the low and high P(O) conditions at either low (−1 *SD*) or high (+1 *SD*) values of the continuous, mean-centered factor RD. Error bars show standard errors. Significances yielded by the Sound Type by P(O) by RD interaction and the analyses of the simple effects conducted to resolve it are indicated by . $p < .1$, **$p < .01$., ***$p < .001$

window. The model fit revealed a significant main effect of Sound type, $F(1, 27) = 33.88$, $p < .001$, with smaller amplitudes for Act-sounds compared to Ext-sounds ($b = -6.97$, $p < .001$). The Sound type by P(O) condition interaction was also significant, $F(1, 114) = 15.56$, $p < .001$ (see the right scatterplot in Figure 5). A simple slope analysis to resolve the interaction revealed significantly reduced amplitudes in the high P(O) compared to the low P(O) condition only for Act-sounds, $F(1, 34.7) = 9.54$, $p = .004$, $b = -1.46$, but not for Ext-sounds, $F(1, 34.7) = 0.13$, $p = .724$. None of the other main effects or interactions were significant (all *ps* > .077).

4 | DISCUSSION

In this study, we investigated the influence of the IoC on reductions of neuronal responses to self-generated compared to external sounds. In a within-subject design, we presented participants with two versions of a classic two-button choice task (Jenkins & Ward, 1965) in order to induce either a strong or weak IoC. Explicit ratings of perceived control in the high P(O) and the low P(O) confirmed that our IoC induction worked. Each IoC induction was followed by the same self-generation paradigm, in which participants performed regular button presses in one condition, with 50% of the button presses eliciting a sound, and listened to externally generated sounds in another condition. ERPs time-locked to sound onset were measured during this self-generation paradigm. For the N1 component, an amplitude reduction was found for Act-sounds relative to Ext-sounds, without modulation by IoC. P2 amplitudes were reduced under high versus low IoC, but only for Act-sounds and not for Ext-sounds. Additionally, the strength of IoC induction (measured as the difference in control ratings between the two IoC conditions) modulated the P2 reduction for Act-sounds. The results of the present study thus support our hypothesis that different levels of IoC would affect the later stages of the processing of self-generated sounds.

4.1 | Dissociation between early and late processing stages for self-generated sounds

In the original concept of forward models it has been proposed that one function of the comparison of their predictions with sensory input is to determine if stimuli are self-generated or not (Wolpert & Flanagan, 2001), finally resulting in agency attributions (Picard & Friston, 2014). Research on the reduced N1 amplitude for self-generated tones, which has been linked to forward model predictions, seem to support this assumption. For example, the typical symptoms in schizophrenic patients like passivity experiences and auditory hallucinations have been interpreted as misattributions of agency, resulting from impaired efference copies and consequent failures of matching predicted and experienced sensory input (Feinberg & Guazzelli, 1999; Spering et al., 2013; Swiney & Sousa, 2014; Synofzik et al., 2010). Indeed, diminished auditory N1 reductions were found in schizophrenia patients for sounds elicited by button presses (Ford et al., 2013) as well as for speech (Ford et al., 2007; Perez et al., 2012) compared to externally generated stimuli. Importantly, deficient auditory N1 reductions during an active speech in schizophrenia were found to correlate with errors in judging the source of the voice as one's own (Heinks-Maldonado et al., 2007), but mixed results have been reported for correlations of N1 reductions and schizophrenic symptom severity (Ford et al., 2007, 2013).

Moreover, our result pattern adds to evidence gained in previous studies that N1 reductions are not directly related to the judgment of agency (Kühn et al., 2011; Timm et al., 2016; Weller et al., 2017). Kühn et al. (2011), for example, reported comparable N1 reductions for self-generated sounds which were explicitly rated as self- or externally generated, and Timm et al. (2016) found unaltered N1 reductions in a condition with reduced agency. Corroborating this evidence, our manipulation of perceived control over sound production, which probably affected agency via effects on perceived contingency (Thompson et al., 2007), did not modulate the N1 amplitude. These results seem to suggest that the mere detection of a match between sensory input and motor-information-related predictions, which likely leads to the N1 reduction, is not sufficient for the formation of agency attributions. However, no firm conclusions can be drawn from negative findings, and further hints for interpreting the early component can come by considering it in relation to the later component(s).

Most previous studies on auditory ERPs for self-generated sounds have only distinguished between N1 and P2. The reported P2, however, appeared in different time windows (Horváth et al., 2012; Knolle et al., 2013; Sowman et al., 2012; Timm et al., 2014; Weller et al., 2017), and some did not exclude the possibility that another component, an early P3a (Baess et al., 2011; Ghio et al., 2018; Polich, 2007), contributes to stimulus processing in this paradigm. Furthermore, P2/3a reductions, unlike N1 reductions, were consistently reported also in paradigms and conditions in which predictions based on a forward model played no or only a minor role. For example, intact P2/3a reductions, but diminished N1 reductions, were found in healthy participants in the absence of own actions, for example, for sounds that were visually cued (Sowman et al., 2012), or predictable due to a button press by an observed person (Ghio et al., 2018). Likewise, cerebellar lesion patients with potentially impaired forward model predictions showed an intact P2 reduction, but no N1 reduction for self-generated versus external sounds. These findings indicate that the N1 and the P2/P3a might be functionally dissociated. Considering that both the P2 (Crowley & Colrain, 2004) and the P3a (Polich, 2007) are sensitive to the (un)expectedness of stimuli in oddball paradigms, these findings might suggest that both later components reflect higher level sensory prediction mechanisms that do not, or at least not exclusively, rely on the predictions of forward models.

4.2 | Agency in the early and late processing stages for self-generated sounds

It is interesting to note that Synofzik et al. (2008) conceptually distinguished between an automatic, intuitive feeling of agency about our own authorship, and a conscious judgment about which agent in the environment acted (Synofzik et al., 2008). According to this dual step account, the feeling of agency does not lead to an actual attribution of agency to an agent, but is a purely subjective experience, based on action-related authorship indicators, probably based on a comparison between forward model predictions and sensory feedback. The neural correlates of such a process could thus be reflected in the N1 component. The judgment of agency, moreover, is based on further processing of this feeling by integrating contextual cues and belief states to determine the most likely responsible agent. Evidence for the notion that the P2/3a components reflect this judgment of agency was reported in several studies in which agency was kept ambiguous by varying the timing or quality of a self-generated sound stimulus, requiring a postdictive agency judgment. Kühn et al. (2011) found that P3a reductions in a self-generation paradigm were less pronounced for sounds that were rated as externally generated compared to those that were rated as self-generated. Timm et al. (2016) showed that the magnitude of the P2 reduction in a condition with reduced agency was correlated with the proportion of trials in which no agency was perceived. P2 reductions were also found for sounds generated by observed actions (Ghio et al., 2020), indicating that agency attribution to another actor might be possible. The framework of optimal cue integration (Synofzik et al., 2013) extends the concept of the judgment of agency and proposes that predictive and postdictive sensorimotor and cognitive information are continuously integrated to form agency judgments. Overall, our results concerning the P2/3a can also be interpreted according to this framework, as the prior belief about high action-effect contingencies, which increases the perceived level of control (Thompson et al., 2007) and the agency experience (Moore et al., 2009), resulted in lower P2/3a amplitudes for self-generated tones. Additionally, this effect emerged for participants that perceived a large difference in their personal level of control between the high and low IoC condition, but not for participants experiencing a small difference. We thus showed that P2/3a amplitudes were not simply related to the presented ratio of desired to undesired sounds in the prior IoC training, but to the individual IoC this training induced. Likewise, the effects cannot be explained by differing motor identity associations between the desired sound and button presses induced by the presented ratios, as such a carryover effect should have been found for the N1 as well (Baess et al., 2008; Hughes et al., 2013). For the N1 though, no effect of the IoC was seen, which is also in line with the notion that early processing is more reflective of the classic mechanism of forward model predictions, which is not affected by prior belief information.

4.3 | Distinct late components for self-generated and external sounds?

From the grand-average ERPs obtained in the present study, it was quite obvious that the P2/P3a latency in the Act-sound condition was delayed relative to external sounds, and this difference between conditions was, indeed, significant (see supplementary materials). This is not a common finding in the literature, but an important difference in our compared to previous study designs is that only 50% of button presses were followed by sounds, while most self-generation paradigms entail a 100% probability. The P3a, as well as a later

P2 around 250 ms and thus in the latency range of the positive peak for self-generated sounds in the present study, have been associated with non-target stimuli in oddball tasks (Crowley & Colrain, 2004; Polich, 2007), suggesting that they reflect an orienting response to an unusual or novel stimulus. It is possible that the sounds following button presses in our paradigm prompted such an orientation response, as with a probability of 50% there was maximal uncertainty about whether a button press was followed by a sound or not. The later positive peak we observed for Act-sounds might, therefore, be explained not as a delay of the P2, but as an additional ERP component overlaying the reduced P2 we expected for the processing of Act-sounds. Interestingly, a similar pattern of an additional later component was seen in studies entailing also external sounds that were presented intermixed with the self-generated sound (Baess et al., 2011; Ghio et al., 2018). Specifically, visual inspection suggests that the late positive component for intermixed external sounds, which were presented in irregular intervals between self-generated sounds and less frequently (only in 40% of the trials), had a later peak compared to the more regular external sounds presented in a separate block. Accordingly, these findings could be interpreted in terms of an overlaying, additional ERP component reflecting an orientation response. The notion that the later positive peak for self-generated sounds in the present study reflects a separate ERP component is also supported by the different topography compared to the peak for externally generated sounds.

Considering that sounds in the IoC task were not just task relevant, but also accompanied by a monetary reward or penalty, they might have gained a strong intrinsic motivational significance for participants, which has been shown to enhance P3 amplitudes (Nieuwenhuis et al., 2011). In this respect, our results would imply that the desired sounds in the low P(O) block were associated with a greater personal significance, as P3 amplitudes in the high P(O) block were lower. This could be explained by the lower frequency of monetary rewards in the low P(O) block, which could have led to a high personal significance of the desired sounds as markers of reward, but this line of argument would require an additional control measure. However, the P2/P3 in our paradigm was not exclusively modulated by the frequency of the desired sound, as reflected by the factor P(O), but also by participants' subjective control ratings. It also has to be noted that if personal significance had been established as a property of the sound, Ext-sounds should exhibit a similar P3 as Act-sounds, as both conditions were presented after the IoC task.

While this pattern in the processing of self- and externally generated sounds should be further examined in future studies, it is important to note that the processing differences for self-generated sounds induced by different levels of IoC in the present study were found already between 160 and 200 ms after tone onset, in the time window centered on the P2/P3a peak for the processing of external sounds. It seems that IoC, and possibly agency, affects processing immediately after the typical N1 time window and thus after the efference-copy based comparison between predicted and actual sensory input.

4.4 | Limitations of the study

One concern in interpreting the results of the present study is that we cannot exclude that different expectation strengths for the upcoming tone stimuli caused the differential effect on the P2/3a in the high and low P(O) block. Even though the relationship of button presses and sounds in both IoC training conditions was non-contingent, desired sounds followed button presses much more often in the high IoC than in the low IoC condition. In the high IoC condition, a generally higher expectation of sounds occurring after button presses may thus have led to lower P2/P3a amplitudes, similar to studies in which P2 reductions for self-generated sounds were found when external sounds could be expected based on visual cues. This explanation for our findings seems unlikely, however, because the effect was not found for participants that perceived only small differences in their control over sound production between both conditions. It is of course still possible that interindividual differences in IoC only affected the expectation of sound appearance after button presses in the context of a higher-level prediction process, independently of judgments about the agency. It also must be noted that postdictive agency in our design was never ambiguous, as all sounds presented in the Act-sound condition were generated by the participants, and the cause-effect relationship was never questioned. In future examinations of predictive influences on the agency, perceived control should thus be manipulated in a self-generation paradigm without altering expectedness, while allowing uncertainty about the source of the produced stimuli.

4.5 | Conclusion

We induced either high or low IoC over the production of a sound in order to assess effects on the neural processing of this sound in a subsequent self-generation task, where it was either self-generated or not. We found no effect of the IoC on the N1 reduction for self- versus externally generated sounds, which is in line with the assumption that motor-related prediction mechanisms are reflected in this early processing stage. A reduction of the P2/3a was found when the perceived IoC was high. This suggests that the later processing stage is affected by predictive aspects underlying the judgments of the agency.

ACKNOWLEDGEMENT

Open access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTION

Alexander Seidel: Data curation; Formal analysis; Investigation; Project administration; Software; Supervision; Validation; Visualization; Writing-original draft; Writing-review & editing. **Marta Ghio:** Conceptualization; Investigation; Methodology; Project administration; Supervision; Validation; Writing-original draft; Writing-review & editing. **Bettina Studer:** Conceptualization; Methodology; Writing-review & editing. **Christian Bellebaum:** Conceptualization; Supervision; Writing-review & editing.

ORCID

Alexander Seidel https://orcid.org/0000-0002-9648-4729

REFERENCES

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147–149. https://doi.org/10.3758/BF03334492

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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. *International Journal of Psychophysiology*, *70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *NeuroReport*, *12*(9), 1879–1884. https://doi.org/10.1097/00001756-200107030-00023

Blakemore, S.-J., Smith, J., Steel, R., Johnstone, E., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine*, *30*(5), 1131–1139. https://doi.org/10.1017/S0033291799002676

Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport*, *11*(11), R11–R16. https://doi.org/10.1097/00001756-200008030-00002

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*(6), 237–242. https://doi.org/10.1016/S1364-6613(02)01907-1

Blanco, F., Matute, H., & Vadillo, M. A. (2013). Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. *Learning & Behavior*, *41*(4), 333–340. https://doi.org/10.3758/s13420-013-0108-8

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). A new look at sensory attenuation: Action-effect anticipation affects sensitivity, not response bias. *Psychological Science*, *21*(12), 1740–1745. https://doi.org/10.1177/0956797610389187

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, *9*(8), 587–600. https://doi.org/10.1038/nrn2457

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Feinberg, I., & Guazzelli, M. (1999). Schizophrenia—A disorder of the corollary discharge systems that integrate the motor systems of thought with the sensory systems of consciousness. *The British Journal of Psychiatry*, *174*(3), 196–204. https://doi.org/10.1192/bjp.174.3.196

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology*, *44*(4), 522–529. https://doi.org/10.1111/j.1469-8986.2007.00533.x

Ford, J. M., Palzes, V. A., Roach, B. J., & Mathalon, D. H. (2013). Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia Bulletin*, *40*(4), 804–812. https://doi.org/10.1093/schbul/sbt072

Ghio, M., Egan, S., & Bellebaum, C. (2020). Similarities and differences between performers and observers in processing auditory action consequences: Evidence from simultaneous EEG acquisition. *Journal of Cognitive Neuroscience*, 1–12. https://doi.org/10.1162/jocn_a_01671

Ghio, M., Scharmach, K., & Bellebaum, C. (2018). ERP correlates of processing the auditory consequences of own versus observed actions. *Psychophysiology*, *55*(6), e13048. https://doi.org/10.1111/psyp.13048

Haggard, P., & Whitford, B. (2004). Supplementary motor area provides an efferent signal for sensory suppression. *Cognitive Brain Research*, *19*(1), 52–58. https://doi.org/10.1016/j.cogbrainres.2003.10.018

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of General Psychiatry*, *64*(3), 286–296. https://doi.org/10.1001/archpsyc.64.3.286

Herrmann, C. S., & Knight, R. T. (2001). Mechanisms of human attention: Event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*, *25*(6), 465–476. https://doi.org/10.1016/S0149-7634(01)00027-6

Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Research*, *1626*, 54–65. https://doi.org/10.1016/j.brainres.2015.03.038

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

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *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

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, *79*(1), 1–17. https://doi.org/10.1037/h0093874

Knolle, F., Schroger, E., Baess, P., & Kotz, S. A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of Cognitive Neuroscience*, *24*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schroger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, *49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

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. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition*, *69*(1), 127–137. https://doi.org/10.1016/j.bandc.2008.06.004

Lange, K. (2011). The reduced N1 to self-generated tones: An effect of temporal predictability? *Psychophysiology*, *48*(8), 1088–1095. https://doi.org/10.1111/j.1469-8986.2010.01174.x

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328. https://doi.org/10.1037/0022-3514.32.2.311

Liu, Y., West, S. G., Levy, R., & Aiken, L. S. (2017). Tests of simple slopes in multiple regression models with an interaction: Comparison of four approaches. *Multivariate Behavioral Research*, *52*(4), 445–464. https://doi.org/10.1080/00273171.2017.1309261

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Matute, H., Blanco, F., Yarritu, I., Diaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology*, *6*, 888. https://doi.org/10.3389/fpsyg.2015.00888

Moore, J., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and Cognition*, *17*(1), 136–144. https://doi.org/10.1016/j.concog.2006.12.004

Moore, J. W., Lagnado, D., Deal, D. C., & Haggard, P. (2009). Feelings of control: Contingency determines experience of action. *Cognition*, *110*(2), 279–283. https://doi.org/10.1016/j.cognition.2008.11.006

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/10.31235/osf.io/a5w4u

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162–175. https://doi.org/10.1111/j.1469-8986.2010.01057.x

Palmer, C. E., Davare, M., & Kilner, J. M. (2016). Physiological and perceptual sensory attenuation have different underlying neurophysiological correlates. *Journal of Neuroscience*, *36*(42), 10803–10812. https://doi.org/10.1523/JNEUROSCI.1694-16.2016

Perez, V. B., Ford, J. M., Roach, B. J., Loewy, R. L., Stuart, B. K., Vinogradov, S., & Mathalon, D. H. (2012). Auditory cortex responsiveness during talking and listening: Early illness schizophrenia and patients at clinical high-risk for psychosis. *Schizophrenia Bulletin*, *38*(6), 1216–1224. https://doi.org/10.1093/schbul/sbr124

Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, *83*(12), 1112–1118. https://doi.org/10.1212/WNL.0000000000000798

Pickering, M. J., & Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, *18*(9), 451–456. https://doi.org/10.1016/j.tics.2014.05.006

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

Popa, L. S., & Ebner, T. J. (2018). Cerebellum, predictions and errors. *Frontiers in Cellular Neuroscience*, *12*, 524. https://doi.org/10.3389/fncel.2018.00524

Reznik, D., Henkin, Y., Schadel, N., & Mukamel, R. (2014). Lateralized enhancement of auditory cortex activity and increased sensitivity to self-generated sounds. *Nature Communications*, *5*, 4059. https://doi.org/10.1038/ncomms5059

Reznik, D., Ossmy, O., & Mukamel, R. (2015). Enhanced auditory evoked activity to self-generated sounds is mediated by primary and supplementary motor cortices. *Journal of Neuroscience*, *35*(5), 2173–2180. https://doi.org/10.1523/JNEUROSCI.3723-14.2015

Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia*, *119*, 302–307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and Cognition*, *17*(4), 1219–1227. https://doi.org/10.1016/j.concog.2008.01.003

Schafer, E. W., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science*, *181*(4095), 175–177. https://doi.org/10.1126/science.181.4095.175

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry*, *162*(12), 2384–2386. https://doi.org/10.1176/appi.ajp.162.12.2384

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental Brain Research*, *222*(1–2), 149–157. https://doi.org/10.1007/s00221-012-3204-7

Spering, M., Dias, E. C., Sanchez, J. L., Schutz, A. C., & Javitt, D. C. (2013). Efference copy failure during smooth pursuit eye movements in schizophrenia. *Journal of Neuroscience*, *33*(29), 11779–11787. https://doi.org/10.1523/JNEUROSCI.0578-13.2013

Studer, B., Geniole, S. N., Becker, M. L., Eisenegger, C., & Knecht, S. (2020). Inducing illusory control ensures persistence when rewards fade and when others outperform us. *Psychonomic Bulletin & Review*, *27*(4), 809–818. https://doi.org/10.3758/s13423-020-01745-4

Swiney, L., & Sousa, P. (2014). A new comparator account of auditory verbal hallucinations: How motor prediction can plausibly contribute to the sense of agency for inner speech. *Frontiers in Human Neuroscience*, *8*, 675. https://doi.org/10.3389/fnhum.2014.00675

Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain*, *133*(1), 262–271. https://doi.org/10.1093/brain/awp291

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: An interplay between prediction and postdiction. *Frontiers in Psychology*, *4*, 127. https://doi.org/10.3389/fpsyg.2013.00127

Thompson, S. C., Nierman, A., Schlehofer, M. M., Carter, E., Bovin, M. J., Wurzman, L., Tauber, P., Trifskin, S., Marks, P., Sumner, J., Jackson, A., & Vonasch, A. (2007). How do we judge personal control? Unconfounding contingency and reinforcement in control judgments. *Basic and Applied Social Psychology*, *29*(1), 75–84. https://doi.org/10.1080/01973530701331189

Timm, J., SanMiguel, I., Keil, J., Schroger, E., & Schonwiesner, M. (2014). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of Cognitive Neuroscience*, *26*(7), 1481–1489. https://doi.org/10.1162/jocn_a_00552

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

Timm, J., Schonwiesner, M., Schroger, 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

Vercillo, T., O'Neil, S., & Jiang, F. (2018). Action–effect contingency modulates the readiness potential. *NeuroImage*, *183*, 273–279. https://doi.org/10.1016/j.neuroimage.2018.08.028

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

Voss, M., Ingram, J. N., Haggard, P., & Wolpert, D. M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. *Nature Neuroscience*, *9*(1), 26–27. https://doi.org/10.1038/nn1592

Weller, L., Schwarz, K. A., Kunde, W., & Pfister, R. (2017). Was it me?—Filling the interval between action and effects increases agency but not sensory attenuation. *Biological Psychology*, *123*, 241–249. https://doi.org/10.1016/j.biopsycho.2016.12.015

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.
Supplementary Material

How to cite this article: Seidel A, Ghio M, Studer B, Bellebaum C. Illusion of control affects ERP amplitude reductions for auditory outcomes of self-generated actions. *Psychophysiology*. 2021;58:e13792. https://doi.org/10.1111/psyp.13792

Cognitive, Affective, & Behavioral Neuroscience (2023) 23:1175–1191
https://doi.org/10.3758/s13415-023-01083-7

RESEARCH ARTICLE

My view on your actions: Dynamic changes in viewpoint-dependent auditory ERP attenuation during action observation

Alexander Seidel[1] · **Constanze Weber**[1] · **Marta Ghio**[1] · **Christian Bellebaum**[1]

Accepted: 23 February 2023 / Published online: 22 March 2023

Abstract

It has been suggested that during action observation, a sensory representation of the observed action is mapped onto one's own motor system. However, it is largely unexplored what this may imply for the early processing of the action's sensory consequences, whether the observational viewpoint exerts influence on this and how such a modulatory effect might change over time. We tested whether the event-related potential of auditory effects of actions observed from a first- versus third-person perspective show amplitude reductions compared with externally generated sounds, as revealed for self-generated sounds. Multilevel modeling on trial-level data showed distinct dynamic patterns for the two viewpoints on reductions of the N1, P2, and N2 components. For both viewpoints, an N1 reduction for sounds generated by observed actions versus externally generated sounds was observed. However, only during first-person observation, we found a temporal dynamic within experimental runs (i.e., the N1 reduction only emerged with increasing trial number), indicating time-variant, viewpoint-dependent processes involved in sensorimotor prediction during action observation. For the P2, only a viewpoint-independent reduction was found for sounds elicited by observed actions, which disappeared in the second half of the experiment. The opposite pattern was found in an exploratory analysis concerning the N2, revealing a reduction that increased in the second half of the experiment, and, moreover, a temporal dynamic within experimental runs for the first-person perspective, possibly reflecting an agency-related process. Overall, these results suggested that the processing of auditory outcomes of observed actions is dynamically modulated by the viewpoint over time.

Keywords Auditory ERP · Action observation · Viewpoint · Agency

The phenomenon of attenuated perceptual intensity and reduced neurophysiological responses for self-produced compared with environmental stimuli (Baess et al., 2011; Sato, 2008) is usually attributed to the inherent predictability of these stimuli. There is disagreement, however, about the contribution and role of motor representations in generating such predictions (Dogge et al., 2019; Korka et al., 2021; Picard & Friston, 2014; Reznik & Mukamel, 2019). While in the framework of *predictive coding*, goal-directed motor actions are assumed to contribute more at a cognitive level (i.e., as an intention or goal) to the prediction of sensory action outcomes (Kilner et al., 2007; Picard & Friston, 2014), motor control theory suggests that motor-based *internal forward models* enable this prediction (Blakemore et al., 2001; Miall & Wolpert, 1996; Wolpert & Flanagan, 2001). According to the latter account, copies of motor commands are sent from the supplementary motor areas to the cerebellum, where sensory consequences of the resulting actions are computed, so that these can be considered in perceptual processing (Blakemore et al., 2001).

In the auditory domain, electroencephalography (EEG) studies consistently reported a reduction of the event-related potential (ERP) component N1 and a later positive component (P2/P3a) for self-produced versus external stimuli (Baess et al., 2011; Horváth, 2015; Schafer & Marcus, 1973). The reduction of the auditory N1 amplitude has been associated with motor-based forward model predictions, as it is less pronounced or absent when motor information is lacking, e.g., for visually cued auditory stimuli (Klaffehn et al., 2019; Lange, 2011; Sowman et al., 2012), or when the forward model is assumed to be compromised, as in cerebellar lesion patients (Knolle et al., 2012, 2013a). The N1

Alexander Seidel and Constanze Weber are co-first authors.

✉ Constanze Weber
Constanze.Weber@hhu.de

[1] Institute of Experimental Psychology, Department of Biological Psychology, Heinrich Heine University, Universitätstrasse, 1, 40255 Düsseldorf, Germany

reduction also is less pronounced for involuntary movements (Jack et al., 2021; Timm et al., 2014). On the contrary, in all of these circumstances P2 amplitude reductions have been reported, thus indicating a reliance of P2 modulations on nonmotor predictions. Furthermore, P2 amplitude reductions have been shown to be sensitive to situational context information related to agency, i.e., the sensation of authorship over a stimulus (Kühn et al., 2011; Timm et al., 2016) or the perceived control over stimulus appearance (Seidel et al., 2021).

An aspect that has received less attention so far is the extent to which similar processes that have been proposed to underlie processing of sensations generated by own actions are also shared for sensations caused by observed actions. After the discovery of mirror neurons that discharge both when a goal-directed action is executed and observed (Bonini, 2017; Di Pellegrino et al., 1992; Mukamel et al., 2010; Rizzolatti & Sinigaglia, 2016), it was hypothesized that observed actions also might trigger motor-based forward model predictions, which in turn might modulate sensory processing during action observation, similar to action performance (Wolpert et al., 2003). This hypothesis has been addressed by few electrophysiological studies to date (Ghio et al., 2018, 2021; Poonian et al., 2015). In our previous studies, we found reduced auditory P2 amplitudes for sounds elicited by actions observed on a computer screen (Ghio et al., 2018) and in person (Ghio et al., 2021), whereas N1 amplitudes were reduced only in the former study. One possible explanation for the differential pattern of the N1 may be related to the difference between the studies in stimulus timing. In the study with on-screen observation (Ghio et al., 2018), there was a delay of nearly 200 ms between observed button press onset and tone, whereas in the study with in-person observation (Ghio et al., 2021), there was a delay of approximately 50 ms. In the latter study, the delay may have been too short for action-related information to affect early processing of the action-related sound, as for action observation the motor system is activated later compared with self-action (Sebastiani et al., 2014). Furthermore, study differences in terms of their setting may have contributed to the different result pattern in the sense that motor action animations with standardized, time-controlled visual stimuli (as in the study with on-screen observation, Ghio et al., 2018), in contrast to a naturalistic setting (as in the study with in-person observation Ghio et al., 2021) might have facilitated a motor-based prediction during action observation, possibly in form of an internal forward model, as described above. Similarly as for self-performed action, a reduction of the P2 in both of our previous studies could then be interpreted to reflect a more general predictive mechanism in action observation, which does not necessarily rely on precise motor-related information.

Interestingly, neurophysiological responses by nonhuman primates and humans during action observation also have been shown to be sensitive to the observer's viewpoint (first-person versus third-person). For instance, it has been shown that a large proportion of single cells in the monkey premotor area F5 shows a selective preference, that is, a stronger discharge, for one specific compared with other tested viewpoints (Caggiano et al., 2011; Maranesi et al., 2017). Moreover, a modulation by viewpoint has been shown for local field potentials in the monkey area F5 (Caggiano et al., 2015). Observing motor actions from a first-person compared with a third-person perspective was associated with a significantly stronger power increase in the low-frequency band (2-10 Hz), which also is found during action execution. In humans, mu-rhythm suppression during action observation, which is regarded to reflect "mirror neuron activity" (Pineda, 2005), has been shown to be stronger during first-person compared with third-person observation of reach-to-grasp actions (Angelini et al., 2018; Drew et al., 2015; Fu & Franz, 2014).

This indicates a specialized processing of actions seen in first-person, supporting the notion that correlated visual and motor/proprioceptive experience, which occurs more frequently with a first-person perspective (e.g., when monitoring own actions for correctness during execution) than with a third-person perspective (e.g., when copying the actions of a dance teacher) is key to their neurophysiological coupling (Heyes, 2001). A stronger association between own actions and actions observed from a first-person compared with a third-person perspective also might facilitate sensory predictions relying on motor information and therefore might have (additionally) contributed to the differential pattern of findings of our previous studies, as the animations used by Ghio et al. (2018) showed another person's hand performing button presses in a first-person perspective, while participants in Ghio et al. (2021) observed a person sitting next to them, i.e., from a third-person perspective. A systematic testing of the effects of the observer's viewpoint on sensory predictions, however, is so far missing.

Furthermore, another crucial aspect that has only recently been considered is that motor-based sensory prediction can change within a short period of time, consistently with the adaptive nature of internal models (Miall & Wolpert, 1996). For instance, Kilteni et al. (2019) demonstrated that exposure to a systematic delay (100 ms) between the execution and reception of a self-generated touch led to a decrease of perceptual attenuation for immediately delivered self-initiated touch and an increase of the attenuation for the delayed touch, representing a retuning of the internal (forward) model. As has been pointed out by Dogge et al. (2019), such learning mechanisms may be particularly relevant for predictions concerning environment-related (as opposed to body-related) action outcomes, such as the sounds resulting from button presses in the self-generation paradigm. In one EEG study in the auditory domain, Timm et al. (2016) showed that after exposure to a systematic delay (200 ms) between the action execution and a self-generated sound, amplitude attenuations for self-generated sounds presented immediately and without this delay (versus visually cued external sounds) were reduced, although only for the P2 and not the N1 component. Furthermore, Schneider et al. (2018) showed that mice learn to selectively suppress reafferent auditory cortical responses to auditory sensations that are coupled to their movements via training in an acoustic virtual reality system. Thus, also for observed actions increasing exposure over the course of the experiment could allow for the tuning of an internal model of the observed action and its effect, which would be reflected in dynamic changes in the processing of the action outcome. A possibility to examine such changes over the course of an experiment is to model variables for the time course in multilevel modeling, which is an increasingly popular statistical approach that also has been applied to the analysis of auditory ERPs in recent years (Bolt & Loehr, 2021; Pinheiro et al., 2019; Seidel et al., 2021; for an overview, see Volpert-Esmond et al., 2021). Applied on trial-level data, it allows modeling of trial-to-trial variability and, thus, dynamic changes in the processes underlying ERP components within an experiment. These processes may comprise confounding variables, such as fatigue, as well as variables of interest, such as the above-mentioned learning processes.

The present study addressed the question whether the viewpoint during action observation can affect the processing of auditory consequences of observed motor actions by employing a purely observational version of the standard self-generation paradigm (Horváth, 2015) to test ERP modulations for sounds elicited by actions observed from a first- versus third-person perspective compared with externally generated sounds. Participants watched videos of actors producing sounds by button presses from both viewpoints, while we recorded EEG data and ocular gaze position on the screen. For both viewpoint conditions, sounds were presented around 300 ms after the onset of button press animation, thus enabling a motor prediction available in early auditory processing. To consider potential dynamic changes over time in the processing of auditory consequences of observed motor actions and examine whether the modulatory effect of the viewpoint appears in a time-variant fashion, we analyzed the data with linear mixed-effects models, as an application of multilevel modelling, on trial-level data. By applying this approach, we not only modelled the effects of Sound Type (action- and externally generated sounds) and Viewpoint (first- and third-person) as experimental factors to test viewpoint-dependent auditory ERP attenuation for action-generated versus externally generated sounds during action observation, but we also modelled the temporal structure of the experiment to consider potential learning processes in action observation. More specifically, we included two predictors that model *time* on different levels, that is, a) the predictor Run to account for the division of each experimental condition in two identical experimental runs presented, respectively, in the first and second half of the experiment, and b) the predictor Trial number to model developments with increasing trials within each run for each condition (see *Methods*).

For the N1 component, in line with the hypothesis that the mirror neuron system might be involved in generating action-observation based predictions, at least when the delay between action and its consequence is long enough to enable a motor prediction available in early auditory processing (Ghio et al., 2018), we expect the amplitude to be reduced for observed action- compared to externally generated sounds for both viewpoints. Furthermore, based on evidence for a stronger involvement of the mirror neuron system for a first-person perspective (Angelini et al., 2018; Fu & Franz, 2014), we expected a stronger N1 reduction for sounds resulting from observed actions for the first- than the third-person perspective condition. Effects of the time course that interact with the type of sound are of particular interest, as such interactions could reflect learning processes in action observation. Consistent with the role of learning in predicting environmental action outcomes (Dogge et al., 2019), we expected that the general N1 attenuation for self-generated sounds would become stronger over the course of the experiment. Effects of the time course that, in addition to the type of sound, depend on the viewpoint during observation could indicate a difference in the ease of learning between viewpoints. Because learning a motor-based prediction, as hypothesized for the N1, might be facilitated by a stronger motor involvement as associated with a first-person perspective during action observation, we expected the increase in N1 attenuation over time to be stronger for the first-person than for the third-person perspective.

Regarding the P2 component, previous studies suggested that this component is not sensitive to motor-related, but rather general, context-dependent predictions (Knolle et al., 2013a; Seidel et al., 2021), and thus it is unlikely that differing mirror neuron system activity in the first- versus third-person perspective would affect this component. Because the motor action to be observed cues sound onset for both viewpoints equally and, thus, allows context-dependent predictions, we expected the P2 to be reduced for observed action-generated compared with externally generated sounds irrespective of the viewpoint conditions, but, in contrast to the N1, without a stronger reduction for the first- versus third-person perspective (as this was hypothesized to reflect a stronger involvement of the mirror neuron system in action-based prediction mechanisms not involved here). On the contrary, given that P2 amplitude reductions have been shown to be sensitive to agency ambiguity (Kühn et al., 2011; Timm et al., 2016), we speculated that the first-person perspective may introduce such an ambiguity, as it is similar to looking at one's own

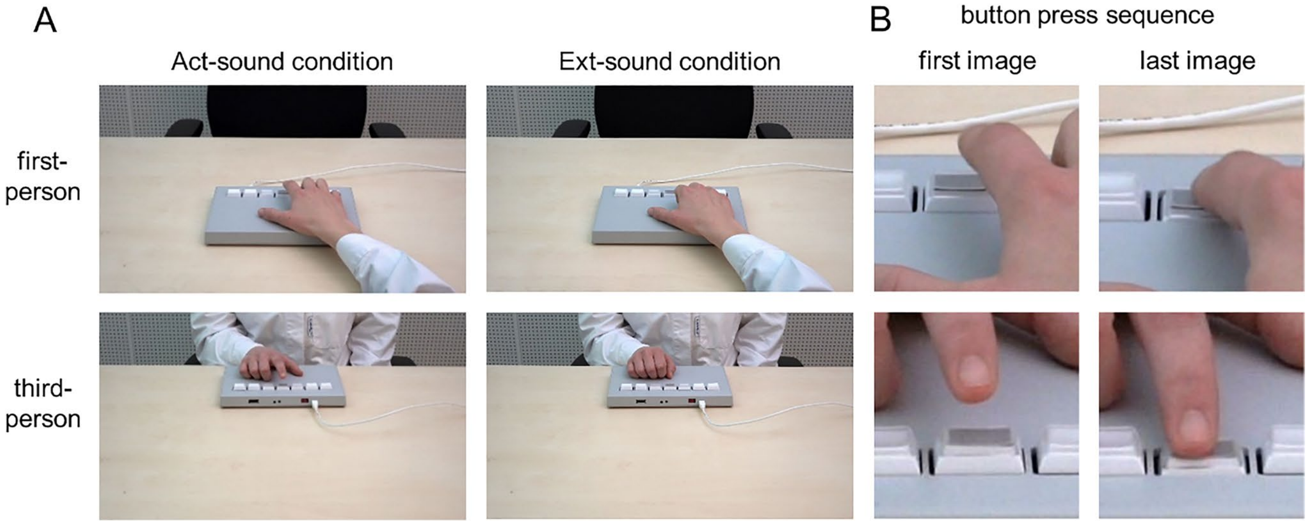

Fig. 1 **A**: Images from the first- and third-person perspective showing the female actor holding the right index finger in the starting position for each button press (Act-sound condition) and resting the right hand on the button box (Ext-sound condition). **B:** Close-up of the first and the last image (of the sequence of 10 images) of the Act-sound condition from the first- and third-person perspective, showing the right index finger of the female actor in the starting position and while fully pressing the button. Please note that the same images used for the button press sequence in the Act-sound condition also were used for the Motor-only condition

hands. This could result in a diminished P2 reduction compared to the third-person perspective, which emphasizes the self-other distinction, and may thus enable an easier agency judgement. Concerning the effects of the time course, we can hypothesize that the suspected advantage in agency attribution for the third-person perspective might become less strong over the course of the experiment, as any initial agency ambiguity for the first-person perspective might dissolve over time, reflected in an increase in P2 attenuation over time only for the first- but not for the third-person perspective.

Method

Participants

Twenty-seven participants (13 females, 14 males, mean age 23.7 years ± 4.4) with normal or corrected-to-normal vision and normal hearing took part in the experiment. The sample size was thus slightly larger compared to our previous studies where we found within-subject modulations of the auditory attenuation effect in groups of 20 participants (Ghio et al., 2018, 2021). Except for one participant, all reported to be right-handed. None of the participants reported a history of neurological disease, mental disorder or current medication affecting the central nervous system. Informed written consent was obtained from each participant before the experiment. Participants received either course credit or money as compensation. This study was approved by the Ethics Committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf, Germany.

Materials

Visual stimuli

As outlined in the Introduction, the purpose of the experiment was to study the processing of sounds that were elicited by observed button presses. We therefore recorded videos showing the hand of either a male or female actor pressing a button on a Cedrus RB-740 response pad (www.cedrus.com) with their right index finger. Furthermore, the viewpoint was varied. The button presses were shown either from the first- or from a third-person perspective, so that there were four versions of the video in total. The observed persons wore a white lab coat, which was identical to the one worn by the participants during the EEG recording (see below), and the videos were recorded in the same EEG-chamber in which the EEG data acquisition was conducted. Figure 1A contains example images showing the female actor (for further information on the video recording, including example images of the male actor, see supplementary material S1).

For visual stimulation during the experiment, however, ten consecutive images (resolution 1920 × 1080 px) were extracted from each video and were shown in succession (see below), so that the ERPs could be time-locked to comparable points in time during the observed button press for each version. The first image of each sequence showed the finger above the button, while in the last image the button was fully pressed (Fig. 1B). During the experiment, the images were shown on a 60-Hz monitor, with consecutive images appearing every second frame, so that each image was shown for approximately 33 ms, which led to the impression of a fluid button press motion (see below for further details). In addition to the images for the observed button press, one image was taken per version (female or male in first- or third-person perspective), which showed the hand of the actor on the button box in a closed fist (Fig. 1A). This image was used for the condition in which no button presses were observed (see below).

Auditory stimulus

The sound played during the experiment was created with MATLAB R2019a (MathWorks Inc., Natick, MA) and delivered via over-ear headphones (Sennheiser HD 201) with the same duration and pitch for all experimental conditions (1000 Hz, 200-ms duration, 20-ms fade in/out).

Experimental design

We adapted the block designed self-generation paradigm (Horváth, 2015), which usually involves active button presses by the participants, in order to create an observational variant of it. Although the participants of the present study only engaged in an observational variant of the self-generation paradigm and never performed actions themselves, we will use the labels for the experimental conditions that are usually used in active versions of the paradigm (see previous studies involving observational versions of the self-generation paradigm, Ghio et al., 2018, 2021). The paradigm involved the observational variant of the three standard conditions of the self-generation paradigm (Horváth, 2015). In one condition, the observed actor performed a sequence of button press actions that elicited sounds (Act-sounds). In another condition, externally generated sounds (Ext-sounds) not preceded by an observed button press were played. Finally, in the Motor-only condition, the observed actor performed button presses without producing sounds. The Motor-only condition merely served to control for effects of movement observation on the ERPs (see below for details), and only the ERPs from the motor-corrected Act- and Ext-sound conditions entered the analysis. Importantly, and in accordance with the main purpose of the present study, the viewpoint of the observed action (first- versus third-person) was added as a further factor to the paradigm, yielding a 2 x 2 experimental design, with the factors Sound Type (Act-sounds, Ext-sounds) and Viewpoint (first-person, third-person) as within-subject factors. The different conditions are explained in detail in the following.

Act-sound condition

Trials of the Act-sound condition started with the presentation of the first image of the button press sequence according to the condition (first- or third-person), showing the finger of a female or male actor (gender-matched to the participant) in the starting position above the button for an average of 1600 ms (± 200 ms random variance). Then, the next nine images in the sequence were shown consecutively to illustrate a button press by the observed person. Eight images were shown for approximately 33 ms each; the last one showed the button fully pressed was presented for ca. 267 ms. Then, the images were shown in reversed order with identical timing to illustrate the button release. The total duration of the observed button press and release was ca. 800 ms. The average interval between observed button presses was 2400 ms (Fig. 2). Importantly, the tone was time-locked to the image showing the button fully pressed, with a delay of approximately 30 ms. It thus appeared ca. 300 ms after the start of the observed button press. Participants were instructed to observe the actions and listen to the sounds attentively. To ensure that participants focused on the button press action, binocular gaze positions were continuously recorded using a dark pupil eye-tracker (see below for details on eye-tracking data acquisition and processing).

Ext-sound condition

During the Ext-sound condition, all sounds from the previous Act-sound condition were replayed with the same timing, but importantly they were not preceded by an observed button press. Instead, the image of the resting hand (again gender-matched to the participant) was continuously presented. In this sense, Viewpoint (first- versus third-person) also was varied in the Ext-sound condition, although no button press actions were performed. Participants were instructed to attentively fixate the button and listen to the sounds.

Motor-only condition

Because the Motor-only condition is usually employed in the self-generation paradigm to account for electrophysiological responses solely driven by the motor action (Horváth, 2015), we applied an analogous correction procedure for observed button presses as we did in our previous work (Ghio et al., 2018, 2021). The visual stimulation in the Motor-only condition was identical to the visual stimulation in the Act-sound condition, but no sounds were presented. Participants were instructed to attentively observe the button press actions. Binocular gaze positions were continuously recorded in this condition as well to ensure that participants focused on the button press action.

Experimental procedure

Participants first completed the consent form and a brief demographic questionnaire. For the EEG acquisition session, during which also eye-movement data were recorded, participants were seated in an electrically and acoustically shielded

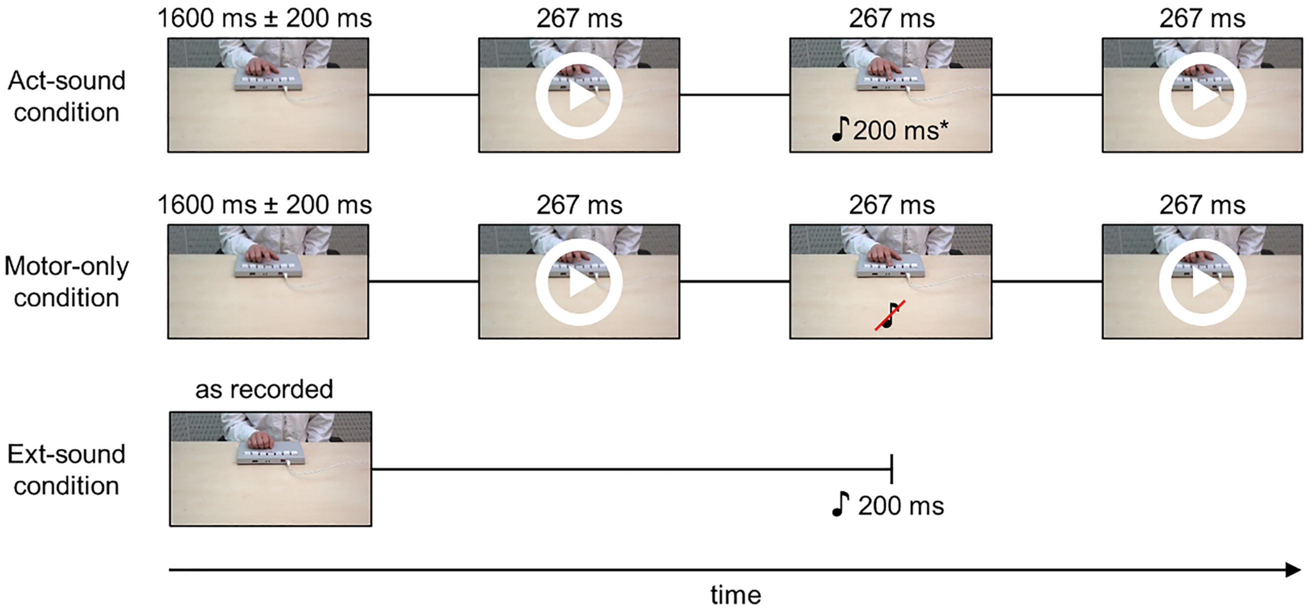

Fig. 2 Experimental sequence for the three conditions of the observational variant of the self-generation paradigm, with example images from the female actor from the third-person perspective. The images overlaid with a white "play" sign represent the start of an animation of 8 images at a rate of 33.3 ms. Presentation times are rounded. *Sound onset approximately 30 ms after image onset

chamber, in front of a 22-inch LCD monitor with a resolution of 1680 × 1050 px. A chin-rest was placed at a preset height for a fixed viewing distance of 62 cm to the screen.

Following the instructions, 11-trial versions of the Act-sound and Ext-sound conditions were presented, once from each viewpoint to familiarize participants with the experimental procedures. This was followed by four experimental runs, each consisting of the three conditions as separated blocks in a fixed order (i.e., Act-sound, Ext-sound, Motor-only). The different runs alternated between showing the observed person from the first- or third-person perspective, with the starting viewpoint counterbalanced between participants. This resulted in two identical runs per Viewpoint, one in each half of the experiment, yielding the two-level predictor Run in the analysis. Before each condition, participants were informed whether they would observe button presses and/or listen to sounds and could take a self-administered break before starting the condition. Fifty-one button presses and/or sounds were presented within each condition in each run, the first of which was disregarded in the analyses, for a total of 100 Act- and 100 Ext-sounds entering analysis for each viewpoint (50 from each run). Stimulus presentation was controlled via Presentation® software (Version 20.3, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). To assure that the time delay between sounds and the corresponding EEG sound markers was minimized, a Sound Blaster Audigy Rx (Creative Technology Ltd., Singapore) in "Bit Accurate Playback" mode was used and the Presentation Mixer was set to "exclusive.".

Eye movement data acquisition and analysis

For each condition involving button presses, binocular gaze positions were continuously recorded using a dark pupil eye-tracker (iView X RED 500, SensoMotoric Instruments) to ensure that participants focused on the button press action. A 9-point calibration of the eye tracker was conducted before each condition. After calibration accuracy was briefly checked by the experimenter, a focus point for observing the actor's finger was established for the upcoming condition block. For this purpose, the first image of the button press sequence for the upcoming condition was presented, and participants were instructed to focus their gaze on the finger for one second and press the space bar. The gaze position detected by the eye tracker at this moment was then displayed and confirmed by the experimenter if it was in the general vicinity of the finger. Otherwise, the focus detection or the entire calibration procedure was repeated. This focus point was saved for each condition for the later analysis of the gaze position data. To familiarize participants with this procedure, it also was performed for the two Act-sound condition trainings, but data were not recorded.

Eye movements were recorded at a sampling rate of 500 Hz with the iView X software (Version 2.8). The raw gaze data (extracted with the IDF Converter 3.0.20, SensoMotoric Instruments) was analyzed offline in MATLAB R2019a. After excluding gaze positions that were detected in only one eye, each trial in the Act-sound and Motor-only condition was checked to determine whether the participants focused on the observed button press. For all gaze positions occurring in the 300 ms after the onset of the observed button press (first frame of the second image of the sequence), and thus until the tone was presented, we calculated the distance to the focus point established after the eye tracker calibration for the respective condition in each block. Trials were excluded from further analyses if less than 75% of gaze positions in the 300-ms interval were within 200 px (approximately 5° viewing angle) of the individual focus point. The corresponding trials in the following Ext-sound condition were excluded as well to ensure that the trials that entered analysis were preceded by identical inter-sound-intervals in both conditions. Two participants for whom more than 25% of trials had to be excluded due to this procedure were excluded from further analyses.

EEG data acquisition and analysis

EEG acquisition

EEG was continuously recorded after the start of the first experimental run with Ag/AgCl passive electrodes positioned according to the 10-20 system at a sampling rate of 1000 Hz and referenced to linked mastoids during acquisition. The signal was amplified using a BrainAmp Standard amplifier and recorded via BrainVision Recorder software (Version 1.21.0402, Brain Products GmbH, Germany). Four electrodes were used to measure the electrooculogram (electrodes F9 and F10 for horizontal, Fp2 and a separate electrode below the right eye for vertical eye movements). AFz position was used for the ground electrode. Impedances were kept below 5 kΩ. The other 28 electrodes were placed at the following positions: F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P7, P3, Pz, P4, P8, PO7, PO3, POz, PO4, and PO8.

Preprocessing

Preprocessing was performed with BrainVision Analyzer software (Version 2.1.2, Brain Products GmbH, Germany) and MATLAB R2019a. After a global direct current de-trend, Butterworth zero-phase filters (low cutoff: 0.3 Hz, 24 dB/oct; high cutoff: 30 Hz, 24 dB/oct), a notch filter (50 Hz), and a semiautomatic independent component analysis (ICA, steps = 512) for the removal of ocular artifacts were applied. The corrected data were segmented into epochs of 700 ms, starting 200 ms before sound onset (a muted sound stimulus was played in the Motor-only condition for this purpose). Segments underwent an automatic artifact rejection (maximal allowed voltage step: 50 μV/ms, maximal allowed difference of values within 100-ms intervals: 100 μV, maximal/minimal allowed amplitude: ± 100 μV, lowest allowed activity within 100-ms intervals: 0.5 μV) and were subsequently baseline corrected using the interval of 200 ms before (muted) sound onset.

Similar to studies in which sounds were actively produced by motor actions (Horváth, 2015), activity evoked by motor observation was removed from the Act-sound segments (similar to Ghio et al., 2018, 2021). The motor correction was applied for each participant, Viewpoint and Run separately. For this purpose, segments of the Motor-only condition were averaged, separately for the first and second run in the experiment, the first- and the third-person perspective and each participant. Then, we subtracted the averaged Motor-only segment (run- and viewpoint-specific) from each individual Act-sound segment (of the corresponding viewpoint and run) to enable analysis based on single-trials (for a visualization of the grand averages of the uncorrected and corrected Act-sounds and Motor-only segments, see supplementary material S3). Visual inspection of these grand averages suggested that the ERPs in the motor-only condition might differ between viewpoints. This was explored in a separate analysis reported in the supplementary material S3. Importantly, since the motor correction was performed separately for the two viewpoints, the motor-corrected Act-sound segments (hereafter referred to as Act-sounds) that entered all further analyses were adjusted for such differences in activity evoked by the motor observation per se and were intended to reflect sound processing only.

To determine the ERP components of main interest, we then created an overall grand average across the Act-sound and Ext-sound conditions for both viewpoints (Fig. 3A). Visual inspection of these grand averages suggested that the signal was modulated by the experimental conditions not only in the N1 and P2 time windows, as expected, but also at a negative peak around 300 ms, which also was explored. In accordance with Sugimoto et al. (2021), we will refer to this component as N2. Analysis of the N2 component reported below are exploratory and not based on a priori hypotheses. We determined Fz, FCz, and Cz as the appropriate electrodes for our analyses, based on the topographical maps shown in Fig. 3B for the overall grand average at the peaks of the components of interest (a similar approach was applied in Seidel et al., 2021).

To prepare the dataset to apply linear mixed effects models on trial-level data, data extraction for each component was performed in three steps. First, we localized peaks in the grand average collapsed over all conditions and participants for each electrode. The latencies of these peaks (averaged between electrodes and rounded) were 94 ms (N1), 167 ms (P2), and 284 ms (N2). Second, we determined peaks in the data averaged separately for each Sound Type, Viewpoint, Participant, Electrode, and Run in a time window of 100 ms around the peaks found in the previous step. Since for some subjects, no peaks were found in these time windows, we extended the time window for the P2 (117–230 ms) and N2 (220–334 ms) detection. The N1 detection time window was slightly shortened (50–144 ms) to avoid the selection of a very early separate peak in only one condition for one participant. In a third step, we calculated a mean amplitude value for each trial, averaging the amplitude values from a 40-ms time

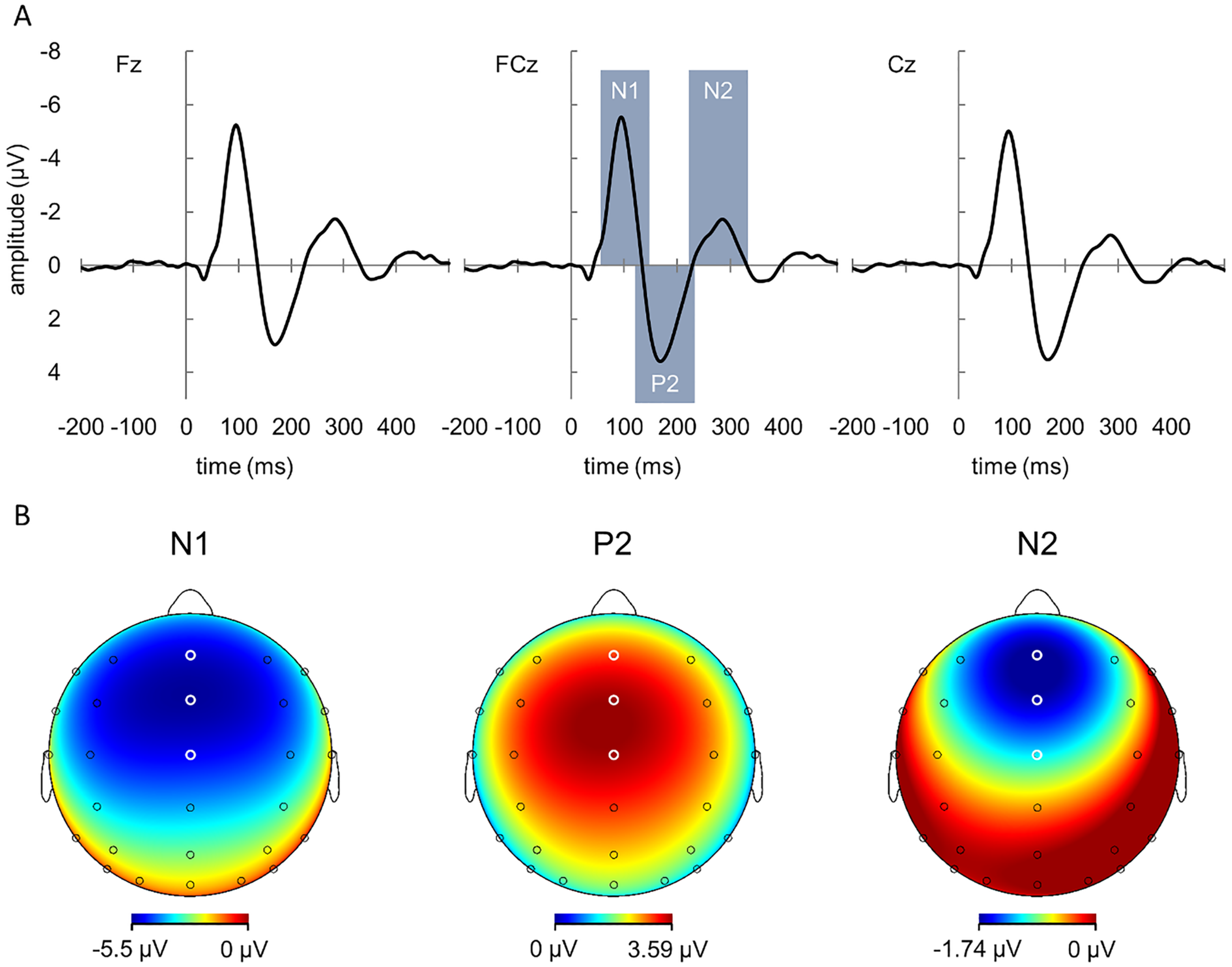

Fig. 3 **A:** Overall sound-related grand average ERPs at Fz, FCz, and Cz across Sound Type and Viewpoint conditions, and the time windows (grey rectangles) used for mean amplitude extraction. **B:** Topographical maps showing scalp potentials at the time of the N1, P2, and N2 peaks from the overall grand average ERPs seen in A. White circles indicate positions for the electrodes Fz, FCz, and Cz (from top to bottom)

window centered on the latency of the previously detected peak for this trial's condition and electrode.

For the purpose of detecting outliers in the single-trial mean amplitudes processed for statistical analysis, means and standard deviations were calculated across trials separately for each Sound Type, Viewpoint, Electrode, and Run combination. Single-trial values were removed if they differed more than 2.5 standard deviations from the respective mean. The resulting number of trials per condition (averaged across participants) after preprocessing and outlier-removal can be found in Table 1.

Statistical analysis

Single-trial mean amplitude data from the N1, P2, and the N2 component were analyzed separately by applying the same procedure. Specifically, each dataset was fitted, using the restricted maximum likelihood approach, to the same linear mixed effects model, which included the simple coded fixed-effect predictors Sound Type (Ext-sounds [−0.5], Act-sounds [0.5]), and Viewpoint (first-person [−0.5], third-person [0.5]) as the experimental factors of main interest. To model the course of the experiment, as suggested by Volpert-Esmond et al. (2021), two additional predictors were added. As each experimental condition was presented in two identical runs of 50 trials (for a total of 100 trials per condition), one in the first half and one in the second half of the experiment, we included the fixed-effect predictor Run (first [−0.5], second [0.5]) to account for the temporal separation of the two sets of trials. We also added the continuous fixed-effect predictor Trialnumber (1-50) to model developments over the 50 trials in each run. This predictor coded the original temporal position of each trial within each condition and accommodated rejected trials. For example,

Table 1 Average number of trials after preprocessing and outlier-rejection

Dataset	Electrode	Act-sounds first-person M (SD)	Ext-sounds first-person M (SD)	Act-sounds third-person M (SD)	Ext-sounds third-person M (SD)
After preprocessing	All	94.6 (6.5)	94.2 (6.5)	94.4 (7.4)	93.9 (7.8)
After outlier-rejection					
N1	Fz	93.3 (6.8)	93.5 (7.5)	93 (6.8)	92.7 (7.8)
P2	Fz	93.4 (6.3)	93.6 (7.2)	92.7 (6.5)	92.8 (7.4)
N2	Fz	93.6 (6.5)	93.5 (7.3)	93 (6.6)	92.5 (7.7)
N1	FCz	93.3 (6.7)	93.4 (7.4)	93.1 (6.7)	92.9 (8)
P2	FCz	93.6 (6.3)	93.5 (7.1)	92.7 (6.5)	92.9 (7.9)
N2	FCz	93.6 (6.3)	93.4 (7.3)	93.2 (6.5)	92.5 (7.9)
N1	Cz	93.1 (6.6)	93.6 (7.5)	93.3 (6.8)	92.9 (8.2)
P2	Cz	93.6 (6.1)	93.4 (7.2)	92.7 (6.6)	92.7 (7.9)
N2	Cz	93.4 (6.2)	93.3 (7.5)	93 (6.6)	92.4 (7.7)

In preprocessing, trials were removed if the action was not observed (according to eyetracker data), and if they did not pass automatic artifact rejection. The final dataset for each component was determined after individual outlier-rejection, resulting in differing trial numbers per component and electrode. The maximum number of trials per condition was 100

if trial 8 was rejected, this did not lead to a numbering from 1-49, but from 1-7 followed by 9-50. It is important to note that linear mixed effects models using maximum likelihood estimation techniques are robust to such unbalanced missing observations (Krueger & Tian, 2004). For all participants, the predictor was then centered around the fixed value of 25.5 instead of the actual mean of trial numbers, because this prevented rejected trials from shifting the centering away from the factual middle of the block. While the predictor Run tested the difference between responses in the first versus second half of the experiment and can thus reveal coarse changes in processing, the predictor Trialnumber can provide information on more fine-grained changes within a Run. The model also contained all possible interactions between all the fixed-effect predictors.

Concerning the random effects, to determine the maximal random-effect structure that still allows the model to converge, we started with random intercepts for participants and random slopes for the predictors Sound Type and Viewpoint and their interaction over participants, and random intercepts for electrodes. The only model that converged when adding random slopes for Run and Trialnumber over participants included random slopes for Run over participants, but no interactions with the other predictors. The final model can therefore be described as:

$$\text{Mean Amplitude} \sim \text{Sound Type} * \text{Viewpoint} * \text{Run} * \text{Trialnumber}$$
$$+ (1 + \text{Sound Type} * \text{Viewpoint} + \text{Run} \mid \text{Participant}) + (1 \mid \text{Electrode})$$

Statistical analyses were conducted in R (Version 4.0.3) using the lme4 package (Version 1.1-26). To test for significant effects, *p* values were calculated with the lmerTest package (Version 3.1-3) with Satterthwaite approximated degrees of freedom. In case of significant interactions, we examined them by performing simple effects analyses. For two-way interactions involving categorical fixed effects (e.g., Sound Type, Viewpoint), we calculated two models, in which one of the two predictors involved in the interaction was dummy-coded (0, 1). One model used the first level of the predictor as the reference level, and the other used the second level. For both models, we then tested the main effect of the second predictor involved in the interaction. For two-way interactions involving the continuous predictor Trialnumber (for which dummy-coding is not possible), when resolving by Trialnumber, we centered this predictor around early trials for one model, and around late trials for the other. The centering values of 13 and 38 were determined by adding/subtracting 12.5 (25% of the possible 50 trials per run) from the centering of 25.5, which was used in the main analyses, yielding values that represented the 25th and 75th percentile of the Trialnumber value. For three-way interactions, we subsequently examined the relevant two-way interaction at each level of the first predictor. In the simple effects analyses, the same random effects were specified as in the main analysis. An α level of 0.05 was considered statistically significant. Interactions are only reported when they involve the predictor Sound Type. Analysis code including output (with parameter estimates of all fixed effects for the three analyzed components) can be found at https://doi.org/10.17605/OSF.IO/FGRB3.

Results

For visualization purposes, Figure S3 in the supplementary material shows grand averages in each of the two runs for the Act- and Ext-sound conditions in the first- and third-person perspective. To visualize potential effects of the predictor Trialnumber, grand averages were calculated separately for two bins of trials for each run (trials 1-25 and 26-50, respectively). Figure 4 provides line plots of the marginal estimated means for both types of sounds derived from the linear mixed effects models separately for early and late trials (corresponding to the values tested in follow-up simple effects analyses) in both runs and each viewpoint for the three ERP components that were analyzed.

N1 component

The model fit for the N1 amplitudes yielded a significant main effect of Sound Type, $F(1, 25.88) = 8.54$, $p = 0.007$, $b = 0.61$, with less negative mean amplitudes for Act-sounds than for Ext-sounds. The main effect of Viewpoint, $F(1, 25.20) = 0.11$, $p = 0.745$, and the Sound Type by Viewpoint interaction did not reach significance, $F(1, 25.86) = 0.36$, $p = 0.556$.

We found a significant interaction of Sound Type by Trialnumber, $F(1, 30046.18) = 20.58$, $p < 0.001$, as well as a three-way interaction of Sound Type by Viewpoint by Trialnumber, $F(1, 30045.19) = 15.41$, $p < 0.001$ (Fig. 4). Simple effects analyses to solve the three-way interaction revealed that the interaction of Sound Type by Trialnumber is only significant for the first-person, $t(30041.49) = 5.98$, $p < 0.001$, but not the third-person perspective, $t(300048.45) = 0.43$, $p = 0.666$. Further simple effects analyses showed that for the first-person perspective, amplitudes for Act-sounds were significantly reduced compared to Ext-sounds in late trials across runs, $t(33.36) = 3.42$, $p = 0.002$, $b = 1.18$, but not in the early trials, $t(32.65) = -0.65$, $p = 0.519$. For the third-person perspective, this reduction for Act-sounds amplitudes were found for early, $t(34.74) = 2.21$, $p = 0.034$, $b = 0.70$, as well as late trials, $t(34.82) = 2.53$, $p = 0.016$, $b = 0.80$. An alternative resolution showed that during first-person observation, amplitudes became significantly less negative with increasing Trialnumber for Act-sounds, $t(30045.76) = 5.91$, $p < 0.001$, $b = 0.04$ (a reduction of 1.96 µV over 50 trials), whereas Ext-sounds amplitudes became significantly more negative, $t(30045.07) = -2.55$, $p = 0.011$, $b = -0.02$ (an increase of 0.85 µV over 50 Trials). During third-person observation, no significant effect of Trialnumber was found, both $p > 0.157$. No other interaction with Sound Type reached significance, all ps > 0.363.

P2 component

Fitting the P2 amplitudes to the model revealed no significant main effects of Sound Type, $F(1, 26.22) = 3.23$, $p = 0.084$, or Viewpoint, $F(1, 26.00) = 0.04$, $p = 0.846$, or Sound Type by Viewpoint interaction, $F(1, 25.98) = 1.11$, $p = 0.302$.

We found a significant Sound Type by Run interaction, $F(1, 30066.87) = 25.58$, $p < 0.001$ (Fig. 4). Follow up simple effects analyses revealed a significant reduction of amplitudes for Act- versus Ext-sounds in the first run, $t(34.35) = -3.49$, $p = 0.001$, $b = -1.01$, but not in the second run, $t(34.49) = -0.13$, $p = 0.901$. All other interactions with Sound Type did not reach significance, all ps > 0.105.

Exploratory analysis: N2 component

The model fit for mean amplitudes from the late N2 time window revealed a significant main effect of Sound Type, $F(1, 26.05) = 18.55$, $p < 0.001$, $b = 0.96$, reflecting less negative amplitudes for Act- compared to Ext-sounds. The main effect of Viewpoint, $F(1, 25.83) = 0.75$, $p = .394$, and the Sound Type by Viewpoint interaction, $F(1, 26.16) = 2.64$, $p < .116$, were not significant.

We found a significant Sound Type by Run interaction, $F(1, 30055.38) = 12.71$, $p < 0.001$. Subsequent simple effects analyses showed significantly smaller amplitudes for Act- compared to Ext-sounds, both for the first run, $t(39.57) = 2.32$, $p = 0.025$, and the second run, $t(39.90) = 5.42$, $p < 0.001$, but parameter estimates revealed that the amplitude difference for the second run ($b = 1.35$) was larger than for the first run ($b = 0.58$).

There were significant interactions between Sound Type and Trialnumber, $F(1, 30038.60) = 12.27$, $p < 0.001$, and between Sound Type, Viewpoint, and Trialnumber, $F(1, 30036.63) = 8.00$, $p = 0.005$. Simple effects analyses to solve the three-way interaction showed that the Sound Type by Trialnumber interaction was only significant for the first-person, $t(30031.68) = 4.48$, $p < 0.001$, but not for the third-person perspective, $t(30040.93) = 0.48$, $p = 0.634$. Subsequent simple effects analyses in first-person perspective revealed significantly reduced Act- compared with Ext-sound amplitudes in early, $t(33.60) = 2.03$, $p = 0.050$, $b = 0.76$, and late trials, $t(34.37) = 5.15$, $p < 0.001$, $b = 1.95$, but with higher parameter estimates for late trials. The analysis for third-person perspective showed no significant Sound Type effect in early trials, $t(36.29) = 1.48$, $p = 0.148$, or in late trials, $t(36.13) = 1.86$, $p = 0.071$. No other interaction with Sound Type reached significance, all ps > 0.116.

Discussion

The present study investigated whether the viewpoint during action observation affects the sensory processing of auditory effects elicited by observed motor actions and whether this hypothesized modulatory effect dynamically changes over time. ERP components associated with auditory processing were compared between sounds generated by actions observed from a first-person and third-person perspective and externally generated sounds during the

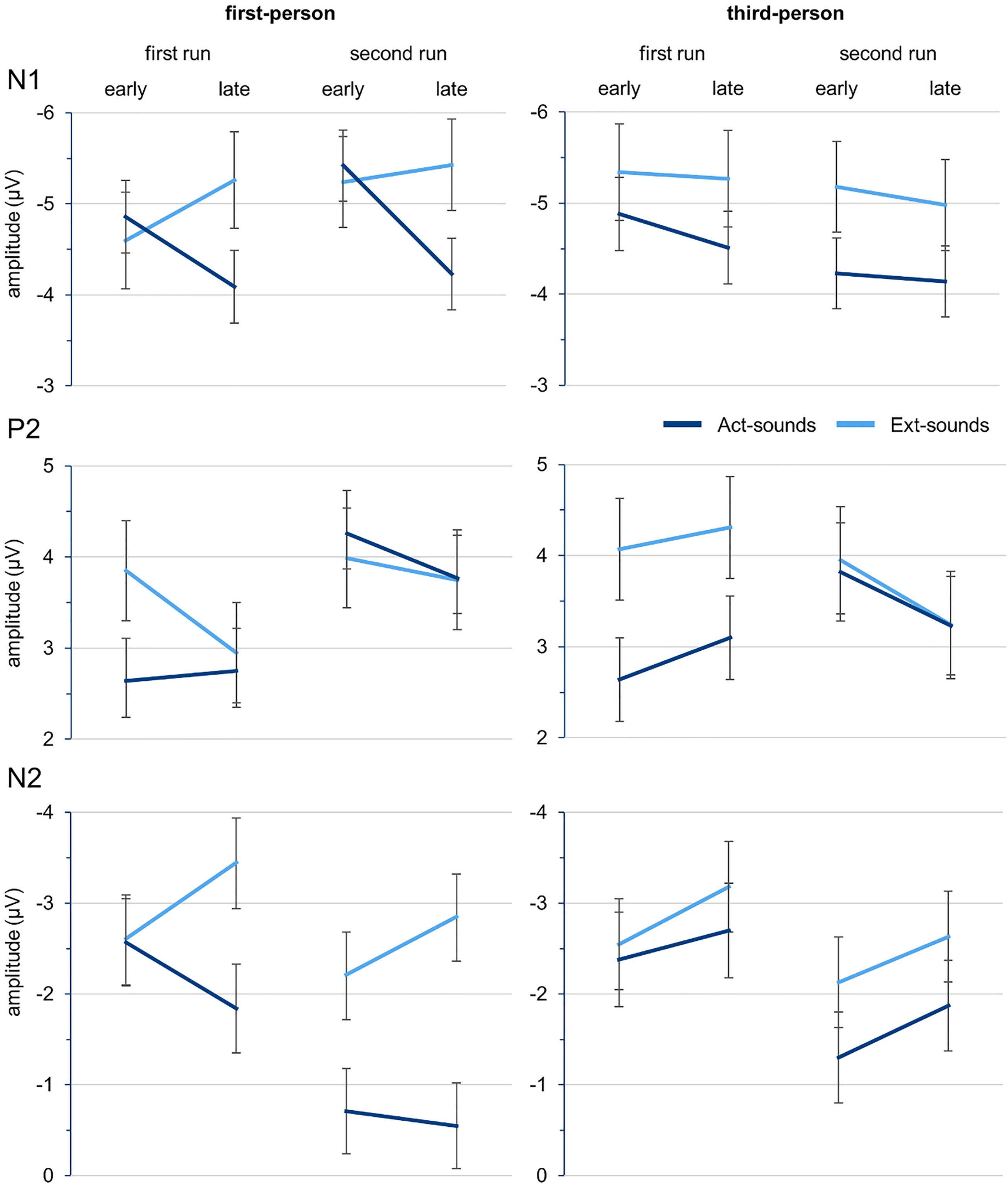

Fig. 4 Line plots of the marginal estimated means for the linear mixed effects models. Each run consists of 50 trials, the "early" and "late" trials displayed are trial 13 (25.5 - 12.5) and trial 38 (25.5 + 12.5) of the runs, corresponding to the simple effects analyses. Error bars represent one standard error

first versus second run of the experiment. The fine-grained temporal dynamics within each run of the experiment also were examined by modelling the temporal position of each individual trial in each condition. By using multilevel modeling on trial-level data (Volpert-Esmond et al., 2018), we found distinct dynamic patterns of amplitude reductions of the N1, P2, and N2 components for the two viewpoints over the course of the experimental session. While a significant reduction of the N1 component in response to sounds generated by actions observed from a third-person perspective was observed over the entire course of the experiment, a significant N1 reduction when observing from a first-person perspective only emerged later in each experimental run, but was not present in the beginning, i.e., it developed with increasing number of trials. For the P2, we observed a viewpoint-independent pattern over the course of the experiment, i.e., a general P2 reduction for sounds elicited by observed actions in the first but not in the second run, regardless of viewpoint. Our exploratory analyses for the N2 revealed distinct effects between and within runs, with only the latter showing a viewpoint-dependent pattern. A stronger reduction in N2 amplitude was found for the second compared to the first run regardless of viewpoint. While the reduction in response to sounds elicited by actions observed from the first-person perspective increased with increasing number of trials, no such temporal dynamic was found for the third-person perspective.

We hereby partially replicate our previous findings that auditory outcomes of observed actions are processed differently compared with externally generated sounds (Ghio et al., 2018, 2021). Furthermore, we show for the first time that this pattern differs dynamically depending on the viewpoint and time course during action observation, namely that the modulatory effect of viewpoint on auditory ERP attenuation seems to appear in a time-variant fashion.

N1 component

We expected an overall N1 reduction for sounds produced by observed actions compared with externally generated sounds. Our analyses indeed revealed an overall N1 reduction for sounds following observed actions, in line with our hypothesis and with a previous study from our lab with similar relative timing of observed action and outcome (Ghio et al., 2018). An N1 reduction for sounds caused by motor actions has been interpreted to reflect forward model predictions, which in case of one's own actions are likely available even before motion onset due to efference copy relayed motor information (Crapse & Sommer, 2008; Reznik et al., 2018). Assuming that, based on the neural substrates of mirror neurons (Bonini, 2017; Di Pellegrino et al., 1992; Mukamel et al., 2010; Rizzolatti & Sinigaglia, 2016), forward model predictions also could be employed during action observation (Wolpert et al., 2003). This also can be an interpretation for an N1 reduction for sounds caused by observed actions.

We also hypothesized that the N1 reduction might be larger during first- versus third-person observation since it has been shown that action observation from a first-person versus third-person perspective evokes stronger mu-rhythm suppression (Angelini et al., 2018), which is regarded to reflect mirror neuron system activity (Pineda, 2005). This hypothesis could not be confirmed, because we did not find a significant interaction between the predictors Sound Type and Viewpoint. With respect to the temporal dynamics, we expected that the N1 reduction for Act-sounds would become more pronounced over the course of the experiment and that this effect would be stronger for the first-person perspective. These hypotheses could only partially be confirmed. We did see an interaction between Sound Type and Trialnumber, hinting at a stronger N1 reduction for Act-sounds in later trials, and this effect was indeed further modulated by Viewpoint. However, during first-person observation, there was no significant N1 reduction at the start of each run, but it developed only toward later trials. This pattern was driven mostly by significantly decreasing Act-sound amplitudes over the course of the experiment but also was amplified by a (relatively smaller) amplitude enhancement for Ext-sounds. For the third-person perspective, in turn, we observed a stable reduction over the entire course of the runs. This pattern of results suggests that, opposite to our hypothesis, the third-person perspective during action-observation facilitates the mechanism underlying the N1 reduction, whereas for the first-person perspective this only develops with time (with increasing Trialnumber). Although an exploratory analysis revealed effects of Viewpoint on the motor-only ERPs before the N1 time window (see supplementary material S2), it seems unlikely that the pattern of results was caused (partially) by differential visual stimulation between viewpoints per se, as we specifically used these viewpoint-specific motor-only control conditions to correct the ERPs of Act-sounds for possible differential activity evoked by movement observation alone (see *Methods*). In the following, we propose three speculative interpretations for our N1 findings.

First, the temporal dynamic for the first-person observation might reflect an increasing precision in the prediction of the observed action-generated sound over the course of the experimental session enabled by ongoing observational motor learning. Observational motor learning can lead to action sequence-specific neural representations in frontoparietal cortex and enhanced performance in action execution similar to physical practice even without an explicit intention to learn (Apšvalka et al., 2018). Because the motor action to be observed in the present study is a simple action which is frequently executed in everyday life (e.g., when typing on a keyboard), a neural representation for the action itself was likely readily available. However, the sensorimotor association between the observed action and its sensory effect was novel and had to be acquired to form a motor-based prediction of the sound. Our finding of a temporal dynamic in the N1 reduction might therefore reflect the formation of this sensorimotor association with ongoing observational practice similarly as shown for physical practice (Burgess et al., 2019). A speculative interpretation for the finding of this temporal dynamic only for first-person observation is that learning from other's actions (e.g., from a dance instructor) typically involves a third-person perspective in real-life scenarios. This therefore might have enabled participants to predict sensory consequences of actions observed from the third-person perspective after very few trials.

A second interpretation can be that the developing N1 reduction during first-person observation reflects a higher demand on the transformation of the sensory information of the observed action into one's own motor and visceromotor representation of the action in question (Fu & Franz, 2014). This transformation—also termed as mirror mechanism (Rizzolatti & Sinigaglia, 2016)—might be impeded by a first-person perspective that usually occurs almost exclusively with the observation of own actions and is therefore unfamiliar during the observation of other's action. The increasing familiarity with the first-person perspective over the course of the experiment may then have facilitated the transformation of the visual input of the observed action into one's own motor and visceromotor representation of the action, tuning an internal forward model (Kilteni et al., 2019) to predict the action's sensory consequence and leading to a stronger N1 reduction.

Along similar lines, and as a third interpretation, the temporal dynamic of the N1 reduction during first-person observation could be explained by an initial failure of sensorimotor integration of the visual, proprioceptive and motor signals within an underlying internal forward model (Wolpert et al., 1995). The visual input for the first-person perspective corresponds to the viewpoint one has on the own hand during own actions. The sensorimotor integration of such a visual signal contradicts proprioceptive and (the lack of actual) motor information and might initially result in an ambiguity in the resulting sensory prediction, and thus, a lack of N1 reduction. Increasing familiarity with this combination of converging signals and an accompanying reliability-based reweighting of the signals in their integration (Boyle et al., 2017), might have enabled increasingly accurate prediction by a fine-tuned internal forward model (with the help of inverse models).

Considering that the involvement of any kind of motor-based prediction mechanism has been questioned by studies reporting N1 reductions simply as a result of temporal predictability (Dogge et al., 2019; Kaiser & Schütz-Bosbach, 2018; Sowman et al., 2012), our results appear to provide evidence to the contrary, at least for action observation. The actor-produced sounds we presented were identical in their temporal predictability and should thus have resulted in comparable N1 reductions if unspecific prediction mechanisms had been at work. Instead, the observation of human actions likely provokes specialized processing and results in predictive mechanisms beyond merely neutral visual stimuli (Klaffehn et al., 2019).

P2 component

For the P2, we also expected a general amplitude reduction for sounds produced by observed actions compared to external sounds since the sound-preceding actions, regardless of viewpoint, allow context-dependent predictions (Knolle et al., 2013a; Seidel et al., 2021). This hypothesis was only partially confirmed. While we did not find a main effect of Sound Type, we observed the hypothesized effect, regardless of viewpoint, for the first experimental run, but the effect disappeared in the second. Our hypothesis that the P2 reduction is generally diminished in the first-person perspective, but increases over the course of the experiment in contrast to the third-person perspective could clearly not be confirmed. While the described interaction pattern between Sound Type and Run suggests a temporal dynamic of the P2 reduction, the effect was not further modulated by Viewpoint, because no interactions involving Sound Type, Viewpoint and any of the two temporal predictors reached significance. Furthermore, the identified viewpoint-independent temporal dynamic appeared in the opposite direction compared to what was expected, i.e., the reduction disappeared instead of becoming stronger over the course of the experiment.

A P2 reduction was consistently reported in our previous studies on auditory consequences of observed actions (Ghio et al., 2018, 2021; van Laarhoven et al., 2021), and it has been associated with the perception of agency in studies examining self-produced sounds (Kühn et al., 2011;Seidel et al., 2021 ; Timm et al., 2016). According to Synofzik et al. (2008), agency can be conceptually split into a feeling of agency, possibly reflected in the N1, and a more conscious judgement of agency, which has been associated with the P2 (Seidel et al., 2021; Timm et al., 2016). At the same time, the P2 has been shown to be attenuated for visually cued sounds (Sowman et al., 2012; but see Harrison et al., 2021) and might therefore at least partly reflect the high temporal predictability that accompanies sounds caused by motor actions, observed and self-performed.

However, these explanations cannot account for the unexpected disappearance of the P2 reduction over time, which we observe in the current study, because both aspects, i.e., agency of the observed action and temporal predictability of its effect, do not change over time. A post-hoc explanation is that a possibly mediating factor that might explain the disappearance of the P2 reduction can be decreasing attention, which is drawn to the visual stimuli over time and might consequently diminish temporal predictability—for both viewpoints equally—and with it the reduction of P2 amplitudes (Sowman et al., 2012). Similarly, relatively heightened selective attention to the consequences of the observed actions (i.e., Act-sounds) in the first run, associated with an ERP termed *processing negativity* (PN) (Näätänen et al., 1978), might have only brought about the P2 reduction in the first place, which then disappeared as selective attention decreased over the course of the experiment. However, this negative shift should have overlapped the potentials in the N1 and N2 time range (Näätänen et al., 1978), which would have manifested as enhanced (i.e. more negative) N1 and N2 amplitudes for Act-sounds and thus less reduced amplitudes compared with Ext-sounds in the first versus second run of the experiment. While we did not observe such an interaction for the N1, we did see an interaction between Sound Type and Run for the N2 that could reflect enhanced Act-sound amplitudes, i.e., a weaker reduction for Act- versus Ext-sounds in the first compared with the second Run.

Another post-hoc explanation for the disappearance of the P2 reduction could be that the processes underlying the P2 reduction in the first experimental run shift in time, i.e., occur earlier or later relative to sound onset, in the second compared with the first run of the experiment. Again, the N2 thereby seems more suitable than the N1, due to the complementary time course of the Sound Type effect compared with the results concerning the P2. Thus, the process underlying the P2 reduction in the first run of the experiment also may have shifted to the N2 time window in the second run (see N2 component section below), rather than both being overlapped by a PN as proposed above. As temporal developments in the reduction of P2 amplitudes in studies examining the processing of self-generated sounds has to our knowledge not yet been analyzed, this might be an interesting avenue for future studies to further characterize similarities and differences during action observation and performance.

N2 component

The N2 is a negative deflection between 200 and 350 ms that has been associated with the detection of deviants in auditory oddball paradigms (Folstein & Van Petten, 2008). Our exploratory analysis of this component showed an overall reduction for Act- compared with Ext-sounds that was stronger in the second run and a difference in the temporal dynamic within the runs, also shown in Fig. 4. Across both runs, the N2 amplitude reduction for Act-sounds increased over time for first-person, but not for third-person observation.

Using a button-press paradigm similar to the action observation version used here, an enhancement of the N2 was observed for self-generated deviants (in terms of pitch) compared with externally generated deviants, which was interpreted to reflect an increased salience of the deviant when specific (i.e., forward model-based) predictions are at work (Knolle et al., 2013b). Similarly, infrequently delaying the onset of self-generated sounds by 250 ms (compared with a 0-, 50-, and 100-ms delay) elicited a significantly larger N2 compared with for intersound intervals controlled externally generated sounds (Pinheiro et al., 2019). In contrast, a reduction of the N2 for sounds in a sequence of tones was recently observed when participants performed continuous actions to modulate the sequence (in terms of pitch and speed) compared with passively listening to the same sequence afterwards (Sugimoto et al., 2021). Taken together, these studies suggest that the N2 reflects the classification of sound features (Ritter et al., 1979), such as the exact temporal occurrence. Enhanced N2 amplitudes thereby appear to reflect the cognitive detection of unpredicted stimulus properties (Näätänen et al., 1982; Ritter et al., 1992). This can be an interpretation for the current unexpected finding of a stronger reduction of the N2 amplitude for sounds generated by observed actions in the second compared with the first run compared with relatively enhanced N2 amplitudes for externally generated sounds that remain less predictable in their precise temporal occurrence. As outlined above, since we observed a complementary interaction between Sound Type and Run for the P2, this also might have been brought about by a negative shift (i.e., a PN), associated with selective attention, spanning the time range of P2 and N2.

On the other hand, the increasing reduction for the N2 from the first to the second experimental run was complemented by a viewpoint-dependent increase with increasing trials within each run that was not observed for the P2. It therefore seems unlikely that an overlapping PN can fully account for the observed effects of the N2. Specifically, the temporal dynamic within runs was only found for first-person observation, even though sounds in both conditions were equally cued and predictable, suggesting that the more pronounced N2 reduction during first-person observation reflects an additional influence. Interestingly, this increase, found only for the first-person perspective, strongly resembles the pattern we hypothesized a priori for the P2, based on a dissolvement of agency ambiguity for the first-person perspective. Thus, the pattern found for the N2 might also reflect a process, hypothesized for the P2, related to a form of (self-) agency, which might exist due to constant self-observation from this viewpoint, possibly also amplified in the second run by a temporal shift from the P2 to the N2 time range (see above). Nevertheless, a clear interpretation of our unexpected result seems premature, especially since findings on relationships between amplitude reductions and (self-) agency have so far only been reported for self-generated and not for observed action effects and therefore need to be clarified in future studies. Notwithstanding, the current observations can hopefully contribute to a better understanding of N2 variations and their temporal development that might be identified in subsequent studies.

Conclusions

Using multilevel modeling on trial-level data, this study shows that the processing of auditory action outcomes during action observation is modulated by the viewpoint in a time-variant fashion. As for self-generated sounds, a reduction of N1 amplitudes for sounds caused by observed actions compared with externally generated sounds was found, emerging for both viewpoints. This indicates that similar prediction mechanisms contribute to early auditory processing of sounds following self-performed and observed actions. However, a temporal dynamic of the N1 reduction for the first-person, but not third-person, perspective (i.e., it only emerged with increasing number of trials and was amplified by a relatively weaker N1 increase for externally generated sounds) suggests that viewpoint-dependent mechanisms might be involved in sensorimotor predictions during action observation. A P2 reduction, as commonly found for self-generated sounds, was found regardless of viewpoint for the first but not the second experimental run. Contrary to the P2 finding, an exploratory analysis of the ensuing N2 component revealed a reduction for sounds caused by observed actions compared with externally generated sounds that increased from the first to the second experimental run, as well as a viewpoint-dependent pattern for this increase within runs. Considered together, we speculated that this might reflect a temporal shift of an agency-related process, which affects the first-person perspective more strongly than the third-person perspective, from the P2 to the N2 range over time. Applying trial-level analyses in future studies on the processing of self-generated sounds can help to elucidate whether the temporal dynamics identified here are specific for action observation or are also seen with self-performed actions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13415-023-01083-7.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors for the submitted work.

Declarations

Ethics approval This study was approved by the Ethics Committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf, Germany.

Consent to participate Informed written consent was obtained from each participant prior to the experiment.

Consent for publication Not applicable (see Open practices statement).

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Angelini, M., Fabbri-Destro, M., Lopomo, N. F., Gobbo, M., Rizzolatti, G., & Avanzini, P. (2018). Perspective-dependent reactivity of sensorimotor mu rhythm in alpha and beta ranges during action observation: An EEG study. *Scientific Reports, 8*(1), 1–11. https://doi.org/10.1038/s41598-018-30912-w

Apšvalka, D., Cross, E. S., & Ramsey, R. (2018). Observing action sequences elicits sequence-specific neural representations in frontoparietal brain regions. *Journal of Neuroscience, 38*(47), 10114–10128. https://doi.org/10.1523/JNEUROSCI.1597-18.2018

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

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport, 12*(9), 1879–1884. https://doi.org/10.1097/00001756-200107030-00023

Bolt, N. K., & Loehr, J. D. (2021). Sensory attenuation of the auditory P2 differentiates self- from partner-produced sounds during joint action. *Journal of Cognitive Neuroscience, 33*(11), 2297–2310. https://doi.org/10.1162/jocn_a_01760

Bonini, L. (2017). The extended mirror neuron network: Anatomy, origin, and functions. *The Neuroscientist, 23*(1), 56–67. https://doi.org/10.1177/1073858415626400

Boyle, S. C., Kayser, S. J., & Kayser, C. (2017). Neural correlates of multisensory reliability and perceptual weights emerge at early latencies during audio-visual integration. *The European Journal of Neuroscience, 46*(10), 2565–2577. https://doi.org/10.1111/ejn.13724

Burgess, J. D., Major, B. P., McNeel, C., Clark, G. M., Lum, J. A., & Enticott, P. G. (2019). Learning to expect: Predicting sounds during movement is related to sensorimotor association during listening. *Frontiers in Human Neuroscience, 13*, 215. https://doi.org/10.3389/fnhum.2019.00215

Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., & Casile, A. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. *Current Biology, 21*(2), 144–148. https://doi.org/10.1016/j.cub.2010.12.022

Caggiano, V., Giese, M., Thier, P., & Casile, A. (2015). Encoding of point of view during action observation in the local field potentials of macaque area F5. *European Journal of Neuroscience, 41*(4), 466–476. https://doi.org/10.1111/ejn.12793

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience, 9*(8), 587–600. https://doi.org/10.1038/nrn2457

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*(1), 176–180. https://doi.org/10.1007/BF00230027

Dogge, M., Custers, R., & Aarts, H. (2019). Moving forward: On the limits of motor-based forward models. *Trends in cognitive sciences.* https://doi.org/10.1016/j.tics.2019.06.008

Drew, A. R., Quandt, L. C., & Marshall, P. J. (2015). Visual influences on sensorimotor EEG responses during observation of hand actions. *Brain Research, 1597*, 119–128. https://doi.org/10.1016/j.brainres.2014.11.048

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

Fu, Y., & Franz, E. A. (2014). Viewer perspective in the mirroring of actions. *Experimental Brain Research, 232*(11), 3665–3674. https://doi.org/10.1007/s00221-014-4042-6

Ghio, M., Egan, S., & Bellebaum, C. (2021). Similarities and differences between performers and observers in processing auditory action consequences: Evidence from simultaneous EEG acquisition. *Journal of Cognitive Neuroscience*, 1–12. https://doi.org/10.1162/jocn_a_01671

Ghio, M., Scharmach, K., & Bellebaum, C. (2018). ERP correlates of processing the auditory consequences of own versus observed actions. *Psychophysiology, 55*(6), e13048. https://doi.org/10.1111/psyp.13048

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

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences, 5*(6), 253–261. https://doi.org/10.1016/s1364-6613(00)01661-2

Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Research, 1626*, 54–65. https://doi.org/10.1016/j.brainres.2015.03.038

Jack, B. N., Chilver, M. R., Vickery, R. M., Birznieks, I., Krstanoska-Blazeska, K., Whitford, T. J., & Griffiths, O. (2021). Movement planning determines sensory suppression: An event-related potential study. *Journal of Cognitive Neuroscience, 33*(12), 2427–2439. https://doi.org/10.1162/jocn_a_01747

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

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Kilteni, K., Houborg, C., & Ehrsson, H. H. (2019). Rapid learning and unlearning of predicted sensory delays in self-generated touch. *Elife, 8*. https://doi.org/10.7554/eLife.42888

Klaffehn, A. L., Baess, P., Kunde, W., & Pfister, R. (2019). Sensory attenuation prevails when controlling for temporal predictability of self-and externally generated tones. *Neuropsychologia, 132*, 107145. https://doi.org/10.1016/j.neuropsychologia.2019.107145

Knolle, F., Schroger, E., Baess, P., & Kotz, S. A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of Cognitive Neuroscience, 24*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schroger, E., & Kotz, S. A. (2013a). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex, 49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

Knolle, F., Schroger, E., & Kotz, S. A. (2013b). Prediction errors in self- and externally-generated deviants. *Biological Psychology, 92*(2), 410–416. https://doi.org/10.1016/j.biopsycho.2012.11.017

Korka, B., Widmann, A., Waszak, F., Darriba, Á., & Schröger, E. (2021). The auditory brain in action: Intention determines predictive processing in the auditory system—A review of current paradigms and findings. *Psychonomic Bulletin & Review, 1-22*. https://doi.org/10.3758/s13423-021-01992-z

Krueger, C., & Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing, 6*(2), 151–157. https://doi.org/10.1177/1099800404267682

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

Lange, K. (2011). The reduced N1 to self-generated tones: An effect of temporal predictability? *Psychophysiology, 48*(8), 1088–1095. https://doi.org/10.1111/j.1469-8986.2010.01174.x

Maranesi, M., Livi, A., & Bonini, L. (2017). Spatial and viewpoint selectivity for others' observed actions in monkey ventral premotor mirror neurons. *Scientific Reports, 7*(1), 1–7. https://doi.org/10.1038/s41598-017-08956-1

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks, 9*(8), 1265–1279. https://doi.org/10.1016/s0893-6080(96)00035-4

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20*(8), 750–756. https://doi.org/10.1016/j.cub.2010.02.045

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

Näätänen, R., Simpson, M., & Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology, 14*(1-2), 53–98. https://doi.org/10.1016/0301-0511(82)90017-5

Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology, 83*(12), 1112–1118. https://doi.org/10.1212/WNL.0000000000000798

Pineda, J. A. (2005). The functional significance of mu rhythms: Translating "seeing" and "hearing" into "doing". *Brain Research Reviews, 50*(1), 57–68. https://doi.org/10.1016/j.brainresrev.2005.04.005

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

Poonian, S. K., McFadyen, J., Ogden, J., & Cunnington, R. (2015). Implicit agency in observed actions: Evidence for N1 suppression of tones caused by self-made and observed actions. *Journal of Cognitive Neuroscience, 27*(4), 752–764. https://doi.org/10.1162/jocn_a_00745

Reznik, D., & Mukamel, R. (2019). Motor output, neural states and auditory perception. *Neuroscience & Biobehavioral Reviews, 96*, 116–126. https://doi.org/10.1016/j.neubiorev.2018.10.021

Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia, 119*, 302–307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028

Ritter, W., Paavilainen, P., Lavikainen, J., Reinikainen, K., Alho, K., Sams, M., & Näätänen, R. (1992). Event-related potentials to repetition and change of auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 83*(5), 306–321. https://doi.org/10.1016/0013-4694(92)90090-5

Ritter, W., Simson, R., Vaughan Jr., H. G., & Friedman, D. (1979). A brain event related to the making of a sensory discrimination. *Science, 203*(4387), 1358–1361. https://doi.org/10.1126/science.424760

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience, 17*(12), 757–765. https://doi.org/10.1038/nrn.2016.135

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and Cognition, 17*(4), 1219–1227. https://doi.org/10.1016/j.concog.2008.01.003

Schafer, E. W., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science, 181*(4095), 175–177. https://doi.org/10.1126/science.181.4095.175

Schneider, D. M., Sundararajan, J., & Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. *Nature, 561*(7723), 391–395. https://doi.org/10.1038/s41586-018-0520-5

Sebastiani, V., De Pasquale, F., Costantini, M., Mantini, D., Pizzella, V., Romani, G. L., & Della Penna, S. (2014). Being an agent or an observer: Different spectral dynamics revealed by MEG. *Neuroimage, 102*, 717–728. https://doi.org/10.1016/j.neuroimage.2014.08.031

Seidel, A., Ghio, M., Studer, B., & Bellebaum, C. (2021). Illusion of control affects ERP amplitude reductions for auditory outcomes of self-generated actions. *Psychophysiology, 58*(5), e13792. https://doi.org/10.1111/psyp.13792

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental Brain Research, 222*(1-2), 149–157. https://doi.org/10.1007/s00221-012-3204-7

Sugimoto, F., Kimura, M., & Takeda, Y. (2021). Attenuation of auditory N2 for self-modulated tones during continuous actions. *Biological Psychology, 166*, 108201. https://doi.org/10.1016/j.biopsycho.2021.108201

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition, 17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Timm, J., SanMiguel, I., Keil, J., Schroger, E., & Schonwiesner, M. (2014). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of Cognitive Neuroscience, 26*(7), 1481–1489. https://doi.org/10.1162/jocn_a_00552

Timm, J., Schonwiesner, M., Schroger, 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

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2021). Suppression of the auditory N1 by visual anticipatory motion is modulated by temporal and identity predictability. *Psychophysiology, 58*(3), e13749. https://doi.org/10.1111/psyp.13749

Volpert-Esmond, H. I., Merkle, E. C., Levsen, M. P., Ito, T. A., & Bartholow, B. D. (2018). Using trial-level data and multilevel modeling to investigate within-task change in event-related potentials. *Psychophysiology, 55*(5), e13044. https://doi.org/10.1111/psyp.13044

Volpert-Esmond, H. I., Page-Gould, E., & Bartholow, B. D. (2021). Using multilevel models for the analysis of event-related potentials. *International Journal of Psychophysiology, 162*, 145–156. https://doi.org/10.1016/j.ijpsycho.2021.02.006

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology, 11*(18), R729–R732. https://doi.org/10.1016/s0960-9822(01)00432-8

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Open practices statement Analysis code including model output is available at https://doi.org/10.17605/OSF.IO/FGRB3. The data cannot be made publicly available, because the consent forms signed by the participants do not cover public availability and permanent archiving of the data, even if fully anonymized. Materials (i.e. full sets of visual stimuli, the auditory stimulus and Presentation® code) are available on request from the corresponding author (constanze.weber@hhu.de).

Conflicting motor plans and sensory attenuation: Evidence from event-related potentials for sounds generated by pro- and antisaccades

Alexander Seidel, Christian Bellebaum

Institute of Experimental Psychology, Department of Biological Psychology, Heinrich Heine University, Düsseldorf, Germany

Running head: Auditory ERP attenuation after pro- and antisaccades

Corresponding author: Alexander Seidel, Heinrich Heine University, Institute of Experimental Psychology, Department of Biological Psychology, Universitätstrasse, 1, 40255, Düsseldorf, Germany.

Email: Alexander.Seidel@hhu.de

Abstract

The reduction of neural responses to self- versus externally-generated stimuli has been ascribed to predictions based on an efference copy of motor commands. However, general predictive mechanisms not specific for movements may also play a role. For antisaccades, that is, eye movement in the opposite direction of a target stimulus, a reflexive prosaccade has to be suppressed, which may lead to conflicting efference copy signals, as an efference copy is likely created also for the prosaccade. If efference copies for the suppressed and executed saccade are in conflict with each other, prediction mechanisms based on their information are potentially disturbed, which may affect the processing of saccade-generated stimuli. We measured the reductions of the N1 and P2 components for pro- and antisaccade-generated sounds compared to visually cued external sounds, and found differing temporal dynamics of both components during the course of the experiment, depending on the saccade type. An N1 reduction was found for pro- but not antisaccade-generated sounds that slowly developed over the course of the experiment. The P2 for prosaccade-generated sounds decreased over time as well, until it was not different from visually cued sounds, while the P2 for antisaccade-generated sounds remained elevated over time. These findings suggest that both early (N1) and late (P2) processing of saccade-generates sounds is affected by conflicting efference copies, with the early effect probably reflecting forward model predictions and the later effect indicating agency perception based on these predictions.

Keywords: auditory ERP, efference copies, saccades, antisaccades

Conflicting motor plans in sensory attenuation: Evidence from auditory N1 reductions for sounds generated by pro- and antisaccades

For self-produced stimuli, studies have consistently shown reduced perceptual intensity and neurophysiological responses compared to externally produced stimuli (Baess et al., 2011; Blakemore et al., 1998; Sato, 2008; Schafer & Marcus, 1973), referred to as sensory attenuation. In the auditory domain, electroencephalography (EEG) studies typically report reductions of the amplitudes of the event related potential (ERP) components N1 and P2 (Baess et al., 2011; Knolle et al., 2013; Sowman et al., 2012), reflecting a neurophysiological sensory attenuation effect. This has been suggested to reflect predictive mechanisms with respect to the sensory consequences of actions, incorporating internal signals and context information to match re-afferent sensory stimuli with their predictions (for a review, see Horváth, 2015). But some studies suggest a functional dissociations between these two components: While the P2 has been shown to be sensitive to contextual factors, like the perceived control (Seidel et al., 2021) and agency (Kühn et al., 2011; Timm et al., 2016) over sound production and is attenuated also for visually cued externally-generated sounds (Sowman et al., 2012), the N1 was not affected in these studies. Thus, the attenuation for the two components might rely on predictions based on different types of information.

While the P2 attenuation has been ascribed to general, motor-independent mechanisms (e.g. Baess et al., 2011; Ghio et al., 2018; Knolle et al., 2013), the prevalent account for explaining the N1 attenuation suggests that cerebellar feed-forward models employ efference copies of motor commands to start the generation of predictions concerning their sensory consequences right after the motor planning stage, enabling a matching process with the actual sensory consequences as early as in the N1 time window, around 100 ms after stimulus onset (Blakemore et al., 2001; Horváth, 2015; Pickering & Clark, 2014; Popa & Ebner, 2018; Reznik et al., 2018; Vercillo et al., 2018; Wolpert & Flanagan, 2001). This view is supported by the findings obtained in Timm et al. (2014) showing an N1 attenuation only for voluntary actions and not when actions were externally induced via transcranial magnetic stimulation. Examining the supposed cerebellar contribution to the prediction of sensory action consequences, Knolle et al., 2012, 2013 reported a reduced N1 attenuation for cerebellar lesion patients compared to healthy controls, supporting the claim that efference-copy-driven cerebellar forward models underly the N1 attenuation. At the same time, the P2 was not affected by cerebellar lesions, supporting its independence from cerebellar forward models.

It has been argued, however, that motoric signals are not necessary to generate the predictions underlying the N1 attenuation, because action execution offers sufficient cues for temporal prediction of action-generated sound onsets, while onsets of external sounds are not predictable (Hughes et al., 2013). In fact, comparing cued self- and externally-generated sounds, and thus matching temporal predictability, Kaiser and Schütz-Bosbach (2018) and Harrison et al. (2021) found no (additional) N1 attenuation for self-produced sounds. Klaffehn et al. (2019) on the other hand, reported an N1 attenuation for self-produced sounds even when controlling for temporal predictability, suggesting that motor-specific N1 attenuation effects may exist beyond unspecific prediction effects.

Dogge et al. (2019) also argued that forward models are unlikely to underlie N1 attenuation for environment-related predictions such as sounds following button presses (as opposed to body-related prediction such as when touching your left hand with your right hand), which are used in most studies, as work in animals has revealed that the tuning of motor-based forward models is quite slow and studies typically entail only short trainings. Button presses such as on computer keyboards or phones are, however, likely commonplace in the every-day life of most humans, and forward models are probably well trained for their auditory feedback. In an experimental setting, learning new, specific button-press-sound associations may thus require only minimal training. The study by Mifsud et al. (2016), suggested that experience with specific associations between actions and their sensory effects indeed plays a role. They reported smaller N1 attenuation when sounds were produced by saccadic eye movements (compared to button presses), which are typically not associated with auditory consequences in every-day life. At the same time this study shows that even for unusual action-sensory effect combinations, the N1 attenuation can be found.

Moreover, single-trial-based linear mixed effects analyses allow to explore the temporal dynamic of ERP amplitude changes over the course of an experiment by adding the trialnumber as a predictor (Volpert-Esmond et al., 2021), thereby modelling effects of practice or experience. Applying this technique thus allows to model training or experience effects. In a study on sensory attenuation in action observation we found a change in N1 attenuation over time for an uncommon first-person observer viewpoint, but not for a common third-person viewpoint (Seidel et al., 2023). It is thus conceivable that similar temporal dynamics emerge for uncommon action-sensory effect associations such as sound generating saccadic eye movements.

In the present study, we explore the temporal dynamics of N1 and P2 amplitude attenuation for sounds following saccadic eye movements and thus for an untrained action-effect association. In addition, we explore the role of conflicting efference copy signals on N1 and P2 attenuation and their temporal dynamics by comparing the processing of sounds generated by anti- and prosaccades. The two conditions are comparable in their motor requirements, but differ in the motor planning and possibly the relayed efference copies. For (pro)saccade-generated sounds, we expect a significant N1 attenuation (see Mifsud et al. [2016]), which becomes stronger over time with increasing experience, as the forward model needs to be tuned for this unusual action-sensory effect association. For correct antisaccade execution, it is assumed that a reflexive saccade to the appearing target has to be suppressed before the antisaccade is performed (Coe & Munoz, 2017; Munoz & Everling, 2004). We speculate that an efference copy of the motor plan for this suppressed reflexive saccade could still be generated and conflict with the efference copy for the executed antisaccade. This should result in a disruption of the proposed cerebellar feedforward model relying on this signal, which is expected to affect sensory attenuation as reflected in the N1 amplitude. Consequently, the attenuation of the N1 amplitude is expected to be stronger for pro- than for antisaccades.

As the P2 is independent from cerebellar forward models (Knolle et al., 2012, 2013), we assumed neither an influence of efference-copy-based predictive mechanisms nor any temporal dynamics. Considering that Mifsud et al. (2016) did not find a P2 attenuation for saccade-generated sounds, we expected neither a general amplitude reduction for saccade-generated compared to externally-generated sounds, nor differences in P2 amplitudes between pro- and antisaccade-generated sounds or amplitude changes over time.

Method

Participants

38 participants (25 women, mean age 24.8 years, SD = 4.5 years) took part in this study for either course credit or monetary compensation. The sample size was larger than in some previous studies on button press elicited sounds (Baess et al., 2011; Ghio et al., 2021; Ghio et al., 2018; Klaffehn et al., 2019), but comparable to the study by Mifsud et al. (2016), who first reported an N1 attenuation for saccade-elicited sounds and tested 36 participants. As we expected similar effect sizes in our study we aimed for a comparable sample size. All participants reported normal hearing and normal or corrected-to-normal vision, as well as no history of neurological or mental illness, or use of medication affecting the nervous system. All but three participants were right-handed. Written informed consent was given by all participants before participation. The study was approved by the Ethics Committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf, Germany.

Experimental paradigm

Stimuli.

In all conditions, including the trainings, three black rings with a diameter of 2° visual angle (40 px/° visual angle) and a line thickness of approximately 0.1° visual angle were continuously displayed on the screen on a gray background (R=G=B=191) to indicate the potential locations in which the target stimuli (see below) could appear during the experiment. This was done to enable easier fixation of target positions after target disappearance and prevent unwanted eye movements during the following time window for ERP analysis. One ring was positioned at the center, the others each 10° visual angle to the left and right. All described target and fixation dots appeared in the left, central or right ring. Black dot stimuli with a diameter of 1° visual angle were used as visual target stimuli in all conditions of the experiment and in the accuracy check during calibrations. A 1000 Hz sinus tone with a duration of 200 ms (fade in/out of 20 ms) was used as the self- and externally-generated sound in the different experimental conditions.

Experimental conditions

In this study we adapted the classic contingent-paradigm (Horváth, 2015), in a similar way as Mifsud et al. (2016), by letting participants produce sounds by saccadic eye movements. In contrast to the study by Mifsud et al. (2016), our main interest was in the comparison of sounds elicited by visually guided (pro)saccades towards a visual target and sounds elicited by antisaccades, which are directed away from a target stimulus. The paradigm thus entailed two conditions in which sounds were produced by actions (act-sound conditions), one with pro- and one with antisaccades. As both types of saccades are performed as a reaction to an appearing visual target stimulus, the conditions with external sounds, which were not associated with any movement, entailed identical cue stimuli to control for effects of predictability based on visual cues (cue-sound conditions). Importantly, the relative timing of visual cue to saccade, and thus to sound onset, in the act-sound conditions depended on saccade latency. As antisaccades typically have longer latencies than prosaccades (Munoz & Everling, 2004), the paradigm contained two separate cue-sound conditions, one in which the relative timing of cue and sound was matched to the act-sound condition with prosaccades and one in which the timing was matched to the act-sound condition with antisaccades.

In addition to these experimental conditions, act-only conditions were also used in the present study, in which the same movements were performed as in the act-sound conditions without producing sounds, to correct for motion-induced ERPs in the act-sound condition, one with pro- and one with antisaccades (see Horvath et al., 2015 for a description of the rationale of the act-only condition in the contingent paradigm). Accordingly, we also included conditions to control for the visual stimulation in the cue-sound conditions. In these so-called cue-only conditions a visual cue was shown without presenting sounds (see Figure 1 for an overview of all conditions).

Act-sound condition for prosaccades. Each trial started with a fixation dot in the central ring for 1400 ms (± 150 ms random variance, counterbalanced). 50 ms after the fixation dot had disappeared from the screen, a target dot appeared in the left or right ring (counterbalanced per condition) for 50 ms. Participants were instructed to fixate the location of each appearing dot and not move their gaze until the next dot appeared, and were informed that performed saccades to the left or right dot position would cause a sound. During the experiment eye position was continuously monitored and processed by the program for stimulus presentation. 170 ms after the horizontal eye position reached the position halfway between fixation point and saccade target position (see below for details), but only if the saccade was aimed in the correct direction, a 200 ms sound was played. Reaching the position between fixation point and target was considered as indicating a saccadic eye movement. The delay ensured that the saccade could be completed before the sound was played. Following this there was another delay of 800 ms until the start of the next trial. If no sound was played, the timing from the detection of the crossing of the halfway-point to the start of the next trial was identical to trials with sound presentation (170 + 200 + 800 ms).

Act-only condition for prosaccades. The act-only condition was identical to the act-sound condition, but neither correct nor incorrect saccades caused a sound. Before the condition started participants were instructed accordingly.

Act-sound condition for antisaccades. The visual stimulation in the antisaccade version of the act-sound condition was identical to that of the prosaccade version, but participants were instructed to aim their gaze at the ring opposite to the one in which the target dot appeared. Only if correct antisaccades in accordance with the instruction were detected, sounds were presented time-locked to the antisaccade. As in the prosaccade condition, sounds were elicited 170 ms after the eye position reached the position halfway between fixation point and target position. Note that the target position was on the opposite side relative to the visual target in the act-sound condition for antisaccades.

Act-only condition for antisaccades. The act-only condition for antisaccades was identical to the act-sound condition for antisaccades, but no sounds were presented. Before the condition started participants were instructed accordingly.

Cue-sound condition for prosaccades. To ensure comparability between act- and cue-sound conditions, the visual stimulation in the cue-sound conditions was kept as similar as possible to the two act-sound conditions described above, while not requiring saccadic eye movements. Stimulus timing in each trial of the cue-sound condition, that is, the duration of fixation dot presentation, visual cue onset and sound onset time relative to trial start, was determined by the timing of these events in the corresponding trial in the previous act-sound condition for prosaccades. The cue-sound condition for prosaccades can thus be considered as a replay of the preceding act-sound condition for prosaccades. Trials started with the fixation dot. In contrast to the act-sound conditions, the fixation dot was presented in the left or right ring where also the visual cue would appear later during the trial, depending on the intended saccade direction of the corresponding trial in the previous act-sound condition. This was done to provide the same visual stimulation as in the act-sound condition, without inducing reflexive saccades, and to ensure that eye position during sound presentation was the same in the act- and cue-sound conditions. 50 ms after the fixation dot disappeared, an identical target dot appeared in the same circle as the fixation dot for 50 ms. After a delay depending on the time between target and sound, and thus on saccade latency, in the corresponding trial in the previous act-sound condition, a 200 ms sound was presented, followed by the next trial 800 ms later. Participants were instructed to fixate dot stimuli as they appeared and focus their gaze on the position until the presentation of the next dot. They were also informed that no saccades were required in this condition. As the cue-sound condition was a replay of the preceding act-sound condition, no sounds were played if the participant had made a saccade direction error in the corresponding trial of the act-sound condition. Instead, the time interval until the next trial started was 1000 ms, to account for the sound duration and keep trial timing consistent, as in the act-sound condition.

Cue-sound condition for antisaccades. The cue-sound conditions for antisaccades only differed from the one for prosaccades with respect to the data on which trial timings, target positions, and sound presentation was based. As was pointed out above, we expected both higher error rates and longer latencies for antisaccades, which would be reflected in more sound omissions and longer cue-sound intervals for the cue-sound condition for antisaccades compared to prosaccades. The cue-sound condition for antisaccades was therefore based on the preceding antisaccade act-sound condition in the same block. Fixation and target dot position in a given trial in the antisaccade version of the cue-sound condition were determined by the required saccade direction in the corresponding trial of the preceding act-sound condition for antisaccades. If, for example, participants in trial x of the act-sound condition for antisaccades were required to look to the left (because the target dot appeared on the right), then fixation and target dot in trial x of the cue-sound condition for antisaccades appeared on the left.

Cue-only condition for prosaccades. The visual stimulation in the cue-only condition for prosaccades was identical to the visual stimulation in the preceding prosaccade cue-sound condition, including stimulus timings and positions (as recorded in the preceding act-sound condition). Importantly, no sounds were played in this condition and participants were informed that no sounds would occur.

Cue-only condition for antisaccades. As with the cue-sound condition, the cue-only conditions for pro- and antisaccades only differed with respect to the data on which trial timings, stimulus position, and sound presentation was based. The visual stimulation in the cue-only condition for antisaccades was thus identical to the visual stimulation in the preceding antisaccade cue-sound condition, including stimulus timings and positions (as recorded in the preceding act-sound condition for antisaccades). No sounds were played in this condition and participants were informed that no sounds would occur.

Contingency training. Before the experimental conditions started, participants underwent a two-part training to establish an act-sound contingency between saccades and sounds. In the first part, the three circles were presented together with a fixation dot in the center circle for 1500 ms. After an additional delay of 1500 ms a target dot was presented randomly in the left or right circle for 50 ms and participants were asked to perform a saccade to the location of the target dot. As soon as a saccade was detected (see information on saccade detection below), a sound was played with a delay of 170 ms. 2000 ms later, the next target dot was shown in the opposite circle, and participants performed another saccade which caused again a sound. This was repeated until 10 sounds had been produced in this manner.

The second part of the training started in the same way as the first part, but after presentation of the fixation dot ended, participants were instructed to alternate their gaze between the two lateral circles in a rhythm of their choice. Each saccade from the left circle to the right or vice versa prompted a sound with a 170 ms delay and participants listened to the sounds their saccades produced. No further target dots were shown. The second part of the training ended after 50 sounds had been produced.

Procedure

After signing the consent form and completing the demographic questionnaire, the participants started with the two training tasks to establish a saccade-sound contingency, followed by short versions of the act-sound condition (20 trials), once requiring pro- and once antisaccades. If participants used visual aids they removed them before the tasks to improve eye tracking. Participation was only possible, however, when they confirmed that the stimuli during calibrations and tasks were perceptible. The main experiment consisted of four experimental blocks each containing 40-trial-versions of the four conditions act-sounds, cue-sounds, act-only, cue-only in this fixed order. Each block was either a pro- or antisaccade block, that is, it either contained the pro- or antisaccade versions of these conditions. The four blocks alternated between pro- and antisaccade blocks, and the type of the first block was counterbalanced between participants. Because of this, the first, and the second half of the experiment contained one block of each pro- or antisaccade condition, and the experimental halves were entered as separate runs into the analysis (see below). To improve performance and minimize the loss of trials, each block started with a short 10 trial training version of the upcoming act-sound condition with the respective saccade type. Every block started with a pictured instruction of the upcoming condition. Via button press the participant could start the block, providing the option for a self-administered break. The act-sound and act-only conditions, in which saccades were assessed, additionally started with a nine-point calibration of the eye tracker, followed by a short accuracy check.

The experiment was performed using Presentation software (Version 20.3, Neurobehavioral Systems, Inc., Berkeley, USA) on a Windows 10 desktop computer and a 22” monitor with a resolution of 1680 × 1050 px. Sounds were presented via a Sound Blaster

Audigy Rx (Creative Technology Ltd., Singapore) using bit accurate playback connected to Sennheiser HD 201 headphones in Presentation’s exclusive sound mode at a fixed volume. EEG sound markers and the sound signal were sent with a timing difference consistently measured below 1 ms with a Tektronix TDS 210 oscilloscope (Tektronix, Inc., Beaverton, USA).

Online eye tracking data capture and analysis

A SensoMotoric Instruments Red 500 eye tracking system (using dark pupil tracking at 500 Hz) was mounted underneath the monitor used for stimulus presentation and connected to a Windows 7 laptop running iView X (Version 2.8.43, SensoMotoric Instruments, Teltow, Germany). A chin rest was used to position participants at a distance of approximately 62 cm from the screen, which resulted in 40 px per degree visual angle.

Saccade detection in all conditions was implemented as a continuous online check of the measured gaze position. A saccade to a left or right target was considered as detected if the x-axis coordinate of the gaze position was further than 5° visual angle (half the distance to a left or right target) from the center of the screen for five consecutive measurements, corresponding to 10 ms. This means that saccades in the wrong direction were detected as well (but counted as an error, see below), except for the contingency training, in which the saccade detection process waited until a saccade in the correct direction was made. In the contingency training, a saccade was considered as detected if five consecutive gaze positions were found in the half of the screen opposite to the last focused circle.

EEG data recording and processing

Twenty-eight Ag/AgCl passive ring electrodes were used to record EEG data continuously at 1000 Hz with BrainVision Recorder software (1.21.0402) and a BrainAmp amplifier (Brain Products, GmbH, Germany). An elastic cap (EasyCap, Brain Products) was used for positioning according to the international 10-20 System at F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P7, P3, Pz, P4, P8, PO7, PO3, POz, PO4 and PO8. The signal at linked mastoids was used for signal referencing, with a ground electrode at AFz. Horizontal eye movements were recorded at F9 and F10, vertical eye movements at Fp2 and the corresponding position below the right eye. Impedances were kept below 5 kΩ.

Data preprocessing was conducted with Brain Vision Analyzer 2.2.0.7383 (Brain Products). A global direct current de-trend, a Butterworth zero phase filter (low cutoff: 0.3 Hz, order 4; high cutoff: 30 Hz, order 34) and a notch filter (50 Hz) were applied. By means of an independent component analysis (ICA, steps = 512, infomax restricted biased) components corresponding to blinks in the Electrooculogram channels were excluded before applying an inverted ICA.

Markers for sounds were used to segment data into 800 ms epochs from -200 ms to 600 ms after sound onset. The conditions in which no sound was played also contained a sound marker, as the sound was muted in these trials. The segmentation could thus be performed in the same way as for the conditions with sounds. After a baseline correction (-100 to 0 ms), the automatic artifact rejection of Brain Vision Analyzer 2.2 was employed to reject noisy segments, with the following parameters: maximal allowed voltage step = 50 µV/ms, maximal allowed difference of values within 100-ms intervals = 100 µV, maximal/minimal allowed amplitude = ±100 µV, lowest activity of 0.5 µV within 100 ms intervals. Act-sound sound (and the corresponding cue-sound and cue-only) segments for trials with erroneous saccades, in which no sounds were played (see above), were not considered for the analysis, as well as trials with erroneous saccades in the act-only condition. To prevent a large variance in inter-sound-intervals, trials that directly followed trials without sound were removed for the act- and cue-sound condition as well.

Continuing in MATLAB (R2018a, The MathWorks, Inc., Natick, MA), act-sound and cue-sound segments were subsequently corrected for their motoric and visual activity. As in Seidel et al. (2023), we performed this procedure on each segment to enable further single-trial analysis of the amplitude values. For this, averaged act-only and cue-only ERPs were first exported separately from Brain Vision Analyzer for each block and saccade type for each participant. These averaged signals were then subtracted from each single corresponding act-sound and cue-sound segment from the same block and saccade type to correct for motor and visual activity, respectively. In the following, the terms act- and cue-sound ERPs will refer to these corrected ERPs.

To extract single trial amplitudes for the components of interest, we first localized the peaks of the N1 and P2 in the overall grand average signal (see Figure 2) at electrodes Fz, FCz and Cz, for which all trials from the act- and cue-sound conditions over all participants were averaged. The mean N1 and P2 peak latencies over all three electrodes were determined as 88 ms and 162 ms, respectively. For each component, we then used a 100 ms time window around these peaks (N1: 38-137 ms, P2: 112-211) to determine participant-specific peaks in the averaged signal for each condition, separately for each saccade type, run and electrode. To avoid that very early additional negative peaks were scored as N1, and in order to capture rather late P2 peaks, as they occurred in some participants, the windows were shortened for the N1 (48-137 ms) and extended for the P2 (112-222 ms) in a second step. In a last step, we collected mean amplitudes in a 40 ms time window around these peaks in the corresponding single trial data, resulting in three mean amplitude values per trial (one per electrode) for each component. To remove intrasubject extreme values for each condition within each participant, we excluded single trial mean amplitude values deviating more than 2.5 SD from the mean, separately for every participant/electrode/condition/saccade type/run combination.

Statistical analysis

EEG data. Components N1 and P2 were analyzed separately, by fitting the amplitude data for each to the same linear mixed effects model. This model included the predictors Sound Type (cue-sounds [-0.5], act-sounds [0.5]) and Saccade Type (prosaccades [-0.5], antisaccades [0.5]) as the experimental factors. Similar to Seidel et al. (2023), and following suggestions by Volpert-Esmond et al. (2021), we modeled the course of time over the experiment, employing two further variables. The predictor Run (first [-0.5], second [0.5]) was used to differentiate data from the blocks in the first half of the experiment from those in the second half, while Trialnumber (1-40) accounted for the order of trials in each run. Instead of numbering the available trials per run and condition after exclusions of error trials and EEG artifacts, this latter predictor contained the original trial numbers before exclusions, to retain the accurate temporal position of each trial. Since missing trials would cause the shifting of the mean when centering this continuous predictor, centering was done using the theoretical mean of 20.5. The model also included all interactions between the four predictors. For random effects we included a random intercept and random slopes (Sound Type, Saccade Type, Run and all their interactions) for the participants, and additionally a random intercept for the electrodes. The final model formula was:

$$\begin{gathered}\text{Mean Amplitude} \sim \text{Sound Type} * \text{Saccade Type} * \text{Run} * \text{Trialnumber} \\ + \ (1 + \text{Sound Type} * \text{Saccade Type} * \text{Run} \mid \text{Participant}) + (1 \mid \text{Electrode})\end{gathered}$$

R (Version 3.6.3) was used for statistical analysis, including the lme4 package (Version 1.1-23) and lmerTest package (Version 3.1-2) to test for significant effects with Satterthwaite approximated degrees of freedom. Significant interactions, were examined by performing simple effects analyses. Interactions of categorical fixed effects (e.g., Sound Type, Viewpoint), were examined by fitting two models that differed in their dummy-coding (0,1) of one involved predictor. The reference level was set to the first level in one model, and to the second level in the other, and the remaining predictors (or interactions for multiple predictors) involved in the interaction were subsequently checked for significance. When resolving for the continuous predictor Trialnumber, we re-centered it to the beginning (1) and end (40) of the run, similar to Volpert-Esmond et al. (2021). Random effects were not changed for any simple effect analysis. An α level of .05 was considered as statistically significant. Main effects are reported for the two experimental predictors, Sound Type and

Saccade Type, interactions are only reported when they involve the predictor Sound Type. Full results are included in the R markdown file in addition to predictor coding and analyses at https://doi.org/10.17605/OSF.IO/BX8FU. Estimated marginal means of both models for the two components are visualized in the line plots in Figure 3.

Behavioral data. To examine behavioral differences between the pro- and antisaccade task, we separately fitted the reaction time and number of errors (concerning saccade direction) data of the two act-sound conditions to the following linear mixed models:

$$\text{Reaction Time} \sim \text{Saccade Type} + \ (1 + \text{Saccade Type} \mid \text{Participant})$$

$$\text{number of errors} \sim \text{Saccade Type} + (1 \mid \text{Participant})$$

Analyzing the reaction time from single trial data allowed us to include the factor Saccade Type as a random effect. The number of errors was instead summed up for each saccade type, and thus aggregated. The reaction times of all correct trials were included in the analysis, differing from the analysis of the ERP data, in which further trials were excluded due to artefacts. An α level of .05 was considered as statistically significant.

Results

Behavioral Data

For reaction times, the model fit revealed a significant main effect of Saccade Type, *F*(1, 36.8) = 68.71, *p* < .001, with longer reaction times in the antisaccade (*M* = 303.4 ms, *SD* = 116.5 ms) than the prosaccade task (*M* = 236.4 ms, *SD* = 140.3 ms). This main effect also reached significance for the number of saccadic errors, *F*(1, 37) = 32.75, *p* < .001, with more errors in the antisaccade (*M* = 13.9, *SD* = 11.4) then the prosaccade task (*M* = 4.1, *SD* = 6.3).

N1 component

The model fit for the N1 revealed no significant effect of Sound Type, *F*(1, 36.2) = 0.31, *p* = .579, or Saccade Type, *F*(1, 36.2) = 0.07, *p* = .797. We found a significant

Sound Type by Trialnumber interaction, $F(1, 28486.7) = 18.46$, $p < .001$, and simple effects analyses showed that amplitudes for act-sounds became less negative over time, $t(28483.7) = 4.94$, $p < .001$, $b = 0.03$ (1.33 µV over 40 trials), but those for cue-sounds did not, $t(28473.1) = -1.14$, $p = .254$, $b = -0.01$. The alternative resolution showed that amplitudes for act-sounds were increased compared to those for cue-sounds at the beginning of runs, $t(65.8) = -2.71$, $p = .009$, $b = -0.97$, but not at the end, $t(68.3) = 1.72$, $p = .089$, $b = 0.62$.

The Sound Type by Saccade Type by Run by Trialnumber four-way interaction reached significance as well, $F(1, 28485.7) = 4.29$, $p = .038$, and a first resolution by Saccade Type showed that the underlying three-way interaction was only significant for prosaccades, $t(28449.2) = 2.70$, $p = .007$, and not antisaccades, $t(28435.7) = -0.39$, $p = .699$. Instead, for antisaccades a significant Sound Type by Trialnumber interaction was found, $F(1, 28439.9) = 2.76$, $p = .006$, for which the same pattern emerged as for the overall Sound Type by Trialnumber interaction: One resolution showed decreasing amplitudes for act-sounds, $t(28423.3) = 3.14$, $p = .002$, $b = -0.03$ (1.27 µV over 40 trials), but not cue-sounds, $t(28351.2) = -0.76$, $p = .445$, $b = -0.01$, the other showed increased amplitudes for act-sounds at the start, $t(63.3) = -2.33$, $p = .023$, $b = -1.27$, not the end of runs, $t(66.5) = 0.49$, $p = .628$, $b = 0.27$.

Resolving the Sound Type by Run by Trialnumber interaction for prosaccades revealed a significant Sound Type by Trialnumber interaction in the second run, $t(28436.7) = 4.32$, $p < .001$, but not in the first run, $t(28440.7) = 0.47$, $p = .637$. Further simple effects analyses showed that at the beginning of the second run, amplitudes for act-sounds were significantly higher than those for cue-sounds, $t(63.1) = -2.08$, $p = .041$, $b = -1.41$, while the opposite effect was significant at the end of the run $t(63.8) = 2.31$, $p = .024$, $b = 1.57$, which is also visible in the marginal estimates means (see Figure 3). In an alternative resolution, the pattern seen in the overall Sound Type by Trialnumber interaction is confirmed, with decreasing amplitudes for act-sounds, $t(28436) = 4.90$, $p < .001$, $b = 0.06$

(2.45 µV over 40 trials), but not for cue-sounds, *t*(28434.3) = -1.20, *p* = .230, *b* = -0.01. In the first prosaccade run, neither the amplitudes for act-sounds nor for cue-sounds showed a significant change over time, both *p* > .510. The remaining interactions including Sound Type did not reach significance, all *p*s > .135.

P2 component

The model fit revealed a significant main effect of Sound Type, *F*(1, 36.6) = 27.32, *p* < .001, with higher amplitudes for act- compared to cue-sounds, *b* = 1.65. The main effect of Saccade Type also reached significance, *F*(1, 34.7) = 12.46, *p* = .001, and the parameter estimate indicated higher amplitudes for antisaccade- compared to prosaccade-generated sounds, b =1.63.

We again found a significant Sound Type by Saccade Type by Run by Trialnumber four-way interaction, *F*(1, 28420.6) = 12.63, *p* < .001. Resolving this interaction by Saccade Type, simple effect analyses revealed a significant underlying three-way interaction for antisaccades, *t*(28222.8) = -3.37, *p* < .001, but not prosaccades, *t*(28422.5) = 1.55, *p* = .122. Instead, a significant Sound Type by Trialnumber interaction for prosaccades, *t*(28419.6) = -2.15, *p* = .032, indicated significantly higher amplitudes for act-sounds than cue-sounds at the start of each run, *t*(58.9) = 3.13, *p* = .003, *b* = 1.76, but not at the end, *t*(59.8) = 1.13, *p* = .263, *b* = 0.64.

The Sound Type by Run by Trialnumber interaction for antisaccades was resolved by Run, and simple effect analysis showed significant Sound Type by Trialnumber interaction for the first, *t*(28402.8) = 2.16, *p* = .031, and the second Run, *t*(27389.6) = -2.60, *p* = .009. In both antisaccade runs, amplitudes for cue-sounds did not change significantly over time, both *p* > .092, while those for act-sounds increased over time in the first, *t*(28403.3) = 2.19, *p* = .028, *b* = 0.03 (1.31 µV over 40 trials), and decreased in the second Run, *t*(28103.9) = -2.00, *p* = .046, *b* = -0.03 (-1.22 µV over 40 trials), and this pattern can also be seen in Figure 3. Accordingly, the alternative resolution for the first run shows no significant Sound Type effect at the start of the run, *t*(51.3) = 0.54, *p* = .591, but significantly higher amplitudes for act-sounds than cue-sounds at the end, *t*(53.6) = 2.41, *p* = .020, *b* = 2.29, and for both the start, *t*(69.9) = 5.12, *p* < .001, *b* = 3.91, and end, *t*(73.8) = 2.19, *p* = .031, *b* = 1.70, of the second run.

Discussion

In this study, we compared the neurophysiological sensory attenuation effects for self-generated auditory stimuli, reflected in the N1 and P2 amplitudes, that were produced by pro- or antisaccades in order to examine a possible influence of interfering efference copies-on forward model predictions concerning sensory consequences of actions. Participants performed either visually guided prosaccades to a target, or antisaccades in the opposite direction, for which efference copies might be disturbed because of suppressed reflexive prosaccades (Coe & Munoz, 2017; Munoz & Everling, 2004). ERPs in response to the saccade-generated sounds were compared to those for visually cued externally-generated sounds. Mixed effect modeling of single-trial data revealed similar temporal dynamics of N1 amplitudes over the course of the experiment for both saccade types, with amplitudes for saccade-generated sounds being increased relative to amplitudes for visually cued sounds at the beginning of each run and then decreasing over time (except for prosaccades in the first run). A sensory attenuation effect in the sense of reduced N1 amplitudes for saccade-generated relative to visually cued sounds was, however, only seen for prosaccades at the end of the experiment, while the N1 amplitude for antisaccades was never lower than for cued external sounds. For the P2, on the other hand, distinct temporal patterns for the two types of saccades emerged. A small P2 enhancement for prosaccade-generated sounds was found at the start of each half of the experiment, which vanished over time. P2 amplitudes for antisaccade-generated sounds increased in the first half of the experiment, and decreased in the second. Throughout the experiment, P2 amplitudes for sounds following antisaccades were enhanced relative to those for external sounds. Considering the identical temporal predictability for both sound eliciting actions, the results demonstrate an influence of a disturbed or conflicting efference copy signal on the N1 and the P2.

N1 component

Based on the finding by Mifsud et al. (2016) we expected an N1 attenuation for prosaccade-generated sounds compared to cued externally-generated sounds, and a diminished or missing attenuation for antisaccade-generated sounds. This hypothesis could not clearly be confirmed, as we did not find a significant interaction between Sound Type and Saccade Type. Moreover, we expected different developments of N1 amplitudes over time for sounds following pro- and antisaccades, and this was indeed reflected in a four-way interaction of both experimental predictors with the two predictors encoding the course of time during the experiment. Using the two-level factor Run and the continuous factor Trialnumber (within Run) we found a similar temporal dynamic of N1 amplitudes in both runs for antisaccades, and the second run for prosaccades, namely decreasing amplitudes over time for saccade-generated sounds, while those for cued externally-generated sounds remained stable. The four-way interaction stems from the fact that only for prosaccades in the second run the N1 amplitudes decrease enough to be significantly reduced compared to cued externally-generated sounds at the end of the run (see Figure 3). N1 amplitudes for antisaccade-generated sounds also decreased over time, but only until they did not significantly differ from cued externally-generated sounds anymore. Thus, the pattern emerging at the end of the experiment is in line with our first hypothesis to find an attenuation for pro- but not antisaccades, while the stronger amplitude decrease over time for pro- than antisaccades is in line with our second hypothesis.

The reduced N1 for sounds elicited by prosaccades is in line with the result by Mifsud et al. (2016), and also extends this finding, as they employed uncued externally-generated sounds, while we found an N1 attenuation at the end of the experiment relative to visually cued external sounds. The prevalent account for the N1 attenuation is that sensory predictions are generated by feed forward models employing efference copy information right after motor planning to enable the early matching of predictions with actual sensory input, resulting in an attenuation of the N1 (Blakemore et al., 2001; Horváth, 2015; Pickering & Clark, 2014). Alternatively, this attenuation has been considered as an effect of temporal predictability, which is inherent in self-generated stimuli compared to unpredictable externally-generated stimuli. Although predictability of sounds has been reported to lower N1 amplitudes (Harrison et al., 2021; Kaiser & Schütz-Bosbach, 2018), the results of the present study cannot be ascribed to general predictability effects. First, we compared the processing of self-generated sounds with the processing of cued and thus temporally predictable external sounds. And second, the comparison of sounds following pro- and antisaccades cannot be affected by temporal predictability, as their motoric execution (and their predictability) is comparable. Thus, the result pattern for the N1 suggest that the suppression of reflexive prosaccades in the antisaccade condition (Coe & Munoz, 2017), leads to conflicting efference copy signals that disturb motor-based sensory predictions mechanisms. In this way our result suggests an involvement of motoric signals in the mechanism attenuating the N1 amplitude that is unrelated to the reduction ascribed to temporal predictability.

The fact that this attenuation (for prosaccades) was only found at the end of the experiment could be interpreted as a tuning of the feedforward model to establish the novel action-effect contingency of eliciting sounds by saccadic movement. Dogge et al. (2019) questioned the involvement of forward models in the N1 amplitude reduction, because their tuning should take longer than the short training sessions that are typically applied. In previous studies employing button press elicited-sounds, however, the acquisition of an action-effect contingency might have been accelerated, because of the common occurrence of button-press-generated sounds in everyday life. For saccades such experience in every-day life is unlikely, as they usually do not elicit sounds. Our results, and especially the specific changes in N1 amplitudes over time for saccade-generated sounds might thus be evidence for the slow tuning process of forward models for novel action-effect contingencies, as suggested by Dogge et al. (2019). This would also be in line with our previous report of a similar temporal dynamic of decreasing N1 amplitudes for sounds generated by observed actions from a first-person perspective, which is also not commonly occurring in every-day life (Seidel et al., 2023). For third-person observation, which is well-trained on the other hand, no such temporal dynamic was found, and N1 amplitudes were attenuated from the start. The results of both studies suggest that action-effect associations that resemble those from everyday life and are thus familiar can be established more quickly than completely novel associations.

The temporal dynamic in our study also revealed increased N1 amplitudes for saccade-generated sounds at the beginning of each run (except for the first prosaccade run) compared to the external sounds. In contrast, action-related sound amplitudes in our previous study started off at the level of external sound amplitudes and decreased from there (Seidel et al., 2023). This difference is likely due to the comparison with cued external sounds in the present study, as N1 amplitudes have been shown to decrease with temporal predictability (Harrison et al., 2021). The fact that enhanced N1 amplitudes were again seen at the start of the second run, even though amplitudes had reached the level of cued sounds in the first run, might be explained by the order of conditions. Between each condition in which sounds were generated by saccades there was one condition in which saccades were performed without generating any sounds. This might have interrupted the tuning of the forward model, especially for novel action-effect associations as applied in this study.

P2 component

For the P2 we did not expect to find an amplitude reduction, neither for pro-, nor for antisaccade-generated sounds compared to external sounds, as reported by Mifsud et al. (2016). Instead we found P2 amplitudes for saccade-generated sounds to be enhanced to differing degrees in both saccade types. For prosaccades, P2 amplitudes were enhanced at the beginning of runs, but not at the end. Amplitudes for antisaccade-generated sounds increased over time in the first run, and decreased in the second, mostly being significantly enhanced compared to externally-generated sounds, except for the start of the first run. These enhancements, which were also represented by a main effect of the Sound Type, can be explained by the fact that amplitudes were compared with those for visually cued sounds, for which reduced P2 amplitudes have been shown (Sowman et al., 2012), while Mifsud et al. (2016) used uncued external sounds in the comparison condition for saccade-generated sounds.

Because the P2 amplitude attenuation has been associated with more contextual influences that are independent of the motor-based forward models (Knolle et al., 2013; Seidel et al., 2021; Sowman et al., 2012; Timm et al., 2016), we did not expect to find differing patterns for pro- and antisaccade-generated sounds, which only vary in their motor planning. But while just a slight temporal dynamic of P2 amplitudes for prosaccade-generated sounds emerged, that was only significant in the comparison with cued sounds, amplitudes for antisaccade-generated sounds showed significant changes over time, with very high amplitudes at the start of the second run. Considering previous results showing increased P2 amplitudes under a perceived loss of agency or control over sound production (Seidel et al., 2021; Timm et al., 2016), this might indicate a diminished belief to generate these sounds in the first run. This uncertainty could be exacerbated by performing the Act-only condition (without eliciting sounds) directly after the first Act-sound condition run, leading to the highly increased amplitudes observed at the start of the second run. The significant decrease of the P2 for antisaccade-generated sounds in the second run might then reflect slightly increasing levels of agency or control over the course of the second run. However, P2 amplitudes for antisaccade-generated sounds continued to be elevated. In comparison, the amplitudes for prosaccade-generated sounds were not enhanced (compared to externally-generated sounds) at the end of the runs, which would imply a higher perceived level of agency or control compared to antisaccades. Such a difference in agency between pro- and antisaccades might stem from the lack of experience in performing antisaccades in everyday life. It is thus possible that the subjective feeling of having performed an antisaccade, which is only initiated after strong top-down regulation, is comparable or even weaker than the feeling of having performed the (suppressed) reflexive prosaccade, that one intuitively wants to perform. The conflicting efference copy signals- that we expected for antisaccades and that may be reflected in the result pattern for the N1 might also hinder a clear agency assignment in this condition.

Conclusions

This study provides evidence for the integration of motoric signals in the formation of predictions for auditory action-effects, by examining N1 and P2 amplitudes for sounds generated by prosaccades and antisaccades in comparison to cued external sounds. Mixed linear model analysis on trial-level data revealed an N1 reduction for prosaccade-, but not antisaccade-generated sounds, that developed over the course of the experiment. This reduction is a replication of findings for button-press-generated sounds, but the slow temporal dynamic might indicate that unusual action-effect contingencies, like sounds generated by saccades, might involve a slower tuning process of internal forward models, while only fine-tuning is necessary when similar action effects are encountered in everyday life, such as sounds following button-presses. For antisaccade-generated sounds, the missing N1 reduction hints at conflicting efference copy information relayed to predictive models, as antisaccades require the suppression of automatically generated prosaccades. We also report a P2 for prosaccade-generated sounds that decreased over time, until no significant difference to cued externally-generated sounds was found, unlike the P2 for antisaccades, that remained enhanced in comparison. This is potentially an effect of reduced perceived control or agency over sound generation when performing antisaccades.

References

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, *48*(9), 1276-1283. https://doi.org/10.1111/j.1469-8986.2011.01196.x

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, *1*(7), 635. https://doi.org/10.1038/2870

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, *12*(9), 1879-1884. https://doi.org/10.1097/00001756-200107030-00023

Coe, B. C., & Munoz, D. P. (2017). Mechanisms of saccade suppression revealed in the anti-saccade task. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160192. https://doi.org/10.1098/rstb.2016.0192

Dogge, M., Custers, R., & Aarts, H. (2019). Moving forward: On the limits of motor-based forward models. *Trends Cogn Sci*. https://doi.org/10.1016/j.tics.2019.06.008

Ghio, M., Egan, S., & Bellebaum, C. (2021). Similarities and Differences between Performers and Observers in Processing Auditory Action Consequences: Evidence from Simultaneous EEG Acquisition. *Journal of cognitive neuroscience*, *33*(4), 1-12. https://doi.org/10.1162/jocn_a_01671

Ghio, M., Scharmach, K., & Bellebaum, C. (2018). ERP correlates of processing the auditory consequences of own versus observed actions. *Psychophysiology*, *55*(6), e13048. https://doi.org/10.1111/psyp.13048

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

Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain research*, *1626*, 54-65. https://doi.org/10.1016/j.brainres.2015.03.038

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *139*(1), 133. https://doi.org/10.1037/a0028566

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

Klaffehn, A. L., Baess, P., Kunde, W., & Pfister, R. (2019). Sensory attenuation prevails when controlling for temporal predictability of self-and externally generated tones. *Neuropsychologia*, *132*, 107145. https://doi.org/10.1016/j.neuropsychologia.2019.107145

Knolle, F., Schroger, E., Baess, P., & Kotz, S. A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of cognitive neuroscience*, *24*(3), 698-706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schroger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of self-initiated sounds. *cortex*, *49*(9), 2449-2461. https://doi.org/10.1016/j.cortex.2012.12.012

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

Mifsud, N. G., Beesley, T., Watson, T. L., & Whitford, T. J. (2016). Attenuation of auditory evoked potentials for hand and eye-initiated sounds. *Biological Psychology*, *120*, 61-68. https://doi.org/10.1016/j.biopsycho.2016.08.011

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218. https://doi.org/10.1038/nrn1345

Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends Cogn Sci*, *18*(9), 451-456. https://doi.org/10.1016/j.tics.2014.05.006

Popa, L. S., & Ebner, T. J. (2018). Cerebellum, Predictions and Errors. *Frontiers in cellular neuroscience*, *12*, 524. https://doi.org/10.3389/fncel.2018.00524

Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia*, *119*, 302-307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and cognition*, *17*(4), 1219-1227. https://doi.org/10.1016/j.concog.2008.01.003

Schafer, E. W., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science*, *181*(4095), 175-177. https://doi.org/10.1126/science.181.4095.175

Seidel, A., Ghio, M., Studer, B., & Bellebaum, C. (2021). Illusion of control affects ERP amplitude reductions for auditory outcomes of self-generated actions. *Psychophysiology*, *58*(5), e13792. https://doi.org/10.1111/psyp.13792

Seidel, A., Weber, C., Ghio, M., & Bellebaum, C. (2023). My view on your actions: Dynamic changes in viewpoint-dependent auditory ERP attenuation during action observation. *Cognitive, Affective, & Behavioral Neuroscience*, 1-17. https://doi.org/10.3758/s13415-023-01083-7

Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental brain research*, *222*(1-2), 149-157. https://doi.org/10.1007/s00221-012-3204-7

Timm, J., SanMiguel, I., Keil, J., Schroger, E., & Schonwiesner, M. (2014). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of cognitive neuroscience*, *26*(7), 1481-1489. https://doi.org/10.1162/jocn_a_00552

Timm, J., Schonwiesner, M., Schroger, 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

Vercillo, T., O'Neil, S., & Jiang, F. (2018). Action–effect contingency modulates the readiness potential. *Neuroimage*, *183*, 273-279. https://doi.org/10.1016/j.neuroimage.2018.08.028

Volpert-Esmond, H. I., Page-Gould, E., & Bartholow, B. D. (2021). Using multilevel models for the analysis of event-related potentials. *International Journal of Psychophysiology*, *162*, 145-156. https://doi.org/10.1016/j.ijpsycho.2021.02.006

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current biology*, *11*(18), R729-732. https://doi.org/10.1016/s0960-9822(01)00432-8

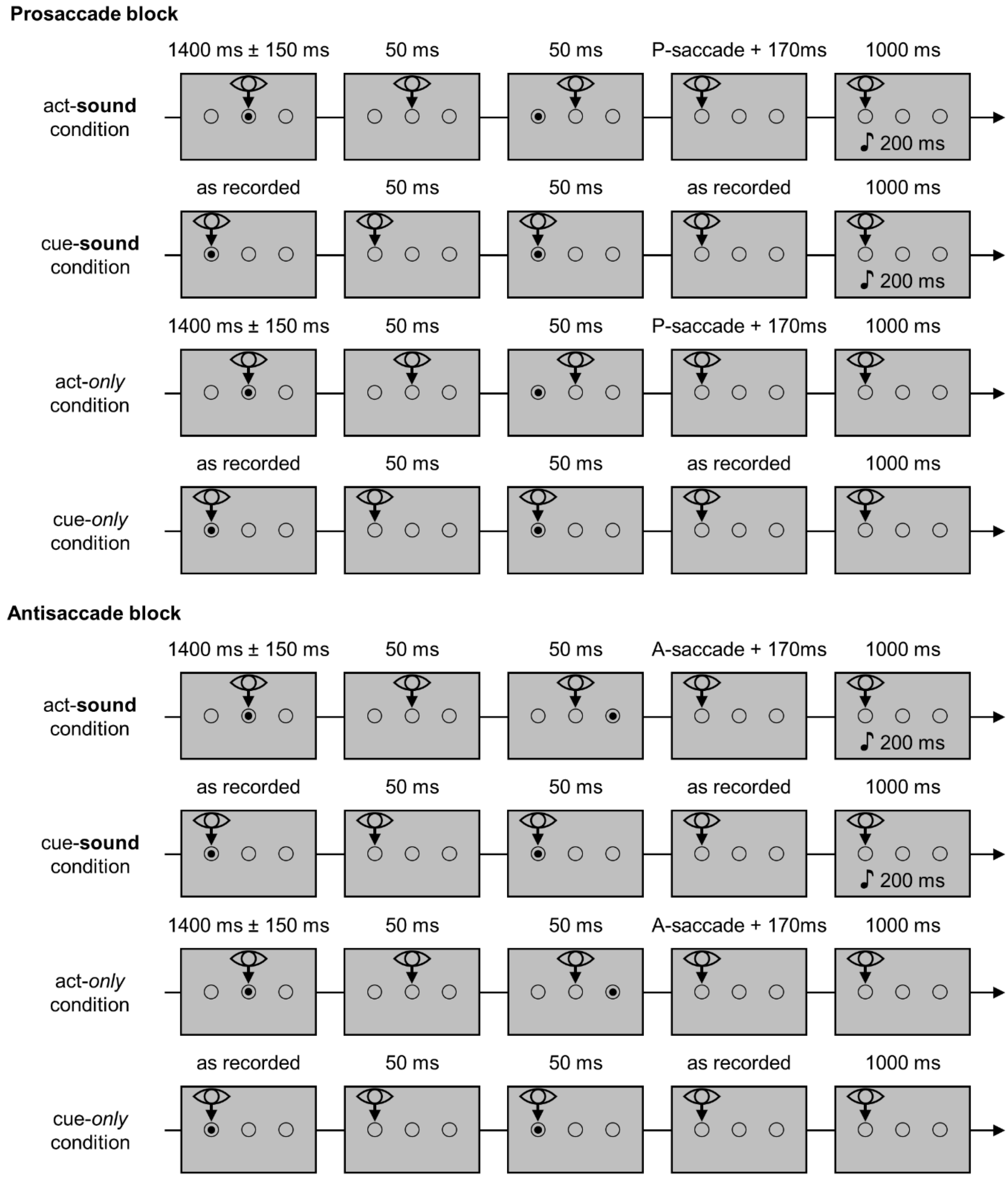

Figure 1. Experimental sequence in the different conditions of one pro- and one antisaccade block. Cue-sound and cue-only conditions used timings and target stimulus positions as they were recorded in the act-sound condition at the beginning of each block. In antisaccade blocks, the target position was recorded as the side the antisaccade was aimed at, not the displayed stimulus position. In the act-sound conditions, sounds were only played if a saccade

in the correct direction was detected, and this was replicated in the corresponding cue-sound condition as well. The order of conditions in each block was fixed.

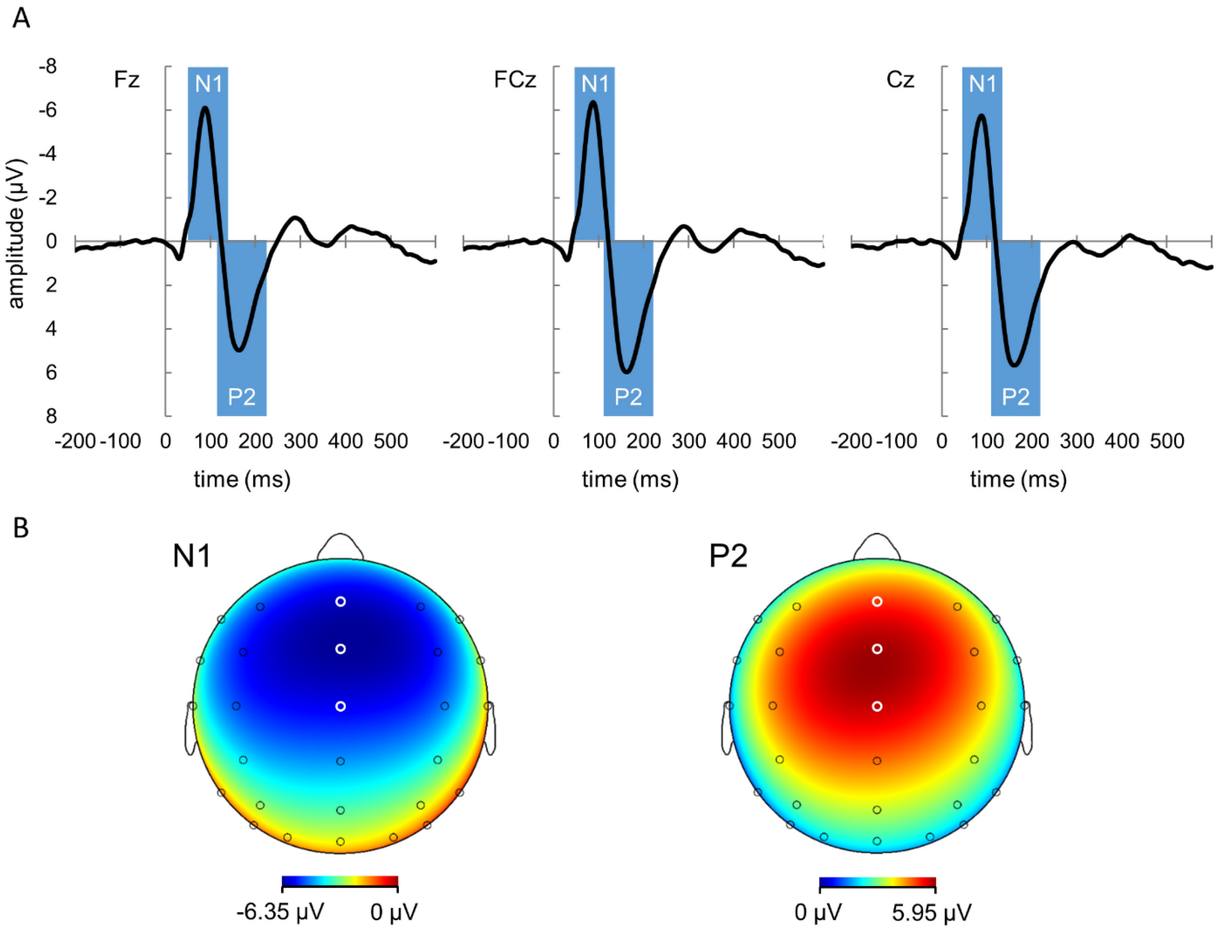

Figure 2. A: Overall sound-related grand average ERPs (from all act- and cue-sound conditions) at Fz, FCz, and Cz of all trials (act- and cue-sounds). Blue bars show the time windows used to locate participant-specific peaks in each condition. B: Topographical maps of scalp potentials at the time of the N1 and P2 peaks from the overall grand average ERPs in A. Electrode positions Fz, FCz, and Cz (from top to bottom) are marked in white.

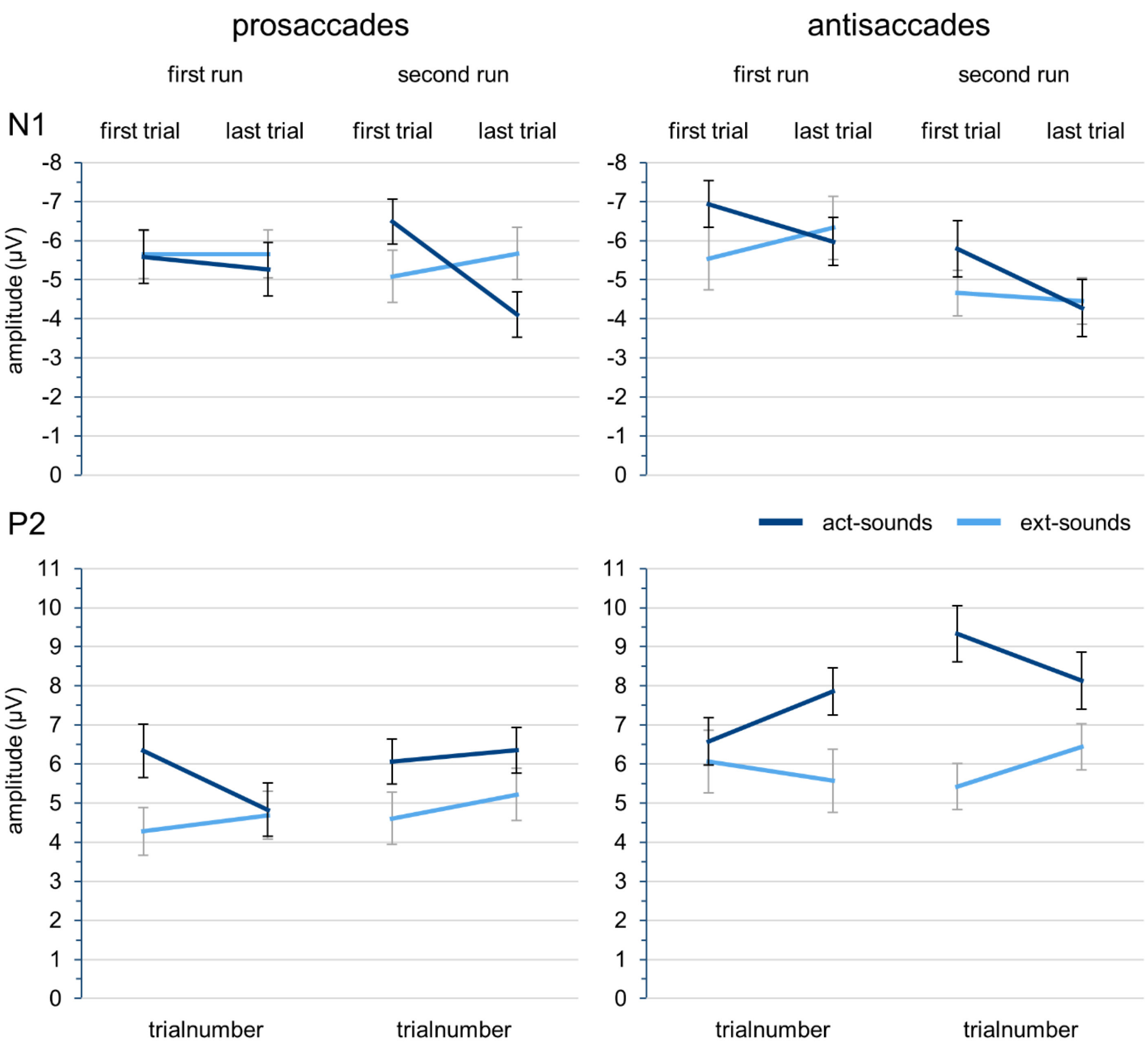

Figure 3. Line plots of the marginal estimated means for the linear mixed effects models. Lines represent the estimate over all 40 trials of each run. Error bars show one standard error at the start and end of each run, corresponding to the simple effects analyses.