

The multifaceted nature of self-generation effects: Independent and interactive effects of actions, predictability, and delay on sensory processing and memory encoding of sounds.

Nadia Paraskevoudi¹ and Iria SanMiguel²

¹Institute of Neurosciences, University of Barcelona

²Universitat de Barcelona Institut de Neurociències

May 6, 2023

The multifaceted nature of self-generation effects: Independent and interactive effects of actions, predictability, and delay on sensory processing and memory encoding of sounds.

Nadia Paraskevoudi^{a,b} & Iria SanMiguel^{a,b,c}

Affiliations:

^aInstitut de Neurociències, Universitat de Barcelona, Barcelona, Spain

^bBrainlab-Cognitive Neuroscience Research Group, Departament de Psicologia Clínica i Psicobiologia, Universitat de Barcelona, Barcelona, Spain

^cInstitut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain.

Corresponding author:

Iria SanMiguel

Department of Clinical Psychology and Psychobiology

P. Vall d'Hebron 171

08035 Barcelona, Spain

Tel: (0034) 93 312 5050

Email: isanmiguel@ub.edu

Declarations of interest: none.

Abstract

Actions attenuate responses to self-produced sounds, but the underlying mechanism remains under debate. In fact, factors such as action-sound predictability and enhanced neuromodulation during movement seem to play a role. Meanwhile, the self-generation effects on memory – and their possible interactions with predictability – remain also largely unexplored. Here, we recorded behavioural, EEG, and pupil responses during a combined self-generation and memory paradigm to assess the effects of self-generation, predictability, and their interactions, as a function of action-sound delay, on the sensory processing and memory of sounds. Predictability was controlled such that self-generated (“motor-auditory”; MA) and externally-generated (“Auditory-only”; A) sounds were equally predictable (or unpredictable) in identity, timing, and probability of occurrence. In each trial, participants memorized predictable or unpredictable MA or A sounds presented with a delay of 150 or 750 ms after the action (MA) or visual cue (A). The same sounds were later tested for recall. Results showed better memory for MA sounds, irrespective of predictability, and larger pupil diameter during actions especially when actions triggered predictable sounds. Self-generation attenuated N1, P2, and Tb responses with short delays, but only in the predictable session. With longer delays, self-generation attenuated N1 (irrespective of predictability), while predictability attenuated P2 and Tb (irrespective of self-generation). Crucially, sensory attenuation correlated with pupil diameter when the effects were strongest (predictable MA sounds presented after short delays). Collectively, we show that actions boost memory and we propose that motor inputs along with neuromodulatory influences shape auditory responses depending on action-stimulus predictability and delay.

1. Introduction

The ability to distinguish between the two possible sources of sensory input (self or environment) is crucial for all organisms. Such distinction allows one to perceptually prioritize the newsworthy unpredictable – and potentially threatening – information (Blakemore et al., 2000; Poulet & Hedwig, 2006; Press et al., 2020) and to shape our sense of agency (Gallagher, 2000), and it highlights that rather than being a veridical reproduction of our external world, perception is scaffolded by our own predictions about the identity and timing of any upcoming event (Press et al., 2020). Such predictions can easily be afforded by our own actions: We cannot tickle ourselves, but we can be tickled by others, and similarly, we will be alerted by the sound of a horn, unless we are the ones generating it. This perceptual attenuation of self-generated reafferent input has been already reported (Sato, 2008, Paraskevoudi & SanMiguel, 2021) and agrees with neurophysiological data showing that indeed responses to self-generated input are suppressed compared to identical, albeit passively presented signals (Schafer & Marcus, 1973; Schröger et al., 2015).

Strikingly, the exact mechanisms driving the attenuation effects are still a matter of debate (for reviews see Schröger et al., 2015; Hughes et al., 2013), and evidence suggests that sensory responses are dampened when we can predict the sensory consequences of our actions (e.g., Baess et al., 2011), but also when we cannot (i.e., attenuation occurs also for stimuli merely coinciding with movement, Horváth et al., 2012). Moreover, predictability alone – irrespective of its source – has also been shown to either dampen or enhance sensory responses (Press et al., 2020). Meanwhile, there is also evidence on effects of actions on memory processes (e.g., the production effect; MacDonald & MacLeod, 1998; Brown & Palmer, 2012), but again in these studies the action and predictability effects are usually conflated. These findings raise the need to disentangle the effects of predictability and self-generation on sensory processing and memory encoding, which constitutes the overarching aim of the present study. Specifically, we contrasted the sensory modulation and memory effects under predictable and unpredictable contexts, to isolate the contribution of motor acts on sensory processing and memory encoding and its interaction with predictability. In the following paragraphs, we provide an overview of the literature, summarizing the findings related to the action-induced modulations in sensory responses and memory performance as a function of predictability.

1.1. Sensory processing of self-produced stimuli

In the auditory domain, the self-generation effects on sensory processing have been mainly assessed using a *contingent* paradigm, where participants either listen passively to sounds presented by the computer or perform button presses that always result in the generation of a fully predictable sound (e.g., Baess et al., 2011). Most studies have reported attenuated auditory N1 and P2 event-related potential amplitudes and – inspired by early animal physiology work (Sperry, 1950; von Holst & Mittelstaedt, 1950) and computational models of motor control (Miall & Wolpert, 1996; Wolpert et al., 1995) – have explained the effect under the dominant cancellation models (also known as the comparator model; Blakemore et al., 1998; Frith et al., 2000; Wolpert & Flanagan, 2001). According to this model, the attenuation effects result from the operation of an internal forward model that allows one to predict the sensory consequences of her own action (corollary discharge) using a copy of the motor command (efference copy) (Blakemore et al., 2000; Wolpert & Flanagan, 2001). The resulting

motor-driven predictions of sensory reafference are compared to the actual sensory consequences of one's actions, and subsequently, only the difference between the two (i.e., prediction error) is sent to higher stages of the neuronal hierarchy for further processing (Friston, 2005), effectively cancelling out responses to predictable, self-generated signals (Blakemore et al., 1998).

Inevitably, according to the cancellation model one would expect that suppression of responses would occur only when the stimulus can be predicted by the action and that the suppression effects would reflect modulations in sensory-specific areas. There is indeed evidence supporting the stimulus-specificity of the effects: For example, attenuation seems to be largest when the stimulus is fully predictable (Fu et al., 2006; Houde et al., 2002; Baess et al., 2008), and the locus of the effect has been found to be in areas within the auditory cortex (Martikainen et al., 2004; Aliu et al., 2009). Although these studies have provided strong support to the specificity of the effects, they assessed the effects only under fully predictable contexts for the self-generated stimulation, but not necessarily for the externally-generated one. That is, while the self-generated sounds were fully predictable in their identity, timing, and probability of occurrence, the timing and probability of occurrence for the externally-generated sounds could not be predicted (since they were presented passively without any cue). Therefore, the majority of these contingent paradigms have consistently conflated self-generation and predictability, and have also confounded different types of predictions (i.e., predictions related to the identity, timing, and probability of occurrence of the stimulus; Hughes et al., 2013), that can possibly exert distinct effects on sensory processing.

However, there is also evidence that sensory attenuation prevails even in the absence of a contingent action-sound relationship. For example, sensory responses are suppressed for stimuli that could not be predicted by the action (Horváth et al., 2012; Horváth, 2013a, 2013b; Paraskevoudi & SanMiguel, 2022; Hazemann et al., 1975; Makeig et al., 1996; Numminen et al., 1999), in line with the stimulus-unspecific generalized attenuation during certain actions reported outside of the domain of self-generation (e.g., saccadic suppression and somatosensory gating on moving body parts, Crapse & Sommer, 2008; Ross et al., 2001; Williams et al., 1998). Further support to the contribution of stimulus-unspecific processes in the suppression effects comes from work showing that part of the attenuation effects, specifically the auditory N1 attenuation, might not reflect a genuine modulation within the auditory cortex (SanMiguel et al., 2013). Specifically, SanMiguel and colleagues showed that instead of attenuating the sensory-specific N1 components (i.e., N1 at the mastoids and the "T complex" that are generated by tangentially oriented sources in the auditory cortex and radial sources in the superior temporal gyrus, respectively; Näätänen & Picton, 1987; Tonnquist-Uhlen et al., 2003; Wolpaw & Penry, 1975), actions reduced mainly the unspecific component of N1, which is proposed to be the cortical projection of a reticular process facilitating motor activity and related to the orienting response (Näätänen & Picton, 1987). Altogether, these findings raise the possibility of the self-generation effects not being solely a consequence of stimulus-specific predictions as proposed by cancellation theories, but they might rather reflect the operation of a broader stimulus-unspecific mechanism (Korka et al., 2021a). This mechanism would allow the system to expect some action-related consequence, without necessarily generating a prediction on the specific sensory effect of the action, and would manifest as a modulation of sensory responses to stimuli presented in close temporal proximity to the action. However, to date, no attempts have been made to assess this possibility by

manipulating all aspects of predictability (identity, timing, and probability of occurrence) in one experimental paradigm.

Meanwhile, despite the converging evidence pointing to stimulus-unspecific processes behind the action-induced modulations of sensory processing (Press et al., 2020; Press & Cook, 2015), the exact mechanism mediating unspecific effects remains largely unexplored. It is possible that actions create a halo of neuromodulation around them which might unspecifically gate auditory processing for stimuli presented in close temporal proximity with the action. This would mean that stimuli presented close in time with the action would be stronger modulated by the neuromodulatory halo than stimuli presented further apart from the motor act. Neuromodulatory influences are a plausible candidate for mediating – at least partly – the action-induced attenuation effects given that actions initiate a series of neuromodulatory processes (Vinck et al., 2015; Eggerman et al., 2014; McGinley et al., 2015), that are known to send inputs to the auditory cortex (sometimes even overlapping with inputs from motor areas; Nelson & Mooney, 2016; for a review see Schneider & Mooney, 2018). Specifically, a possible candidate for creating a halo of neuromodulation that could mediate stimulus-unspecific effects during movement could be the locus coeruleus norepinephrine system (LC-NE). This subcortical structure provides diffuse and widespread neuromodulatory inputs to the entire cortex, and is part of the ascending reticular activating system, which has been shown to modulate arousal and cortical responsivity to sensory stimulation (Aston-Jones & Cohen, 2005). In fact, the LC-NE has connections with both the auditory (Budinger et al., 2008; Foote et al., 1975) and primary motor cortex (Foote & Morrison, 1987), and many studies have shown a close association between pupil diameter – a proxy of LC-NE activity (Aston-Jones & Cohen, 2005; Vinck et al., 2015; Murphy et al., 2014; Joshi et al., 2016) – and actions (e.g., whisking or button press; McGinley et al., 2015; Vinck et al., 2015; Lubinus et al., 2022; Paraskevoudi & SanMiguel, 2022). Based on this evidence, the link between action-induced attenuation and increased neuromodulation might seem plausible, however, the only study that has assessed this relationship only showed a trend, but not a clear link between the two (Paraskevoudi & SanMiguel, 2022).

1.2. Memory encoding for self-produced stimuli

Given the differential sensory processing of self-generated stimuli, one would expect effects of self-generation also on memory. This assumption seems valid considering evidence reporting modulatory effects of movement on hippocampal and parahippocampal activity (Halgren, 1991; Mukamel et al., 2010; Rummell et al., 2016), hippocampal involvement in predicting upcoming events (Davachi & DuBrow, 2015; Hindy et al., 2016), as well as sensitivity to mismatches between expected and observed information (Fonken et al., 2020; Kumaran & Maguire, 2009). However, the direction of the effects of movement on memory remain mixed. One line of evidence points to a beneficial effect of actions on memory encoding: Memory enhancements have been observed for *fully* predictable self-initiated stimuli such as spoken words and played melodies compared to words read silently and passively heard melodies (i.e., production effect, MacDonald & MacLeod, 1998; Brown & Palmer, 2012). In stark contrast with the production effect studies, predictive coding theories postulate that the prediction error elicited by surprising stimuli results in better memory performance than memory performance for *fully* predictable inputs (Bar, 2009; Pine et al., 2018), reminiscent of the von Restorff effect (von Restorff, 1933). Indeed, evidence points to a positive relationship between prediction error

and hippocampus activity at encoding for unpredictable inputs, which ultimately results in better memory performance (Gagnepain et al., 2011; Henson & Gagnepain, 2010; Pine et al., 2018). Thus, from this perspective, worse memory would be expected for self-generated stimuli which are inherently more predictable than externally-generated stimuli. To our knowledge, our recent study is the only one providing supporting evidence for a link between sensory processing and memory encoding in the domain of self-generation (Paraskevoudi & SanMiguel, 2022): Sensory attenuation for sounds that coincided with button presses correlated with worsened memory performance for these stimuli. Interestingly, memory performance also correlated negatively with pupil diameter, suggesting that sensory suppression and pupil dilation for self-initiated stimulation independently relate to memory weakening. However, in all the above-mentioned studies, self-generated stimuli were to some extent more predictable than the passive comparisons (e.g., increased temporal predictability in Paraskevoudi & SanMiguel, 2022 and fully predictable in identity, timing, and probability of occurrence in production effect studies; MacDonald & MacLeod, 1998; Brown & Palmer, 2012), which raises the need to disentangle predictability- and action-effects also on memory encoding.

1.3. The present study

As we outlined in the first two sections of the introduction, the paradigms attempting to explore the effects of actions on sensory processing and memory differed in the relationship between the action and the sound, usually conflating self-generation with predictability, but also in the types of actions performed. In many paradigms, actions were self-paced and goal-directed to generate the sounds and they provided either stimulus-specific identity predictions (“what” stimulus will be generated), or only temporal predictions (“when” the stimulus will be generated; for reviews see Hughes et al., 2013; Schröger et al., 2015). Meanwhile, many other paradigms presented the sounds close in time to, but independently of, the actions and showed that movement dampens sound processing in an indiscriminate unspecific manner (as in human coincidence studies; Horváth et al., 2012; Horváth, 2013a, 2013b), which may point to a diffuse temporal expectation tied to the action (since actions typically have some auditory consequences). Additionally, despite the evidence of the effects of actions on memory, the two dominant lines of research (i.e., production effect studies and predictive coding framework) have observed these effects (either enhancement or weakening, respectively) only in fully contingent designs, leaving unexplored the possibility of other factors (i.e., action-related temporal control over the stimulus or unspecific temporal expectation about the stimulus) affecting the encoding of sounds in memory.

The present study constitutes a first attempt to orthogonally control for predictability and self-generation in order to assess whether the self-generation effects on sensory processing and memory encoding of sounds are due to stimulus-specific predictions, due to stimulus-unspecific effects of action, or a mixture of both. Specifically, the aim of this study was twofold: First, we aimed to assess the effects of actions and predictability, as well as the interactions between the two, on sensory responses (i.e., measured by auditory evoked potentials) and subcortical neuromodulation mediated by the LC-NE system (i.e., measured by pupil diameter), and assess whether sensory attenuation and subcortical neuromodulation during actions are related. Second, we sought out to examine in what way semi self-paced actions affect the encoding of predictable and unpredictable sounds in memory, and

whether the effects of actions on memory performance are related to the sensory attenuation effects and/or the neuromodulatory processes during movements.

Related to our first aim, we hypothesized that electrophysiological responses to self-generated sounds would be attenuated, with the attenuation being strongest when all aspects of the sound (i.e., identity, timing, and probability of occurrence) are predictable (e.g., Baess et al., 2008). We also hypothesized that button presses would increase pupil diameter (e.g., Paraskevoudi & SanMiguel, 2022), but we remained agnostic as to whether this effect would interact with predictability. Finally, we hypothesized that sensory attenuation and subcortical neuromodulation would be linked. Related to our second aim, namely the effects of actions and predictability on memory performance, we expected differences in recall rates between self- and externally-generated sounds but given the mixed evidence we did not have a specific hypothesis about the direction of the memory effects: Based on production effect studies, we would expect better memory for the self-generated sounds, however, predictive coding theories would predict the opposite effect (i.e., higher memory performance for externally-generated sounds). Critically, these two alternative hypotheses stemming from different lines of memory research only apply when the self-generated stimuli are more predictable than the externally-generated ones. Therefore, we aimed to assess whether indeed such differences would interact with predictability or whether the mere effect of temporal proximity to an action and temporal control over a stimulus (irrespective of predictability) can also modulate memory encoding.

To tackle these questions, we manipulated the predictability (predictable vs. unpredictable sessions) and source (self- vs. externally-generated) of sounds in a paradigm, where participants had to encode the sounds and later recall them. In the predictable session, self- and externally-generated sounds were predictable in a) identity (i.e., fixed mapping between action and sound category), b) timing (all sounds were presented after a fixed delay), and c) probability of occurrence (i.e., all actions resulted in a sound). In contrast, in the unpredictable session, the self-generated sounds were unpredictable in all these aspects. Critically, in our design, externally-generated sounds were equally predictable or unpredictable in all these aspects (see Methods), allowing us to isolate the effects of motor prediction (in the predictable session) or motor proximity (in the unpredictable session) on sensory processing and memory encoding.

2. Methods

2.1. Participants

Twenty-six healthy, normal-hearing subjects, participated in the present study. Participants were typically undergraduate university students at the University of Barcelona. Data from six participants had to be excluded due to technical problems, inability to comply with the task instructions, or excessive artifacts in the EEG recording, leaving data from twenty participants (10 women, $M_{\text{age}} = 24.55$, age range: 18-43, 18 right-handed). None of them had any hearing impairments, had suffered from psychiatric disorders or had taken substances affecting the central nervous system the 48 hours prior to the experiment. All participants gave written informed consent for their participation after the nature of the study was explained to them and they were monetarily compensated (10 euros per hour).

Additional materials included a personal data questionnaire and a data protection document. The study was accepted by the Bioethics Committee of the University of Barcelona.

2.2. Experimental design

The experiment consisted of a memory task where each trial had an encoding, a retention and a retrieval phase (Figure 1). In each trial, at encoding, participants were presented with a series of sounds of different categories (animals, musical instruments, environmental sounds, and human-related sounds), that were either self-generated (motor-auditory, MA) or externally-generated (auditory-only, A). At retrieval, the same sounds, along with a set of new ones, were passively presented and participants had to indicate whether the sound was presented during encoding ('Old') or not ('New'). We manipulated predictability during the encoding phase, by employing an unpredictable and a predictable session, which were performed on separate days (2-4 days apart and order counterbalanced). Predictability for self-generated sounds was afforded by the action chosen by the participant, while predictability for the externally-generated sounds was afforded by externally-controlled visual cues. The sounds to be remembered were either fully predictable or fully unpredictable with respect to their identity (i.e., *what* sound category would be presented), their timing (i.e., *when* they would be presented after the action or cue) and probability of occurrence (i.e., *whether* they would be presented after the action or cue).

Encoding phase

In both sessions, every trial started with an encoding phase that consisted of 20 events of different types, each type of event presenting a different combination of presence or absence of actions, visual cues, and auditory stimuli (Table 1). Specifically, of the 20 events of the encoding phase of each trial, 10 included sounds (5 coinciding with an action, MA and 5 without action, A). The remaining 10 events were secondary control events (5 Motor-only (M) and 5 Visual-only (V) events in the unpredictable session, and 10 Empty (E) events in the predictable session). All event types and their roles are described in detail below. The different event types were presented in random order within each trial.

Table 1

Study design

	Predictable session <i>Fixed sound delay</i> <i>Fixed mapping</i> <i>100% contingency</i>	Unpredictable session <i>Random sound delay</i> <i>Random mapping</i> <i>50% coincidence</i>
Press events <i>"P" cue</i>	MA Press + circle fill + sound <i>5 events</i>	MA Press + circle fill + sound <i>5 events</i>
		M Press + circle fill <i>5 events</i>
Listen events <i>"L" cue</i>	A Circle fill + sound	A Circle fill + sound

	<i>5 events</i>	<i>5 events</i>
		V Circle fill <i>5 events</i>
Empty events <i>“L” cue</i>	E - <i>10 events</i>	<i>Corresponds to M and V events</i>

Note. The predictable and unpredictable conditions were performed in different sessions. In each trial, 20 events of different types were presented in randomized order. Every trial of the predictable session contained 5 Motor-auditory (MA) events, 5 Auditory (A) events and 10 empty (E) events. Every trial of the unpredictable session contained 5 MA events, 5 motor-only (M) events, 5 A events and 5 visual-only (V) events. In the predictable session, there was a fixed delay between the filling of the circle and the sound presentation, which alternated block-wise between 150 and 750 ms; and there was a fixed mapping between the button pressed, the circle filled, and the sound category played. Finally, in the predictable session, there was 100% contingency in that an action for MA and the filling of the circle for A events always generated a sound. In the unpredictable session, the delay between the filling of the circle and the sound presentation was either 150 or 750 ms, chosen randomly for every event; and the mapping between the button pressed, the circle filled and the sound category played was random. Moreover, in the unpredictable session, there was a 50% coincidence of an action or a filled circle generating a sound (for MA and A, respectively).

Each event started with the presentation of the letter “P” or “L” centered on the screen along with four unfilled circles of different colours below it (Figure 1). When the letter “P” (i.e., “PRESS”) was presented, participants had to press one of the four pre-specified buttons. Button pressing was (semi-)self-paced (within a 1-3s window) and self-chosen, similar to previous self-generation studies (e.g., Baess et al., 2011), thus giving a stronger feeling of temporal control or agency. Once participants pressed the button, one of the circles was filled immediately and the letter “P” disappeared. In the unpredictable session, the button press and colour-filling of the circle was followed by a sound only half of the times (5 unpredictable MA events), while in the other half no sound was presented (5 M events). Additionally, there was a random mapping between the button press, circle to be colour-filled, and sound category. Thus, in the unpredictable session, the sound category was randomly chosen and could not be predicted by the button pressed or the circle filled, and the probability of occurrence of the sound after the press was 50%. In contrast, in the predictable session, the button press and colour-filling of the circle was always followed by a sound (5 predictable MA events) and there was a fixed mapping between all three: For example, if the left most button was pressed, the left most circle on the screen would be colour-filled, and a human sound would be presented. Thus, in the predictable session, sound category for the MA sounds could be predicted by the button that was self-chosen and the circle that was subsequently filled, and there was a 100% probability of hearing a sound after the press.

When the letter “L” (i.e., “LISTEN”) was presented along with the four unfilled circles, participants had to remain still without pressing any button. In the unpredictable session, after a random delay of up to 3 seconds (drawn from the button press delays from the “P” trials within session), one of the circles – chosen randomly – was colour-filled. Immediately, the letter “L” disappeared and in half of the occasions a sound was presented – chosen randomly among the four possible sound categories (5

unpredictable A events), while in the other half of the occasions no sound was presented after the colour-filling of the circle (5 V events). Thus, similar to the MA sounds of the unpredictable session, sound category for the A sounds could not be predicted by the circle that was colour-filled, and the probability of occurrence of the sound after the visual cue was 50%. In contrast, in the predictable session, after the presentation of the letter “L” on the screen, a circle was not always filled. This was necessary to achieve 100% contingency between cues or actions and sounds in this session, while keeping the trials equal in length and memory load (i.e., same sounds per second ratio) as in the unpredictable session (that contained also V and M events in which no sounds were presented after the cues or actions). Thus, in the predictable session, only 5 events starting with the letter “L” within each trial were similar to the unpredictable session: the letter “L” was presented and after a random delay of up to 3 seconds (drawn from the button press delays from the “P” trials of this session) one of the circles was filled. Immediately, the letter “L” disappeared, but here a sound from the sound category corresponding to the filled circle was always presented (5 predictable A events). Thus, similar to the MA sounds of the predictable session, sound category for the A sounds could be predicted by the circle that was colour-filled and there was a 100% probability of hearing a sound after the visual cue. The remaining 10 events within each trial of the predictable session started with the letter “L”, but no circle was filled and no sound was played. Thus, the full cue-sound contingency was not disturbed by these empty events (10 E events).

With this design, we could manipulate identity predictability, by providing fixed or random mapping between button press, circle to be colour-filled, and sound category, and probability of occurrence, by having 100% or 50% contingency between button presses and sounds and colour-filled cues and sounds. We additionally manipulated temporal predictability and proximity, by introducing two delays (150 and 750 ms) between the button press and/or the colour-filling of the circle and the sound, that were intermixed in a random fashion within each trial in the unpredictable session, while they alternated block-wise in the predictable session. Thus, the button press or filling of the circle predicted the exact timing of the sound in the predictable session, while the sound timing (either 150 or 750 ms delay) could not be predicted in the unpredictable session. Finally, to make sure that the event duration was similar for all event types, for the no-sound events (i.e., M, V, and E), we presented an empty sound of the same duration as the sounds presented as A or MA. The interval between events (e.g., offset of sound or no-sound and start of the next P or L letter on the screen) was set to 350 ms. Therefore, each event consisted of the interval between letter appearance and button press or cue-filling (1–3 s), the delay between button press/cue-filling and sound (150 or 750 ms), the duration of the sound (or no sound for M, V, E events; 500 ms), and the inter-event interval (350 ms), resulting in a total event duration in the range between 2 and 4.6 s, and a total encoding phase duration of 40–92 s, depending on the above.

Retention and retrieval phases

Following the encoding phase, a short retention phase of 3 seconds followed, where participants were presented with a fixation cross on the screen (Figure 1b). Following the retention phase, the retrieval phase started (Figure 1c). A series of 20 sounds were presented passively to the participants. During the presentation of the sound (500 ms), a fixation cross was shown in the screen. Each sound was followed 350 ms after sound offset by the question O/N (i.e., Old/New?) and participants had to

perform an Old/New judgment. Half of the sounds at retrieval were new (i.e., they were not presented at the encoding phase of the trial), and the other half were old (i.e., they were presented in the preceding encoding phase). Among the old sounds, half of them had been encoded as MA, and the other half as A, and of each half with 150 and half with 750 ms delay. Response buttons were counterbalanced across participants and the response window was 1 second. Once participants responded, or the response window was over, the question disappeared, a fixation cross was displayed for 1,150 ms and the message “NEXT” was presented for 2 seconds (inter-trial interval) informing participants that the following sound was about to start.

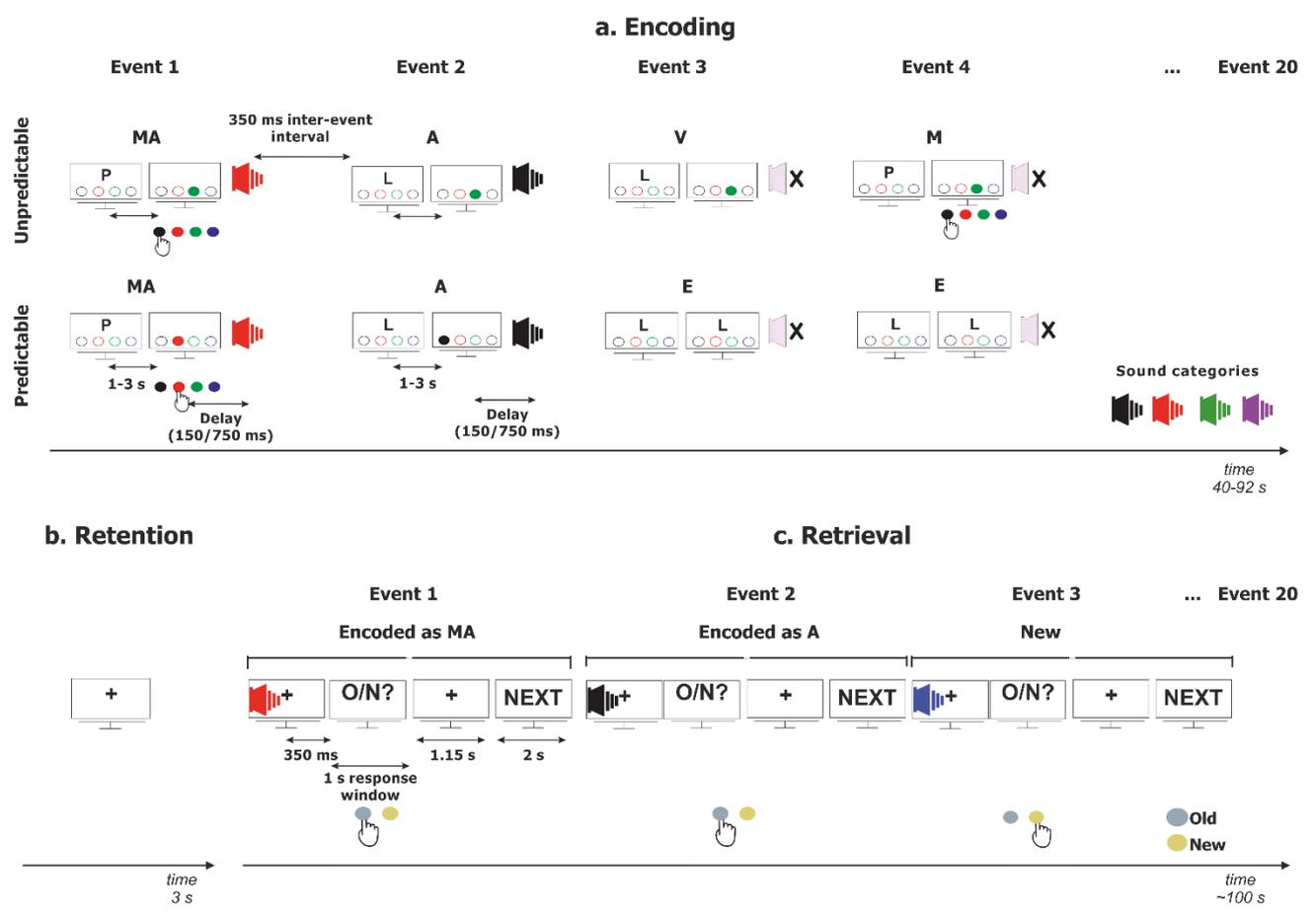


Figure 1. Schematic representation of the design. Two sessions were employed (predictable and unpredictable) that differed only in the encoding. Each trial within each session consisted of three phases: encoding, retention, and retrieval. **a)** At encoding, participants were presented with 20 events starting with the letter “P” (press) or “L” (listen), along with four unfilled circles, and they had to press one of the four predefined buttons (1–3 s) or not perform any button pressing and wait for a circle to be colour filled. In the unpredictable session, only half of the button presses and colour-filled cues resulted in unpredictable Motor-auditory (MA) and Auditory-only (A) sounds presented 150 or 750 ms (in mixed fashion) after the cue or button press. The other half of the button presses and colour-filled cues did not result in a sound (i.e., an empty sound was presented to match the duration with the sound events; Motor-only and Visual-only events; M and V). In the predictable session, fully predictable Motor-auditory (MA) and Auditory-only (A) sounds were presented 150 or 750 ms (in blocked fashion) after the button press and the colour-filling of the cue. The sound category (animal, human, environmental, or musical) was contingent on the button that participants pressed and the circle

that was colour-filled. Empty events (i.e., no colour-filling of the cue and no sound presentation; E) were added to make the predictable and unpredictable trials equal in length and had the same timing as the rest of the events. The inter-event interval (i.e., sound offset to upcoming letter) was set to 350 ms. **b)** At retention, in both predictable and unpredictable sessions, participants were presented with a fixation cross and were instructed to maintain the previously presented sounds in memory. **c)** At retrieval, a series of 20 sounds (10 old and 10 new) were presented passively to the participants. Among the old sounds, half of them had been encoded as MA, and the other half as A (either with 150 or 750 ms delay). During sound presentation, a fixation was shown on the screen. The question O/N (i.e., Old/New?) appeared 350 ms after sound offset and subjects had 1 s to reply. Following the end of the response window, a fixation cross appeared for 1,15 s, followed by the cue “NEXT” (2 s) indicating that the next sound was about to be presented.

2.3. Auditory stimuli

The auditory stimuli consisted of a pool of identifiable sounds from four categories (animals, musical instruments, environmental sounds, and human-related sounds), which allowed us to have well-established identity mappings between button type and sound category for the predictable session. These sounds were first drawn from several freely available sound repositories (the Adobe and FreeSound databases, and those of Norman-Haignere et al., 2015; Gygi & Shafiro, 2010; Hocking et al., 2013; Belin et al., 2000), and were then edited to have a 500 ms duration including 0.01 s exponential ramps and a sampling rate of 44.1 kHz, and to be played at 16 bits, mono and 75 dB intensity. Sound identifiability was assessed by the ratings of three volunteers that were presented with a series of sounds and indicated whether each sound could be easily assigned to one of the four categories. An additional option was provided to them for the sounds that they could not assign to any of the above-mentioned categories. The most identifiable sounds, as assessed by this rating, were used as the auditory stimulation in the present work. In each predictability session, we used 340 sounds at encoding (A and MA that were also presented at retrieval) and 340 additional and different sounds that were used as New at retrieval. The same sounds were used in both predictability sessions. Each sound was repeated twice as an encoding sound in each session.

2.4. Apparatus

The visual stimuli were presented on an ATI Radeon HD 2400 monitor. The auditory stimuli were presented via Sennheiser KD 380 PRO noise cancelling headphones. To record participants' button presses and behavioural responses, we used a Korg nanoPAD2. The buttons of this device do not produce any mechanical noise when pressed, and, thus, do not interfere with our auditory stimuli. The presentation of the stimuli and recording of participants' button presses and responses were controlled using MATLAB R2017a, the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007), and the Eyelink add-in toolbox for eyetracker control.

EEG activity was acquired at a 500 Hz sampling rate with a Neuroscan SynAmps RT amplifier (NeuroScan, Compumedics, Charlotte, NC, USA), using the CURRY 8 Neuroscan software. We recorded continuously with Ag/AgCl electrodes from 64 standard locations according to the 10% extension of the International 10–20 system (Chatrian et al., 1985; Oostenveld & Praamstra, 2001)

mounted in a nylon cap (Quick-Cap; Compumedics, Charlotte, NC, USA). An additional electrode was placed at the tip of the nose (serving as online reference). The vertical electrooculogram (EOG) was measured with two electrodes placed above and below the left eye, and the horizontal EOG with two electrodes placed on the outer canthi of the eyes, all of them referenced to the common reference. The ground electrode was placed at AFz. All impedances were kept below 10 k Ω during the whole recording.

Concurrently with the EEG recording, horizontal and vertical gaze position, as well as the area of the pupil, were recorded using EyeLink 1000 desktop mount (SR Research, sampling rate: 1,000 Hz; left eye recordings). Visual stimulation was presented at the center of the screen (i.e., within central vision; visual angle 2.3°) to eliminate any effects of gaze position on pupil diameter (Gagl et al., 2011).

2.5. Procedure

Participants were seated in an electrically and acoustically shielded room and were asked to place their head in a chinrest at approximately 60 cm from the screen. Eye-tracker calibration was performed at the start of the experiment. In order to familiarize themselves with the task and the button pressing, participants completed three training blocks at the start and in the middle of the experiment. During the first two training blocks participants were presented with the letter “P” and four unfilled circles and were instructed to press a button. They had to produce a uniform distribution of letter-press delays between 1 and 3 s, as well as a uniform distribution of the presses over the 4 buttons. Once they pressed, a circle was colour-filled and a sound was presented either based on a fixed button-circle-sound category mapping (predictable session) or based on a random one (unpredictable session). The delay between button press and sound was either 150 or 750 ms, presented randomly intermixed in the unpredictable session and in a blocked fashion in the predictable session. The third training block consisted of two trials that had the same structure as the experimental trials within predictability session, with the only difference being that the sounds were drawn from a different pool so that during training participants would not be exposed to the sounds used in the main experiment.

The main experiment consisted of 34 trials per predictability session, divided in eight experimental blocks (six blocks of four trials and two blocks of five trials). At the end of each block, a message appeared informing participants about the number of missed responses, the mean letter-press delay, and the percentage of pressing each one of the four buttons for this block. Participants took a break of approximately 5 minutes every 2 blocks to prevent fatigue. Each session of the experiment lasted for approximately 2 hours excluding the EEG preparation.

2.6. Data analysis

Given our hypotheses, for the EEG and pupillometric data, we focused only on the encoding data. The retrieval data was used only for the behavioural analysis.

2.6.1. Behavioural analysis

To test for differences in memory performance, we calculated the percent correct for the sounds at retrieval and we ran a 3x2 ANOVA with factors Delay (150 ms vs. 750 ms), Predictability (Predictable vs Unpredictable), and Source (Auditory vs. Motor-auditory).

2.6.2. EEG preprocessing

EEG data was analyzed with EEGLAB (Delorme & Makeig, 2004) and Fieldtrip (Oostenveld et al., 2011) and plotted with EEProbe (ANT Neuro). Data were high-pass filtered (0.5 Hz high-pass, Kaiser window, Kaiser β 5.653, filter order 1812), manually inspected so as to reject atypical artifacts and identify malfunctioning electrodes, and corrected for eye movements with Independent Component Analysis, using the compiled version of runica (binica) that uses the logistic infomax ICA algorithm (Onton & Makeig, 2006). Components capturing eye movement artifacts were rejected by visual inspection and the remaining components were then projected back into electrode space. Data was then low-pass filtered (30 Hz low-pass, Kaiser window, Kaiser β 5.653, filter order 1812), remaining artifacts were rejected by applying a 75 μ V maximal signal-change per epoch threshold, and malfunctioning electrodes were interpolated (spherical interpolation).

We were interested in comparing auditory responses between externally generated (A) and self-generated (MA) sounds at encoding as a function of predictability, separately for each delay condition. Thus, to analyze the ERPs, we defined epochs time-locked to the sound presentation. However, for the events with a 150 ms delay between press or cue and sound, these epochs included not only auditory responses elicited by the MA and A sounds, but also visual responses elicited by the colour-filling of the circle (in both MA and A events), and sensorimotor responses related to the planning and execution of the button press (in MA events only), both occurring 150 ms prior to the sound. Typically, in self-generation studies, to correct for motor activity differences, the comparison between self- and externally-generated sounds is done after subtracting motor-only ERPs obtained in a separate block, from the motor-auditory ERPs (for a review see Schröger et al., 2015). Here, motor-only events were included only in the unpredictable session and could have been used to perform this correction. However, motor activity differs as a function of predictability context (Neszmelyi & Horváth, 2017), therefore using these events to perform motor correction in the predictable session would have contaminated the predictability contrasts of interest. Therefore, instead, to correct the 150 ms delay events, we used the data from the 750 ms delay events within each session. In these events, the visual and sensorimotor responses took place 750 ms prior to sound onset, and the signal returned to baseline prior to onset of the auditory responses (see Supplementary materials, figure S1). Thus, for each predictability session, to extract the auditory responses from the 150 ms delay events, first we created subjects' averages locked to 150 ms delay sounds using an epoching window of -250 to $+550$. Second, we created control averages locked to the button press and/or colour-filling of the circle in the 750 ms delay events, with an epoching window -100 to $+700$ ms. Thus, both these averages had an overall 800 ms duration and time point 0 in the control averages aligned with time point -150 ms in the 150 ms delay event averages, both coinciding with the button press and/or colour filling of the circle. No baseline correction was applied in either average. Finally, we subtracted the control averages from the 150 ms delay events averages. In each predictability session, 150 ms delay MA events were corrected using 750 ms delay MA events (thus subtracting both visual and sensorimotor ERPs) and 150 ms delay A events were corrected using 750 ms delay A events (thus subtracting visual ERPs), always from the

corresponding predictability session. To extract the auditory responses from the 750 ms delay events, we created epochs locked to the 750 ms delay sounds in the same window as for the 150 ms events (i.e., -250 to +550), but no correction was required with the long delay. Finally, we calculated the grand averages for each event of interest (corrected or uncorrected A and MA sounds at encoding for each predictability and delay condition). After trial rejection, each individual average for each of the event types at encoding was obtained averaging a mean of 69.88 epochs ($SD = 14.97$).

2.6.3. Targeted analysis of auditory ERP components of interest

To assess the self-generation effects on auditory sensory responses at encoding as a function of predictability we compared responses to MA sounds with the responses elicited by the A sounds, separately for each Delay condition. All statistical contrasts between self- and externally-generated sounds were performed on the corrected waveforms (see previous section, *EEG preprocessing*). Specifically, we examined the following auditory responses: N1 and P2 at vertex, N1 at the mastoids ($N1_{mast}$), and the N1 subcomponents Na and Tb at temporal sites. The windows were defined after visual inspection of the data by locating the highest negative or positive (depending on the component of interest) peak in the usual latencies for each component as reported by previous work (SanMiguel et al., 2013). Time windows for N1, $N1_{mast}$, P2, Na, and Tb were defined on the grand-average waveforms of the auditory-only sounds in the predictable session. Na and Tb were identified as the first and second negative peaks, respectively, identifiable after sound onset on electrodes T7 and T8, as recommended by Tonnquist-Uhlen et al. (2003). N1 and P2 were identified as the negative and positive peaks occurring in the window ~70 to 150 ms, and ~150 to 250 ms after stimulus onset on Cz, respectively. Conversely, $N1_{mast}$ was identified as the positive peak in the window ~70 to 150 ms after stimulus onset on electrodes M1 and M2. Given variations in peak latencies across the conditions, the width of the windows was defined such that it could capture the peak of the MA sound waveform as well, and it was proportional to the width of the component. Thus, the time windows were centered on the identified peaks and spanned ± 12 (N1), 15 ($N1_{mast}$), 25 (P2), 10 (Na), and 15 (Tb) ms. For each Delay condition, we performed a 2x2 ANOVA with factors Source (A vs. MA) and Predictability (Predictable vs. Unpredictable) on the mean amplitude measured in the defined windows and electrodes. For the components identified in two electrodes (i.e., Na, Tb, and $N1_{mast}$) an additional factor Laterality was introduced in the ANOVAs (i.e., T7 vs. T8 and M1 vs. M2).

Moreover, we aimed to address the possible three-way interactions between Source, Predictability and Delay. Therefore, we calculated the self-generation (i.e., Source) effects (A - MA difference waves) and assessed how they were modulated by Predictability and Delay. Additionally, we calculated the predictability effects (Unpredictable - Predictable difference waves) and assessed how they were modulated by Source and Delay. To this end, we performed 2x2 ANOVAs (Predictability x Delay on self-generation effects and Source x Delay on predictability effects) on our components of interest (Na, Tb, N1, $N1_{mast}$, P2). For components identified in more than one electrode, an additional factor of Laterality was introduced in the model.

Finally, in addition to the auditory ERP analyses, our design partly offered us the possibility to assess whether the effects of self-generation and predictability interact on visual ERPs. This question was out of the scope of our experimental paradigm, however, given that the discrepancy between the effects of

action on auditory vs. visual responses has been discussed in previous work (Shimaoka et al., 2018; Parker et al., 2020), we decided to include this preliminary analysis in our Supplementary Material.

2.6.4. *Mass univariate ERP analysis*

In addition to specifically analyzing targeted auditory ERP components, we opted for a data-driven analysis in order to explore the possibility of further effects of self-generation and predictability on sensory processing during encoding beyond specific modulations of targeted components. To this end, we performed a mass univariate analysis across time-points and electrodes, using cluster-based non-parametric correction to account for multiple comparisons. Using the entire epoch of our data (i.e., -250 +550), we tested for main effects of Source (A vs. MA) and Predictability (predictable vs. unpredictable) as well as for possible interactions between the two factors. Specifically, for the interaction, the cluster-permutation test was performed comparing the difference waves ([A predictable – MA predictable] vs [A unpredictable – MA unpredictable]). Each cluster was constituted by the samples that consecutively passed a p value of 0.05. The cluster statistics was chosen as the sum of the paired t values of all the samples in the cluster. For the significant interactions, we performed post-hoc comparisons within the window of significance. For each statistical test, this procedure was performed by randomly switching labels of individual observations between these paired sets of values. We repeated this procedure 10,000 times and computed the difference between the group means on each permutation. The obtained p value was the fraction of permutations that exceeded the observed difference between the means (i.e., two-sided dependent samples tests).

2.6.5. *Pupillometry analysis*

Pupil preprocessing followed the procedure reported in Paraskevoudi & SanMiguel 2022. After interpolating missing data and blinks, the interpolated pupil time series were bandpass filtered using a 0.05–4 Hz third-order Butterworth filter. First, we estimated the effect of blinks and saccades on the pupil response through deconvolution and removed these responses from the data using linear regression following a procedure detailed in previous work (Knapen et al., 2016; Urai et al., 2017). The residual bandpass filtered pupil time series was used for the evoked analyses (van Slooten et al., 2019). After zscoring per trial, we epoched the data, baseline corrected each trial by subtracting the mean pupil diameter 500 ms before onset of the event and resampled to 100 Hz. We locked the data to the colour-filling of the circle (epoching window -0.5 to 2 s post-event) that coincided with the button press in MA events, since we were interested in the effect of action on the pupil response.

For each participant, we first obtained the average evoked response for the main events of interest. Specifically, we obtained the averages for the A and MA events at encoding locked to the colour-filling of the cue, separately for each predictability and delay condition. We used non-parametric permutation statistics to test for the group-level significance of the individual averages, following the same procedure as for the EEG data (see section 2.6.4. *Mass univariate ERP analysis*) except that for pupillometry the clusters were defined based on the time dimension only. For each sound delay, we aimed to test for possible main effects of Source (A vs. MA events) and Predictability (Predictable vs. Unpredictable sessions), as well as for possible interactions between the two. For the main effects of Source and Predictability, the permutation statistics were performed between A and MA events

irrespective of the predictability session and between predictable and unpredictable events irrespective of their source (A or MA). To test for possible interactions, the cluster-permutation test was performed comparing the difference waves (e.g., [A predictable – MA predictable] and [A unpredictable – MA unpredictable]).

2.6.6. Correlations

We hypothesized that the electrophysiological and neuromodulatory effects at encoding (i.e., sensory suppression and pupil dilation for MA events) might be driving any memory encoding differences between A and MA sounds, and that neuromodulation might be behind the suppression of ERP responses to MA sounds. To assess these relationships, we tested for possible correlations between the behavioural, electrophysiological and neuromodulatory (i.e., pupil diameter) effects of actions, separately for each delay condition and predictability session. Only those differences between MA and A events that were found to be significant in the previous analyses were introduced in the correlation analyses. For all the behavioural and the electrophysiological effects, we first calculated the difference by subtracting the MA from A values (i.e., difference in memory performance and ERP amplitude for each component of interest between A and MA, separately for each delay condition and predictability session). Regarding the ERP components identified in two electrodes (e.g., Na, Tb, and N1_{mast}), we calculated the mean amplitude across the two (T7/T8 and M1/M2, respectively). For the pupil data, we used the peak of the difference wave between A and MA events at encoding for each condition separately (150/750 ms and predictable/unpredictable session). We then submitted these values to a Pearson correlation coefficient to test for correlations between a) self-generation effects on ERPs at encoding and on memory performance, b) the neuromodulatory effects of actions at encoding and self-generation effects on memory performance, and c) the self-generation effects on the ERPs and the neuromodulatory effects of actions at encoding. In all correlations, for the ERPs, this results in larger attenuation effects being reflected as more negative values for the negative (N1, Na, Tb) components, and more positive values for the positive components (N1_{mast}, P2). Conversely, for the pupil and the behavioural data, the more negative the value, the larger the pupil diameter and the worse the memory performance for MA sounds.

We complemented these analyses with corresponding Bayesian correlation analyses. For all Bayesian correlations, the Bayes factor (BF₁₀) for the alternative hypothesis (i.e., correlation is not equal to zero and can be either positive or negative) was calculated (using the function *correlationBF* of the *BayesFactor* package in R). The Bayes factor, and its interpretation, was based on Jeffreys' (1961) test for linear correlation. Specifically, Bayes factors below 1 are thought to provide evidence against the alternative hypothesis, while weak, moderate, strong, and very strong evidence for the alternative hypothesis is provided by Bayes factors in the ranges 1 – 3, 3 – 10, 10 – 30, 30 – 100, respectively (Jeffreys, 1961). Bayes factors above 100 provide extreme evidence in support of the alternative hypothesis.

3. Results

All statistical analyses were performed using R (version 3.6.0). For all the post-hoc comparisons following significant ANOVAs, we used the Bonferroni correction.

3.1. Behavioural performance

For the analysis of the behavioural data, we calculated the percent correct (i.e., memory performance) for sounds that were encoded as MA or A, separately for each Predictability session and Delay condition. The 3x2 ANOVA (Source: A vs. MA, Predictability: Predictable vs. Unpredictable, Delay: 150 vs 750 ms) showed only a main effect of Source, $F(1, 19) = 8.98, p = .007, \eta_p^2 = .32$, with better performance for sounds encoded as MA compared to those encoded as A ($M_{MA} = .72, M_A = .69, SD_{MA} = .12, SD_A = .14$). The rest of the effects did not reach significance (all $ps > .05$, see Figure 2).

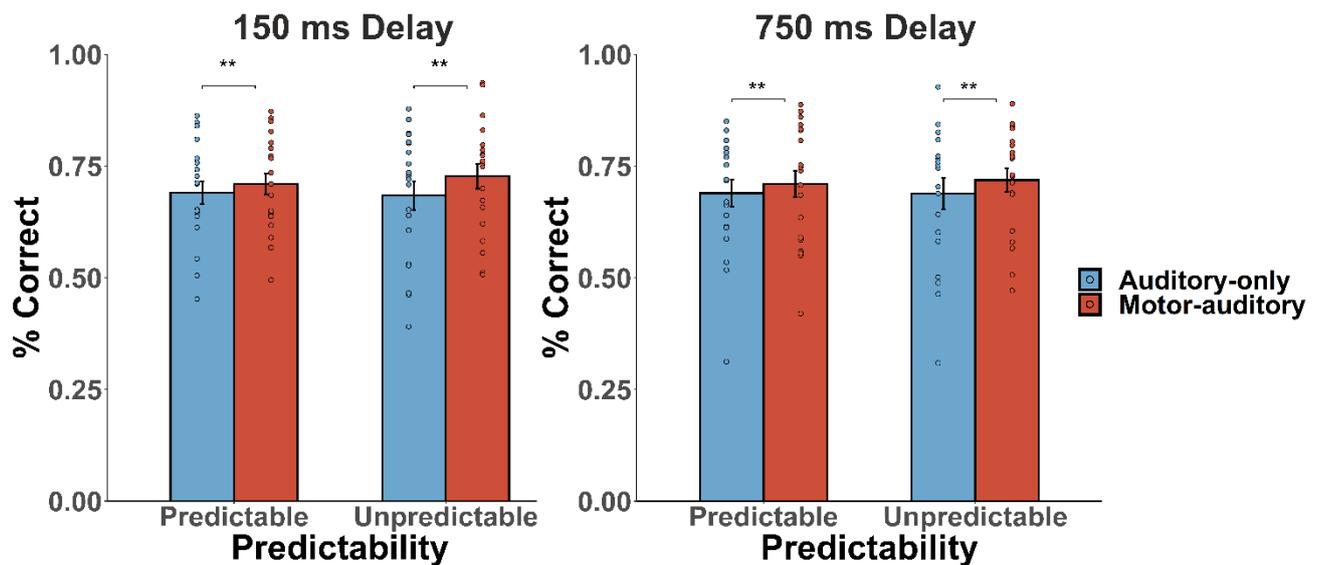


Figure 2. Summary of the behavioural findings showing a significant main effect of Source, with better performance for sounds that were encoded as MA compared to those encoded as A, irrespective of predictability and delay. Error bars depict the standard error of the mean. Individual data points are shown for each predictability, source, and delay condition.

3.2. Electrophysiological responses

3.2.1. Auditory responses as a function of source and predictability for each delay

Separately for each sound delay, we assessed the effects of Source, Predictability and their interactions across the whole epoch and all scalp locations, and also in a more restrictive approach for specific auditory components of interest. Figures 3 and 4 provide an overview of the results of the targeted component analysis and the mass univariate analysis, respectively. The time windows defined for each component were the following: Na 70–90 ms, Tb 120–150 ms, N1 92–116 ms, N1_{mast} 88–118 ms, P2 180–230 ms. Table 2 summarizes the main effects and interactions obtained in the component analyses presented in this section. The mean amplitudes per component and condition are provided in the supplementary material (Tables S1-S3).

— Auditory-only Predictable Auditory-only Unpredictable
 — Motor-Auditory Predictable Motor-Auditory Unpredictable

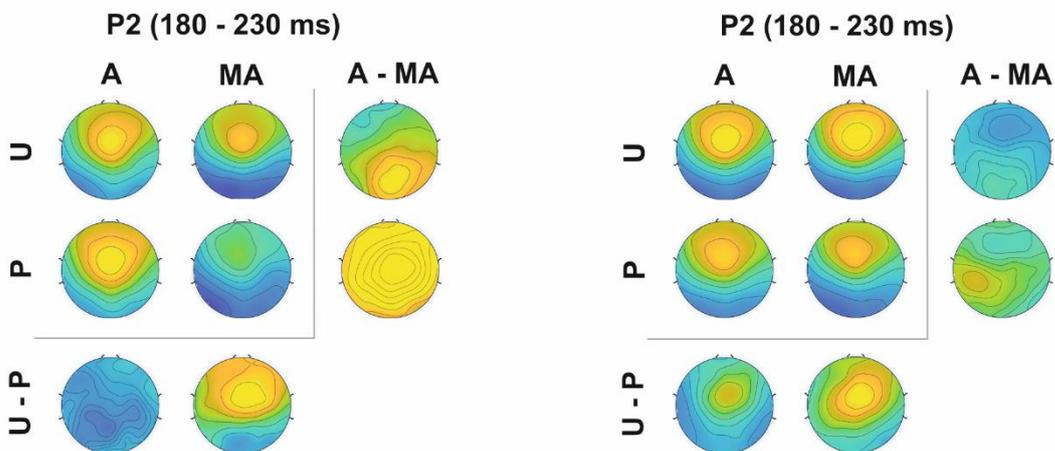
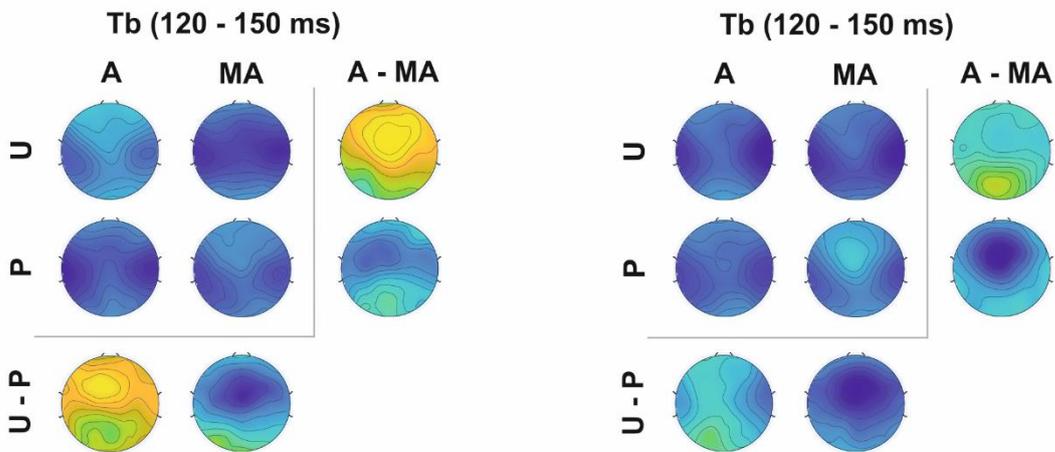
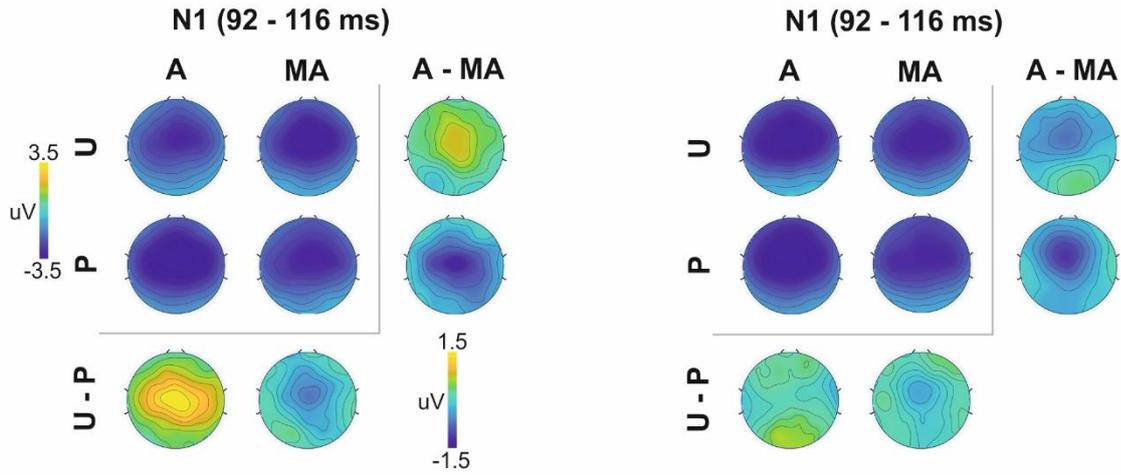
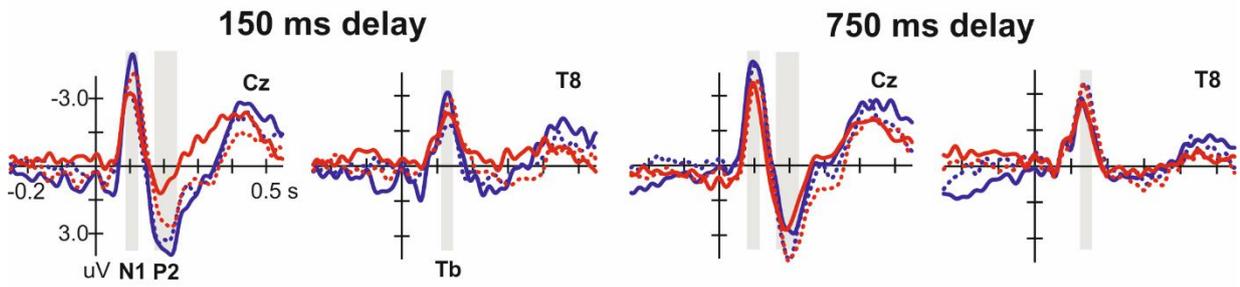


Figure 3. Summary of results yielded by the targeted component analysis. Top row: Group-average event-related potentials on Cz and T8 for sounds presented 150 ms (left) and 750 ms (right) after the button press and/or colour-filling of the cue, separately for the motor-auditory (red) and auditory-only (blue) sounds, in the predictable (lines) and unpredictable (dashed lines) sessions. The time windows analyzed for the N1, Tb and P2 components are shown in gray shading. Below, scalp potential maps are shown separately for the 150 ms delay sounds (left) and 750 ms delay sounds (right) for each of the three components. For each component, separate scalp maps are depicted for the auditory-only (A, first columns), motor-auditory (MA, second columns), unpredictable (U, first rows) and predictable (P, second rows) conditions. The scalp distribution of the self-generation effects (A – MA difference waves) and the predictability effects (U – P difference waves) are shown in the third columns and third rows, respectively.

150 ms delay

First, we analyzed the responses to sounds presented 150 ms after the press and/or colour-filling of the circle. Main effects of Source were found for the P2 component only (Table 2), which was attenuated for MA sounds compared to A sounds. This P2 effect was encompassed within a more extensive significant cluster, reflecting an increased positivity in A compared to MA sounds across central-parietal electrodes between 152 – 312 ms post-stimulus ($p = .004$). A second significant cluster reflected an enhanced frontal negativity for A compared to MA sounds between 396 and 548 ms ($p = .011$). We found no significant main effects of predictability neither in the mass-univariate analysis nor in the targeted-component analysis. However, source and predictability had interactive effects on the vertex N1, the Tb and the P2 components (Table 2). Specifically, only when both self- and externally-generated sounds were fully predictable, self-generation suppressed vertex N1 ($t(19) = -1.91, p = .036, d = -0.43$; one-sided t-test with the hypothesis of observing an attenuation based on the well-established finding of attenuation for predictable self-generated sounds), P2 ($t(19) = 4.85, p < .001, d = 1.08$; one-sided t-test) and Tb ($t(39) = -2.01, p = .026, d = -0.32$; one-sided t-test); while when both sounds were equally unpredictable the opposite effect was found for the Tb component (i.e., increased Tb for MA compared to A sounds; $t(39) = 3.66, p < .001, d = -0.71$; two-sided t-test). Moreover, predictability enhanced vertex N1 ($t(19) = -2.92, p = .009, d = -0.65$; two-sided t-test) and Tb ($t(39) = -4.48, p < .001, d = -0.71$; two-sided t-test) responses to externally-generated sounds only, while it suppressed the P2 elicited by self-generated sounds ($t(19) = -2.80, p = .011, d = -0.6$; two-sided t-test). These interactive effects on the N1 and Tb components were not captured by any significant clusters in the mass-univariate analysis. However, we found a significant cluster for the interaction between Source and Predictability spanning between 196 – 390 ms ($p = .001$). This cluster thus partly encompassed the P2 component, and post-hoc tests in the significant window of interaction yielded similar results as for the targeted component analysis on the P2: a decreased central positivity for MA compared to A sounds in the predictable session only between 196 – 380 ms ($p = .001$), and a decreased fronto-central positivity in predictable compared to unpredictable MA sounds only between 198 – 390 ms ($p < .001$). The mass-univariate analysis also revealed interactive effects of Source and Predictability in the pre-stimulus interval, between -178 and -16 ms ($p = .002$). Post-hoc tests revealed a more negative response over central-parietal locations for MA compared to A sounds in the predictable session only between -178 and -20 ms pre-stimulus ($p = .001$). Further, only MA sounds yielded a more negative signal in the predictable compared to the unpredictable session on occipital

electrodes between -150 and -22 ms pre-stimulus ($p = .003$). These pre-stimulus effects probably reflected interactive effects of Source and Predictability on the visual responses elicited by the colour-filling of the circle immediately after the button press (i.e., 150 ms prior to sound onset, see Supplementary Materials, section *Effects of self-generation and predictability on visual responses*).

In sum, we found that source and predictability had interactive effects on the vertex N1, the Tb and the P2 components, while they did not affect the Na and the N1_{mast}. Specifically, self-generation suppressed vertex N1, P2 and Tb only when both self- and externally-generated sounds were fully predictable, while the opposite effect was found for the Tb component when both sounds were equally unpredictable. Predictability enhanced vertex N1 and Tb responses to externally-generated sounds only, while it suppressed P2 elicited by self-generated sounds.

750 ms delay

Next, we analyzed the responses to sounds presented 750 ms after the press and/or colour-filling of the circle. Overall, here, Source and Predictability independently modulated different auditory components and did not interact (see Table 2). In the targeted component analysis, significant effects of Source were found only for the vertex N1, which was suppressed in MA compared to A sounds (Table 2). This effect was not captured by the mass univariate analysis which however yielded additional main effects of Source in a later time-window: A sounds elicited more negative responses over central-parietal areas between 456 and 522 ms post-stimulus than MA sounds ($p = .04$). Regarding the main effects of Predictability, the targeted-component analysis revealed only suppressed P2 and Tb responses in predictable compared to unpredictable sounds (Table 2). Again, the Tb effects were not captured by the mass univariate analysis, but the P2 effects appeared to be part of a more extensive cluster showing more positive responses for unpredictable compared to predictable sounds over central sensors between 184 and 386 ms post-stimulus ($p = .01$).

The mass univariate analysis again revealed additional effects in the pre-stimulus window. Specifically, a main effect of Source ($p = .01$) was found between -250 and -110 ms pre-stimulus. However, a significant interaction between Source and Predictability was found covering this same window (negative cluster, -250 to -46 ms pre-stimulus, $p < .001$). The post-hoc comparisons in the window of interaction showed that the Source effects were only significant in the predictable session, where an enhanced negativity was found for predictable MA compared to predictable A sounds between -250 and -46 ms pre-stimulus ($p < .001$) initially observable over occipito-parietal areas and later over central-parietal sensors on the right hemisphere. Further post-hoc tests assessing the effects of predictability separately for each source revealed a significant cluster for A sounds only between -250 and -118 ms pre-stimulus ($p > .05$), reflecting more negative waveforms over central and parietal areas for unpredictable A compared to predictable A sounds.

In sum, we found that with a longer delay, source and predictability independently modulated different components and did not interact in the post-stimulus responses. Specifically, self-generation suppressed vertex N1, while predictability suppressed P2 and Tb.

Table 2

Summary of the significant main effects (Source, Predictability) and interactions separately for each delay and ERP component.

ERPs	150 ms delay
N1	Source x Predictability: $F(1, 19) = 7.03, p = .016, \eta_p^2 = .27$
P2	Source: $F(1, 19) = 18.32, p < .001, \eta_p^2 = .49$ Source x Predictability: $F(1, 19) = 9.38, p = .006, \eta_p^2 = .33$
N1 _{mast}	Laterality: $F(1, 19) = 12.08, p = .003, \eta_p^2 = .39$
Na	Non-significant.
Tb	Source x Predictability: $F(1, 19) = 7.67, p = .012, \eta_p^2 = .29$
750 ms delay	
N1	Source: $F(1, 19) = 17.36, p < .001, \eta_p^2 = .48$
P2	Predictability: $F(1, 19) = 7.05, p = .016, \eta_p^2 = .27$
N1 _{mast}	Laterality: $F(1, 19) = 11.11, p = .003, \eta_p^2 = .37$
Na	Laterality: $F(1, 19) = 6.58, p = .019, \eta_p^2 = .26$
Tb	Predictability: $F(1, 19) = 6.46, p = .02, \eta_p^2 = .25$

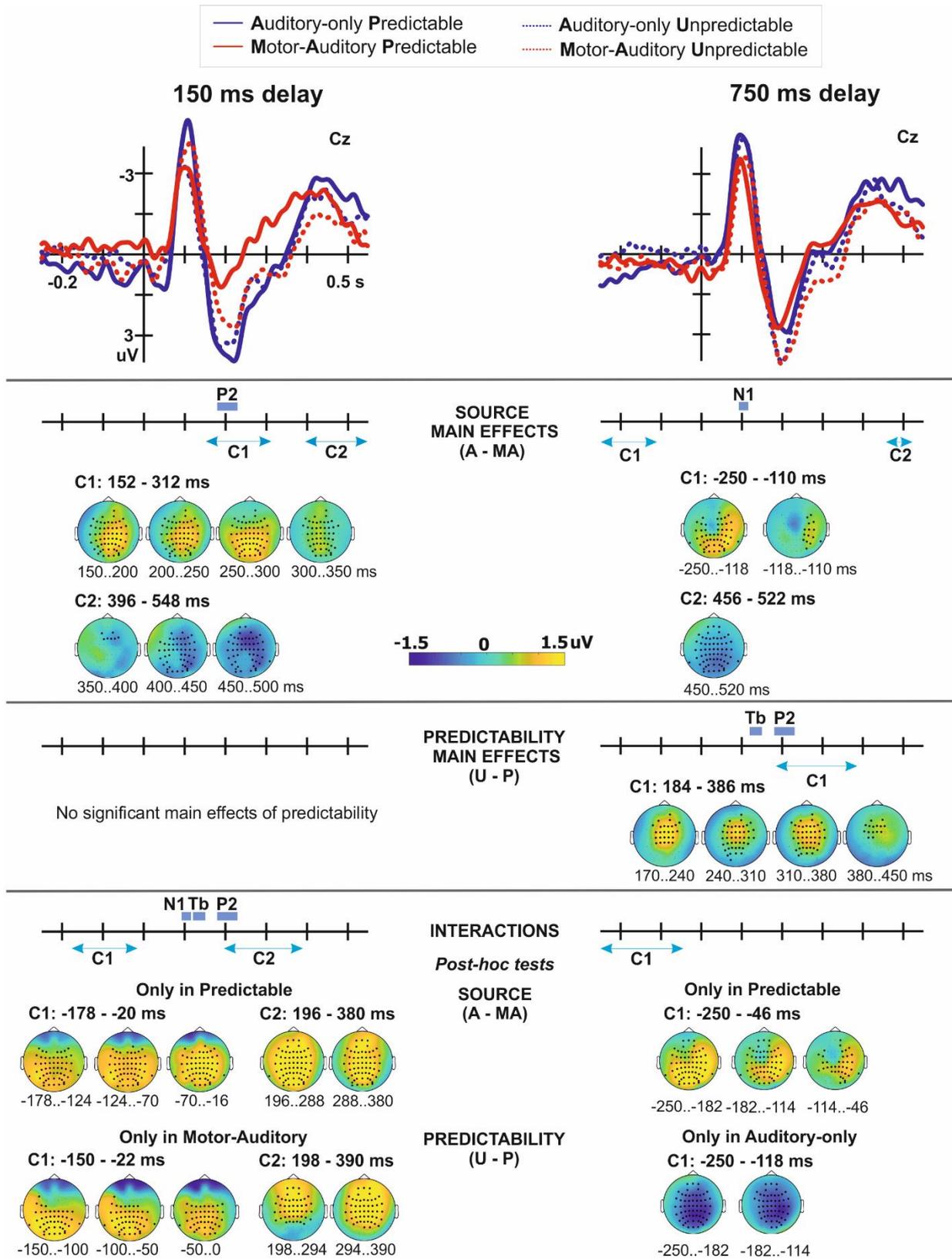


Figure 4. Summary of electrophysiological findings. Group-average event-related potentials on Cz across 20 participants for sounds presented 150 ms (left) and 750 ms (right) after the button press and/or colour-filling of the cue, separately for the motor-auditory (red) and auditory-only (blue) sounds, in the predictable (lines) and unpredictable (dashed lines) sessions. The findings obtained by

the targeted analysis of ERP components of interest and mass univariate ERP analyses are organized by effect (e.g., main effect of source, main effect of predictability, interaction between source and predictability). For each effect, blue boxes above the timeline represent the components of interest that were found to show significant differences. Significant clusters are marked with blue arrows below the timeline, and scalp potential maps of the effect are depicted for each cluster. Electrodes belonging to the significant cluster are highlighted with black dots. For the interaction effects, the scalp plots of the significant clusters yielded by the post-hoc analyses are shown. C: Cluster; MA: Motor-auditory; A: Auditory-only; P: Predictable; U: Unpredictable.

3.2.2. Comparison of the self-generation and predictability effects across delays

We also addressed the possible three-way interactions between Source, Predictability and Delay (see Methods). Specifically, we performed 2x2 ANOVAs (Predictability x Delay on self-generation effects and Source x Delay on predictability effects) on the same targeted components: Na: 70–90 ms, Tb: 120–150 ms, N1: 92–116 ms, N1_{mast}: 88–118 ms, P2: 180–230 ms). For components identified in more than one electrode, an additional factor of Laterality was introduced in the model. Figure 5 shows N1 and P2 amplitudes separately for each source, predictability and delay (a); and the effects of self-generation (b) and predictability (c) represented as the difference A – MA and Unpredictable – Predictable, respectively, in violin plots. A summary of the main effects and interactions obtained in the analyses of this section is reported in Table 3.

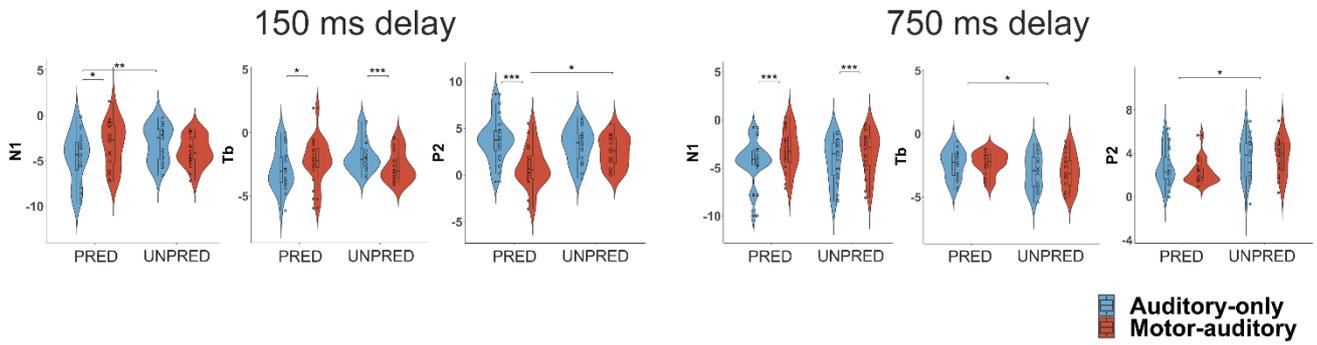
Self-generation effects as a function of predictability and delay

Delay and Predictability interactively modulated self-generation effects on N1, Tb, and P2. The post-hoc comparisons for all three components showed that within the 150 ms delay condition, action-driven suppression was significantly more pronounced for the predictable vs. the unpredictable session (N1: $t(19) = -2.72, p = .014, d = -0.61$, Tb: $t(19) = -3.67, p < .001, d = -0.58$, P2: $t(19) = 3.80, p = .001, d = 0.85$) while N1-, P2-, and Tb-suppression for self-generated sounds did not vary as a function of predictability in the 750 ms delay (all $ps > .05$). Additionally, the magnitude of the suppression of these components for self-generated sounds differed between the delays as a function of predictability: For N1 and Tb, motor-driven suppression appeared significantly larger in the 750 ms compared to the 150 ms delay, but only in the unpredictable session ($t(19) = 3.90, p < .001, d = 0.87$ and $t(19) = 4.04, p < .001, d = 0.64$) in which for the 150 ms delay we found a significant Tb-enhancement and a trend towards N1-enhancement rather than suppression for self-generated sounds (see above). In contrast, for P2 the motor-driven suppression effects were larger in the 150 compared to the 750 ms delay only in the predictable session ($t(19) = 3.54, p = .002, d = 0.79$), but not in the unpredictable one, in which motor-driven P2 attenuation was not found in either delay (see above). As for the N1_{mast} and Na, we did not find significant effects. In sum, all these interactions confirm that self-generation effects were modulated by predictability at short delays only, with self-generation effects being largest for all components when sounds were predictable and presented at a 150 ms delay, while when sounds were unpredictable and presented at 150 ms delay, there was a general trend towards sensory enhancement rather than suppression.

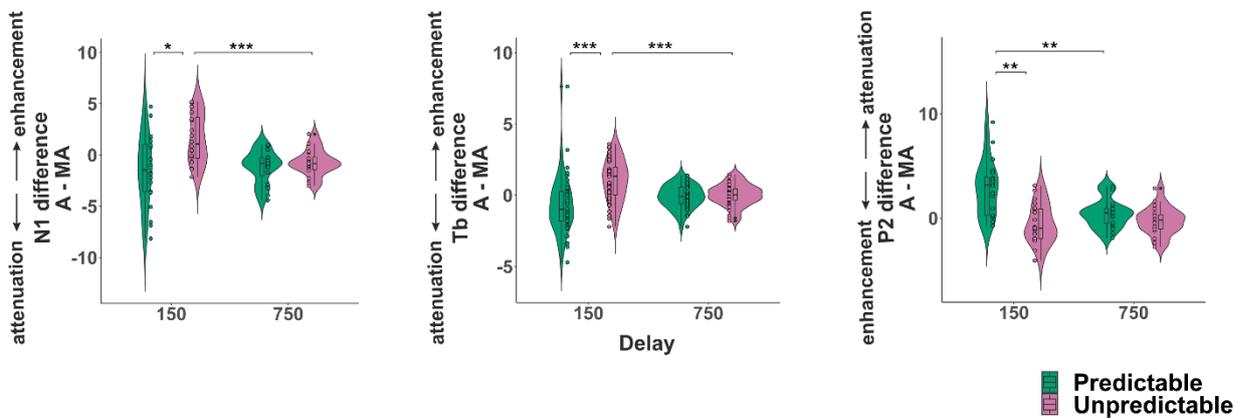
Predictability effects as a function of source and delay

The source of the predictability (a self-chosen action in MA sounds or visual cue in A sounds) modulated predictability effects on N1, P2 and Tb (see Table 3 and Figure 5). We also found a main effect of Delay on predictability effects for the Tb. However, the effects of Source and Delay on predictability interacted for N1 and Tb, while they did not interact for P2 (Table 3). The main effect of Source on P2 reflected larger P2-suppression for predictable compared to unpredictable sounds when predictability was given by a self-chosen action (i.e., MA sounds), compared to when predictability was afforded by a visual cue (i.e., A sounds). For the N1 and Tb, however, predictability effects were modulated by source in the 150 ms delay only, where predictability afforded by a visual cue enhanced responses, while in comparison predictability afforded by the action had no significant effects (N1: $t(19) = 2.65$, $p = .016$, $d = 0.59$; Tb: $t(39) = 3.53$, $p = .001$, $d = 0.56$). Conversely, in the 750 ms delay, predictability effects on the N1 and Tb were not modulated by Source (both $ps > .05$). Recall that at the 750 ms delay, predictability effects consisted of a significant Tb suppression for predictable sounds (main effect over A and MA of Predictability, see above) while predictability did not show any significant effects on N1 (see Table 2 above). Additionally, on the N1 and Tb components, the effects of predictability afforded by visual cues (A sounds) were significantly different at the short compared to the longer cue-sound delay (N1: $t(19) = 2.44$, $p = .025$, $d = 0.55$; Tb: $t(39) = 5.87$, $p < .001$, $d = 0.93$). For both the N1 and Tb, predictability enhanced responses for A sounds at the 150 ms Delay compared to the responses for A sounds 750 ms Delay. In contrast, predictability effects afforded by motor actions (MA sounds) on the N1 and Tb components (which were in general not significant except for the main effect of predictability on Tb at the 750 ms Delay, see Table 2 above) were not modulated as a function of Delay (both $ps > .05$).

a. N1, Tb, and P2 responses by source and predictability



b. Self-generation effects across predictability and delay



c. Predictability effects across source and delay

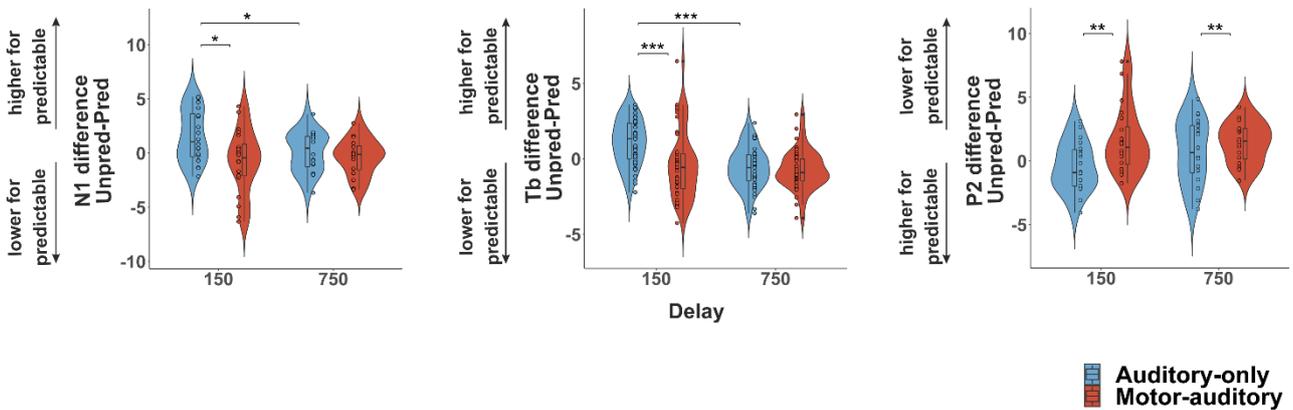


Figure 5. Violin plots representing N1, Tb, and P2 amplitudes (a), and self-generation (b) and predictability (c) effects on the N1, Tb, and P2 components. The width of each curve corresponds with the approximate frequency of data points in each region. In the middle of each density curve is a small box plot, with the rectangle showing the ends of the first and third quartiles and central dot the median. Asterisks (*, **, ***) indicate *p* values smaller than .05, .01, and .001, respectively. **a)** N1, Tb, and P2 responses are shown separately for the conditions of interest, e.g., for predictable and unpredictable A and MA sounds, for the 150 ms (left) and the 750 ms (right) delay conditions. **b)** Self-generation effects (A – MA) as a function of predictability and delay: Larger attenuation effects correspond to more negative values in the y-axis for the N1 and Tb, and more positive values for the

P2. **c)** Predictability effects (Unpredictable – Predictable) as a function of source and delay: Enhancing effects of predictability are represented with more positive values on the y-axis for the N1 and Tb components, and more negative values for the P2 component. For all plots, Tb is presented as the mean amplitude between T7 and T8.

Table 3

Summary of the significant main effects and interactions for each ERP component when comparing the self-generation and predictability effects across delays.

ERPs	Self-generation effects as a function of predictability and delay
N1	Predictability: $F(1, 19) = 7.12, p = .015, \eta_p^2 = .27$ Delay: $F(1, 19) = 6.61, p = .019, \eta_p^2 = .26$ Predictability x Delay: $F(1, 19) = 5.53, p = .030, \eta_p^2 = .23$
P2	Predictability: $F(1, 19) = 19.11, p < .001, \eta_p^2 = .50$ Delay: $F(1, 19) = 12.66, p = .002, \eta_p^2 = .40$ Predictability x Delay: $F(1, 19) = 6.64, p = .018, \eta_p^2 = .26$
N1 _{mast}	Non-significant.
Na	
Tb	Delay: $F(1, 19) = 6.18, p = .022, \eta_p^2 = .25$ Predictability x Delay: $F(1, 19) = 9.00, p = .007, \eta_p^2 = .32$
Predictability effects as a function of source and delay	
N1	Source: $F(1, 19) = 6.46, p = .020, \eta_p^2 = .25$ Source x Delay: $F(1, 19) = 4.52, p = .047, \eta_p^2 = .19$
P2	Source: $F(1, 19) = 11.36, p = .003, \eta_p^2 = .37$
N1 _{mast}	Non-significant.
Na	
Tb	Source: $F(1, 19) = 5.86, p = .026, \eta_p^2 = .24$ Delay: $F(1, 19) = 10.14, p = .005, \eta_p^2 = .35$ Source x Delay: $F(1, 19) = 8.39, p = .009, \eta_p^2 = .31$

In sum, these interactions show that predictability effects were modulated by Source and Delay, but in different ways for the N1 and Tb components on the one hand and the P2 on the other. For the N1 and Tb, again the results confirm that Source and Predictability interact at the 150 ms Delay only, where visually-driven predictability, but not motor-driven predictability enhanced these components. The findings are more inconclusive regarding the P2 component, since the interaction reported in this section (i.e., Source x Delay on predictability effects for P2) indicates that motor-driven predictability led to higher P2-suppression than visually-driven predictability irrespective of the delay; however, a significant interaction between source and predictability was only found in the 150 ms delay and not in the 750 ms delay when analyzed separately (see above). All in all, the data shows a general tendency towards suppressed responses, especially the P2, for more predictable sounds that were self-generated and/or presented at the longer delay; whereas there is a generalized tendency towards enhanced responses, especially the N1 and Tb, for more predictable sounds that were externally-generated and presented at the short delay.

3.3. Pupil responses

Cluster-based permutation statistics were used to test for possible differences in pupil diameter between the conditions of interest (Figure 6). We locked the data to the button press and/or the colour-filling of the circle (for the motor-auditory events and for the auditory-only events, respectively) since we were mainly interested in examining the motor-driven modulation of the pupil response, rather than the response to the sound (cf. Paraskevoudi & SanMiguel, 2022). To this end, we tested for possible main effects of Source (A vs. MA) and Predictability (Unpredictable vs. Predictable), as well as for interactions between these two on the pupil responses, separately for each delay.

150 ms delay

We obtained a significant main effect of Source starting before the button press and/or colour-filling of the cue (window: -490 to 1,630 ms, $p < .05$), showing a remarkable increase of pupil diameter after button press. We also obtained a significant main effect of Predictability (window: -200 to 760 ms, $p < .05$), with larger pupil diameter for predictable vs unpredictable events. However, this effect seems to be driven by the highly significant interaction (window: -160 to 640 ms, $p < .05$). Specifically, the post-hoc comparisons of the interaction reflected the main effect of Source by showing pupil dilation after button presses compared to the passive events both in the predictable and the unpredictable sessions (both $ps < .05$). These comparisons also showed that while predictability context did not affect pupil responses to A-events (i.e., when merely observing the colour-filling of the cue), it modulated responses following button presses with larger diameter in the predictable vs. the unpredictable session.

750 ms delay

For the 750 ms data, we obtained similar results as in the 150 ms delay. First, we observed a main effect of Source starting before the button press and/or colour-filling of the cue (window: -490 to 1,830 ms, $p < .05$), showing again increased pupil diameter following a button press. We also observed a significant main effect of Predictability (window: -180 to 1,170 ms, $p < .05$), with larger pupil diameter for predictable compared to the unpredictable events. Similar to the analyses for the 150 ms data, this effect seems to be driven by the significant interaction (window: -150 to 1,170 ms $p < .05$), that showed that button presses elicited larger pupil diameter compared to passive events in both predictability contexts (both $ps < .05$), and that predictability affected pupil diameter but only when an action has been performed (i.e., larger diameter for predictable vs. unpredictable events following a button press, but no differences between passive events as a function of predictability).

Finally, we also aimed to test whether the source and predictability effects on pupil were modulated as a function of delay. To this end, we used the difference A – MA, separately for each predictability session and delay to test for a possible interaction (i.e., ([A – MA in predictable – 150 ms] – [A – MA in predictable – 750 ms]) compared to ([A – MA in unpredictable – 150 ms] – [A – MA in unpredictable – 750 ms])). This analysis did not show a significant interaction between predictability and delay on the source pupil effects (A – MA; all $ps > .05$).

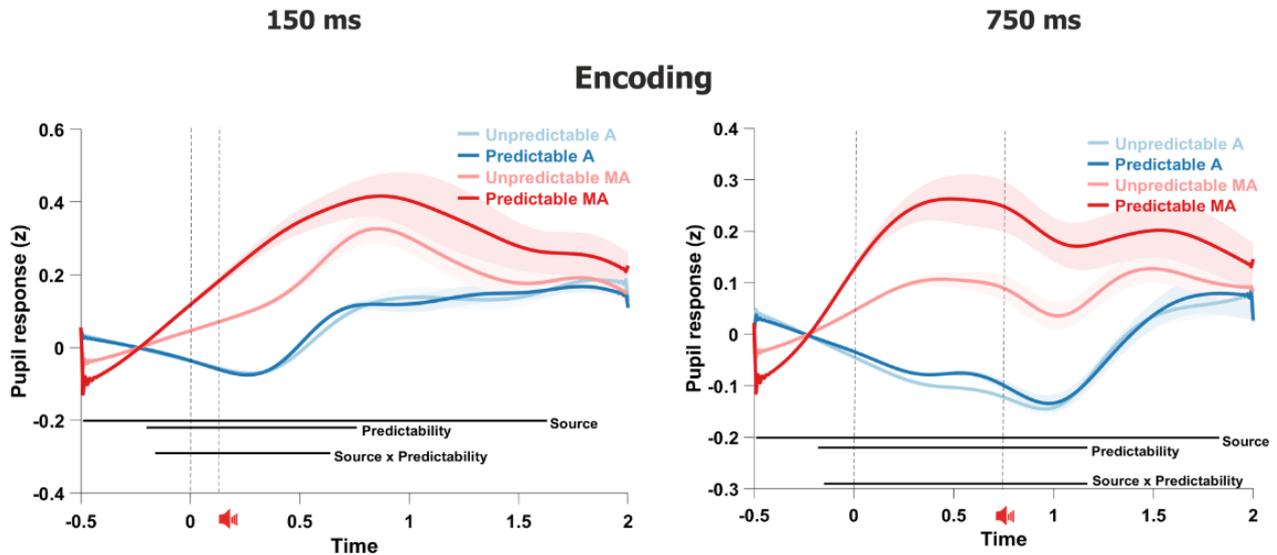


Figure 6. Pupil responses separately for the 150 and 750 ms delay. Black bars indicate significant effects, $p < .05$ (cluster-based permutation tests). Responses were locked to the button press and/or the colour-filling of the cue. Dashed grey lines represent the time of the button press and/or the colour-filling of the cue and the sound onset (150 or 750 ms after the press, respectively).

3.4. Correlations

Next, we tested for possible correlations between the behavioural performance, pupillometric and electrophysiological data. For the correlation analyses, we focused on the significant neurophysiological effects at encoding (i.e., ERPs and pupil diameter) and the significant behavioural production effect. The effects were introduced in the correlation analyses as the difference between A and MA events separately for each delay condition and predictability session (see Methods). For the significant effects on Tb, we calculated the mean amplitude across the T7 and T8 electrodes. For the pupil data, we calculated the peak of the difference wave (A – MA) within the window of the significant interaction for each delay (i.e., -160 to 640 ms for the 150 ms delay and -150 to 1,170 ms for the 750 ms delay). All the planned correlations are reported in Table 4.

First, we tested whether the significant self-generation effects at encoding (on N1, P2, and Tb amplitudes) correlated with the significant production effect on memory performance, separately for each delay condition and predictability session (Table 4 (a)). Second, we assessed whether the difference in pupil diameter between auditory-only and motor-auditory events was related to memory performance, separately for each delay condition and predictability session (Table 4 (b)). Third, we tested for possible links between the self-generation effects obtained in the ERP analyses (i.e., N1, P2, and Tb) and the larger pupil diameter for motor-auditory events (within the window of the significant interaction for each delay), separately for each delay condition and predictability session (Table 4 (c)). As reported in Table 4, the only significant correlations obtained were limited to the predictable session and with short action/cue-sound delays. First, we obtained a significant positive correlation between pupil diameter and memory performance ($r = 0.48$, $p = 0.034$, $BF_{10} = 2.79$), suggesting that the larger the pupil diameter for button presses triggering MA sounds, the worse the memory performance. Second, we obtained a significant negative correlation between Tb attenuation and pupil dilation in

the predictable session with short action-sound delays ($r = -0.48$, $p = 0.031$, $BF_{10} = 2.93$), suggesting that the larger the pupil diameter, the stronger the attenuation effects for MA sounds. Both significant correlations are in line with our previous work (Paraskevoudi & SanMiguel, 2022) and their corresponding Bayes Factors provide supporting, albeit weak, evidence for the alternative hypotheses (i.e., correlation between pupil diameter and memory, and pupil diameter and Tb attenuation). The rest of the correlations did not reach significance (all $ps > .05$).

Table 4.

Correlations between the self-generation effects, for each combination of predictability and delay

		<i>Correlations between</i>	<i>r</i>	<i>p</i>	<i>BF₁₀</i>	
Predictable 150	(a) Memory performance	N1 (attenuation)	-0.094	0.69	0.51	
		Tb (attenuation)	-0.047	0.84	0.48	
		P2 (attenuation)	0.3	0.2	0.89	
	(b) Memory performance	Pupil dilation	0.48	0.034*	<u>2.79</u>	
		(c) Pupil diameter	N1 (attenuation)	-0.38	0.1	<u>1.38</u>
			Tb (attenuation)	-0.48	0.031*	<u>2.93</u>
	P2 (attenuation)	0.011	0.96	0.47		
Predictable 750	(a) Memory performance	N1 (attenuation)	0.31	0.18	0.97	
		Tb (n.s.)	-	-	-	
		P2 (n.s.)	-	-	-	
	(b) Memory performance	Pupil dilation	0.19	0.43	0.61	
		(c) Pupil diameter	N1 (attenuation)	-0.44	0.052	<u>2.10</u>
			Tb (n.s.)	-	-	-
	P2 (n.s.)	-	-	-		
Unpredictable 150	(a) Memory performance	N1 (n.s.)	-	-	-	
		Tb (enhancement)	0.25	0.28	0.76	
		P2 (n.s.)	-	-	-	
	(b) Memory performance	Pupil dilation	-0.044	0.85	0.48	
		(c) Pupil diameter	N1 (n.s.)	-	-	-
			Tb (enhancement)	0.2	0.41	0.62
	P2 (n.s.)	-	-	-		
Unpredictable 750	(a) Memory performance	N1 (attenuation)	-0.42	0.063	<u>1.85</u>	
		Tb (n.s.)	-	-	-	
		P2 (n.s.)	-	-	-	
	(b) Memory performance	Pupil dilation	0.024	0.92	0.48	
		(c) Pupil diameter	N1 (attenuation)	0.044	0.85	0.48
			Tb (n.s.)	-	-	-
	P2 (n.s.)	-	-	-		

Note. For each one of the ERP components, we specify the direction of the self-generation effect, that is attenuation, enhancement, or no effect (i.e., n.s.). Correlations were performed only for the significant self-generation effects. Pearson's coefficients (r), their corresponding p values, and the Bayes Factors for the alternative hypothesis (BF_{10}) are reported. Significant correlations are

highlighted and indicated by an asterisk. Bayes Factors providing evidence (i.e., values > 1) in favour of the alternative hypothesis are underlined. **a)** auditory responses at encoding (N1, P2, and Tb amplitudes) and memory performance, **b)** neuromodulatory effects at encoding (pupil diameter) and memory performance, **c)** auditory responses (N1, P2, and Tb amplitudes) and neuromodulatory effects (pupil diameter) at encoding.

4. Discussion

In this study, we assessed whether the self-generation effects on sensory processing and memory encoding of sounds are due to stimulus-specific predictions, due to stimulus-unspecific effects of action proximity, or a mixture of both. Specifically, we examined the effects of self-generation and predictability, as well as the potential interactions between the two on a) sensory processing and subcortical neuromodulation and b) memory encoding of sounds. To this end, we employed a paradigm that controlled for all aspects of predictability (*what, when, whether*) and recorded behavioural, electrophysiological, and pupil responses to predictable or unpredictable self- and externally-generated sounds that were either presented shortly after or farther away from the action or the visual cue, respectively.

4.1. Effects of actions and predictability on sensory processing and subcortical neuromodulation

The first aim of the present study was to assess the effects of actions and predictability on auditory responses and subcortical neuromodulation, but also the possible relationships between the two. Related to the effects of actions and predictability on auditory responses, we first showed that actions affected sensory responses in both short and long action-sound delays; however, when the sound is presented close in time with the action, self-generation interacts with predictability, that is the attenuation effects on N1, P2, and Tb were limited to the predictable session. Yet, self-generation and predictability no longer interacted when the sound was farther away from the action, affecting N1, Tb, and P2 in differential ways: N1 attenuation was driven by self-generation, while Tb and P2 modulations were driven by predictability. Second, predictability modulated sensory responses, but these modulations were dependent on the presence or absence of an action and the action-sound delay: With short delays, predictability interacted with self-generation such that it enhanced N1 and Tb responses to predictable compared to unpredictable sounds in the absence of action (but no differences as a function of predictability in the presence of action), while it suppressed P2 responses for predictable compared to unpredictable stimuli triggered by an action (but no differences as a function of predictability in the absence of action). In contrast, with longer delays, predictability suppressed only the Tb and P2 components irrespective of the presence of action, while it did not exert any effects on N1. Finally, the data-driven cluster-based permutation analyses replicated previous work (Korka et al., 2021b; Pinheiro et al., 2020; Reznik et al., 2018) by showing that actions modulate activity already before sound onset and showed strong interactions between source and predictability in the cue/pres-sound interval, suggesting that the pre-stimulus modulations caused by action-driven predictability are distinct than the modulations triggered by visually-driven predictability. In what follows, we discuss the implications of these findings as for the specificity of the attenuation effects.

Support to the specificity of the effects could be provided by the significant interaction between self-generation and predictability with short action-sound delays: Attenuation of N1, P2, and Tb was only observed for predictable self-generated sounds, supporting previous evidence that the suppression effects are specific to the predicted stimulus (e.g., Houde et al., 2002; Fu et al., 2006) and pointing to a critical difference between motor-based and sensory-based predictions, in that attenuation occurs for motor-auditory predictions (i.e., self-generated predictable sounds) but not for the visual-auditory ones (i.e., externally-generated predictable sounds that were cued by a visual stimulus). Nevertheless, contrary to previous work (Martikainen et al., 2004; Aliu et al., 2009), we did not find clear evidence that the N1 attenuation reflects modulations within auditory areas which would be observable at the sensory specific component on the mastoids (thought to partly originate from a supratemporal generator; Vaughan & Ritter, 1970); however, the lack of polarity reversal does not mean that the specific component of N1 is not modulated; rather, it might suggest that attenuation of the supratemporal component of N1 might overlap with the unspecific component that can be simultaneously modulated during self-generation (Horváth et al., 2012).

Meanwhile, our findings also point to the involvement of unspecific processes in the self-generation effects. We show that N1 is attenuated for both predictable and unpredictable sessions in the long delay condition, suggesting that – at least for longer action-sound delays – N1-suppression is not solely driven by specific motor-predictions (Horváth et al., 2012; SanMiguel et al., 2013). One possibility is that with longer action-sound delays, differences in temporal expectations (i.e., expressed as the hazard function) were minimized between predictable and unpredictable sessions, since the conditional probability of an event occurring at a given time given that it has not yet occurred (Nobre et al., 2007) evidently increases as time passes (Janssen & Shadlen, 2005; Yang & Shadlen, 2007). This could possibly have allowed for predicting the temporal onset of the stimulus irrespective of the predictability context. Critically, given that equal predictability was provided for both A and MA sounds in the present study, the prevalence of the suppression effects suggests that actions afforded more robust temporal predictions – irrespective of whether they are fully predictive of the upcoming sound or not – compared to the temporal predictions afforded by sensory cues, thereby giving rise to a N1 suppression effect that is known to prevail even when temporal predictability is controlled (Klaffehn et al., 2019) and when actions are not fully predictive of the sound (Paraskevoudi & SanMiguel, 2022).

The second important finding related to the effects of actions on sensory processing is that subcortical neuromodulatory processes, possibly mediated by LC-NE (i.e., reflected in pupil diameter; Aston-Jones & Cohen, 2005, Murphy et al., 2014; Joshi et al., 2016), were robustly triggered by actions (also reported in rodents: McGinley et al., 2015; Vinck et al., 2015, monkeys: Bornert & Bouret, 2021, and humans: Yebra et al., 2019; Lubinus et al., 2022). Specifically, pupil dilation started already ~500 ms before action initiation (cf. Aston-Jones & Cohen, 2005; Reimer et al., 2016) and remained dilated for more than 1 s compared to the passive comparison events. Critically, the pattern of activation was remarkably consistent across the two delay conditions, in line with evidence suggesting that phasic LC-NE activation is more tightly coupled with behaviour (e.g., motor response) than sensory stimulation (Aston-Jones & Cohen, 2005). More importantly, we show that pupil diameter was interactively modulated by self-generation and predictability, evident by a striking difference between actions generating highly predictable auditory stimulation compared to those that could not predict the

identity, timing, and probability of occurrence of a sound. An intriguing possibility is that pupil responses might have been affected by differences between predictable and unpredictable contexts in action preparation and execution. Indeed, there is some evidence that action-induced LC-NE activity might depend on the type of action, differing between goal-directed actions and general active states representing more automatic behaviours such as grooming (Aston-Jones & Bloom, 1981), while work from the self-generation domain has shown that action preparation and execution are affected by the action-stimulus contingency of the paradigm, being modulated as a function of whether the action systematically triggers a stimulus or not (Neszmélyi & Horváth, 2017; Neszmélyi & Horváth, 2018). Collectively, we argue that differences in movement parameters, such as action-stimulus contingency, might contribute to the magnitude of the action-induced pupil dilation effects; yet, more work is needed to replicate these interactive effects of actions and predictability on pupil diameter and isolate which aspects of predictability and movement parameters might drive differences in pupil responses during actions.

Moreover, we also hypothesized that these subcortical neuromodulatory processes – reflected in pupil diameter – might be related to the stimulus-unspecific effects of actions on the auditory evoked responses. Indeed, pupil dilation correlated with the Tb attenuation effects, such that the larger the pupil diameter for button presses, the larger the suppression of the Tb component to sounds generated by them. However, this link was only found with fully contingent relationships and short delays between the action and sound (i.e., predictable session with 150 ms delay) where the attenuation effects were strongest, suggesting that neuromodulatory processes do not suppress sensory processing in an indiscriminate way, since such link was not found with longer action-sound delays nor in the unpredictable session (see also Paraskevoudi & SanMiguel, 2022 for similar findings). Therefore, we argue that our results have two highly important implications: First, along with the differences in pupil response as a function of predictability (increased pupil diameter following actions that had predictable sensory consequences), we showed that the more pronounced the attenuation effects, the largest the pupil diameter for self-triggered and predictable inputs, suggesting that the link between action-induced sensory dampening and pupil dilation can be captured when the effects are maximized, namely when the action is fully predictive of the upcoming stimulus. Second, this link was only obtained when the sound was presented shortly after the action, but not with the longer 750 ms delay. This might provide some interim evidence in favour of the hypothesis that a halo of neuromodulation relates to the attenuation effects when the action is in close temporal proximity with the sensory refference. Crucially, the effects of such halo on sensory responses seem to fade out with longer action-sound delays. Therefore, based on these findings and the implications outlined above we propose that the link between sensory attenuation and subcortical neuromodulation might depend on the combination of two processes operating simultaneously. These two processes are outlined in the following paragraphs and depicted in Figure 7.

The first process might result from the LC-NE projections on motor areas. In fact, there is already evidence that LC-NE innervates the primary motor cortex (Chandler et al., 2014) and that norepinephrine release in primary motor areas result in increased firing of motor neurons (Schiemann et al., 2015). Therefore, if LC-NE activity increases motor excitability, this might increase the inhibitory effect of motor engagement on auditory responses (Reznik & Mukamel, 2019), thereby attenuating activity in the auditory cortex. Combining these lines of evidence, it is likely that the

enhancing effects of LC-NE activation on motor excitability further increase the inhibitory effect of motor cortex on auditory neurons, thereby providing a plausible explanation for the links between subcortical neuromodulation and sensory suppression during movement. More importantly, this mechanism would yield stronger attenuation effects under fully predictable contexts, given our data showing larger LC-NE activity under predictable vs. unpredictable actions (see left and middle panel in Figure 7) which would result in higher motor excitability and therefore larger inhibitory effects of the motor cortex on auditory areas with action-sound contingency.

The second process we propose is related to the effects of motor and neuromodulatory inputs on auditory areas. There is already evidence showing that motor and neuromodulatory inputs overlap onto the same auditory neurons (Nelson & Mooney, 2016; for a review see Schneider & Mooney, 2018). This might suggest that despite the global effect of neuromodulation in the neocortex, neuromodulatory processes would differentially affect those cells that also receive top-down motor-driven predictions (which would be the case only in conditions that stimulus-specificity mechanisms can operate, that is in highly predictable contexts). These neuromodulatory influences could possibly act as an attentional and orienting filter (Aston-Jones & Cohen, 2005) that facilitates or inhibits sensory responses (depending on the receptor mediating the effects; Vitrac & Benoit-Marand, 2017) for stimulation falling close in time with the action. Given that norepinephrine application in sensory areas has been found to suppress the firing of the vast majority of neurons, but meanwhile increases firing of a small neural population (Castro-Alamancos & Bezdudnaya, 2015), we assume that the net effect on auditory areas would be of inhibitory nature. Therefore, although the entire auditory cortex will receive LC-NE input during movement, only neurons that receive convergent and simultaneous input from motor areas would be the ones modulated. This mechanism would consequently result in greater differences in unspecific modulation between self- and externally-generated sounds, because the latter would lack any neuromodulatory and motor input due to the absence of movement (see right panel in Figure 7). Thus, we propose that ascending neuromodulatory information might alter auditory cortical responsiveness (Schneider & Mooney, 2018; Reimer et al., 2014), specifically for the neural populations that receive inputs from both motor and neuromodulatory processes.

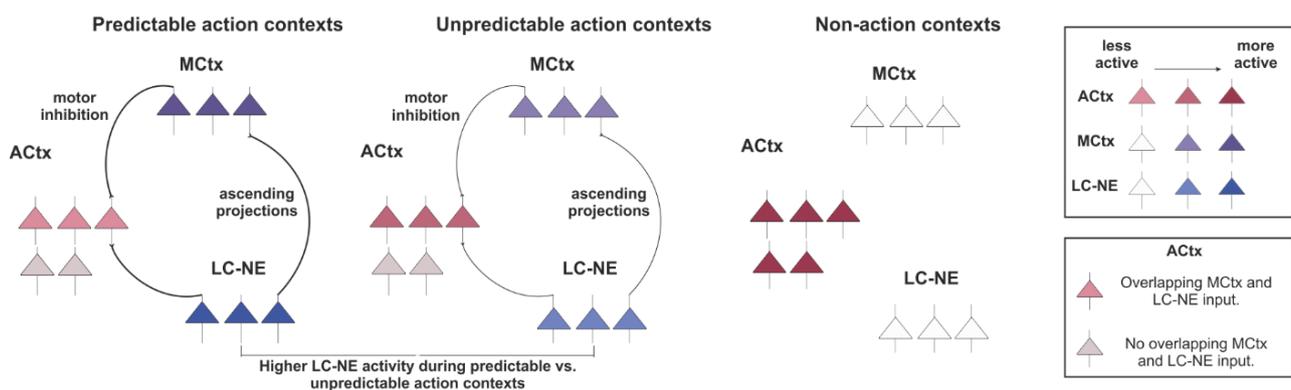


Figure 7. Schematic representation of the proposed mechanism behind the links between sensory attenuation and subcortical neuromodulation during movement. In action contexts (left and middle panel), movement will preferentially modulate activity in neurons that receive overlapping motor and neuromodulatory input. In highly predictable action contexts (left panel), LC-NE activity increases,

thereby boosting motor excitability through ascending projections (thick black lines) of LC-NE to motor cortex (i.e., MCtx), which in turn increases the inhibitory effect of MCtx on auditory cortex (i.e., ACtx), resulting in larger attenuation effects. In unpredictable action contexts (middle panel), LC-NE activity increases as well, but to a lesser extent than in predictable action contexts. These increases will boost MCtx activity, but again to a lesser extent than in predictable action contexts (thin black lines). Importantly, although motor engagement will still attenuate ACtx responses, the attenuation effects in this case will be smaller than in predictable action contexts. Finally, in the absence of movement (non-action contexts; right panel), LC-NE and MCtx activity remains silent, and thus, responses in ACtx are merely driven by the sensory input, resulting in stronger responses than responses to stimulation *predicted by* or *presented close in time with* an action (as depicted in the action contexts panels).

4.2. *Effects of actions on memory encoding*

The second aim of the present study was to examine whether memory encoding of identifiable sounds is affected by semi self-paced actions (and its possible interactions with predictability and/or action-sound delay), and whether any possible effects on memory correlate with sensory suppression and/or subcortical neuromodulation during encoding. We obtained a significant enhancement for sounds that were encoded as self-generated, irrespective of predictability (and delay), providing support to the production effect studies highlighting the beneficial impact of production on memory recall of – typically fully predictable – stimuli (MacDonald & MacLeod, 1998; Brown & Palmer, 2012). This behavioural enhancement for self-generated stimulation seems at odds with predictive coding accounts of memory (Henson & Gagnepain, 2010), but also with previous work on the effects of self-generation on memory showing the opposite effect (i.e., behavioural enhancement for the more surprising externally-generated stimuli; Paraskevoudi & SanMiguel, 2022). In this section, we provide an interim interpretation of this discrepancy in the effects of actions on memory processing.

One possibility is that the direction of the action effects on memory encoding might depend on factors related to the type of action and the temporal control afforded by it, the confounding factor of predictability, and the type of the sounds to be encoded. First, contrary to Paraskevoudi & SanMiguel (2022), the present study allowed for a variety of action choices (i.e., choosing which action to perform) that is known to yield beneficial effects on memory (Murty et al., 2015). Second, temporal control over the upcoming self-generated stimulus was higher in both predictability sessions of the present study than in Paraskevoudi & SanMiguel (2022): Here, actions were semi self-paced and generated a sound 100% or 50% of the time after either 150 or 750 ms for the predictable and unpredictable session, respectively. This might have provided slightly higher temporal control over self-generated stimulation than over the externally-generated inputs, where participants simply waited for the circle to be filled on its own. Third, the present study controlled for all aspects of predictability contrary to Paraskevoudi & SanMiguel (2022) that did not (i.e., self-generated sounds afforded better temporal predictions). Finally, auditory stimulation differed between the two studies (i.e., non-identifiable noise stimuli in Paraskevoudi & SanMiguel (2022) and identifiable human, animal, environmental, and musical sounds in the present study). There is already evidence that different types of sounds (e.g., noise, pure tones, speech, and music) exhibit different patterns of activity (for a meta-analysis see Samson et al., 2011), and it is also known that noise stimuli have inherently less structure

and therefore afford less predictability than identifiable sounds (Agus et al., 2010). Thus, the employment of non-identifiable stimuli in Paraskevoudi & SanMiguel (2022) – along with the confounding factor of temporal predictability – might have triggered saliency-detection mechanisms (Agus et al., 2010; Itti et al., 1998) that boosted the memory of the more surprising externally-generated sounds, as proposed by predictive coding accounts of memory (Henson & Gagnepain, 2010).

In sum, we argue that when self-generation affords better predictions (as in Paraskevoudi & SanMiguel, 2022), memory will be weakened as proposed by predictive coding accounts (Henson & Gagnepain, 2010), probably resulting from surprise-driven generalized predictive mechanisms that operate when predictability is confounded with self-generation. This idea is supported by the fact the magnitude of attenuation effects correlates with weakened memory only when temporal predictability is not identical between self- and externally-generated inputs (Paraskevoudi & SanMiguel, 2022). In contrast, the present study did not show significant correlations between sensory attenuation and memory enhancement for self-generated stimulation, which suggests that the behavioural production effect is not linked to the action-induced sensory attenuation, thereby implicating that surprise-driven mechanisms (as reflected in prediction error responses) cannot be eliciting the effects on memory in the present study. Instead, here we showed that when predictability is controlled and actions provide temporal control over the self-generated stimulus, memory will be boosted, irrespective of whether the action is fully predictive of the sound or not. One possibility is that intentionality and choice-related processes (i.e., freely choosing which action to perform and when) might drive learning and memory (Herwig et al., 2007; Waszak et al., 2005; Murty et al., 2015), possibly by increasing the feeling of control over action outcomes (Wenke et al., 2010) that might allow for better processing of the self-triggered stimulus when an action can be selected among several alternatives compared to when action is limited to one alternative (Barlas et al., 2017). Collectively, the findings we obtained here along with our previous work (Paraskevoudi & SanMiguel, 2022) suggest that self-generation and predictability confounds might act in opposite directions, but in the presence of both, predictability might win over the action effects on memory, even more so when the action is not self-chosen and does not provide temporal control over the upcoming stimulation (as in Paraskevoudi & SanMiguel, 2022).

Additionally, and in line with previous work in rodents (McGinley et al., 2015) and humans (Paraskevoudi & SanMiguel, 2022), we show that the larger the pupil diameter for self-generated events, the worse the memory performance for these sounds at retrieval, but only when the action-induced pupil dilation effects are maximized, that is with fully contingent action-sound relationships and short action-sound delays. Although our correlation analyses could only test for linear relationships between memory performance and pupil diameter, this finding could possibly fit with the idea of an inverted U-shaped dependence of behaviour on arousal (Yerkes & Dodson, 1908) – indexed by pupil diameter (McGinley et al., 2015) – suggesting that performance drops during high-arousal states such as movement, while it is optimal at intermediate arousal levels (i.e., states of quiet wakefulness).

5. Conclusion

In sum, the overarching aim of the present work was to control for all aspects of predictability (*what, when, whether*) in order to assess whether the self-generation effects on sensory processing and memory encoding of sounds are due to stimulus-specific predictions, due to the unspecific effects of actions, or a mixture of both. Sensory attenuation for predictable self-generated sounds presented shortly after the action prevailed even when providing identical predictability for the externally-generated ones, albeit with no clear evidence of attenuation in primary auditory areas (i.e., no attenuation on the mastoids), thereby providing only partial support to the specificity of the effects. Meanwhile, N1 attenuation also occurred for both predictable and unpredictable sounds presented farther away from the actions, pointing to some sort of diffused and generalized attenuation mechanism that acts independent of the action-sound contingency. Simultaneously with these effects, subcortical neuromodulatory processes were also triggered by actions that were reflected in increases in pupil diameter during actions. Critically, pupil dilation correlated with the attenuation effects when they were more strongly elicited, providing evidence that unspecific neuromodulatory processes may partly underly the action-induced sensory suppression. Finally, we provide strong evidence that when providing identical predictability for the externally-generated sounds, the presence of self-chosen and semi self-paced actions boosts memory encoding for the self-generated ones, pointing to effects of opposite direction when self-generation is conflated with predictability compared to when predictability is orthogonally controlled. Taken together, the present work shows that actions trigger a cascade of stimulus-specific and unspecific processes – presumably driven by subcortical neuromodulatory areas – that collaboratively shape auditory processing, and that predictability confounds in self-generation paradigms might be critical in determining the direction of the effects of actions on behaviour.

CRedit authorship contribution statement

Nadia Paraskevoudi: Conceptualization, Formal Analysis, Methodology, Software, Investigation, Writing - Original draft preparation, Visualization. **Iria SanMiguel:** Conceptualization, Formal Analysis, Methodology, Software, Writing - Original draft preparation, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work is part of the project PSI2017-85600-P, funded by MCIN/AEI/[10.13039/501100011033](https://doi.org/10.13039/501100011033) and by “ERDF A way of making Europe”; it has been additionally supported by the MDM-2017-0729-18-2M Maria de Maeztu Center of Excellence UBNeuro, funded by MCIN/AEI/[10.13039/501100011033](https://doi.org/10.13039/501100011033), and by the Excellence Research Group 2017SGR-974 funded by the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya. ISM was supported by grant RYC-2013-12577, funded by MCIN/AEI/[10.13039/501100011033](https://doi.org/10.13039/501100011033) and by “ESF Investing in your future.” NP was supported by predoctoral fellowship FI-DGR 2019 funded by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the European Social Fund.

References

- Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, *66*(4), 610–618. <https://doi.org/10.1016/j.neuron.2010.04.014>
- Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802. <https://doi.org/10.1162/jocn.2009.21055>
- Aston-Jones, G., & Bloom, F. E. (1981). Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *The Journal of Neuroscience*, *1*(8), 876–886. <https://doi.org/10.1523/jneurosci.01-08-00876.1981>
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study: Selective suppression of self-initiated sounds. *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>
- Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. <https://doi.org/10.1098/rstb.2008.0310>
- Barlas, Z., Hockley, W. E., & Obhi, S. S. (2017). The effects of freedom of choice in action selection on perceived mental effort and the sense of agency. *Acta Psychologica*, *180*, 122–129. <https://doi.org/10.1016/j.actpsy.2017.09.004>

- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. <https://doi.org/10.1038/35002078>
- 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. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640. <https://doi.org/10.1038/2870>
- Bornert, P., & Bouret, S. (2021). Locus coeruleus neurons encode the subjective difficulty of triggering and executing actions. *PLOS Biology*, *19*(12), e3001487. <https://doi.org/10.1371/journal.pbio.3001487>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brown, R. M., & Palmer, C. (2012). Auditory–motor learning influences auditory memory for music. *Memory & Cognition*, *40*(4), 567–578. <https://doi.org/10.3758/s13421-011-0177-x>
- Budinger, E., Laszcz, A., Lison, H., Scheich, H., & Ohl, F. W. (2008). Non-sensory cortical and subcortical connections of the primary auditory cortex in Mongolian gerbils: Bottom-up and top-down processing of neuronal information via field AI. *Brain Research*, *1220*, 2–32. <https://doi.org/10.1016/j.brainres.2007.07.084>
- Castro-Alamancos, M. A., & Bezdudnaya, T. (2015). Modulation of artificial whisking related signals in barrel cortex. *Journal of Neurophysiology*, *113*(5), 1287–1301. <https://doi.org/10.1152/jn.00809.2014>
- Chandler, D. J., Gao, W.-J., & Waterhouse, B. D. (2014). Heterogeneous organization of the locus coeruleus projections to prefrontal and motor cortices. *Proceedings of the National Academy of Sciences*, *111*(18), 6816–6821. <https://doi.org/10.1073/pnas.1320827111>
- Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG Technology*,

25(2), 83–92. <https://doi.org/10.1080/00029238.1985.11080163>

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>

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends in Cognitive Sciences*, 19(2), 92–99.
<https://doi.org/10.1016/j.tics.2014.12.004>

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>

Eggermann, E., Kremer, Y., Crochet, S., & Petersen, C. C. H. (2014). Cholinergic signals in mouse barrel cortex during active whisker sensing. *Cell Reports*, 9(5), 1654–1660.
<https://doi.org/10.1016/j.celrep.2014.11.005>

Fonken, Y. M., Kam, J. W. Y., & Knight, R. T. (2020). A differential role for human hippocampus in novelty and contextual processing: Implications for P300. *Psychophysiology*, 57(7).
<https://doi.org/10.1111/psyp.13400>

Foote, S. L., Freedman, R., & Oliver, A. P. (1975). Effects of putative neurotransmitters on neuronal activity in monkey auditory cortex. *Brain Research*, 86(2), 229–242.
[https://doi.org/10.1016/0006-8993\(75\)90699-X](https://doi.org/10.1016/0006-8993(75)90699-X)

Foote, S. L., & Morrison, J. H. (1987). Extrathalamic modulation of cortical function. *Annual Review of Neuroscience*, 10(1), 67–95. <https://doi.org/10.1146/annurev.ne.10.030187.000435>

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>

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1404), 1771–1788. <https://doi.org/10.1098/rstb.2000.0734>

- Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., Yágüez, L., van Haren, N. E. M., Matsumoto, K., & McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969–977. <https://doi.org/10.1093/cercor/bhj039>
- Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods*, *43*(4), 1171–1181. <https://doi.org/10.3758/s13428-011-0109-5>
- Gagnepain, P., Henson, R., Chételat, G., Desgranges, B., Lebreton, K., & Eustache, F. (2011). Is neocortical–hippocampal connectivity a better predictor of subsequent recollection than local increases in hippocampal activity? New insights on the role of priming. *Journal of Cognitive Neuroscience*, *23*(2), 391–403. <https://doi.org/10.1162/jocn.2010.21454>
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. [https://doi.org/10.1016/S1364-6613\(99\)01417-5](https://doi.org/10.1016/S1364-6613(99)01417-5)
- Gygi, B., & Shafiro, V. (2010). Development of the database for environmental sound research and application (DESRA): Design, functionality, and retrieval considerations. *EURASIP Journal on Audio, Speech, and Music Processing*, *2010*, 1–12. <https://doi.org/10.1155/2010/654914>
- Halgren, E. (1991). Firing of human hippocampal units in relation to voluntary movements. *Hippocampus*, *1*(2), 153–161. <https://doi.org/10.1002/hipo.450010204>
- Hazemann, P., Audin, G., & Lille, F. (1975). Effect of voluntary self-paced movements upon auditory and somatosensory evoked potentials in man. *Electroencephalography and Clinical Neurophysiology*, *39*(3), 247–254. [https://doi.org/10.1016/0013-4694\(75\)90146-7](https://doi.org/10.1016/0013-4694(75)90146-7)
- Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. <https://doi.org/10.1002/hipo.20857>
- Herwig, A., Prinz, W., & Waszak, F. (2007). Two modes of sensorimotor integration in intention-based and stimulus-based actions. *Quarterly Journal of Experimental Psychology*, *60*(11),

1540–1554. <https://doi.org/10.1080/17470210601119134>

- Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667. <https://doi.org/10.1038/nn.4284>
- Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for environmental sound stimuli. *PLoS ONE*, *8*(9), e73382. <https://doi.org/10.1371/journal.pone.0073382>
- Horváth, J. (2013a). Attenuation of auditory ERPs to action-sound coincidences is not explained by voluntary allocation of attention: Action-sound coincidence effect is not attentional. *Psychophysiology*, *50*(3), 266–273. <https://doi.org/10.1111/psyp.12009>
- Horváth, J. (2013b). Action-sound coincidence-related attenuation of auditory ERPs is not modulated by affordance compatibility. *Biological Psychology*, *93*(1), 81–87. <https://doi.org/10.1016/j.biopsycho.2012.12.008>
- 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
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–1138. <https://doi.org/10.1162/089892902760807140>
- 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>
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. <https://doi.org/10.1109/34.730558>
- Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque

area LIP. *Nature Neuroscience*, 8(2), 234–241. <https://doi.org/10.1038/nn1386>

Jeffreys, H. (1961). *Theory of Probability* (3rd ed., Ser. Oxford Classic Texts in the Physical Sciences.). Oxford Univ. Press.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1), 221–234. <https://doi.org/10.1016/j.neuron.2015.11.028>

Kimura, M. (2021). Prediction, suppression of visual response, and modulation of visual perception: Insights from visual evoked potentials and representational momentum. *Frontiers in Human Neuroscience*, 15, 730962. <https://doi.org/10.3389/fnhum.2021.730962>

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>

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psyctoolbox-3. *Perception*, 36(14), 1-16.

Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016).

Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLOS ONE*, 11(5), e0155574. <https://doi.org/10.1371/journal.pone.0155574>

Korka, B., Widmann, A., Waszak, F., Darriba, Á., & Schröger, E. (2021a). The auditory brain in

action: Intention determines predictive processing in the auditory system—A review of current paradigms and findings. *Psychonomic Bulletin & Review*.

<https://doi.org/10.3758/s13423-021-01992-z>

Korka, B., Schröger, E., & Widmann, A. (2021b). The encoding of stochastic regularities is facilitated by action-effect predictions. *Scientific Reports*, 11, 6790.

Kumaran, D., & Maguire, E. A. (2009). Novelty signals: A window into hippocampal information processing. *Trends in Cognitive Sciences*, 13(2), 47–54.

<https://doi.org/10.1016/j.tics.2008.11.004>

Lubinus, C., Einhäuser, W., Schiller, F., Kircher, T., Straube, B., & van Kemenade, B. M. (2022).

Action-based predictions affect visual perception, neural processing, and pupil size, regardless of temporal predictability. *NeuroImage*, *263*, 119601.

<https://doi.org/10.1016/j.neuroimage.2022.119601>

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica*, *98*(2–3), 291–310. [https://doi.org/10.1016/S0001-6918\(97\)00047-4](https://doi.org/10.1016/S0001-6918(97)00047-4)

Makeig, S., Müller, M. M., & Rockstroh, B. (1996). Effects of voluntary movements on early auditory brain responses. *Experimental Brain Research*, *110*(3), 487–492.

<https://doi.org/10.1007/BF00229149>

Martikainen, M. H., Kaneko, K., & Hari, R. (2004). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*, *15*(3), 299–302.

<https://doi.org/10.1093/cercor/bhh131>

McGinley, M. J., David, S. V., & McCormick, D. A. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron*, *87*(1), 179–192.

<https://doi.org/10.1016/j.neuron.2015.05.038>

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](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>

Murphy, P. R., O’Connell, R. G., O’Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*,

35(8), 4140–4154. <https://doi.org/10.1002/hbm.22466>

- Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *Journal of Neuroscience*, *35*(16), 6255–6264.
<https://doi.org/10.1523/JNEUROSCI.4181-14.2015>
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>
- Nelson, A., & Mooney, R. (2016). The basal forebrain and motor cortex provide convergent yet distinct movement-related inputs to the auditory cortex. *Neuron*, *90*(3), 635–648.
<https://doi.org/10.1016/j.neuron.2016.03.031>
- Neszmélyi, B., & Horváth, J. (2017). Consequences matter: Self-induced tones are used as feedback to optimize tone-eliciting actions: Self-induced tones used as feedback for actions. *Psychophysiology*, *54*(6), 904–915. <https://doi.org/10.1111/psyp.12845>
- Neszmélyi, B., & Horváth, J. (2018). Temporal constraints in the use of auditory action effects for motor optimization. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(11), 1815–1829. <https://doi.org/10.1037/xhp0000571>
- Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470. <https://doi.org/10.1016/j.conb.2007.07.006>
- Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281–1296. <https://doi.org/10.1016/j.neuron.2015.11.035>
- Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, *265*(2), 119–122.
[https://doi.org/10.1016/S0304-3940\(99\)00218-9](https://doi.org/10.1016/S0304-3940(99)00218-9)
- Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. In *Progress in Brain Research* (Vol. 159, pp. 99–120). Elsevier. <https://doi.org/10.1016/S0079->

6123(06)59007-7

- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 1–9. <https://doi.org/10.1155/2011/156869>
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112(4), 713–719. [https://doi.org/10.1016/S1388-2457\(00\)00527-7](https://doi.org/10.1016/S1388-2457(00)00527-7)
- Paraskevoudi, N., & SanMiguel, I. (2022). Sensory suppression and increased neuromodulation during actions disrupt memory encoding of unpredictable self-initiated stimuli. *Psychophysiology*. <https://doi.org/10.1111/psyp.14156>
- Parker, P. R. L., Brown, M. A., Smear, M. C., & Niell, C. M. (2020). Movement-related signals in sensory areas: Roles in natural behavior. *Trends in Neurosciences*, 43(8), 581–595. <https://doi.org/10.1016/j.tins.2020.05.005>
- Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. *Nature Communications*, 9(1), 1673. <https://doi.org/10.1038/s41467-018-03992-5>
- Pinheiro, A. P., Schwartz, M., Gutiérrez-Domínguez, F., & Kotz, S. A. (2020). Real and imagined sensory feedback have comparable effects on action anticipation. *Cortex*, 130, 290–301.
- Poulet, J. F. A., & Hedwig, B. (2006). The cellular basis of a corollary discharge. *Science*, 311(5760), 518–522. <https://doi.org/10.1126/science.1120847>
- Press, C., & Cook, R. (2015). Beyond action-specific simulation: Domain-general motor contributions to perception. *Trends in Cognitive Sciences*, 19(4), 176–178. <https://doi.org/10.1016/j.tics.2015.01.006>
- Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, 24(1), 13–24. <https://doi.org/10.1016/j.tics.2019.11.003>

- Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., & Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, *84*(2), 355–362. <https://doi.org/10.1016/j.neuron.2014.09.033>
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*(1), 13289. <https://doi.org/10.1038/ncomms13289>
- 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.
- Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*(2), 113–121. [https://doi.org/10.1016/S0166-2236\(00\)01685-4](https://doi.org/10.1016/S0166-2236(00)01685-4)
- Rummell, B. P., Klee, J. L., & Sigurdsson, T. (2016). Attenuation of responses to self-generated sounds in auditory cortical neurons. *The Journal of Neuroscience*, *36*(47), 12010–12026. <https://doi.org/10.1523/JNEUROSCI.1564-16.2016>
- Samson, F., Zeffiro, T. A., Toussaint, A., & Belin, P. (2011). Stimulus complexity and categorical effects in human auditory cortex: An activation likelihood estimation meta-analysis. *Frontiers in Psychology*, *1*. <https://doi.org/10.3389/fpsyg.2010.00241>
- SanMiguel, I., Todd, J., & Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP: Auditory N1 suppression: N1 components. *Psychophysiology*, *50*(4), 334–343. <https://doi.org/10.1111/psyp.12024>
- 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. P., & 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>

Schiemann, J., Puggioni, P., Dacre, J., Pelko, M., Domanski, A., van Rossum, M. C. W., & Duguid,

I. (2015). Cellular mechanisms underlying behavioral state-dependent bidirectional

modulation of motor cortex output. *Cell Reports*, 11(8), 1319–1330.

<https://doi.org/10.1016/j.celrep.2015.04.042>

Schneider, D. M., & Mooney, R. (2018). How movement modulates hearing. *Annual Review of*

Neuroscience, 41(1), 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>

Shimaoka, D., Harris, K. D., & Carandini, M. (2018). Effects of arousal on mouse sensory cortex

depend on modality. *Cell Reports*, 22(12), 3160–3167.

<https://doi.org/10.1016/j.celrep.2018.02.092>

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual

inversion. *Journal of Comparative and Physiological Psychology*, 43(6), 482–489.

<https://doi.org/10.1037/h0055479>

Tonnquist-Uhlen, I., Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2003). Maturation of

human central auditory system activity: The T-complex. *Clinical Neurophysiology*, 114(4),

685–701. [https://doi.org/10.1016/S1388-2457\(03\)00005-1](https://doi.org/10.1016/S1388-2457(03)00005-1)

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision

uncertainty and alters serial choice bias. *Nature Communications*, 8(1), 14637.

<https://doi.org/10.1038/ncomms14637>

- van Slooten, J. C., Jahfari, S., Knapen, T., & Theeuwes, J. (2019). Correction: How pupil responses track value-based decision-making during and after reinforcement learning. *PLOS Computational Biology*, *15*(5), e1007031. <https://doi.org/10.1371/journal.pcbi.1007031>
- Vaughan, H. G., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology*, *28*(4), 360–367. [https://doi.org/10.1016/0013-4694\(70\)90228-2](https://doi.org/10.1016/0013-4694(70)90228-2)
- Vinck, M., Batista-Brito, R., Knoblich, U., & Cardin, J. A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, *86*(3), 740–754. <https://doi.org/10.1016/j.neuron.2015.03.028>
- Vitrac, C., & Benoit-Marand, M. (2017). Monoaminergic modulation of motor cortex function. *Frontiers in Neural Circuits*, *11*. <https://doi.org/10.3389/fncir.2017.00072>
- von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, *37*(20), 464–476. <https://doi.org/10.1007/BF00622503>
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, *18*(1), 299–342. <https://doi.org/10.1007/BF02409636>
- Waszak, F., Wascher, E., Keller, P., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental Brain Research*, *162*(3), 346–356. <https://doi.org/10.1007/s00221-004-2183-8>
- Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, *115*(1), 26–38. <https://doi.org/10.1016/j.cognition.2009.10.016>
- Williams, S. R., Shenasa, J., & Chapman, C. E. (1998). Time course and magnitude of movement-related gating of tactile detection in humans. I. Importance of stimulus location. *Journal of Neurophysiology*, *79*(2), 947–963. <https://doi.org/10.1152/jn.1998.79.2.947>

- Wolpaw, J. R., & Penry, J. K. (1975). A temporal component of the auditory evoked response. *Electroencephalography and Clinical Neurophysiology*, *39*(6), 609–620.
[https://doi.org/10.1016/0013-4694\(75\)90073-5](https://doi.org/10.1016/0013-4694(75)90073-5)
- 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](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>
- Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080. <https://doi.org/10.1038/nature05852>
- Yebra, M., Galarza-Vallejo, A., Soto-Leon, V., Gonzalez-Rosa, J. J., de Berker, A. O., Bestmann, S., Oliviero, A., Kroes, M. C. W., & Strange, B. A. (2019). Action boosts episodic memory encoding in humans via engagement of a noradrenergic system. *Nature Communications*, *10*(1), 3534. <https://doi.org/10.1038/s41467-019-11358-8>
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.
<https://doi.org/10.1002/cne.920180503>