Neuropsychologia 218 (2025) 109242

Contents lists available at ScienceDirect

Neuropsychologia

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

Motor-based and memory-based predictions distinctively modulate sensory processes

Xinjing Li [a,b], Qian Chu [d,e,f], Yuhan Lu [a,b], Yuqi Su [g], Xing Tian [a,b,c,*]

[a] *Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China*
[b] *NYU-ECNU Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai, 200062, China*
[c] *Shanghai Frontiers Science Center of Artificial Intelligence and Deep Learning, Division of Arts and Sciences, New York University Shanghai, Shanghai, 200122, China*
[d] *Max Planck - University of Toronto Centre for Neural Science and Technology, Toronto, ON, M5S 2E4, Canada*
[e] *Neural Circuits, Consciousness, and Cognition Research Group, Max Planck Institute for Empirical Aesthetics, 60322, Frankfurt am Main, Germany*
[f] *Krembil Brain Institute, Institute of Biomedical Engineering, University of Toronto, Toronto, ON, M5T 1M8, Canada*
[g] *School of Biological Sciences, The University of Hong Kong, 999077, Hong Kong, China*

ARTICLE INFO

Keywords:
Action-induced suppression
Associative memory
Internal forward model
Predictive coding
Dual-Stream prediction model

ABSTRACT

Action suppresses the neural responses to its sensory feedback. The phenomenon, termed action-induced suppression, highlights the predictive processes in sensorimotor integration but remains controversial regarding the underlying mechanisms. The predictive coding framework posits that action-induced suppression is a general, non-action-specific process driven by predictions. In contrast, the Dual-Stream Prediction Model (DSPM) argues that motor-based and memory-based predictions are mediated by distinct processes — motor predictions rely on precise action-perception mappings and temporal synchrony, whereas memory predictions are based on learned associations. To test these competing theories, we compared auditory ERP responses elicited by self-initiated keypresses (motor-based) and visually cued auditory events (memory-based) in a matching judgment task. Results revealed significant suppression at the P2 component, when the prediction matched the auditory feedback only in the motor-auditory task but not in the visual-auditory task. The findings qualitatively replicated common observations of action-induced suppression; the suppression effects are at a later component rather than N1, indicating the interaction between prediction and perception at a higher level, such as syllable categorization in the current experimental design. Surprisingly, we observed N1 enhancement to the auditory probe in both conditions, with greater enhancement in the motor-auditory task compared to the visual-auditory task. The enhancement effects likely reflect a prediction-induced attentional-like modulation at an early auditory processing stage, potentially driven by the demands of the matching judgment task. Together, these findings support the DSPM by demonstrating functional dissociable mechanisms of motor-based and memory-based predictions.

Action-induced suppression refers to the reduced intensity of neural perceptual responses to sensory stimuli resulting from self-generated actions. Such attenuation of perceptual responses caused by action has been observed across various sensory modalities, including vision, hearing, and touch (Cardoso-Leite et al., 2010; Weiss et al., 2011; Shergill et al., 2003). The suppressive effects are hypothesized as an important neural indicator to differentiate self-produced sensory feedback from external stimuli, preventing sensory overload and supporting a stable perceptual experience (Frith et al., 2000).

Audition is a representative domain for investigating the action-induced suppression with a commonly used contingent paradigm (McCarthy and Donchin, 1976; see also Horváth, 2015 for a review). The paradigm, by comparing button-press-induced-sound condition with button-press-only and sound-only (passive listening) conditions, isolates action-induced suppression to auditory responses specifically induced by the action of button-press. Human non-invasive electrophysiological studies have consistently found the attenuation in early auditory responses of event-related potential (ERP), namely N1/P2 components in EEG and M100/M200 in MEG (e.g., Aliu et al., 2009; Bäß et al., 2008; Bolt and Loehr, 2021; Cao et al., 2017; Han et al., 2022; Harrison et al.,

---

* Corresponding author. Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China.
E-mail address: xing.tian@nyu.edu (X. Tian).

https://doi.org/10.1016/j.neuropsychologia.2025.109242
Received 22 February 2025; Received in revised form 20 July 2025; Accepted 26 July 2025
Available online 4 August 2025
0028-3932/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2021; Mifsud et al., 2016; Poonian et al., 2015; Timm et al., 2013; Weller et al., 2017).

Although action-induced suppression is consistently observed and is commonly interpreted within the framework of action-specific prediction theories (Bäß et al., 2008; Aliu et al., 2009; Timm et al., 2014, 2016; see also Horváth, 2015 for a review), some researchers argue that the attenuation may reflect a more general, non-action-specific predictive mechanism (Sanmiguel et al., 2013; Dogge et al., 2019a; Korka et al., 2022). For example, several lines of evidence reveal that the attenuation effect may be due to the predictable timing between the action and the onset of the sounds (e.g., Horváth et al., 2012; Kaiser and Schütz-Bosbach, 2018). Kaiser and Schütz-Bosbach (2018) found that by introducing a countdown, the external cues boost the timing predictability and lead to a stronger attenuation to other-induced sounds than that of self-induced sounds, suggesting a possibly non-action-specific predictive mechanism.

This inconsistency in explaining action-induced suppression highlights a theoretical debate about predictive mechanisms in a broader context. The predictive coding mechanism posits that the brain continuously generates predictions of sensory input and compares these predictions with actual sensory input to update internal models. According to this theory, action and memory predictions are governed by a unified predictive mechanism. When predictions align with actual input, sensory suppression occurs, reducing the processing of redundant information (Friston, 2005, 2010). This view supports the idea that sensory suppression is a general, non-action-specific process driven by predictions. In contrast, the Dual-Stream Prediction Model (DSPM) suggests that action and memory predictions are independent and complementary mechanisms (Tian and Poeppel, 2013; Tian et al., 2016). In this model, action predictions are based on precise action-perception mappings and correspond to the sensory consequences of voluntary actions, with a requirement for temporal synchronization between actions and their sensory outcomes (Tian and Poeppel, 2015). Memory predictions, on the other hand, depend on experience and associations between sensory inputs. According to DSPM, action-induced suppression is mainly realized by motor-based prediction (Wolpert and Ghahramani, 2000; Shadmehr et al., 2010), while learning-modulated sensory processing (such as the association between certain visual and auditory stimuli) is underlined by memory-based prediction (Kok et al., 2012a; Todorovic et al., 2011; see also Chu et al., 2023 for a review).

Therefore, contrasting motor-based predictions with memory-based predictions under comparable conditions is crucial to differentiate the two theories and to investigate the nature of predictive processes. In the present study, we compared ERP responses to self-action-induced sounds (action-auditory congruence) and visual-cue-induced sounds (visual-auditory congruence), and examined the changes in auditory responses caused by different types of predictions. If both motor-based and memory-based predictions are mediated by a common predictive mechanism, as hypothesized by the predictive coding theory, one would expect similar attenuated responses in both cases. However, according to the Dual-Stream Prediction Model (DSPM), motor-based and memory-based predictions are governed by distinct mechanisms, action-auditory congruence and visual-auditory congruence should yield different suppression effects.

1. Method

*Partici*pants. A total of 27 participants (20 females; age $M = 22.37$ years ($SD = 2.68$), range = 18–29 years) were enrolled in the study. We conducted an a priori power analysis for our primary paired-samples *t*-test using the "solve_power" method (Seabold and Perktold, 2010). Drawing on Weller et al. (2017; $t(15) = 2.67$, $p = 0.017$), who reported Cohen's $d = 0.67$ for P2 attenuation over Cz in a Motor-Auditory vs. Auditory contrast, we set $\alpha = 0.05$ and $1 - \beta = 0.90$. Solving the t-distribution power equation yielded a required total sample of approximately 25.4 to ensure >90 % power to detect an effect of this size. All participants had normal hearing and normal or corrected-to-normal vision, with no self-reported neurological deficits, and all were right-handed. They received monetary compensation for their participation. Written informed consent was obtained from all participants. The study was approved by the Institutional Review Board at New York University Shanghai.

1.1. *Materials and experimental design*

Two audible syllables (/ba/and/ga/) were used as auditory stimuli. All stimuli were synthesized using the Neospeech web engine (www.neospeech.com) at a sampling rate of 44.1 kHz in a male voice (c.f., Li et al., 2020). Each stimulus had a duration of 400 ms. Auditory presentations were binaural, with an intensity normalized to 70 dB SPL using Praat. The sound stimuli were delivered via plastic air tubes connected to foam earplugs (ER-3C Insert Earphones; Etymotic Research). The study employed a Cedrus Response Pad RB-740, a high-precision device with 7 keys and a 2–3 ms response time resolution.

The experiment included four task sessions using a block design.

All participants completed the four blocks in the same fixed order: 1) Motor-Auditory (practice: 40 trials → test: 144 trials); 2) Visual-Auditory (practice: 40 trials → test: 144 trials); 3) Passive Listening (144 trials); 4) Motor-Only (144 trials).

Motor-Auditory task: During the practice session, which consists of 40 trials, participants first observed a central fixation on the screen before choosing to press either the third or the fifth key on the response pad, respectively using the index fingers of both hands. Each keypress triggered an auditory stimulus,/ba/or/ga/after a maximum delay of approximately 16 ms. Participants were trained to associate each sound with a specific key. All participants learned the associations by the end of the practice session, as confirmed by their verbal description of the key-sound association. The subsequent test session, consisted of 144 trials. Each trial began with the display of a fixation cross. Participants were free to press one of the two keys, followed by a sound. In five-sixths of the trials, the sound was congruent with the key pressed (motor-match condition), whereas in one-sixth of the trials, the sound was incongruent (motor-mismatch condition). 500 ms after sound offset, a visual prompt was presented and asked participants to determine whether the sound matched the keypress. Participants responded by pressing one of the two keys on the response pad with their middle fingers. The association between keys and sounds was counterbalanced across participants.

Visual-Auditory task: During the practice session, which consists of 40 trials, participants were trained to associate visual symbols (a circle or a square) with auditory stimuli,/ba/or/ga/. Mastery of these associations was confirmed by every participant through direct questioning at the end of the practice. In the test session, which consists of 144 trials, one of the visual symbols was randomly selected and presented, followed by a syllable. In five-sixths of the trials, the syllable was congruent with the visual symbols (memory-match condition), whereas in one-sixth of the trials, the sound was incongruent (memory-mismatch condition). The duration of the visual symbol was determined by the average response times from the previous motor-auditory task, excluding any outliers more than three standard deviations from the mean. 500 ms after sound offset, a visual prompt was presented and asked participants to determine whether the sound matched the keypress. Participants responded by pressing one of the two keys on the response pad with their middle fingers. The associations were counterbalanced.

Passive Listening: Participants passively listen to sounds. To ensure attentive participation, in one out of six trials, a question follows the sound display, asking whether the auditory stimulus was/ba/or/ga/. Participants responded by pressing one of the two keys on the response pad with their middle fingers. This session consists of 144 trials.

Motor-Only task: During this session, participants perform key presses, choosing either the third or the fifth key on the response pad at intervals of 2–3 s. Earplugs are worn to ensure no auditory feedback influences their responses, maintaining a purely motor-focused task. This session also consists of 144 trials.

Practice for each block was administered immediately before its test trials. The Motor-Auditory block was run first because the average keypress latency measured in this block determined the cue-to-sound interval for both the Visual-Auditory and Passive Listening tasks. The Visual-Auditory block preceded Passive Listening so that participants would form reliable temporal predictions during its practice phase. This fixed order was held constant across all participants. We placed the Motor-Only block last because it is an active yet simple task and ensures that efference-copy predictions formed during motor-auditory trials would not be weakened by the Motor-Only block in which simple keypress was not associated with any sound output.

1.2. *EEG data acquisition and preprocessing*

EEG signals were recorded using a 32-channel Brain Products actiCHamp system. The electrodes were positioned according to the 10–20 International electrode placement system. To monitor ocular activity, the electrooculogram (EOG) was recorded using four additional electrodes: the vertical EOG (VEOG) was monitored with electrodes placed above and below the left eye, while the horizontal EOG (HEOG) was recorded using a bipolar montage with electrodes placed on the right and left outer canthi. Electrode impedances were kept below 10 kΩ. The Cz electrode was used as the online reference.[1] An online low-pass filter with a 200 Hz cutoff and a notch filter at 50 Hz were applied. The EEG signals were digitized at a sampling rate of 1000 Hz.

EEG data preprocessing was performed using the EEGLAB toolbox in MATLAB. The data were down-sampled to 500 Hz and bandpass filtered between 0.1 Hz and 30 Hz. The data were then re-referenced to the average of all scalp electrodes. Bad channels were visually identified and repaired using spherical spline interpolation (Perrin et al., 1989). Epochs with particularly large fluctuations were removed within the time window from −200 ms to 2000 ms relative to the auditory probe onset. Independent component analysis (ICA) was performed on segmented data (−200 ms–2000 ms relative to the auditory probe onset) to remove artifacts, followed by manual inspection. Epochs of 700 ms were segmented from −100 ms to +600 ms relative to the onset of the auditory probe, with the first 100 ms serving as a pre-stimulus baseline. Baseline correction was applied using the 100 ms pre-stimulus period. Epochs were rejected if the amplitude exceeded a threshold criterion of ±100 μV. Epochs containing missing or incorrect responses were excluded. The average rejection rate was 9.05 %.

1.3. *Data analysis*

1.3.1. *Behavior*

For the motor-auditory and visual-auditory tasks, response times (RTs) were calculated as the interval between the onset of the matching prompt and the participant's response.[2] RTs were treated as a behavioral index of prediction error because unexpected events typically incur additional processing costs and slow responses (Hsu et al., 2013; Parmentier et al., 2022). For each participant and each condition, outlier trials of ±3 standard deviations away from the mean were first removed. The mean RT of the remaining trials was then calculated separately for Match and Mismatch trials in both the motor–auditory and visual-auditory tasks. A two-way repeated-measures analysis of variance (ANOVA) was conducted with the factors of Task (motor-auditory vs. visual-auditory) and Congruency (match vs. mismatch) to assess differences in RT. When significant main effects or interactions were found, Bonferroni-corrected paired-sample t-tests were used for post hoc comparisons. Effect sizes were reported using Cohen's *d* for t-tests and partial eta squared (η^2) for the ANOVAs.

1.3.2. *EEG*

To determine the ERP components of interest, we used the collapsed localizer approach (Luck and Gaspelin, 2017) to define unbiased time windows for each component. First, we collapsed the ERP waveforms across all conditions and participants to identify the peak latencies. Based on this collapsed waveform, we defined the following time windows: the N1 was measured from 0.071 to 0.101 s, the P2 from 0.155 to 0.185 s, the P3a from 0.285 to 0.335 s, and the P3b from 0.465 to 0.515 s. We confined our N1/P2 analysis to the "match" trials in which auditory stimuli were consistent with prediction. This selection was to first isolate neural responses and effects of fulfilled predictions. Mismatch trials (prediction-violation) were excluded in this analysis because they evoke additional error-related signals (e.g., mismatch negativity; Garrido et al., 2009) and attention-related signals (e.g., early P3a; Escera et al., 2000; Polich, 2007) that could temporally overlap with auditory ERPs and complicate interpretation. For electrode selection—taking into account the scalp topography, relevant literature (e.g., Bednark et al., 2015; Han et al., 2022), and our 32-channel electrode setup—we extracted the N1 and P2 components from electrodes Fz and Cz, which provided a clear representation of these components. P3a indexes rapid attentional orienting to unexpected events and peaks at fronto-central sites (Cz, Fz/Cz) (Polich, 2007; Schröger et al., 2015; Korka et al., 2019; Baess et al., 2011). We therefore extracted P3a from Fz and Cz as the frontal counterpart to P3b, which we measured at Pz (Pz/POz), where amplitudes rise monotonically from Fz→Cz→Pz and peak over posterior parietal sites (Polich, 2007; Comerchero and Polich, 1998; Bednark et al., 2015). This anterior-posterior ROI pairing maximizes SNR and illustrates the frontal-to-parietal cascade of prediction-error processing.

To remove the influence of action-related potentials on the auditory-evoked responses in the motor–auditory condition, we followed these steps. First, in the motor–auditory condition, participants pressed a button that triggered an auditory stimulus, and data were recorded with time zero aligned to the actual onset of the sound. Next, in the "only motor" condition, participants executed the same button-press movement without producing any sound (i.e., headphones were disconnected). However, a "virtual" auditory onset was recorded via a trigger box, ensuring that the latency between the button press and the virtual auditory marker was matched to that in the motor–auditory condition. Finally, we subtracted the "only motor" data—which contained only movement-related potentials—from the motor–auditory data, yielding difference waveforms that primarily reflect the auditory-evoked component (see Appendix Fig. 1 for a detailed depiction of the ERP waveforms and topographic maps across the three conditions: the uncorrected motor–auditory condition, the motor-only condition, and the motor–auditory condition minus the motor-only condition).

For the ERP analyses, paired-sample t-tests were used to compare the mean amplitudes of the ERP components between tasks (e.g., motor–auditory vs. passive listening; visual–auditory vs. passive listening). In addition, a 2 × 2 repeated-measures ANOVA with the factors Task (motor-auditory vs. memory-auditory) and Congruency (match vs. mismatch) was applied to the P3b amplitude to assess later-

---

[1] We chose Cz as the online reference because its central location on the scalp minimizes impedance and motion-related noise compared to alternatives like the nose tip or mastoids (Yao et al., 2019). Although Cz has no separate channel in the raw data, we subsequently re-referenced offline to the average of all scalp electrodes—a "memoryless" monopolar method that preserves all inter-electrode differences without information loss (Yao et al., 2019). This two-step referencing scheme is widely endorsed: Hu et al. (2019) demonstrate that transformations between different monopolar references are equivalent; and Zhang et al. (2020) show that average referencing effectively corrects biases introduced by single-electrode referencing.

[2] Note that two participants were removed from the RT analysis due to a malfunction with the button board. However, since the accuracy of their button presses was recorded, these participants were not excluded from the EEG analysis.

stage processing effects. Bonferroni-corrected post hoc comparisons were performed. Effect sizes were computed using Cohen's d for t-tests and partial η^2 for the ANOVAs.

2. Results

2.1. Behavioral results

Participants demonstrated high accuracy across all tasks: motor-auditory task (97.18 %, $SD = 4.72$ %), visual-auditory task (98.84 %, $SD = 1.31$ %), and passive listening (99.53 %, $SD = 0.79$ %). In the Motor-Auditory task block, the averaged self-initiated key-press time was 0.62 s ($SE = 0.07$). For the following matching judgment, a two-way repeated-measures ANOVA on response times (RTs) with the factors of Task (motor-auditory task vs. visual-auditory task) and Congruency (match vs. mismatch) revealed a significant main effect of Congruency, $F(1, 24) = 9.49$, $p = 0.005$, partial $\eta^2 = 0.28$, indicating that match trials yielded faster responses than mismatch trials. The main effect of the Task was not significant, $F(1, 24) = 0.68$, $p = 0.418$, partial $\eta^2 = 0.03$, suggesting no overall RT difference between action-induced prediction and memory-associative conditions. However, there was a significant Task × Congruency interaction, $F(1, 24) = 11.32$, $p = 0.003$, partial $\eta^2 = 0.32$, indicating that the effect of Congruency depended on the Task. Simple effects analyses showed that, within the motor-auditory task, match trials were significantly faster than mismatch trials, $t(24) = -4.27$, $p\_Bonf < 0.001$ (Bonferroni-corrected), $d = 0.85$. In contrast, within the visual-auditory task, the difference between match and mismatch trials was not significant, $t(24) = -0.56$, $p\_Bonf = 0.580$, $d = 0.11$. Thus, only the motor-auditory task exhibited a significant RT cost for mismatches (see Fig. 1).

2.2. ERP results

Fig. 2a illustrates the ERP responses across three tasks. Typical N1/P2 auditory responses with clear topographies were observed in all tasks. The statistical tests on the N1 component revealed significant differences among tasks (Fig. 2b). Specifically, the motor-auditory task, after accounting for motor ERP, exhibited a significantly more negative mean amplitude compared to the passive listening task, ($t(26) = -6.24$, $p < 0.001$, $d = -1.21$). Additionally, the visual-auditory task showed a significantly more negative mean amplitude relative to the passive listening task, ($t(26) = -2.74$, $p = 0.011$, $d = -0.57$). Moreover, the motor-auditory condition elicited a significantly larger (more enhanced) N1 amplitude than the visual-auditory condition ($t(26) = -5.25$, $p < 0.001$, $d = -1.01$). These results indicate that both the motor–auditory and visual-auditory tasks elicited an N1 enhancement effect when the sound was consistent with expectations. Contrary to previous findings of N1 suppression, the enhancement effects could be driven by the demands of the matching judgment task that was unique in this experiment. The greater enhancement in the motor-auditory task compared to the visual-auditory task suggests that motor-based prediction could induce stronger attentional-like modulation on early auditory processing than memory-based prediction, demonstrating quantitative differences between the two types of predictions.

As shown in Fig. 2a and b, for the P2 component, significant differences were observed between the motor-auditory task and passive listening. The motor-auditory task demonstrated a significantly less positive mean amplitude compared to the passive listening, ($t(26) = -5.16$, $p < 0.001$, $d = -0.99$). In contrast, no significant difference was found between the visual-auditory task and the passive listening, ($t(26) = 0.10$, $p = 0.918$, $d = 0.02$). These results demonstrated that motor-based predictions induced the suppression effects, whereas memory-based predictions did not. Note that compared to commonly observed action-induced suppression at the N1 component, we observed significant suppression at a later P2 component when motor-based predictions matched the auditory feedback. The latency shift could be caused by the manipulation at the syllable level.

We extracted P3a amplitudes (see Fig. 2d–f) from an anterior-

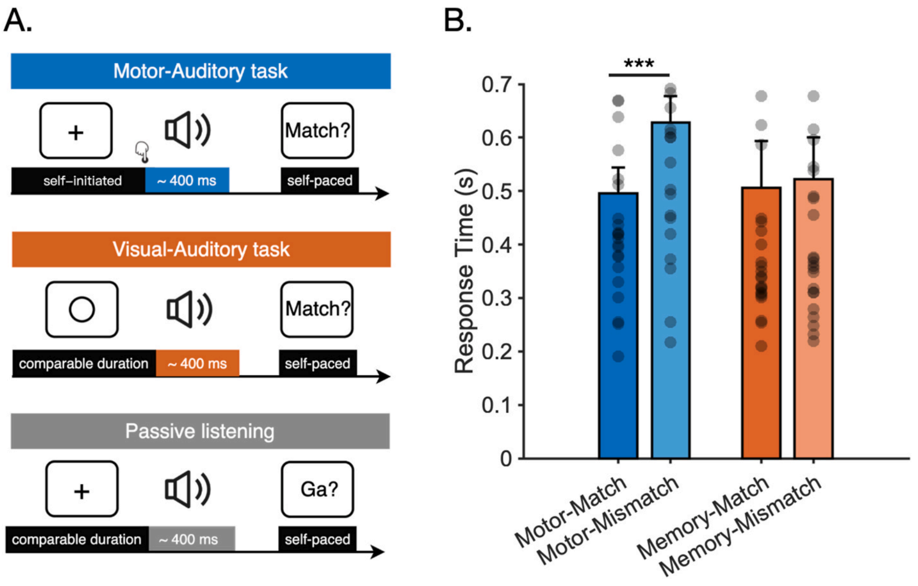

Fig. 1. Experimental procedures and behavioral results. A) Experimental procedures. The top row depicts a trial in the motor-auditory task where participants learned the associations between two syllable sounds (/ba/and/ga/) and two keys before the experiment. In each trial, after a fixation cross was presented, participants were free to press one of two keys. A sound was played after pressing a key, followed by a question asking whether the played sound matched the one associated with the key press. The middle row illustrates the visual-auditory task, where participants knew the associations between two visual symbols and syllable sounds before the experiment. In each trial, one of the two symbols was randomly presented followed by a sound. The lag between the visual cue onset and sound onset was the same duration as the average time of initiating key-press in the motor-auditory task. The similar duration among conditions aims to equalize temporal prediction caused by preceding stimuli and sounds. A question was prompted after the sound, asking whether the played sound matched the one associated with the visual cue. The bottom row represents the passive listening task. After a fixation appeared on the screen with a similar duration as the other two conditions, participants listened to a syllable sound. In one out of six trials, they were asked to identify the sound. B) Behavioral results. Response time in four conditions: motor-match, motor-mismatch, memory-match, and memory-mismatch. Each dot represents the average response time for an individual participant. ***$p < 0.001$.

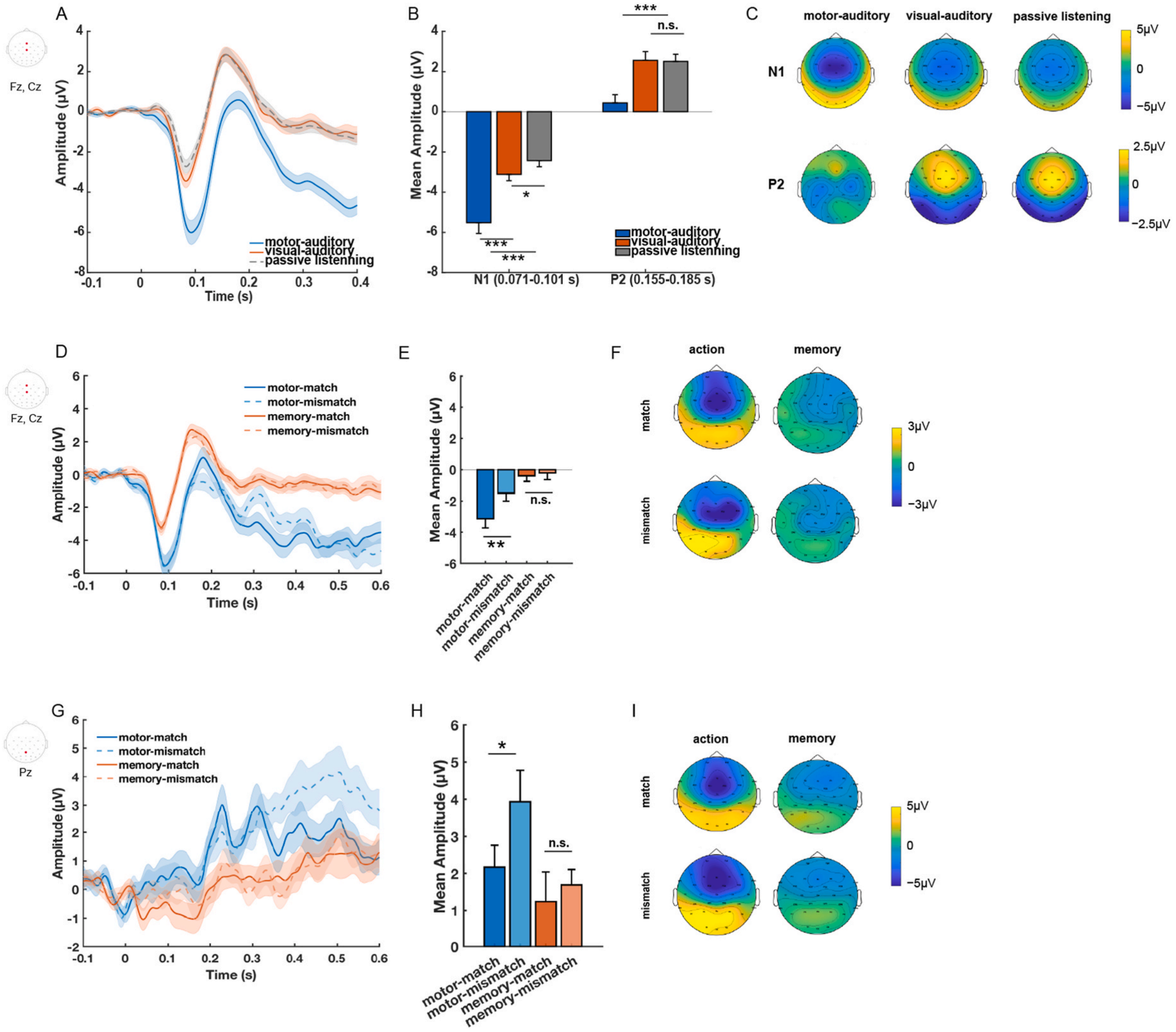

Fig. 2. ERP results. A) Grand-averaged ERP waveforms at Fz and Cz for the motor-auditory (blue), visual-auditory (orange) and passive-listening (grey) tasks. Shaded areas denote ±1 SEM. B) Mean N1 (71–101 ms) and P2 (155–185 ms) amplitudes in those three tasks (averaged across Fz and Cz); error bars = ±1 SEM. C) Scalp topographies showing the average voltage distribution over the entire N1 window (71–101 ms; top row) and P2 window (155–185 ms; bottom row) for each task. D) Grand-average ERPs at fronto-central electrodes (averaged across Fz and Cz) for the four conditions: motor-match (blue solid), motor-mismatch (blue dashed), memory-match (orange solid) and memory-mismatch (orange dashed); Shaded areas denote ±1 SEM. E) Mean P3a amplitude (285–335 ms) for each condition; error bars = ±1 SEM. F) Scalp topographies of the 285–335 ms window for match (top row) and mismatch (bottom row) trials. G) Grand-averaged ERP waveforms at Pz for the four experimental conditions: motor-match (blue solid), motor-mismatch (blue dashed), memory-match (orange solid) and memory-mismatch (orange dashed). Shaded areas denote ±1 SEM. H) Mean P3b (465–515 ms) amplitudes for those four conditions (averaged across Pz); error bars = ±1 SEM. I) Scalp topographies showing the average voltage distribution over the entire P3b window (465–515 ms) for match (top row) and mismatch (bottom row) trials. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; n.s., not significant.

midline ROI (Fz, Cz) and ran a 2 × 2 repeated-measures ANOVA with factors Task (motor-auditory vs. visual-auditory) and Consistency (Match vs. Mismatch). The analysis revealed a significant main effect of Task, $F(1, 26) = 24.77$, $p < 0.001$, *partial* $\eta^2 = 0.49$; a significant main effect of Consistency, $F(1, 26) = 9.90$, $p = 0.004$, *partial* $\eta^2 = 0.28$; and a marginal Task × Consistency interaction, $F(1, 26) = 4.04$, $p = 0.055$, *partial* $\eta^2 = 0.13$. Bonferroni-corrected simple effects showed that in the motor-auditory task, P3a for Mismatch ($M = -1.50$ μV, $SE = 0.50$ μV) was significantly less negative (i.e. larger) than for Match ($M = -3.14$ μV, $SE = 0.58$ μV), $t(26) = -3.65$, $p\_Bonf = 0.002$, $d = -0.70$; whereas in the visual-auditory task there was no reliable difference between Mismatch ($M = -0.22$ μV, $SE = 0.42$ μV) and Match ($M = -0.40$ μV, $SE = 0.35$ μV), $t(26) = -0.37$, $p\_Bonf = 1.000$, $d = -0.07$. These results suggest that, in the P3a time window, motor-based predictions drive an enhanced P3a response to mismatches, whereas memory-based predictions do not.

We further examined the later P3b component to quantify the possible differences among tasks and congruency in later effects. A 2 × 2 repeated-measures ANOVA was conducted to assess the effects of Task (motor-auditory vs. visual-auditory) and Congruency (Match vs. Mismatch) on the mean amplitude within the P3b time window (see Fig. 2g–i)). The results revealed significant main effects of Task, $F(1, 26) = 8.56$, $p = 0.007$, partial $\eta^2 = 0.25$, and Congruency, $F(1, 26) = 5.64$, $p = 0.025$, partial $\eta^2 = 0.18$. Additionally, the interaction was significant, $F(1, 26) = 4.51$, $p = 0.043$, partial $\eta^2 = 0.15$. Bonferroni-corrected simple effects analyses showed that, in the motor-auditory task, the amplitude of the P3b elicited by Mismatch auditory stimuli ($M = 3.76$, $SE = 0.65$ μV) was significantly larger than that elicited by Match stimuli ($M = 2.01$, $SE = 0.40$ μV), $t(26) = -3.81$, $p\_Bonf = 0.002$, $d = -0.73$. In contrast, in the visual-auditory task, no significant difference was observed between Match ($M = 1.47$, $SE = 0.52$ μV) and Mismatch ($M = 1.56$, $SE = 0.37$ μV), $t(26) = -0.15$, $p\_Bonf = 1.000$, $d = -0.03$. These results suggest that, in the P3b time window, the motor-based prediction is more sensitive to manipulations of Consistency, whereas memory-based prediction is not.

3. Discussion

The current study systematically compared the effects of keypress-induced sounds and cross-modal visual-auditory associations on auditory processing. By examining ERPs and corresponding behavioral measures, our findings revealed that both motor-induced sounds and cross-modal memory cues, when aligned with expectations, elicit enhanced N1 components, indicative of an enhancement effect. Moreover, only motor-based predictions gave rise to the classic P2 suppression, which manifests as reduced perceptual processing of self-generated sounds. Consistent with these electrophysiological patterns, response times and P300 amplitudes in the motor-auditory task were also more sensitive to match-mismatch distinctions. In contrast, although cross-modal memory predictions can heighten early perceptual sensitivity, they did not exert notable suppression or resource reallocation effects on the later components (P2 and P300). Taken together, these observations support the DSPM, indicating the coexistence of motor-based and memory-based predictive mechanisms.

Contrary to previous reports of attenuated N1 amplitudes in self-generated sound paradigms (Baess et al., 2011; Ford et al., 2014; Horváth et al., 2012; SanMiguel et al., 2013; Knolle et al., 2012), our results demonstrated significant suppression at the P2 component in the motor-auditory task. One potential explanation for this discrepancy of effect latency is differences in the hierarchical level of predictions engaged by the stimuli. Hierarchical predictive-coding models (Tian and Poeppel, 2013, 2015) suggest that an initial processing stage around 100 ms is sensitive to basic acoustic features, whereas a later stage around 200 ms encodes more abstract linguistic categories. Consistent with this model, Tian and Poeppel (2015) observed pitch-related modulations at the latency of the N1 component (~100 ms), while Tian and Poeppel (2013) reported syllable-level effects at the latency of the P2 component (~200 ms). Notably, most seminal studies reporting N1 suppression employed simple pure tone as stimuli—primarily 1000 Hz sine waves (Baess et al., 2011; Ford et al., 2014; Horváth et al., 2012; SanMiguel et al., 2013) or a 600 Hz tone (Knolle et al., 2012)—which predominantly engage early acoustic processing stages. In contrast, in this study, we used speech syllables (/ba/and/ga/) as auditory probes. The prediction could extend to the abstract phonological level and interact with the perceptual responses, and hence yield the suppression at a later P2 component. This finding aligns with emerging evidence that P2, rather than N1, more directly indexes self-specific attenuation, pointing to a later stage of stimulus evaluation and categorization in the auditory hierarchy (Bolt and Loehr, 2021; Cao et al., 2017; Crowley and Colrain, 2004; Han et al., 2022; Mifsud et al., 2016; see also Korka et al., 2022 for a review).

In the visual-auditory task, we observed no reduction in P2 amplitude, whereas in the motor-auditory task a clear P2 attenuation emerged. Another possible explanation for this discrepancy is that it reflects the influence of temporal predictability on P2 amplitude. For example, Behroozmand et al. (2016) showed that self-generated sounds with predictable timing evoked reduced P2 amplitudes, whereas unpredictable timing evoked enhanced P2. Similarly, Pinto et al. (2019) reported that providing explicit temporal cues (i.e., reliable “when” information) attenuates the N1/P2 complex; removing temporal predictability abolishes this attenuation. In our experiment, although we approximately equalized the lag between the cue and auditory feedback in both conditions, the motor-auditory task featured near-instantaneous synchronization between keypress and auditory feedback, providing an additional cue for high temporal predictability; whereas in the visual-auditory task the fixed delay between the visual cue and sound onset remained less precise compared to the immediate feedback in the action-auditory condition. Thus, the high predictability in the motor-auditory task could underlie the P2 attenuation, whereas the lower predictability in the visual-auditory task may have diminished or even reversed that effect. Future studies should take temporal aspects into consideration and match temporal predictability as closely as possible, for example using the same fixed or jittered cue-feedback intervals, so that any observed P2 effects reflect modality rather than timing differences.

Surprisingly, we observed N1 enhancement in both tasks, which diverges from the commonly observed action-induced N1 suppression (e.g., Baess et al., 2011; Ford et al., 2014; Horváth et al., 2012; Knolle et al., 2012; SanMiguel et al., 2013). This reversal of the expected N1 pattern likely reflects an interaction between attentional gain and predictive suppression. Predictive coding typically suppresses neural responses to expected stimuli, enhancing the salience of unexpected events (Friston, 2005, 2010). However, this suppression can be modulated or reversed when attention proactively targets the anticipated sensory input (Kok et al., 2012b). Supporting this, Timm et al. (2013) showed that explicit auditory attention universally enhances auditory N1 subcomponents. Moreover, attentional processes may partially superimpose the self-generation effect on the N1 (Saupe et al., 2013; Korka et al., 2022). In the current study, the incorporation of a sound-keypress matching task markedly increased auditory attentional demands, prompting continuous monitoring and rapid congruency assessment. This intensified attention may counteract or surpass traditional predictive suppression driven by motor- or memory-based expectations, thereby producing the observed N1 enhancement. Consistent with our hypothesis, in a tactile study, Thomas et al. (2022) found that when active tasks involving probabilistic prediction were introduced, expected tactile events were perceived as more intense than unexpected ones, suggesting that prediction enhances sensory perception of expected stimuli. Future studies could thus consider methodological improvements such as: (1) employing passive paradigms to isolate predictive effects, (2) independently manipulating attentional load and prediction factors within factorial designs, and (3) examining various cross-modal or non-linguistic stimuli. These approaches would allow clearer separation of attentional and predictive influences on early ERP components, refining our understanding of predictive coding and motor efference copy mechanisms.

From the theoretical aspects of the interaction between prediction and attention, both Predictive Coding (Friston, 2005, 2010) and the Dual Stream Predictive Model (DSPM; Tian and Poeppel, 2013) propose that neural responses—especially the N1/P2 complex—should be attenuated when sensory feedback matches internal predictions. Our observation of an N1 enhancement thus reveals a possible boundary condition: when the task explicitly elevates auditory attention (e.g., through a matching judgment), attention-driven gain mechanisms can override or even reverse classical predictive suppression. This suggests that both theoretical models require refinement to explicitly incorporate interactions between prediction and attention, and that predictive suppression might predominantly emerge under low-attention or passive paradigms. Building on our findings, it is instructive to consider how motor-based versus memory-based predictions differentially engage attentional gain mechanisms. For example, Li et al. (2020) showed that showed that preparing to speak enhanced the auditory responses to the expected speech output. Our recent studies show that manual movement could alleviate the attentional blink effect, implying a motor-driven gain mechanism (unpublished). These findings posit that efference copy or corollary discharge in the motor-based prediction may serve as an endogenous source of attentional amplification, selectively boosting the sensory gain of predicted action outcomes. In the study, we observed that the N1 enhancement in the action condition was significantly larger than in the memory condition, consistent with the view that motor-based prediction may carry a possible gain control mechanism. These results indicated a task-modulated precision difference: action-driven prediction via efference copy serves as an online “active” channel for enhancing basic auditory feature processing, whereas memory-driven prediction operate based on less precise past experience (Chu et al., 2023; Tian and Poeppel, 2013; Tian et al., 2016; Ma and

Tian, 2019).

Dogge et al. (2019b) employed a two-phase paradigm in which participants first learned associations between keypresses or visual symbols and pure tones (low and high frequencies) and then performed a two-alternative forced-choice loudness discrimination task. They found that identity-specific motor prediction effects were weak and influenced by learning history—suggesting that sensory attenuation may primarily reflect a general predictive mechanism rather than one that is strictly motor-based. In contrast, participants in our study learned associations between keypresses or visual symbols and speech syllables (/ba/and/ga/) and then used a matching judgment task where participants actively determined whether the played sound corresponded to the pre-learned association. This approach may have more effectively engaged efference copy mechanisms—particularly in the keypress condition—and recruited a more complex speech-processing network, resulting in robust P2 suppression that did not emerge in the visual (memory) condition. Differences in stimulus complexity (speech syllables vs. pure tones), task demands (matching judgment vs. loudness discrimination), and attentional focus during learning and testing likely contribute to the divergent neurophysiological outcomes between our study and Dogge et al.'s.

The current study further explored behavioral and neurophysiological differences under conditions of prediction consistency versus inconsistency. Our results indicate that only in the motor-auditory task did mismatches lead to increased P300 amplitudes, whereas no such significant changes were observed in the visual-auditory task. These findings are consistent with Bednark et al. (2015), who reported that the P300 component is sensitive to prediction-inconsistent information when actions are predictable and accompanied by identity-specific cues. That is, stimuli that violate identity-based predictions evoke higher P300 amplitudes. Motor-based internal forward models and their associated efference copy mechanisms provide an explanation (e.g., Ford et al., 2007; Wolpert and Kawato, 1998). When intended actions do not align with the sensory feedback, the brain must update the contextual model—recruiting additional cognitive resources in the process. This resource-intensive update is reflected in both elevated P300 amplitude and prolonged response times (Feldman and Friston, 2010; Fogelson, 2015; Polich, 2007). By contrast, memory-based predictions may not elicit the same degree of late-stage cognitive engagement or resource allocation—possibly because the two syllables used in our study differed sufficiently—thus showing limited or no effects on P300 amplitude and behavioral performance.

Furthermore, our data revealed that self-initiated sounds elicited higher P3b amplitudes than externally triggered audiovisual stimuli. These observations are consistent with Bednark et al.'s findings that self-triggered inputs induce more pronounced P3b responses. A likely contributing factor is the sense of agency: when individuals initiate actions themselves, the brain generates efference copies to predict sensory outcomes, thereby strengthening the coupling between action and perception (Blakemore et al., 1999, 2000; Waszak and Herwig, 2007; Wolpert and Ghahramani, 2000). When these predictions are violated, additional cognitive resources must be marshaled—evidenced by elevated P3b activity—to reconcile discrepancies between expected and actual sensory input. Taken together, these findings suggest that the coupling of action and sensation is pivotal in determining how the brain allocates late-stage attentional and cognitive resources, ultimately shaping the electrophysiological signatures of prediction processing.

Taken together the effects in different ERP components underscore the hierarchical nature of predictive processing: although expectation-congruent trials produce enhanced N1 in both motor-based and memory-based conditions, significant P300 amplification emerges mainly in the motor condition during mismatches. This suggests that while forward prediction and self-specific inhibition (e.g., P2 attenuation) may be more relevant to successfully predicted actions, substantial updates to internal models, evidenced by P300, are particularly triggered when self-generated predictions fail to align with actual outcomes.

We analyzed reaction time (RT) as a behavioral proxy of surprise based on predictive-coding frameworks, wherein unexpected events incur additional processing costs and thus prolong RTs. Match trials served as a baseline measure of processing speed, ensuring that any RT cost for mismatch trials reflected true prediction violations rather than general task difficulty. Classic oddball studies consistently show longer RTs for rare, expectation-violating tones (Hsu et al., 2013; Parmentier et al., 2022), and Coy et al. (2024) demonstrated that facilitation effects for repeated deviants vanish when prediction certainty is low. Quantitative links between surprise caused by violation of prediction, amplitude of ERP P300 component, and RT metrics have been observed (Quintela-Vega et al., 2023; Auksztulewicz et al., 2018). We observed a significant RT cost for mismatch trials only in the motor-auditory task, indicating that precise, action-based predictions were violated. By contrast, the visual-auditory (memory-based) task showed no RT difference. We attribute this to the inherently lower temporal precision and reduced precision-weighting of memory-driven predictions (Summerfield and Egner, 2009; Ma and Tian, 2019), which often do not slow RT unless supported by multiple converging cues (Pieszek et al., 2013). Moreover, our P300 results qualitatively mirrored the RT pattern—both neural and behavioral indices showed match-match effects in the motor-auditory task but not in the memory-based task, underscoring their concordance as markers of surprise caused by violation of prediction.

While our ERP results demonstrate functional dissociations between motor- and memory-based predictions, the current evidence cannot test the anatomical hypothesis of the underlying neural circuits (Chu et al., 2023). We are currently using sEEG (stereoelectroencephalography) to explore the neural sources of these predictive mechanisms. Future studies could also employ neural modulation methods, targeting at sensorimotor versus association cortices, to establish causal links and to map anatomical segregation between motor- and memory-driven predictive circuits.

In this study, we compared auditory ERP responses under self-initiated keypresses (motor predictions) and visually cued events (memory predictions) to reveal functionally dissociable mechanisms of suppression and enhancement. We found that significant P2 suppression occurred only in the motor-auditory task when feedback matched the prediction, whereas no such effect was observed in the visual-auditory task. Unexpectedly, N1 amplitudes were enhanced in both conditions—with a markedly greater enhancement in the motor-auditory condition—suggesting that predictive processes can evoke attentional-like gain modulation during early auditory processing. These findings demonstrate that motor- and memory-based predictions modulate sensory processing via functionally and temporally distinct mechanisms, supporting the DSPM.

CRediT authorship contribution statement

Xinjing Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Qian Chu:** Writing – review & editing, Visualization, Methodology, Data curation, Conceptualization. **Yuhan Lu:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Yuqi Su:** Writing – review & editing, Data curation. **Xing Tian:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition.

Acknowledgments

This study is supported by the National Natural Science Foundation of China 32271101, Program of AI-Driven Initiative to Promote Research Paradigm Reform and Empower Disciplinary Advancement by Shanghai Municipal Education Commission (SMEC), Program of

Introducing Talents of Discipline to Universities, Base B16018, and NYU Shanghai Boost Fund to X.T.

We would like to express our gratitude to Keyu Liu, Gefei Guan, and Yao Yi for their assistance with data collection, and we thank Yuchunzi Wu and Zhiyuan Xu for their invaluable help with the experimental procedures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2025.109242.

Data availability

Data will be made available on request.

References

Aliu, S.O., Houde, J.F., Nagarajan, S.S., 2009. Motor-induced suppression of the auditory cortex. J. Cognit. Neurosci. 21 (4), 791–802.

Auksztulewicz, R., Schwiedrzik, C.M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A.C., Friston, K.J., Melloni, L., 2018. Not all predictions are equal: "What" and "When" predictions modulate activity in auditory cortex through different mechanisms. J. Neurosci. 38 (40), 8680–8693.

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

Bednark, J.G., Poonian, S.K., Palghat, K., McFadyen, J., Cunnington, R., 2015. Identity-specific predictions and implicit measures of agency. Psychology of Consciousness: Theory, Research, and Practice 2 (3), 253–268.

Behroozmand, R., Sangtian, S., Korzyukov, O., Larson, C.R., 2016. A temporal predictive code for voice motor control: evidence from ERP and behavioral responses to pitch-shifted auditory feedback. Brain Res. 1636, 1–12.

Blakemore, S.J., Frith, C.D., Wolpert, D.M., 1999. Spatio-temporal prediction modulates the perception of self-produced stimuli. J. Cognit. Neurosci. 11 (5), 551–559.

Blakemore, S.J., Wolpert, D., Frith, C., 2000. Why can't you tickle yourself? Neuroreport 11 (11), 11–16.

Bolt, N.K., Loehr, J.D., 2021. Sensory attenuation of the auditory P2 differentiates self- from partner-produced sounds during joint action. J. Cognit. Neurosci. 33 (11), 2297–2310.

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

Cao, L., Veniero, D., Thut, G., Gross, J., 2017. Role of the cerebellum in adaptation to delayed action effects. Curr. Biol. 27 (16), 2442–2451.

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., Waszak, F., 2010. A new look at sensory attenuation: action-effect anticipation affects sensitivity, not response bias. Psychol. Sci. 21 (12), 1740–1745.

Chu, Q., Ma, O., Hang, Y., Tian, X., 2023. Dual-stream cortical pathways mediate sensory prediction. Cerebr. Cortex 33 (14), 8890–8903.

Comerchero, M.D., Polich, J., 1998. P3a and P3b from typical auditory and visual stimuli. Cogn. Brain Res. 7 (1), 41–50.

Coy, N., Bendixen, A., Grimm, S., Roeber, U., Schröger, E., 2024. Conditional deviant repetition in the oddball paradigm modulates processing at the level of P3a but not MMN. Psychophysiology 61 (6), e14545.

Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin. Neurophysiol. 115 (4), 732–744.

Dogge, M., Custers, R., Aarts, H., 2019a. Moving forward: on the limits of motor-based forward models. Trends Cognit. Sci. 23 (9), 743–753.

Dogge, M., Hofman, D., Custers, R., Aarts, H., 2019b. Exploring the role of motor and non-motor predictive mechanisms in sensory attenuation: perceptual and neurophysiological findings. Neuropsychologia 124, 216–225.

Escera, C., Alho, K., Winkler, I., Näätänen, R., 2000. Neural mechanisms of involuntary attention to acoustic novelty and change. J. Cognit. Neurosci. 12 (3), 540–551.

Feldman, H., Friston, K.J., 2010. Attention, uncertainty, and free-energy. Front. Hum. Neurosci. 4, 215.

Fogelson, N., 2015. Neural correlates of local contextual processing across stimulus modalities and patient populations. Neurosci. Biobehav. Rev. 52, 207–220.

Ford, J.M., Palzes, V.A., Roach, B.J., Mathalon, D.H., 2014. Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. Schizophr. Bull. 40 (4), 804–812.

Ford, J., Gray, M., Faustman, W., Roach, B., Mathalon, D., 2007. Dissecting corollary discharge dysfunction in schizophrenia. Psychophysiology 44, 522–529.

Friston, K., 2005. A theory of cortical responses. Phil. Trans. Biol. Sci. 360 (1456), 815–836.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11 (2), 127–138.

Frith, C.D., Blakemore, S.J., Wolpert, D.M., 2000. Abnormalities in the awareness and control of action. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 355 (1404), 1771–1788.

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

Han, N., Jack, B.N., Hughes, G., Whitford, T.J., 2022. The role of action-effect contingency on sensory attenuation in the absence of movement. J. Cognit. Neurosci. 34 (8), 1488–1499.

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

Horváth, J., 2015. Action-related auditory ERP attenuation: paradigms and hypotheses. Brain Res. 1626, 54–65.

Horváth, J., Maess, B., Baess, P., Tóth, A., 2012. Action-sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. J. Cognit. Neurosci. 24 (9), 1919–1931.

Hsu, Y.-F., Hämäläinen, J.A., Waszak, F., 2013. Temporal expectation and spectral expectation operate in distinct fashion on neuronal populations. Neuropsychologia 51 (13), 2548–2555.

Hu, S., Yao, D., Bringas-Vega, M.L., Qin, Y., Valdes-Sosa, P.A., 2019. The statistics of EEG unipolar references: Derivations and properties. Brain Topogr. 32, 696–703.

Kaiser, J., Schütz-Bosbach, S., 2018. Sensory attenuation of self-produced signals does not rely on self-specific motor predictions. Eur. J. Neurosci. 47 (11), 1303–1310.

Knolle, F., Schröger, E., Baess, P., Kotz, S.A., 2012. The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. J. Cognit. Neurosci. 24 (3), 698–706.

Kok, P., Jehee, J.F., De Lange, F.P., 2012a. Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75 (2), 265–270.

Kok, P., Rahnev, D., Jehee, J.F., Lau, H.C., De Lange, F.P., 2012b. Attention reverses the effect of prediction in silencing sensory signals. Cerebr. Cortex 22 (9), 2197–2206.

Korka, B., Schröger, E., Widmann, A., 2019. Action intention-based and stimulus regularity-based predictions: same or different? J. Cognit. Neurosci. 31 (12), 1917–1932.

Korka, B., Widmann, A., Waszak, F., Darriba, Á., Schröger, E., 2022. The auditory brain in action: intention determines predictive processing in the auditory system—a review of current paradigms and findings. Psychon. Bull. Rev. 29 (2), 321–342.

Li, S., Zhu, H., Tian, X., 2020. Corollary discharge versus efference copy: distinct neural signals in speech preparation differentially modulate auditory responses. Cerebr. Cortex 30 (11), 5806–5820.

Luck, S.J., Gaspelin, N., 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology 54 (1), 146–157.

Ma, O., Tian, X., 2019. Distinct mechanisms of imagery differentially influence speech perception. eNeuro 6 (5).

McCarthy, G., Donchin, E., 1976. The effects of temporal and event uncertainty in determining the waveforms of the auditory event related potential (ERP). Psychophysiology 13 (6), 581–590.

Mifsud, N.G., Oestreich, L.K., Jack, B.N., Ford, J.M., Roach, B.J., Mathalon, D.H., Whitford, T.J., 2016. Self-initiated actions result in suppressed auditory but amplified visual evoked components in healthy participants. Psychophysiology 53 (5), 723–732.

Parmentier, F.B., Gallego, L., Micucci, A., Leiva, A., Andrés, P., Maybery, M.T., 2022. Distraction by deviant sounds is modulated by the environmental context. Sci. Rep. 12 (1), 21447.

Perrin, F., Pernier, J., Bertrand, O., Échallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72 (2), 184–187.

Pieszek, M., Widmann, A., Gruber, T., Schröger, E., 2013. The human brain maintains contradictory and redundant auditory sensory predictions. PLoS One 8 (1), e53634.

Pinto, S., Tremblay, P., Basirat, A., Sato, M., 2019. The impact of when, what and how predictions on auditory speech perception. Exp. Brain Res. 237 (12), 3143–3153.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148.

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

Quintela-Vega, L., Morado-Díaz, C.J., Terreros, G., Sánchez, J.S., Pérez-González, D., Malmierca, M.S., 2023. Novelty detection in an auditory oddball task on freely moving rats. Commun. Biol. 6 (1), 1063.

Sanmiguel, I., Todd, J., Schröger, E., 2013. Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. Psychophysiology 50 (4), 334–343.

Saupe, K., Widmann, A., Trujillo-Barreto, N.J., Schröger, E., 2013. Sensorial suppression of self-generated sounds and its dependence on attention. Int. J. Psychophysiol. 90 (3), 300–310.

Schröger, E., Marzecová, A., SanMiguel, I., 2015. Attention and prediction in human audition: a lesson from cognitive psychophysiology. Eur. J. Neurosci. 41 (5), 641–664.

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python. SciPy 7 (1), 92–96.

Shadmehr, R., Smith, M.A., Krakauer, J.W., 2010. Error correction, sensory prediction, and adaptation in motor control. Annu. Rev. Neurosci. 33 (1), 89–108.

Shergill, S.S., Bays, P.M., Frith, C.D., Wolpert, D.M., 2003. Two eyes for an eye: the neuroscience of force escalation. Science 301 (5630), 187.

Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. Trends Cognit. Sci. 13 (9), 403–409.

Thomas, E.R., Yon, D., de Lange, F.P., Press, C., 2022. Action enhances predicted touch. Psychol. Sci. 33 (1), 48–59.

Tian, X., Poeppel, D., 2013. The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. J. Cognit. Neurosci. 25 (7), 1020–1036.

Tian, X., Poeppel, D., 2015. Dynamics of self-monitoring and error detection in speech production: evidence from mental imagery and MEG. J. Cognit. Neurosci. 27 (2), 352–364.

Tian, X., Zarate, J.M., Poeppel, D., 2016. Mental imagery of speech implicates two mechanisms of perceptual reactivation. Cortex 77, 1–12.

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

Timm, J., SanMiguel, I., Saupe, K., Schröger, E., 2013. The N1-suppression effect for self-initiated sounds is independent of attention. BMC Neurosci. 14, 1–11.

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

Todorovic, A., van Ede, F., Maris, E., de Lange, F.P., 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J. Neurosci. 31 (25), 9118–9123.

Waszak, F., Herwig, A., 2007. Effect anticipation modulates deviance processing in the brain. Brain Res. 1183, 74–82.

Weiss, C., Herwig, A., Schütz-Bosbach, S., 2011. The self in action effects: selective attenuation of self-generated sounds. Cognition 121 (2), 207–218.

Weller, L., Schwarz, K.A., Kunde, W., Pfister, R., 2017. Was it me?–Filling the interval between action and effects increases agency but not sensory attenuation. Biol. Psychol. 123, 241–249.

Wolpert, D.M., Ghahramani, Z., 2000. Computational principles of movement neuroscience. Nat. Neurosci. 3 (11), 1212–1217.

Wolpert, D.M., Kawato, M., 1998. Multiple paired forward and inverse models for motor control. Neural Netw. 11 (7–8), 1317–1329.

Yao, D., Qin, Y., Hu, S., Dong, L., Bringas Vega, M.L., Valdés Sosa, P.A., 2019. Which reference should we use for EEG and ERP practice? Brain Topogr. 32 (4), 530–549.

Zhang, L., Wang, P., Zhang, R., Chen, M., Shi, L., Gao, J., Hu, Y., 2020. The influence of different EEG references on scalp EEG functional network analysis during hand movement tasks. Front. Hum. Neurosci. 14, 367.