Motor-based prediction during preparation of hand movement modulates auditory processing in two distinct directions

Xinjing Li [a,b], Xiaodan Zheng [a,b], Yuchunzi Wu [a,b] [ID], Hao Zhu [a,b] [ID], Yunying Shu [b] [ID], Ruiqi Tong [b], Xing Tian [a,b,c,*] [ID]

[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

ARTICLE INFO

ABSTRACT

Differentiating self-generated from externally induced sounds is crucial for survival. Predictions can be generated based on action-outcome contingency and suppress neural responses to sensory reafference for distinguishing the origin of stimuli. The action-outcome contingency can be flexible or relatively fixed (e.g., keypress could trigger various sounds vs. articulatory gestures generate corresponding speech sounds) and can be available during the entire course of action (including stages of intention, preparation and execution). Are motor-based predictions created equally based on different types of action-outcome contingency and during distinct stages of action? We conducted three EEG experiments to determine how motor preparation modulates auditory processing using a delayed keypress paradigm in which participants prepared to press a key to trigger a sound without knowing what key to press. In Experiment 1, keypress preparation showed overall enhanced N1 responses (~100 ms), largest for syllables, but did not reveal any suppression effects. Experiment 2 replicated N1 enhancement and showed significant P2 suppression (~200 ms) in response to auditory syllables, when participants were pianists who had extensive keypress-sound mapping experience. Experiment 3, when pianists were in their unfamiliar pairing of lab keys and familiar piano tones, again showed N1 enhancement, but the P2 suppression was absent. Together, these results suggest that preparatory motor prediction in an optional mapping can influence auditory processing in multiple directions and motivate a two-stage gain-to-attenuation hypothesis that may depend on the precision (reliability) of action-outcome associations.

1. Introduction

Every action—from vocalization and walking to the manipulation of objects—inevitably generates sensory outcomes, necessitating a neural mechanism that can distinguish self-produced from externally generated stimuli with high reliability, a capacity conserved across species (Blakemore et al., 1998; Crapse and Sommer, 2008; Miall and Wolpert, 1996). Through experience with the contingency between an action and its sensory consequence, the brain can establish a motor-based prediction mechanism (also termed *an internal forward model*), whereby the sensory outcome of an action is predicted from the programmed motor command. Such predictions are thought to suppress neural responses to sensory reafference, thereby labelling the feedback as self-generated (Blakemore et al., 2000; Li et al., 2020; Miall and Wolpert, 1996; Rao,

2024; Zheng et al., 2022). This sensory suppression is widely regarded as a hallmark of predictive processing for self-generated sensory events.

Different types of contingencies exist between actions and their sensory consequences. For instance, pressing a key may result in different sensory events depending on the device and setup, illustrating *optional sensorimotor mapping,* where the action-outcome relationship is flexible. In contrast, speech production involves a more rigid mapping, with each articulatory gesture consistently producing a specific sound—an example of *mandatory sensorimotor mapping.* Do motor-based predictions suppress sensory reafference in the same way across these different types of sensorimotor mappings? Addressing this question with keypress-auditory mappings will clarify whether such predictive modulation is a general principle across systems.

Studies have consistently shown that neural responses to self-

* Corresponding author.
  E-mail address: xing.tian@nyu.edu (X. Tian).

generated sounds – whether triggered by vocalizations or keypresses – are significantly suppressed compared to passive listening to the same sounds. In vocal tasks, suppression effects have been observed approximately 100 to 200 ms after vocal onset (e.g. Heinks-Maldonado et al., 2006; Niziolek et al., 2013; Curio et al., 2000; Houde et al., 2002; Gonzalez et al., 2024), using multiple electrophysiological methods, including single-neuron recordings (e.g. Eliades and Wang, 2003, 2008), intracranial EEG (e.g. Greenlee et al., 2011; Khalilian-Gourtani et al., 2024; Lorenz et al., 2025) and scalp EEG/MEG (e.g., Houde et al., 2002; Ford, 2010). In keypress tasks, a similar action-induced suppression phenomenon has been observed (e.g., Aliu et al., 2009; Baess et al., 2013; Ford et al., 2014; SanMiguel et al., 2013; Schneider et al., 2018; Weiss et al., 2011; Weller et al., 2017). These converging observations of action-induced suppression across tasks involving distinct effector systems suggest the presence of shared motor-based prediction mechanisms underlying both optional and mandatory sensorimotor mappings, although such effects may not be uniform across all contexts.

Evidence for variability comes from reports of deviations from the typical action-induced suppression pattern, particularly in optional sensorimotor mapping tasks involving keypress-auditory associations. In some cases, suppression is reduced, such as when attention is explicitly directed to the action-triggered sounds or when the properties of the resulting sound vary unpredictably. For example, Bäß et al. (2008) and Knolle et al. (2012) reported that when the frequency of self-generated pure tones varied unpredictably across trials, the magnitude of N1 suppression was weakened. Under certain conditions, enhancement rather than suppression has been observed. Ott and Jäncke (2013) found that keypress-triggered piano tones elicited larger N1 amplitudes than passively heard tones. Recent work further reveals dynamic effects, with early enhancement followed by later suppression (Rittershofer et al., 2025; Yon and Press, 2017). Using time-resolved EEG decoding, Rittershofer et al. demonstrated that expected action outcomes dominate neural representations early (including pre-stimulus), but ~200 ms after onset, unexpected outcomes become more strongly represented. These deviations are more common in keypress paradigms, likely due to the flexibility and manipulability of action-sound mappings. In contrast, vocalization tasks involve mandatory mappings, where articulatory gestures have rather fixed and predictable auditory consequences (Houde and Jordan, 1998). The sensorimotor system is highly adapted to these stable contingencies, leading to robust suppression effects (Curio et al., 2000; Eliades and Wang, 2003; Flinker et al., 2010). While suppression effects occur in both vocalization and keypress tasks, findings from keypress paradigms show greater variability depending on attention, stimulus predictability, and task demands. This variability suggests that the underlying motor-based prediction mechanisms may operate differently in optional versus mandatory sensorimotor mappings.

One potential source of the difference between optional and mandatory sensorimotor mappings may lie in the strength of association between an action and its sensory consequence. Here, we operationally define association strength as how reliably a type of action has been linked with specific sensory consequences. Long-term ecological experience is one of the factors that strengthen the association. In the case of manual action such as keypress, the association strength can be reinforced by extensive experiences of the keypress action triggering sounds in a specific setting, such as pianists' overlearned such action–sound couplings. In the current study, we manipulated the association strength by recruiting different groups of participants with distinct degrees of experience in keypress-sound mapping and varied the settings. Experiment 1 establishes a low association strength baseline in novices. Experiment 2 instantiates high, long-term association strength by testing pianists with mappings consistent with their extensive experience on keypress-triggering-sound. Experiment 3 varied the strength by pairing the tapping on the touchpad with piano tones, which created an apparent contrast with the pianists' familiar pairing of piano keys with tones.

Differences in association strength could be difficult to detect at the execution stage, where the motor plan is already fully specified, and the sensory outcome is either tightly coupled to the action or made explicit by task design, thereby "flattening" potential contrasts between the two mapping types. In contrast, the preparation stage—when an action is being planned but not yet executed—may offer a more sensitive window into how strongly a specific sensory consequence is anticipated. This stage is particularly informative for optional mappings, where the outcome of an action may remain uncertain or underspecified until just before execution.

Evidence from speech production supports this view. In one study, preparing to speak without knowing which syllable to produce elicited suppression of auditory responses relative to passive listening (Li et al., 2020). A follow-up study further revealed that this preparation-induced suppression is graded: strongest for speech syllables with fixed one-to-one articulatory-acoustic mappings, intermediate for coughs whose motor patterns and resulting sounds are more variable, and weakest for pure tones, which were represented vocally by a hum but bear only a weak association between the vocal action and the resulting sound (Zheng et al., 2022). This gradient suggests that stronger action-outcome associations yield greater preparation-induced suppression. Whether the preparation-induced suppression extends to keypress tasks is informative to illustrate the potential differences in the motor-based predictions between optional and mandatory sensorimotor mappings.

The current study tested whether the strength of action-outcome association influences preparation-induced suppression in an optional sensorimotor mapping and related these findings to previous results from mandatory mappings such as vocalization. By doing so, we sought to advance understanding of the functional characteristics of motor-based prediction mechanisms and their dependence on the nature of the sensorimotor mapping. Across three EEG experiments, we adapted the delayed articulation paradigm used in Li et al. (2020) and Zheng et al. (2022) into a delayed keypress paradigm, in which participants were prompted to prepare for a keypress action without knowing which key to press. Auditory probes were presented during the preparation stage, and changes in auditory responses relative to a passive listening condition were used to index predictive modulation effects of keypress preparation.

Experiment 1 (low-association baseline) was a keypress analogue of Zheng et al.'s study, designed to test whether keypress preparation would produce suppression effects and, if so, whether they would show the same gradient across sound categories as in vocal preparation. Non-musicians learned an arbitrary key-sound mapping with minimal prior experience; thus, any action–sound association was established only through brief laboratory exposure. If association strength plays a decisive role, a clear gradient of suppression would be observed across different sound types. Whereas, if association strength is not the determining factor, no gradient effect would be observed. Experiment 2 (high association via expertise) further examined whether stronger action-sound associations shaped by long-term ecological experience enhance preparatory suppression by involving skilled pianists using mappings consistent with years of piano training. Experiment 3 (specificity of association) used pianists as participants and piano tones as probes but paired them with a novel, non-standard key-tone mapping, dissociating long-term sound familiarity from the habitual motor-sound mapping; we expected that such ecologically invalid pairings would reduce or even eliminate suppression effects. In particular, we treat any later attenuation (e.g., P2) during preparation as a hypothesis-driven effect and evaluate its robustness across mapping contexts. Importantly, Experiments 2 and 3 included the same extensive practice session (144 trials) to equate short-term exposure to the task contingencies. Therefore, the critical test distinguishing long-term ecological association from short-term exposure is the Experiment 2 vs. Experiment 3 comparison, where practice is matched but the action-sound mapping either aligns with or violates pianists' habitual sensorimotor statistics.

The modulatory function of prediction could be beyond a unitary "suppression" that applies to multifaceted computations manifested in rich neural dynamics. For example, prediction can influence sensory gain control on the early auditory response component of N1 (~80–120 ms), whereas the effects on perceptual evaluation, stimulus categorization, and learning or context-dependent updating are usually manifested in the later auditory response component of P2 (~160–240 ms). This distinction is particularly relevant when contrasting mandatory and deterministic sensorimotor mappings, such as speech, with optional and learned mappings, such as keypress-sound associations. In vocalization, the motor-to-auditory mapping is overlearned and intrinsically coupled to the act of speaking, generating precise predictions in the internal forward model once an articulatory plan is specified. In contrast, manual keypress-sound links are more flexible and context-dependent. The prediction generated via the optional mapping may be relatively imprecise (e.g., multiple possible actions/outcomes, weaker or arbitrary action-sound contingencies). Consequently, preparatory motor states may modulate auditory processing in a component-specific manner: early facilitation (N1 gain) may arise when predictive precision is low and sensory evidence is up-weighed, whereas suppression-like attenuation at later stages (P2) may emerge only when action-outcome predictions become sufficiently reliable. This component-resolved perspective motivates analyzing N1 and P2 separately and directly testing how the association strength in optional mappings influences auditory processes during manual action preparation, compared with the deterministic case of mandatory mapping in vocalization.

Motor-based prediction can be viewed from two perspectives that make divergent predictions about the modulatory function on sensory processes. Internal forward models emphasize an efference-copy mechanism that generates a sensory prediction and attenuates the matched reafferent sensory input, yielding reduced neural responses to expected self-generated stimuli and creating larger responses when the reafference mismatches prediction (Miall and Wolpert, 1996; Blakemore et al., 2000). Predictive-coding accounts, in contrast, propose that sensory prediction errors are created by comparing prediction with sensory input and the prediction errors are weighed by their precision (i.e., expected reliability), such that predictable inputs are down-weighed whereas uncertain or unexpected inputs receive increased gain (Friston, 2010; Clark, 2013). This framework predicts that well-learned, high-certainty action outcomes would produce attenuation, whereas novel or low-certainty contexts may yield weaker suppression or even relative enhancement of sensory responses to self-generated stimuli. Recent perspectives have attempted to reconcile these views by positing two complementary processes: an early bias toward expected outcomes followed by increased processing of surprising inputs (Press et al., 2020). Empirically, motor-induced sensory modulation indeed appears to comprise both suppressive and facilitatory components; for instance, early enhancement followed by later suppression has been reported when action outcomes are uncertain (Yon and Press, 2017). Recently, Li et al.(2020) has suggested that internal forward models may incorporate both an inhibitory signal and an enhancing signal, rather than implementing a uniform cancellation of predicted input (Li et al., 2020; Zheng et al., 2022). The present study tests how the nature of the action-outcome mapping (optional keypress-sound vs. mandatory vocalization-sound) and the stage of action (preparation vs. execution) influence auditory processing. By contrasting these conditions, we aim to examine whether the neural modulations caused by motor-based prediction can be varied by the nature of action-sound mapping and hence test whether each theory would apply to different stages of motor-based prediction.

2. Materials and methods

2.1. Participants

The required sample size was determined via a power analysis using the `solve_power` function from the statsmodels library in Python (Seabold and Perktold, 2010). Drawing on the effect size reported in a representative study employing a similar experimental paradigm (Li et al., 2020), where $t(18) = 3.406$, $p = 0.003$ ($dz = 0.78$), a power analysis using a two-tailed test ($\alpha = 0.05$) indicates that approximately 19.3 participants are required to achieve 90 % statistical power. Therefore, we rounded up and recruited at least 20 participants in each experiment.

Experiment 1 involved 22 participants (13 female; M_age = 22.2 years, range = 19–25 years), mostly from East China Normal University. Experiment 2 and Experiment 3 were completed by the same cohort of 24 pianists (12 female; M_age = 23.6 years, range = 19–31 years) in two separate sessions on the same day. For clarity of testing specific hypotheses, we first report Experiment 2 and Experiment 3 separately because they were designed to address distinct questions. We then report an integrative analysis treating Experiment 2 and 3 as two counterbalanced sessions to test whether the effects generalize across sessions. Most of these piano-experienced participants were affiliated with the Shanghai Conservatory of Music.

To provide a clear description of "well-learned mapping" in Experiments 2 and 3, we provide additional details of the pianist group's musical background. Here, "well-learned mapping" refers to long-term, automatized associations between keyboard actions (keypresses/fingerings) and their auditory outcomes (pitch), constituting the motor-auditory linkage relevant to our design. On average, participants began learning piano at 7.58 ($SD = 3.74$) years of age and reported 11.14 ($SD = 6.52$) years of taking formal piano instruction (self-report; excluding periods without lessons). Nineteen participants identified piano as their primary instrument, and 14 had conservatory/professional training (students or graduates), while the remaining participants had long-term extracurricular piano study (see Table S1 for a summary). Formal qualifications were common: 17 of 24 participants had passed at least one graded piano examination, including 11 who reached the highest level (Grade 10 in the Chinese social grading system) and four with advanced performance qualifications (e.g., diploma or degree-level training). Public performance experience was also prevalent (20 of 24 reported ≥1 public performance), and 15 reported participating in piano competitions. Given that recent engagement can fluctuate even among trained musicians, we additionally report practice habits in Table S1: in the past 12 months, 18 of 24 participants reported practicing ≥1 h/week on average, while six reported practicing <1 h/week.

All participants were right-handed, had normal or corrected-to-normal vision, and reported no language, visual, or auditory impairments. Informed consent was obtained from all participants before the experiments, and they received monetary compensation upon completion. The study protocol was approved by the Institutional Review Board at NYU Shanghai.

2.2. Materials

Speech and non-speech sounds were used in three experiments to examine how the preparation of keypresses influences auditory processes. In Experiment 1, three categories of stimuli were employed, including a synthesized auditory syllable /ba/ generated via the Neospeech platform (male voice), a recording of a male cough sound, and a 500 Hz pure tone generated in Matlab. The pure tone was included as a simple, narrow-band non-speech control to elicit a robust onset-evoked auditory ERP while avoiding the spectrotemporal structure that is intrinsic to speech. All stimuli had a sampling rate of 44.1 kHz and were normalized to 70 dB SPL using Praat. The duration of each stimulus was 400 ms. In Experiment 2, two syllables, /ba/ and /ka/, were used as auditory probes. Both were synthesized via Neospeech (male voice), 400 ms in duration, and normalized to 70 dB SPL using Praat. In Experiment 3, piano tones (notes C3 and C5), which were recorded from a tuned YAMAHA piano, each 700 ms in duration and presented at a sampling rate of 44.1 kHz, were introduced. All sound stimuli were delivered via

plastic air tubes connected to foam earplugs (ER-3C Insert Earphones; Etymotic Research). Cedrus Response Pad RB-740 was used to record behavioral responses.

2.3. Procedures

2.3.1. Experiment 1: keypress preparation modulates auditory responses

Experiment 1 employed a within-subjects design using a novel delayed keypress paradigm (Fig. 1). Each of the three sounds (auditory syllable, cough, and pure tone) could be triggered by pressing one of three keys. Before the main experiment, participants completed a training session in which they learned to associate the three keys in the center of the Cedrus response pad with corresponding sounds. This training consisted of a brief practice block of 16 trials (mixed GP/GPns/NP/PL), which familiarized participants with the task structure and the three key-sound mappings; no additional mapping-only practice was administered. The behavioral accuracy was near ceiling (>97 %), suggesting the brief training was sufficient for participants to learn the simple three-key mapping. Table 1 summarizes the design and task parameters across Experiments 1–3.

In the *general preparation with auditory probe* condition (GP), participants first prepared to press a key without knowing which key to press, then pressed a key according to a visual cue and received the corresponding sound. Specifically, each trial began with a fixation presented at the center of the screen for 500 ms, followed by a yellow meaningless

visual symbol "# %" that was displayed for a random duration between 1500 ms and 2000 ms (in 100 ms increments). Participants were asked to prepare the action of keypress during the presentation of a meaningless symbol that did not provide any information about which key to press. During the final 400 ms of the preparation phase, one of the three auditory probes (auditory syllable, cough, and pure tone) was randomly presented. Participants were explicitly instructed that these probe sounds were randomly presented and task-irrelevant; they should ignore the sounds. Importantly, probe identity was selected independently of the subsequent green visual cue (and thus independently of the action-effect sound triggered by the keypress), such that the probe did not predict the identity of the upcoming action-effect sound. After the offset of the auditory probe and a subsequent blank interval of 600–800 ms, one of the green visual cues appeared ("ba", "<-", or "–"), prompting participants to press the corresponding key as quickly and accurately as possible based on the associations learned during training. The corresponding sound was triggered immediately after pressing a given key. Trials were separated by an inter-trial interval (ITI) of 1.25–1.75 s.

Four control conditions were included in the experiment. First, the *general preparation without auditory probe* condition (GPns) used the same procedure as the GP condition, except that no auditory probe was delivered during the preparation stage. The GPns condition neutralizes the strategy of preparing according to the auditory probe and ensures participants are prepared for the keypress during the entire preparation stage. Second, in the *no preparation* condition (NP), the green visual cue
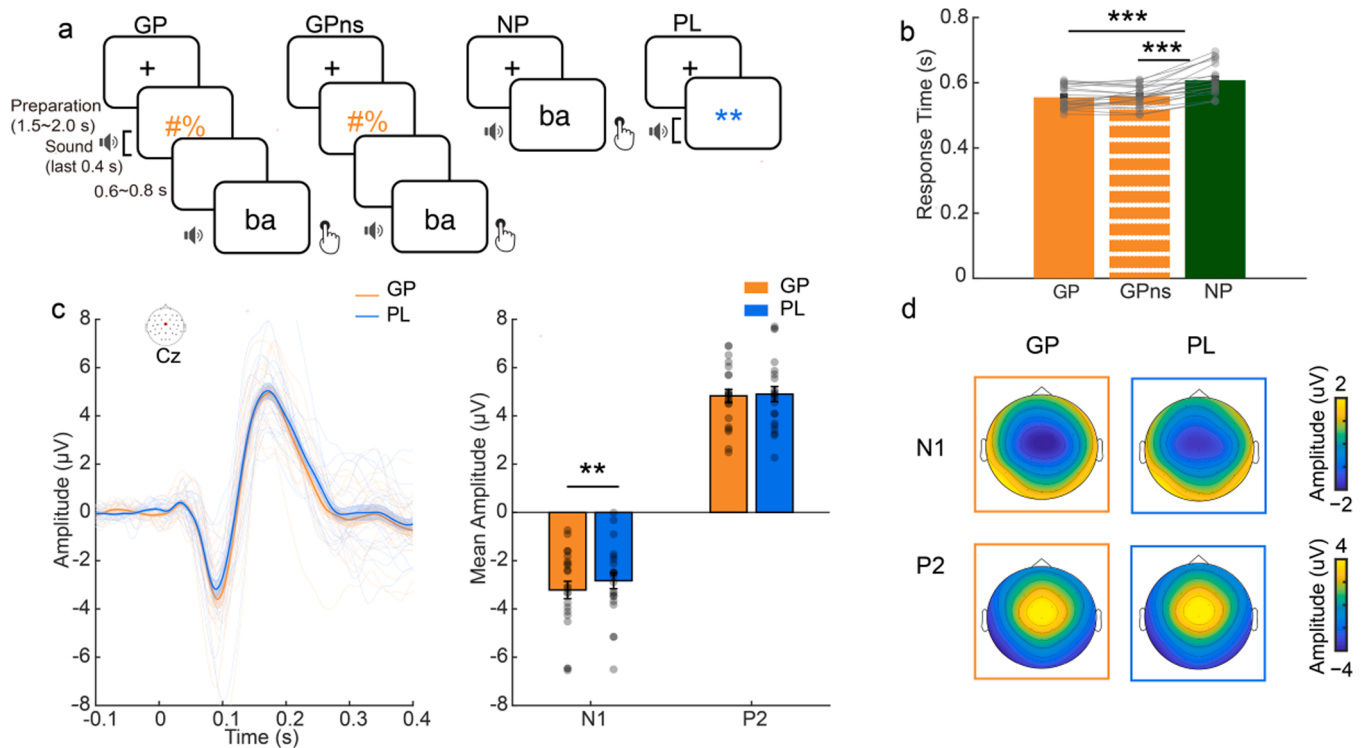

Fig. 1. Experimental paradigm and results of Experiment 1.

a) Experimental procedures for four conditions: General Preparation with auditory probe (GP), General Preparation without auditory probe (GPns), No Preparation (NP), and Passive Listening (PL). Trials in all conditions began with a fixation cross. In GP and GPns conditions, a 1.5–2.0 s keypress preparation period followed, indicated by an orange visual cue ("# %") signaling participants to prepare for keypress without specifying content. In the GP condition, an auditory probe was presented during the final 400 ms of the preparation period, while no auditory probe was presented in GPns. After the visual cue, the screen remained blank for 600–800 ms, followed by a go-cue signaling participants to initiate keypress (e.g., syllable "/ba/"). In the NP condition, there was no preparation phase, and participants executed the keypress immediately after the go-cue. In the PL condition, a blue visual cue ("**") indicated passive listening only. b) Mean response times for GP, GPns, and NP conditions. Error bars denote ±1 SEM; gray lines connect each participant's mean across the three conditions (gray circles). Response times in both GP and GPns were significantly faster than in NP (***p < 0.001). c) Left panel: Grand-average ERPs at Cz elicited by auditory probes under the GP (orange) and PL (blue) conditions. Light orange and light blue lines indicate individual participants' waveforms for GP and PL, respectively; shaded areas represent ±1 SEM. Right panel: Mean amplitudes of N1 (77–107 ms) and P2 (155–185 ms); gray circles represent each participant's amplitude within the corresponding time window. Compared with PL, N1 amplitude in GP was significantly enhanced (**p < 0.01), whereas P2 amplitude did not differ reliably between conditions (n.s.). d) Scalp topographies of N1 (top row) and P2 (bottom row) amplitudes for GP (left, orange border) and PL (right, blue border) conditions. Color bars indicate amplitude (μV).

Table 1
Experimental Design and Task Parameters (Experiments 1–3).

Key element	Experiment 1 (baseline key-sound mapping)	Experiment 2 (pianists; familiar key-speech mapping)	Experiment 3 (pianists; unfamiliar key-tone mapping)
Participants	General volunteers	Piano-experienced volunteers	Piano-experienced volunteers
Prep probes	Syllable, cough, pure tone (3)	/ba/, /ka/ (2)	Piano tones C3, C5 (2)
Response keys	3 keys	2 keys	2 keys
Mapping between keys and sounds	3:3 mapping; counterbalanced across participants	2:2 mapping; counterbalanced across participants	Fixed (not counterbalanced): C3 → left, C5 → right
Prep cue (on screen)	Yellow meaningless symbol "# %"	Circle	Same as Experiment 2
Prep duration	1.5–2.0 s (100 ms steps)	Same as Experiment 1	1.8–2.3 s (+300 ms)
Probe timing	Random within the final 400 ms of prep	Same	Same
Post-probe blank	600–800 ms	200–400 ms	200–400 ms
Go visual cue (keypress)	Green "ba", "←", or "–" (per learned associations)	Visual syllable inside circle	Same as Experiment 2
Task conditions	GP, GPns, NP, PL+ motor-only	Same as Experiment 1	Same as Experiment 2
Blocks & trials	6 × 96; plus motor-only 96 (final separate block)	3 × 96; plus motor-only 96	Same as Experiment 2
Training/ practice	N/A (see Note)	144-trial practice	144-trial practice

Notes. Abbreviations: GP = general preparation (with probe); GPns = general preparation (no probe); NP = no preparation; PL = passive listening; Experiment 1 practice: 16 mixed trials (GP/GPns/NP/PL); no mapping-only practice.

appeared immediately after the fixation, requiring participants to make a direct keypress without prior preparation. The NP condition was used as a behavioral reaction time baseline to compare and quantify the effects of preparation conditions. Third, in the passive listening condition (PL), a blue symbol ("**") was shown for a duration of 1500–2000 ms, and an auditory probe was randomly presented during the final 400 ms. Participants were instructed to passively listen to the sound. The neural responses to passively listening in the PL condition were used as a baseline to quantify the modulation effects of preparation on the auditory probes in the GP condition. Fourth, a motor-only control condition was included: after a keypress following the green visual cue symbol, no auditory feedback was provided. The motor-only control condition was used to isolate the motor execution component from the auditory responses to the sound feedback during the keypress execution phase.

The main experiment was organized into six blocks. Each block contained 96 trials, with an equal number of trials for each of the four conditions (GP, GPns, NP, and PL). Trials were presented in a randomized order. Thus, during the presentation of the visual symbol for action preparation, participants could not know whether an auditory probe would occur during the preparation phase. Short breaks were provided between blocks. The motor-only condition was run in a separate block of 96 trials at the end of the experiment.

Visual cues indicated the experimental conditions. In Experiment 1, color cues were used (yellow = General Preparation (GP/GPns); blue = Passive Listening (PL)). In Experiments 2 and 3, shape cues were used (circle = General Preparation; square = Passive Listening). These cue-condition mappings were fixed across participants.

Block order. In all experiments, the main task blocks (containing intermixed GP, GPns, NP, and PL trials) were completed first, followed by a final silent motor-only block. Participants were explicitly informed before the motor-only block that their button presses would no longer

produce any sounds. This fixed order was chosen to preserve a stable action-sound mapping during the main task and to obtain a motor baseline without disrupting the action–effect coupling (Reznik et al., 2021; Whitford et al., 2017).

2.3.2. Experiment 2: Strengthening the keypress-sound association with pianists

Volunteers with piano experience participated in Experiment 2 to further strengthen the association between keypresses and sounds. The extensive exposure of manual action and consequently triggered sounds could provide a strong foundation of auditory-motor integration in the pairing of specific hand effectors and sounds.

The procedures were similar to those of Experiment 1 with four exceptions (Fig. 3a). First, only two auditory syllables, /ba/ and /ka/, were used in Experiment 2. Second, a circle replaced the "# %" symbol for preparation, with the visual syllable cue displayed inside the circle for the keypress task. In the passive-listening condition, the original "**" marker was replaced with a square to facilitate recall. Third, the blank-screen interval between the end of the general preparation period and the keypress, originally ranging from 600 to 800 ms, was shortened to 200–400 ms to better match pianists' rapid action sequencing and to increase sensitivity to preparation-related modulatory effects. The cue–go delay was shortened in Experiment 2 (200–400 ms vs. 600–800 ms in Experiment 1) to sample motor preparation closer to the imperative signal and to reduce the need to sustain a preparatory set over an extended waiting interval, during which readiness may decay and introduce additional variability (Ito, 1991). This choice was further motivated by foreperiod research showing that shorter preparation intervals are typically associated with faster responses and earlier motor-related timing indices than longer intervals (Niemi and Näätänen, 1981; Tandonnet et al., 2003). Fourth, extensive training was conducted before the main experiment to further strengthen the keypress-sound associations. Participants completed a 144-trial practice session, thoroughly learning the mappings between visual syllable cues, two response keys, and their corresponding auditory syllables. The extensive training was to ensure that the key-sound mappings were fully consolidated. The mean accuracy during the practice phase was 97.17 % (±2.59 %), indicating that participants had successfully and proficiently learned the mappings between response keys and auditory syllables. The associations between syllables and keys were counterbalanced across participants. Additionally, the main experiment consisted of three blocks of 96 trials each, with an equal number of trials allocated to each of the four conditions (GP, GPns, NP, PL). The final motor-only block included 96 trials. The number of trials was appropriately reduced to minimize redundancy while still ensuring sufficient statistical power for each experimental condition.

2.3.3. Experiment 3: Unfamiliar keypress-tone pairing influencing the modulation effects

The procedures of Experiment 3 were identical to those of Experiment 2, except that the auditory stimulus was changed to the piano tones C3 and C5, and the corresponding preparation period was lengthened by 300 ms (1.8–2.3 s, Fig. 4a) to accommodate the relatively longer duration of piano tones. As in Experiment 2, participants completed a 144-trial practice session to learn the response key-tone mappings before the main task. The pairing of lab response keys and piano tones created a discrepancy with pianists' familiar pairing of piano keys and tones, offering another perspective on how weakened action-perception links in an unfamiliar context influence the modulation effects of motor-based prediction during preparation. Because C3 is always located to the left of C5 on the piano keyboard, we did not counterbalance the mapping between the tones and the response keys. The mean accuracy during the practice phase was 98.61 % (±1.81 %).

2.4. EEG data acquisition and preprocessing

EEG signals were recorded using a 32-channel Brain Products acti-CHamp 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 in MATLAB using EEGLAB. The data were downsampled to 500 Hz and bandpass filtered between 0.1 and 30 Hz. Preprocessing and analysis parameters (filtering, epoching, baseline windows, and artifact thresholds) were defined a priori following established lab procedures and were applied consistently across experiments (Li et al., 2025; Zheng et al., 2022). Data were re-referenced to the average of all scalp electrodes. Noisy channels were identified using objective criteria (clean_rawdata; channels with flat-lined signals > 5 s or low correlation with their reconstruction, $r < 0.8$) and were then repaired using spherical spline interpolation (Perrin et al., 1989). Artifact Subspace Reconstruction was applied to attenuate transient high-variance artifacts (burst criterion = 20). Long epochs (−200 to 2000 ms relative to auditory-probe onset) were extracted, and epochs exceeding ±200 μV were discarded to remove gross artifacts before ICA. ICA was then performed on the epoched data, and artifactual components were rejected using ICLabel with predefined probability thresholds (e.g., Eye > 0.80; Muscle > 0.70; Heart > 0.80; Line Noise > 0.85). For ERP quantification, epochs were segmented from −100 to 400 ms relative to auditory-probe onset, baseline-corrected using the −100 to 0 ms period, and rejected if the amplitude exceeded ±100 μV. Although the auditory stimulus in Experiment 3 lasted 700 ms, all quantification of early auditory components (N1/P2) was restricted to 0–400 ms to ensure comparability across experiments. Rejection rates for Experiments 1, 2, and 3 were 8.09 %, 4.05 %, and 4.25 %, respectively.

2.5. Data analysis

Behavior. To assess the behavioral effects of general preparation, reaction times (RT) were quantified as the interval from the onset of the visual cue to the start of the key press. Before the analysis, for each participant and each condition, outliers that deviated by more than three standard deviations from the mean were removed. The RTs were compared across three conditions (GP, GPns, and NP). A one-way repeated-measures ANOVA was performed to examine the differences among these conditions. Post hoc *t*-tests with Bonferroni correction were conducted to make pairwise comparisons between the conditions. Effect sizes were quantified using *dz* for the *t*-tests and partial eta squared (η^2) for the ANOVA. In Experiment 1, one participant's RT file was corrupted by a response pad logging failure; thus, RT analyses were performed on 21 participants.

EEG. Preparation stage. 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. For each component on the collapsed waveform, we

defined a 30-ms time window centered on its peak latency by extending ±15 ms from the peak, and calculated the mean amplitude within that interval: Experiment 1: the N1 was measured from 0.077 to 0.107 s, the P2 from 0.155 to 0.185 s; Experiment 2: the N1 was measured from 0.079 to 0.109 s, the P2 from 0.167 to 0.197 s; Experiment 3: the N1 was measured from 0.079 to 0.109 s, the P2 from 0.161 to 0.191 s. For electrode selection—considering the scalp topography, relevant literature (e.g., Li et al., 2020; Zheng et al., 2022), and our 32-channel electrode setup—we extracted the response amplitude of N1 and P2 components from electrodes Cz, which provided a clear representation of these components. For the ERP analyses, paired-sample *t*-tests were used to compare the mean amplitudes of the ERP components between tasks (e.g., general preparation (GP) vs. passive listening (PL)). Bonferroni-corrected post hoc comparisons were performed. Effect sizes were computed using *dz*. As a robustness check, we also conducted a temporal and spatiotemporal cluster-based permutation analysis of GP vs. PL conditions (Maris and Oostenveld, 2007); these results, reported in Appendix A, were consistent with the N1/P2 findings. We complemented NHST with Bayesian paired *t*-tests (JZS prior, $r = 0.707$) and also report BIC-approximate Bayes factors computed from t and n for reproducibility. Values >3 or <1/3 were interpreted as moderate evidence. To rule out potential CNV and visual-cue baseline differences between GP and PL, we also applied a regression-based baseline control (mean and slope over −100–0 ms; Alday, 2019). The results were unchanged (see Appendix B).

For clarity of testing specific hypotheses, we first report Experiment 2 and Experiment 3 separately because they were designed to address distinct questions. We then report an integrative analysis treating Experiments 2 and 3 as two counterbalanced sessions to test whether the effects generalize across sessions.

Execution phase. This analysis served as a validity check of statistical power to verify that our paradigm reproduced the well-established motor-induced suppression (MIS) of auditory ERPs. The "motor-auditory" condition was defined as all trials in which the participant's keypress triggered a sound (i.e., the sound-feedback epochs following the Go cue in the GP, GPns, and NP conditions pooled together). These were contrasted with trials from the Passive Listening (PL) condition, where the same sounds were heard without any action. 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 motor-only condition, participants executed the same keypress 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 keypress and the virtual auditory marker was matched to that in the motor-auditory condition. Finally, we subtracted the responses in the motor-only condition that contained only movement-related potentials from the motor-auditory data, yielding difference waveforms that primarily reflected the auditory-evoked component (see Fig. 5 for a detailed depiction of the ERP waveforms and topographic maps across conditions: the corrected motor-audio condition, and the passive listening condition). Because the number of these motor-auditory trials differed from that of the passive listening (PL) condition, we randomly selected a subset of motor-auditory trials matching the PL trial count for comparison. In a subset of datasets, the markers used to time-stamp the "virtual" auditory onset were intermittently missing during the motor-only block, precluding reliable time-locking for the subtraction. Consequently, six participants were excluded from Experiment 1 and one participant each from Experiments 2 and 3. For Experiment 1, the N1 was measured from 0.085 to 0.115 s and the P2 from 0.155 to 0.185 s; for Experiment 2, the N1 was measured from 0.093 to 0.123 s and the P2 from 0.165 to 0.195 s; for Experiment 3, the N1 was measured from 0.087 to 0.117 s and the P2 from 0.159 to 0.189 s. For the ERP analyses, paired-sample *t*-tests were used to compare the mean amplitudes of the

---

[1] Because Cz served as the online reference and was therefore absent from the raw recordings, we first appended a zero-initialized Cz channel to the EEG dataset. We then re-referenced all EEG channels (excluding mastoid and EOG electrodes) to the common average using EEGLAB's pop_reref, which subtracts the instantaneous mean across channels at each time point and thereby restores the true Cz waveform.

ERP components between tasks (i.e., motor-audio vs. passive listening).

Trial-by-trial coupling between CNV and N1. To test whether GP-related N1 enhancement is linked to the strength of motor preparation, we analyzed trial-by-trial correlations between the CNV and the auditory N1 in the GP condition. For each participant (in each experiment), we extracted single-trial CNV amplitude (X), N1 peak amplitude measured relative to the visual-cue baseline (Y), and the local pre-auditory baseline voltage immediately before sound onset (Z) from electrode Cz. We then computed within-subject semi-partial and full partial Pearson correlations between X and Y, controlling for Z. In the semi-partial correlation, N1 amplitude was residualized with respect to the pre-sound baseline (i.e. we removed any variance in Y that could be linearly predicted by Z), and this baseline-corrected N1 was correlated with X. In the full partial correlation, we residualized both X and Y with respect to Z (removing any influence of baseline shifts on both CNV and N1) before correlating. These correlations were calculated for each participant and Fisher-z transformed for statistical analysis at the group level.[2]

Time-frequency analysis (auditory-probe-locked). Time–frequency analyses were performed in MATLAB using FieldTrip (release fieldtrip-20,240,309) (Oostenveld et al., 2011). Artifact-free EEG data were epoched relative to auditory probe onset (−0.8 to 1.3 s). Single-trial time-frequency representations were computed with complex Morlet wavelets in ft_freqanalysis (method = 'wavelet'; 6 cycles), evaluated from 8 to 30 Hz in 1-Hz steps and sampled every 10 ms. Power was defined as the squared magnitude of the complex coefficients and baseline-normalized per frequency, using −0.55 to −0.30 s for the α band and −0.50 to −0.25 s for the β band, expressed as dB change relative to baseline ($10 \cdot \log10(P/Pbaseline)$). Analyses focused on α (8–12 Hz) power over an *a* priori auditory ROI (T7, P7, T8, P8) and β (13–30 Hz) power over an a priori motor ROI (FC1, FC2, C3, Cz, C4); band-limited power was obtained by averaging across frequencies within the band and across channels within the ROI. Inter-trial coherence (ITC) was computed from unit-normalized complex coefficients as the length of the across-trial mean phase vector, then averaged within band and ROI. When trial counts differed between conditions within a participant, trials were randomly subsampled to match counts before ITC estimation.[3] GP–PL differences were tested separately for each experiment using two-sided cluster-based permutation tests over time (dependent-samples t statistic; 5000 permutations), with cluster-level correction controlling the family-wise error rate at α = 0.05.

3. Results

3.1. Behavioral results

Participants demonstrated high accuracy across all experiments: Experiment 1 (98.70 %, SD = 1.41 %), Experiment 2 (97.98 %, SD = 1.92 %), and Experiment 3 (98.88 %, SD = 1.00 %), suggesting all participants followed the instructions and fulfilled the task. Next, we analyzed the reaction time data to further test the effects of preparation.

Experiment 1. A repeated-measures one-way ANOVA on the dependent variable of RT revealed a significant main effect of condition (Fig. 1b), $F(2,40) = 34.32$, $p < 0.001$, partial $\eta^2 = 0.632$. Bonferroni-corrected paired-sample *t*-tests indicated that performance in the GP condition was significantly faster than that in the NP condition, $t(20) = -5.52$, $p\_bonf < 0.001$, $dz = -1.20$. Performance in the GPns condition

was also significantly faster than that in the NP condition, $t(20) = -7.49$, $p\_bonf < 0.001$, $dz = -1.63$, demonstrating the facilitation effects of preparation on following manual responses. Moreover, performance in the GP condition did not differ significantly from that in the GPns condition, $t(20) = -0.92$, $p\_bonf = 0.999$, $dz = -0.20$, suggesting that preparation was performed equally according to the visual cue in both conditions.

Experiment 2. A repeated-measures one-way ANOVA on the dependent variable of RT revealed a significant main effect of condition (Fig. 3b), $F(2,46) = 55.37$, $p < 0.001$, partial $\eta^2 = 0.707$. Bonferroni-corrected paired *t*-tests further indicated that performance in the GP condition was significantly faster than that in NP ($t(23) = 8.40$, $p\_bonf < 0.001$, $dz = 1.71$) and that performance in the GPns condition was faster compared to NP ($t(23) = 5.47$, $p\_bonf < 0.001$, $dz = 1.12$), demonstrating the effects of preparation in facilitating following manual responses. The difference between GP and GPns conditions was also significant ($t(23) = -7.62$, $p\_bonf < 0.001$, $dz = -1.56$).

Experiment 3. A repeated-measures one-way ANOVA on the dependent variable of RT revealed a significant main effect of condition (Fig. 4b), $F(2,46) = 28.96$, $p < 0.001$, partial $\eta^2 = 0.557$. Bonferroni-corrected paired t tests further indicated that performance in the GP condition significantly differed from that in NP ($t(23) = 6.38$, $p\_bonf < 0.001$, $dz = 1.30$) and that performance in the GPns condition also significantly differed from NP ($t(23) = 3.58$, $p\_bonf = 0.005$, $dz = 0.73$). The difference between GP and GPns conditions was also significant ($t(23) = -5.18$, $p\_bonf < 0.001$, $dz = -1.06$).

All three experiments exhibited a robust preparation effect: regardless of whether an auditory probe was present, reaction times in the GP and GPns conditions were significantly faster than those in the NP condition, demonstrating that preparation reliably enhanced behavioral efficiency. The significant GP vs. GPns difference emerged only in the two musician groups (Experiments 2 & 3) and was absent in the novice group (Experiment 1), perhaps because years of intensive piano training enable pianists to hear a probe (e.g., "/ba," "/ka," or a piano tone) and immediately recruit the exact fingering needed to produce that sound.

3.2. ERP results

3.2.1. Overview of preparatory auditory modulation (across experiments)

Across three EEG experiments, we compared auditory ERPs to probe sounds during general preparation (GP) versus passive listening (PL). Results revealed a two-stage modulation on the auditory process during action preparation. First, GP reliably increased the response amplitude of N1 relative to PL in all three experiments, indicating an early gain in auditory responsiveness during preparation of manual action. Second, P2 was smaller in GP than PL in Experiment 2, but no reliable P2 modulation was observed in Experiments 1 or 3.

3.2.2. Experiment 1: Preparation of keypress modulating auditory responses

We first tested whether the preparation of manual action would modulate auditory responses and whether the modulation effects would be for specific types of sounds. Fig. 1c illustrates the grand-average ERPs averaged across probe types for the GP and PL conditions; clear N1/P2 auditory responses with distinct topographical distributions were observed in both conditions (Fig. 1d). For the N1 component, the GP condition elicited a larger (i.e. more negative) amplitude than the PL condition, $t(21) = -3.52$, $p = 0.002$, $dz = -0.75$; Bayesian analysis indicated strong evidence for a difference (BF10 = 35.01). This finding suggests enhanced early auditory processing under GP. In contrast, for the P2 component, the GP condition did not differ significantly from the PL condition, $t(21) = -0.33$, $p = 0.746$, $dz = -0.07$; Bayes factors provided moderate evidence for the null (BF01 = 4.43).

Next, we analyzed the ERP responses to each type of probe to further investigate whether preparation would differentially modulate responses to distinct sounds (Fig. 2). For the probe of syllable, the GP

---

[2] Because the epoch spanned 3 s, some participants showed substantial trial loss when requiring artifact-free data across the full window; therefore, one participant in Experiment 2 and three participants in Experiment 3 with fewer than 50% valid trials remaining were excluded from this trial-by-trial coupling analysis.

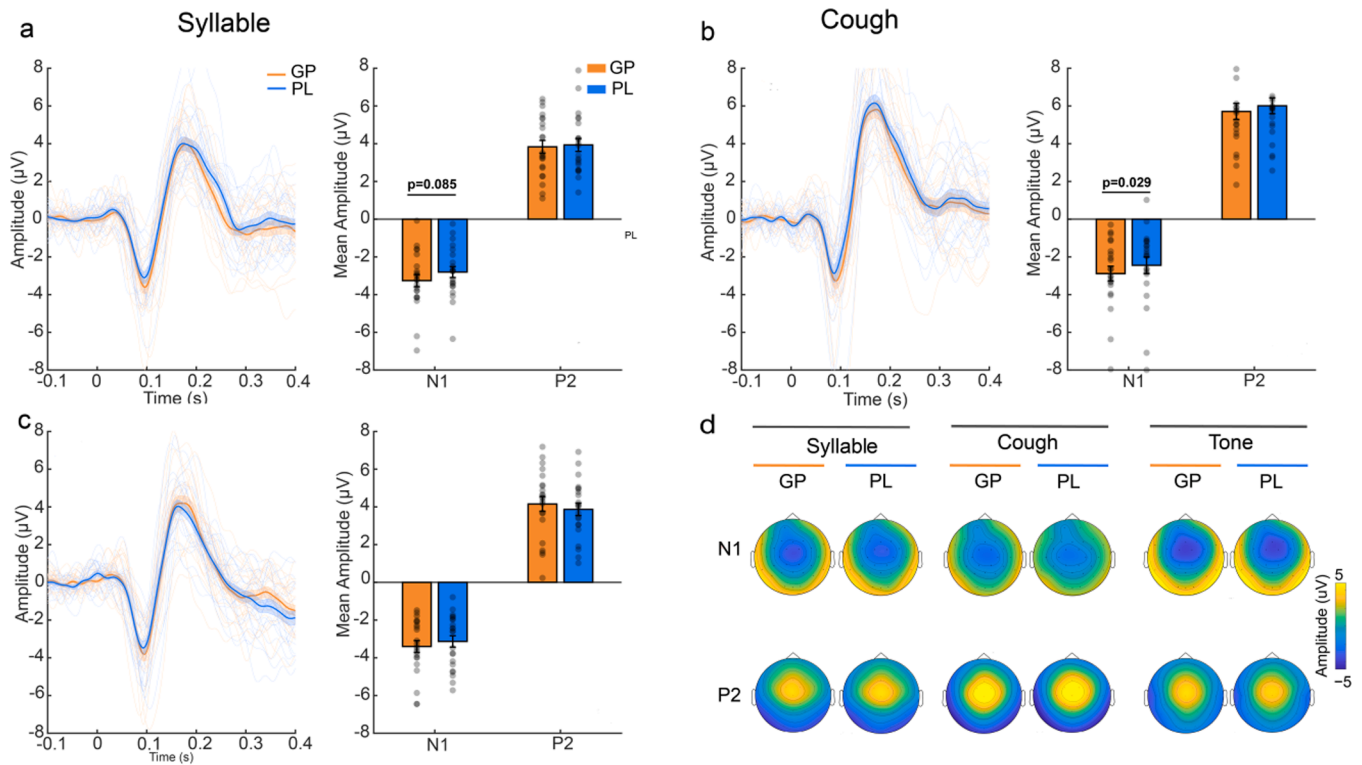[3] Across Experiments 1–3, ITC did not show a consistent GP–PL modulation in either the α or β band.

Fig. 2. . **EEG responses to different sound types in Experiment 1.**
ERP responses to a) the syllable "ba", b) a cough sound, and c) a pure tone. In each panel, the left subpanel shows the grand-average ERP waveforms at Cz under the GP (orange) and PL (blue) conditions. Light orange and light blue lines show individual participants' waveforms for GP and PL, respectively; shaded areas denote ± 1 SEM; The right subpanel plots the mean amplitudes of the N1 and P2 components for GP and PL, with each gray circle representing one participant's value. The N1 amplitude for "ba" showed a similar trend (GP > PL, $p = 0.085$); the cough sound was significantly larger in GP than PL (GP > PL, $p = 0.029$); no significant GP-PL difference was found for the tone. d) Scalp topographies for the N1 (top row) and P2 (bottom row) time windows. Left panels show responses to syllable probes ("ba"), center panels to cough probes, and right panels to pure tones. GP topographies are outlined in orange, and PL topographies are outlined in blue. Color scales indicate amplitude (μV).

condition elicited a marginally larger (i.e. more negative) N1 amplitude than the PL condition, $t(21) = -1.81$, $p = 0.085$, $pbonf = 0.255$, $dz = -0.39$, suggesting a small-to-medium trend toward enhanced early auditory processing under GP; however, BF10 = 1.05 indicated that the data were not diagnostic between the null and alternative. In contrast, for the P2 component, the GP condition did not differ significantly from the PL condition, $t(21) = -0.54$, $p = 0.592$, $pbonf = 1.000$, $dz = -0.12$, and the Bayes factor provided moderate support for the null (BF01 = 4.03), indicating negligible modulation of later auditory processing by GP.

For the probe of a cough sound, the GP condition elicited a larger N1 amplitude than the PL condition, $t(21) = -2.34$, $p = 0.029$, $p\_bonf = 0.087$, $dz = -0.50$, indicating a small-to-medium enhancement of early auditory processing during preparation; BF10 = 2.73, providing anecdotal evidence favoring the alternative over the null. In contrast, for the P2 component, the GP condition did not differ significantly from the PL condition, $t(21) = -1.14$, $p = 0.266$, $p\_bonf = 0.798$, $dz = -0.24$, with anecdotal-to-moderate Bayesian support for no effect BF01 = 2.42, indicating negligible modulation of later auditory processing by motor preparation.

For the probe of a pure tone, the amplitude of the N1 component in GP did not differ significantly from that in PL, $t(21) = -1.15$, $p = 0.265$, $p\_bonf = 0.795$, $dz = -0.24$, with anecdotal evidence favoring the null BF01 = 2.40, suggesting that the modulation effect may be weak and may not reliably extend to artificial sounds. For the P2 component, the GP condition also did not differ significantly from the PL condition, $t(21) = 1.21$, $p = 0.240$, $p\_bonf = 0.720$, $dz = 0.26$, with anecdotal support for the null BF01 = 2.23.

Taken together, preparation selectively enhanced early auditory

responses, with effects varying by sound: largest for cough, smaller for syllables, and absent for pure tones. P2 amplitudes were uniformly unaffected. Thus, in non-musicians with minimal key-sound training, preparation produced a reliable N1 gain (graded across probe types), but no evidence for a preparatory P2 attenuation.

3.2.3. Experiment 2: Strengthening the keypress-sound association with pianists

In Experiment 1, we did not find preparation-induced suppression. The absence of the effects may be because the action of keypress and its consequence of sound are not common in everyday life for typical participants. That is, lack of experience in keypress-sound association could have less power to reveal the suppression effects of motor-based prediction in preparation of manual action. Therefore, in Experiment 2, we strengthened the keypress-sound association by recruiting pianists whose experience of keypress and sound is abundant in everyday experience, and then tested the preparation-induced suppression effects in preparation for manual action. Fig. 3c illustrates the ERP responses recorded in the GP and PL conditions. Typical N1/P2 auditory responses with clear topographical distributions were observed in both conditions (Fig. 3d). For the N1 component, the GP condition elicited a larger (i.e. more negative) amplitude than the PL condition, $t(23) = -3.04$, $p = 0.006$, $dz = -0.62$; Bayesian analysis indicated moderate-to-strong evidence for a difference (BF10 = 7.66), reflecting enhanced N1 amplitude in the GP condition. Conversely, the GP condition elicited a significantly smaller P2 amplitude than the PL condition, $t(23) = -2.13$, $p = 0.044$, $dz = -0.43$; the Bayes factor provided anecdotal evidence for a difference (BF10 = 1.44), consistent with a modest/weak-to-moderate reduction of P2 during preparation.
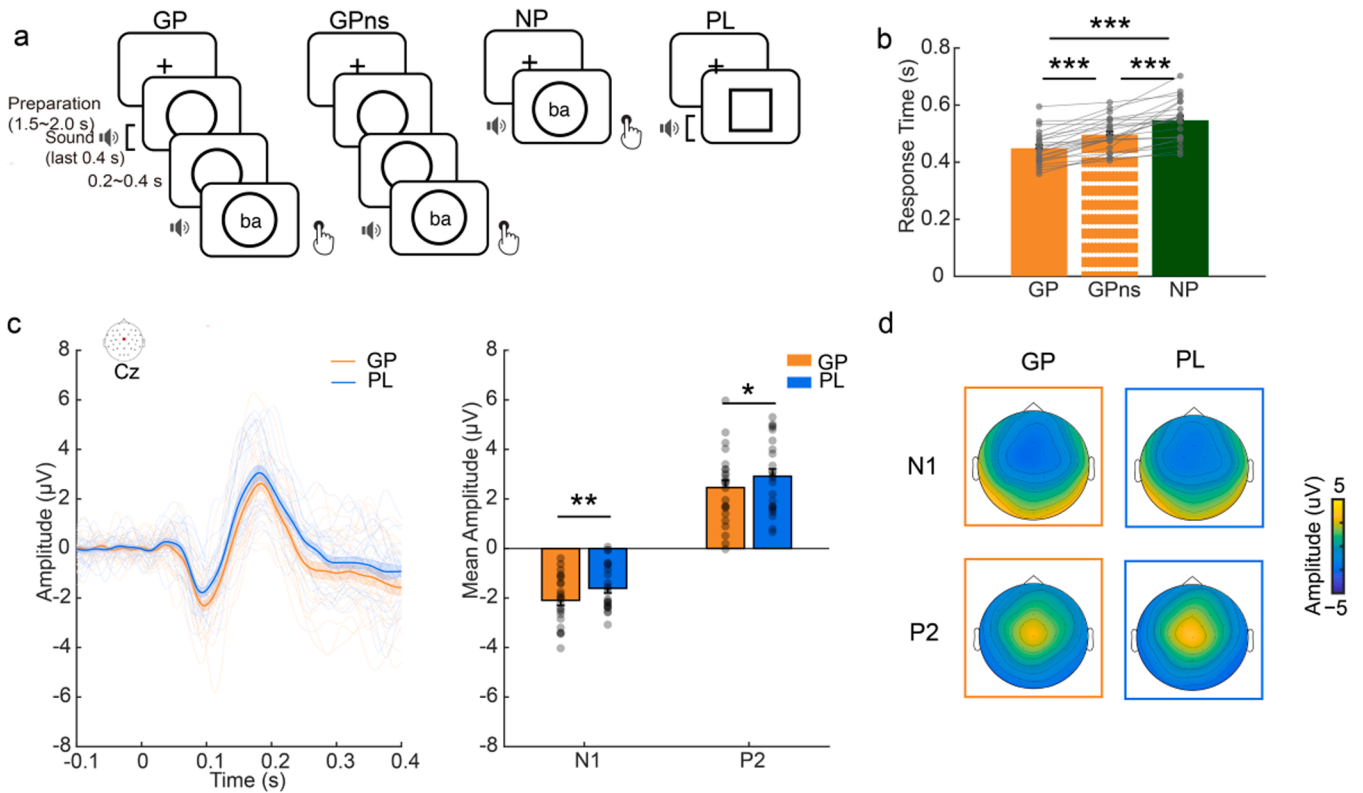
Fig. 3. Experimental paradigm and results for Experiment 2.
a) Schematic of the four task conditions: GP (General Preparation with Auditory Probe), GPns (General Preparation without Auditory Probe), NP (No Preparation), and PL (Passive Listening). In the GP condition, participants viewed a circle, during the final 400 ms an auditory probe ("ba" or "ka") was presented. After a delay of 200–400 ms, "ba" or "ka" appeared inside the circle, prompting a keypress. In GPns, the visual sequence was identical, but no auditory probe was played. In NP, no preparation cue was given; participants pressed the key as soon as "ba" or "ka" appeared. In PL, a square indicated passive listening: participants heard the probe but did not respond. **b)** Mean reaction times for GP, GPns, and NP ($n = 24$). Error bars denote ± 1 SEM; gray lines connect each participant's data across conditions. Reaction times in GP and GPns were both significantly faster than in NP (***$p < 0.001$), with GPns was significantly slower than GP (***$p < 0.001$). **c)** Grand-average ERP waveforms at Cz for GP (orange) and PL (blue; $n = 24$). Light orange and light blue lines show individual participants' waveforms for GP and PL, respectively. Shaded regions represent ± 1 SEM. Bar graphs on the right display mean amplitudes for each participant in the N1 (77–107 ms) and P2 (155–185 ms) time windows. Paired-samples *t*-tests revealed a significantly more negative N1 amplitude in GP compared to PL (**$p < 0.01$), and a significantly smaller P2 amplitude in GP compared to PL(*$p < 0.05$). **d)** Scalp topographies for the N1 (top row) and P2 (bottom row) time windows. GP topographies are outlined in orange; PL in blue. Color scales indicate amplitude (μV).

3.2.4. Experiment 3: Unfamiliar keypress-tone pairing influencing the modulation effects

In Experiment 2, the suppression effect of motor-based prediction was revealed during the preparation of manual action after strengthening the keypress-sound association. In Experiment 3, we further asked how the pairing influenced the modulation effects. In this experiment, the familiar piano tones were paired with the keypress of a lab device that was unfamiliar to pianists. If the suppression effects were driven by specific established associations between action and its auditory consequences, the unfamiliar key-tone pair would reduce or even diminish the suppression effects that were observed in Experiment 2. Fig. 4c illustrates the ERP responses recorded in the GP and PL conditions. Typical N1/P2 auditory responses with clear topographical distributions were observed in both conditions (Fig. 4d). For the N1 component, the GP condition elicited a larger (i.e. more negative) amplitude than the PL condition, $t(23) = -3.87$, $p < 0.001$, $dz = -0.79$; Bayesian analysis indicated strong evidence for a difference (BF10 = 43.79), reflecting enhanced N1 amplitude in the GP condition. In contrast, for the P2 component, there was no significant difference between the GP and PL conditions $t(23) = -0.136$, $p = 0.893$, $dz = -0.03$; the Bayes factor provided moderate evidence for the null (BF01 = 4.62). In the novel lab-key–piano-tone mapping, the N1 gain was replicated, whereas P2 showed moderate evidence for no GP–PL difference.

Because tone identity (C3 vs. C5) was fixed to response hand (left vs.

right) in Experiment 3, we assessed whether the critical context effect (GP vs. PL) differed across the two tone/hand combinations. Although tone identity was confounded with response hand in Experiment 3, additional analyses showed no evidence indicating that the GP–PL ERP difference was modulated by response side (no Tone × Context interaction in N1 or P2 window; see Supplementary Material for details).

3.2.5. Analysis across experiments 2 and 3

For N1, the ANOVA showed a significant main effect of Condition, F $(1, 23) = 14.20$, $p = 0.001$, $\eta p^2 = 0.38$, indicating that GP elicited a more negative N1 than PL across two experiments. There was also a significant main effect of Experiment, F$(1, 23) = 13.44$, $p = 0.001$, $\eta p^2 = 0.37$ (overall N1 was more negative in Experiment 3 than Experiment 2). The Experiment × Condition interaction did not reach significance, F$(1, 23)$ $= 3.26$, $p = 0.084$, $\eta p^2 = 0.124$; thus, we did not find evidence that the GP–PL difference in N1 differed across Experiments. For P2, there was a significant main effect of Experiment, F$(1, 23) = 6.11$, $p = 0.021$, $\eta p^2 = 0.210$, but the main effect of Condition was not significant, F$(1, 23) = 2.39$, $p = 0.136$, $\eta p^2 = 0.094$. The Experiment × Condition interaction is marginally significant, F$(1, 23) = 3.60$, $p = 0.071$, $\eta p^2 = 0.135$. Follow-up paired-samples *t*-tests comparing GP and PL within each experiment showed a significant GP–PL difference in Experiment 2, t$(23) = -2.13$, $p = 0.044$, dz = -0.43, but not in Experiment 3, t$(23) = -0.14$, $p = 0.893$, dz = -0.03. The integrative analysis across experiments revealed a
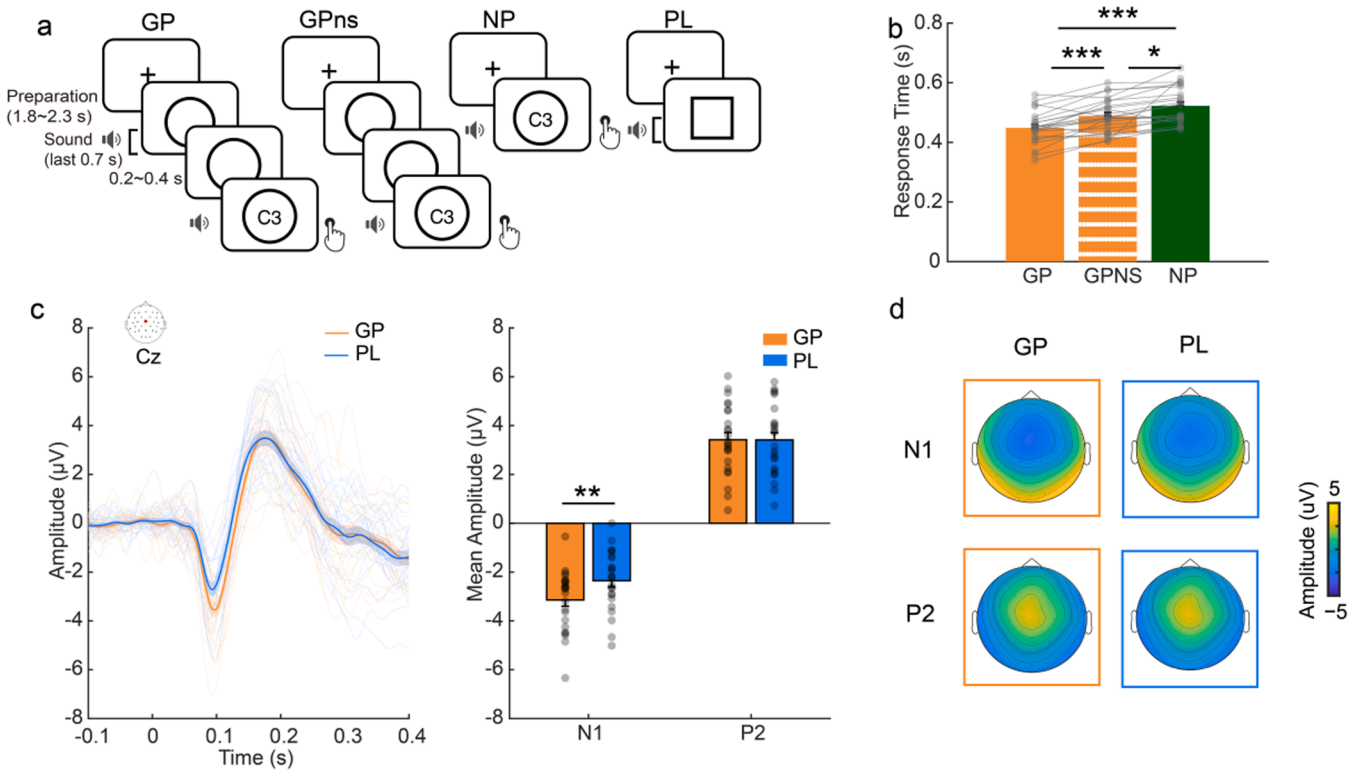
Fig. 4. Experimental paradigm and results for Experiment 3.
a) Trial structure (GP, GPns, NP, PL) was identical to Experiment 2, except that both the auditory probe and the key-press feedback were replaced by 700 ms piano tones at C3 and C5, and the corresponding preparation period was lengthened by 300 ms. b) Mean response times for GP, GPns, and NP ($n = 24$). Error bars show ± 1 SEM; gray lines link data from the same participant. Both GP and GPns were significantly faster than NP (*** $p < 0.001$), and GPns was slower than GP (* $p < 0.05$). c) Grand-average ERPs at Cz for GP (orange) and PL (blue; $n = 24$). Light orange and light blue traces depict individual participants in the GP and PL conditions, respectively; shaded bands indicate ± 1 SEM. Right: Mean N1 (79–109 ms) and P2 (161–191 ms) amplitudes for each participant. N1 was significantly more negative in GP than in PL (** $p < 0.01$), whereas P2 did not differ (n.s.). d) Scalp topographies of N1 (top row) and P2 (bottom row) for GP (orange border) and PL (blue border). Color scales indicate amplitude (μV).

consistent GP-related N1 enhancement across sessions, whereas the P2 modulation differed in Experiment 2, aligning with the results from separate analyses.

3.2.6. Component- and hemisphere-specific tests of GP–PL modulation

A key question raised by our prior pattern of findings was whether the GP–PL modulation is consistently expressed as an N1 enhancement across Experiments 1–3, whereas a P2 suppression emerges more selectively (previously observed only in Experiment 2), and whether any of these effects differ by hemisphere. Mean ERP amplitudes were therefore quantified in two a priori time windows (N1, P2) and extracted from left and right ROIs (left ROI: FC1/C3/CP1; right ROI: FC2/C4/CP2; cf. Pilling, 2009). These electrode sites were selected because auditory ERP components (N1 and P2) typically exhibit maximal amplitudes in these regions, thereby also reducing the number of statistical comparisons (Pilling, 2009; Kaganovich et al., 2016).

For each experiment, amplitudes were entered into a 2 (Condition: GP vs. PL) × 2 (Time Window: N1 vs. P2) × 2 (Hemisphere: left vs. right) repeated-measures ANOVA. We report all main effects and interactions with F, p, and partial ηp^2. To directly evaluate whether the GP–PL difference is present within each time window (N1 and P2), we additionally conducted planned paired comparisons of GP vs. PL within N1 and within P2, collapsing across hemispheres. These two planned contrasts were Bonferroni-corrected (i.e., corrected across the two windows), and effect sizes are reported as dz. (With the current coding of the GP–PL contrast, negative t values indicate GP < PL, i.e., a more negative/less positive amplitude in GP relative to PL.)

Experiment 1

The 2 × 2 × 2 ANOVA revealed significant main effects of Condition,

$F(1, 21) = 15.49$, $p < 0.001$, $\eta p^2 = 0.42$, and Time Window, $F(1, 21) = 220.68$, $p < 0.001$, $\eta p^2 = 0.91$, with no main effect of Hemisphere, $F(1, 21) = 1.01$, $p = 0.327$, $\eta p^2 = 0.05$. No interactions reached significance, including the three-way interaction, Condition × Time Window × Hemisphere, $F(1, 21) = 0.14$, $p = 0.709$, $\eta p^2 = 0.01$, and the two-way interactions (Condition × Time Window: $F(1, 21) = 3.24$, $p = 0.086$, $\eta p^2 = 0.13$; Condition × Hemisphere: $F(1, 21) = 0.01$, $p = 0.943$, $\eta p^2 < 0.01$; Time Window × Hemisphere: $F(1, 21) = 0.62$, $p = 0.441$, $\eta p^2 = 0.03$).

Planned paired comparisons (Bonferroni-corrected across N1 and P2) showed a robust GP–PL difference in the N1 window, $t(21) = -4.25$, $p < 0.001$, $pBonf < 0.001$, $dz = -0.91$, consistent with an N1 enhancement. The corresponding P2 contrast was smaller and did not survive correction, $t(21) = -2.06$, $p = 0.053$, $pBonf = 0.105$, $dz = -0.44$. Thus, under the standardized analysis, Experiment 1 shows a clear N1 GP–PL effect but no corrected evidence for a P2 suppression. No evidence for hemisphere-specific modulation was observed (all Hemisphere-involving interactions ns).

Experiment 2

The ANOVA showed a significant main effect of Condition, $F(1, 23) = 9.64$, $p = 0.005$, $\eta p^2 = 0.30$, and a significant main effect of Time Window, $F(1, 23) = 228.25$, $p < 0.001$, $\eta p^2 = 0.91$, with no main effect of Hemisphere, $F(1, 23) = 0.01$, $p = 0.904$, $\eta p^2 = 0.01$. No interactions were significant, including the three-way interaction, Condition × Time Window × Hemisphere, $F(1, 23) = 2.39$, $p = 0.136$, $\eta p^2 = 0.09$, and the two-way interactions (Condition × Time Window: $F(1, 23) = 0.01$, $p = 0.956$, $\eta p^2 < 0.01$; Condition × Hemisphere: $F(1, 23) = 0.01$, $p = 0.983$, $\eta p^2 = 0.01$; Time Window × Hemisphere: $F(1, 23) = 0.97$, $p = 0.334$, $\eta p^2 = 0.04$).

Planned comparisons again showed a reliable GP–PL difference in the N1 window, $t(23) = -3.12$, $p = 0.005$, $pBonf = 0.010$, $dz = -0.64$. In contrast, the P2 GP–PL difference was smaller but significant, $t(23) = -2.47$, $p = 0.022$, $pBonf = 0.043$, $dz = -0.50$. Thus, Experiment 2 also shows an N1 enhancement, and the P2 suppression pattern reaches significance under the present correction and analysis. As in Experiment 1, there was no evidence that the effect varies by hemisphere (Hemisphere-involving interactions ns)

Experiment 3

The ANOVA yielded significant main effects of Condition, $F(1, 23) = 4.59$, $p = 0.043$, $\eta p^2 = 0.17$, and Time Window, $F(1, 23) = 214.27$, $p < 0.001$, $\eta p^2 = 0.90$, with no main effect of Hemisphere, $F(1, 23) = 0.70$, $p = 0.412$, $\eta p^2 = 0.03$. Importantly, Experiment 3 showed a significant Condition × Time Window interaction, $F(1, 23) = 7.96$, $p = 0.010$, $\eta p^2 = 0.26$, indicating that the GP–PL effect differed across N1 and P2. In contrast, Condition × Hemisphere was not significant, $F(1, 23) = 0.01$, $p = 0.948$, $\eta p^2 < 0.01$, and the Condition × Time Window × Hemisphere interaction was also not significant, $F(1, 23) = 1.84$, $p = 0.188$, $\eta p^2 = 0.07$.

Planned comparisons revealed a clear GP–PL difference in the N1 window, $t(23) = -3.75$, $p = 0.001$, $pBonf = 0.002$, $dz = -0.77$, whereas the P2 comparison was not significant, $t(23) = 0.32$, $p = 0.752$, $pBonf = 1.000$, $dz = 0.07$. Thus, Experiment 3 provides strong evidence for an N1 enhancement and indicates that the GP–PL difference is not reliably present in P2 under the current analysis; there was again no evidence for hemisphere-dependent modulation.

Summary across Experiments 1–3. Under a standardized ROI and time-window framework, the GP–PL modulation was consistently and robustly expressed as an N1 enhancement across Experiments 1–3. In contrast, the P2 suppression emerged reliably only in Experiment 2 (with a trend in Experiment 1 and no effect in Experiment 3). Furthermore, there was no evidence supporting hemispheric differences in these effects.

3.2.7. Trial-by-trial coupling between preparatory CNV and auditory N1 in the GP condition

Experiment 1 (mixed sounds context). CNV amplitude significantly predicted subsequent N1 amplitude in the GP condition. The within-subject semi-partial correlation (CNV vs. N1_peak, controlling for N1 baseline) was mean $r = 0.173$, $t(21) = 9.81$, $p < 0.001$. The full partial correlation (baseline removed from both CNV and N1_peak) was even stronger at mean $r = 0.310$, $t(21) = 10.81$, $p < 0.001$. Importantly, all 22 participants showed a positive CNV–N1 full-partial correlation in GP (see Fig. 5).

Experiment 2 (musical tones context): A similar pattern was found. The semi-partial CNV–N1 correlation was mean $r = 0.19$, $t(22) = 8.93$, $p < 0.001$, and the full partial correlation was mean $r = 0.39$, $t(22) = 10.76$, $p < 0.001$. Again, all participants individually showed a positive correlation.

Experiment 3 (piano chords context): We again observed a robust coupling. The semi-partial correlation was mean $r = 0.14$, $t(20) = 6.03$, $p < 0.001$, and the full partial correlation was mean $r = 0.27$, $t(20) = 6.19$, $p < 0.001$. As in the other experiments, the full partial correlation was positive for every participant in Experiment 3.

3.2.7. Time-frequency analyses

3.2.7.1. α-band activity in the auditory ROI (T7/P7/T8/P8). Experiments 2–3 showed lower α power in GP than in PL. In contrast, Experiment 1 showed no reliable α-band differences between conditions (Figure SN).

Experiment 1. α power did not differ between conditions: no significant positive clusters were observed, and the observed negative cluster did not survive cluster correction (all cluster-corrected $ps > 0.05$).

Experiment 2. α power showed two significant negative clusters (GP
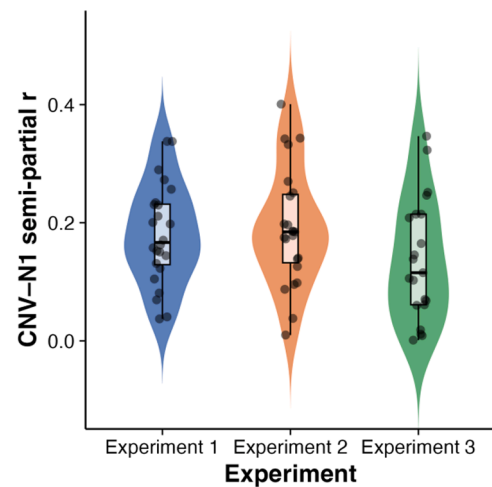

Fig. 5. Participant-level distribution of CNV–N1 correlation in the General Preparation (GP) condition in three experiments. Violin plots summarize the distribution of participant-level correlation coefficients within each experiment in each experiment; boxplots indicate the median and interquartile range, and dots represent individual participants. Positive values indicate that trials with stronger preparatory responses were associated with larger N1 responses.

< PL), reflecting stronger α desynchronization in GP in both early and late post-probe windows: 0.06–0.53 s (cluster-corrected $p = 0.003$) and 0.85–1.01 s (cluster-corrected $p = 0.025$).

Experiment 3. α power again showed a significant negative cluster (GP < PL) spanning −0.08–0.67 s relative to probe onset (cluster-corrected $p = 0.002$), indicating robustly lower α power in GP than PL across an extended interval that began slightly before probe onset and continued through the early post-probe period.

3.2.7.2. β-band activity in the motor ROI (FC1/FC2/C3/Cz/C4). Across all three experiments, β-band power in the motor ROI was consistently lower in GP than in PL (Figure SN). Effects were primarily observed in mid-to-late post-probe intervals, with an additional earlier component in Experiment 3. Lower β power (i.e., stronger β desynchronization) is commonly observed in the motor preparatory state (e.g., Doyle et al., 2005; Tzagarakis et al., 2021).

Experiment 1. β power showed a significant negative cluster (GP < PL) from 0.79 to 1.08 s (cluster-corrected $p = 0.003$), indicating stronger late β desynchronization in GP.

Experiment 2. β power exhibited a single extensive negative cluster from 0.36 to 1.07 s (cluster-corrected $p < 0.001$), reflecting sustained β desynchronization in GP throughout the mid-to-late post-probe period.

Experiment 3. β power showed two significant negative clusters (GP < PL): an early cluster from 0.30 to 0.52 s (cluster-corrected $p = 0.024$) and a late cluster from 0.69 to 1.08 s (cluster-corrected $p = 0.001$), indicating both early and late components of stronger β desynchronization in GP.

3.3. Control analyses: action-induced suppression suggests enough statistical power

Experiment 1. Fig. 6a shows the grand-averaged ERPs elicited in the motor-audio and passive listening conditions. Both conditions produced typical N1/P2 responses with a clear central–parietal scalp distribution. N1 amplitudes did not differ between the two conditions, $t(18) = 0.68$, $p = 0.506$, $dz = 0.16$. In contrast, the motor-audio condition elicited an attenuated P2 response relative to the passive listening condition, $t(18) = -4.82$, $p < 0.001$, $dz = 1.10$. Together, these findings suggest that, in non-musicians, keypress–sound associations predominantly influence the later P2 stage of auditory processing, while leaving the earlier N1 stage unaffected under the present task parameters.
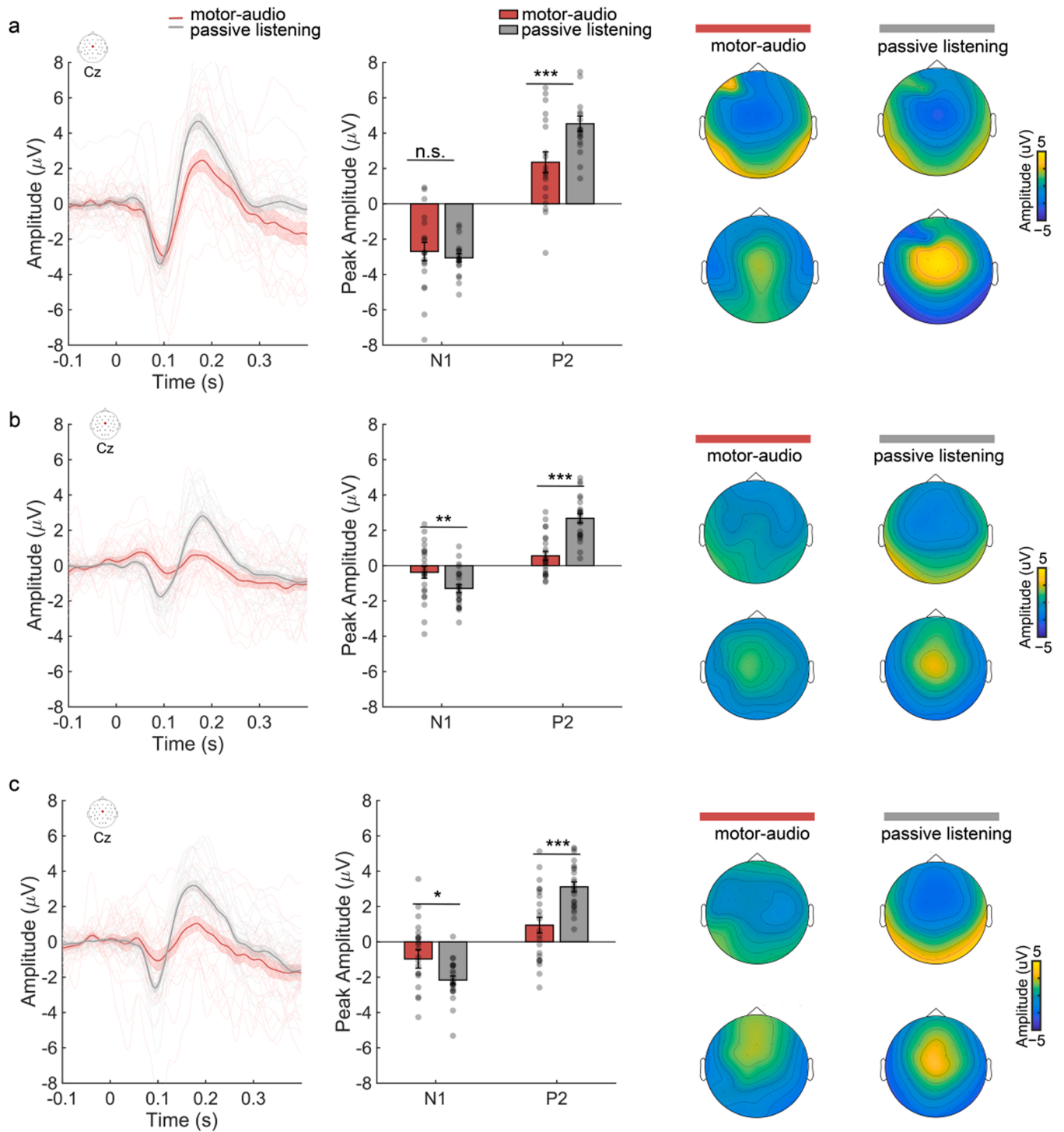
Fig. 6. Action-induced suppression during action execution across all three experiments.
(a) Experiment 1. Motor-auditory (self-generated sound) trials were defined as the keypress-triggered feedback sounds during the execution/go-cue response phase pooled across the GP, GPns, and NP conditions. These were contrasted with Passive Listening (PL) trials, in which the same sounds were presented without any action. Left: ERP waveforms at electrode Cz for motor-auditory (red) and passive listening (gray). Thick lines show group means; thin lines show individual participants. Shading indicates ±1 SEM. Time 0 ms denotes feedback-sound onset. Middle: Mean amplitudes in the N1 (85–115 ms) and P2 (155–185 ms) windows. N1 was not significantly different (n.s.), whereas P2 was significantly suppressed in the motor-auditory condition (***$p < 0.001$). Points indicate individual participant means; error bars denote ±1 SEM. Right: Scalp topographies for the N1 (top) and P2 (bottom) windows for motor-auditory (left) and passive listening (right); color scales show amplitude (μV).
b) Experiment 2. Left: ERP waveforms at Cz electrode, same conventions as in (a). Middle: Mean amplitudes for N1 (93–123 ms, **$p < 0.01$) and P2 (165–195 ms, ***$p < 0.001$) showing significant suppression in the motor-auditory condition. Right: Scalp topographies as described in (a).
c) Experiment 3. Left: ERP waveforms at Cz electrode, same conventions as in (a). Middle: Mean amplitudes for N1 (87–117 ms, *$p < 0.05$) and P2 (159–189 ms, ***$p < 0.001$) showing significant suppression in the motor-auditory condition. Right: Scalp topographies as described in (a).

Experiment 2. In trained pianists, sounds induced by keypress again produced clear N1/P2 responses (Fig. 6b). For the N1 component, the motor-audio condition showed reduced negativity compared with the passive listening condition, $t(22) = 2.92$, $p = 0.008$, $d = 0.61$, indicating a suppression effect and suggesting a stronger top-down modulation of early auditory processing under action-induced prediction. Similarly, the motor-audio condition elicited an attenuated P2 response compared with the passive listening condition, $t(22) = -5.69$, $p < 0.001$, $d = 1.19$, supporting the view that this suppression effect extends to later stages of auditory processing.

Experiment 3. In trained pianists, piano tone induced by keypress again produced clear N1/P2 responses (Fig. 6c). For the N1 component, the motor-audio condition showed reduced negativity compared with the passive listening condition, $t(22) = 2.16$, $p = 0.042$, $d = 0.45$, indicating a suppression effect and suggesting a stronger top-down modulation of early auditory processing under action-induced prediction. Similarly, the motor-audio condition elicited an attenuated P2 response compared with the passive listening condition, $t(22) = -5.43$, $p < 0.001$, $d = 1.13$, supporting the view that this suppression effect extends to later stages of auditory processing.

4. Discussion

The present study examined how the strength of the action-outcome association shapes the modulatory function of motor-based prediction on auditory processing in the context of an optional sensorimotor mapping (manual action triggering sounds). Using a delayed keypress paradigm adapted from delayed articulation designs (Li et al., 2020; Zheng et al., 2022), participants prepared a manual action while auditory probes were presented during the preparation interval. Across three EEG experiments, preparation produced a consistent two-stage profile: early facilitation of auditory processing (enhanced N1) was robust across experiments, whereas evidence for later attenuation (reduced P2) depended on context. In Experiment 1, N1 enhancement showed a stimulus-category gradient (coughs > syllables > tones), resembling the gradient reported for vocal preparation (Zheng et al., 2022). In Experiment 2, pianists showed a smaller P2 in GP than PL under the strongest-association context; Whereas in Experiment 1 and 3, where the strength of action-sound association was weak or ecologically invalid, the P2 suppression was absent. These results contrast with the observations of suppression in previous studies in the context of speaking (Li et al., 2020; Zheng et al., 2022), suggesting that motor-based prediction in optional mappings is not uniformly suppressive, but can manifest as an early, context-general gain with a possible later attenuation that requires highly precise and reliable action-outcome predictions.

4.1. Robust execution-phase motor-induced suppression across auditory categories

In line with classic demonstrations of motor-induced sensory attenuation, self-initiated sounds elicited reduced auditory ERPs at execution, attenuating both N1 and P2 compared to passive listening. This effect was observed in both our non-musician and pianist samples, and it generalized across the auditory categories tested (speech syllables /ba/, /ka/ and piano tones C3, C5), indicating category-general suppression within audition. This attenuation is consistent with forward-model accounts, where an efference copy of the motor command generates a prediction of the impending sound. When the actual feedback matches this prediction, the corresponding auditory input is attenuated (Blakemore et al., 2000; Martikainen et al., 2005; Schneider and Mooney, 2015; Wolpert et al., 1995). Previous studies have reported N1/P2 attenuation for self-generated versus externally generated sounds (e.g., N1/M100: Aliu et al., 2009; Baess et al., 2013; Harrison et al., 2023; Hughes et al., 2013; Knolle et al., 2012; Timm et al., 2013; Whitford et al., 2011; N1, P2: Horváth et al., 2012a; Klaffehn et al., 2019; Weller et al., 2017).

Notably, piano tones that showed N1 enhancement during preparation were nevertheless attenuated during execution. This contrasts with Ott et al. (2013), who reported a facilitation effect for predictable piano tones. We infer that categorical and temporal expectations established during preparation contributed to execution-phase attenuation; by contrast, Ott et al. examined immediate keypress-tone couplings without such an interval. Differences in training intensity (Ott: some practice, but not the high-intensity, handshake-like mapping consolidation of Experiments 2/3) and in the physical properties of the stimuli (Ott: synthesized piano tones, ~360 ms; present study: recorded tones, 700 ms) may also have influenced the results.

We propose that the robust N1/P2 suppression observed during execution reflects category-constrained forward-model predictions formed during preparation. For example, in our tasks, the exact pitch (C3 vs. C5) was unspecified, but the category (piano tone) remained fixed; analogously, syllable identity (/ba/ vs. /ka/) was constrained within the speech category. Such categorical predictability may have supported early expectations, which in turn attenuated early sensory processing. Previous studies suggest that forward-model mechanisms can tolerate parameter uncertainty: Baess et al. (2013) found that N1 suppression persists despite unpredictable tone frequency, and Knolle et al. (2019) demonstrated that accurate category predictions during self-generated speech elicit N1/P2 attenuation, while mismatches between prediction and feedback abolish this effect. By the time of execution, specific expectations regarding both category and timing were likely established, and a precise match between the efference-copy-based prediction and the actual sensory feedback likely underlies the observed N1/P2 suppression. We infer that category-constrained predictions established during preparation contributed to execution-phase N1/P2 attenuation.

Because auditory ERPs in the execution-phase were estimated using a silent motor-only baseline to isolate motor activity from sound-evoked responses, it is important to consider a trade-off in experimental design about the placement of the motor-only block. Specifically, we administered the silent motor-only block at the end of the session. Interleaving silent motor-only trials within the main task could interrupt the action-sound coupling (i.e., actions sometimes do not yield auditory outcomes), potentially weakening the effects of interest. We therefore adopted a fixed order in which all main task blocks (with intermixed GP/ GPns/NP/PL trials) preceded the motor-only block, consistent with prior work (e.g., Reznik et al., 2021; Whitford et al., 2017). This choice may introduce order-related confounds, particularly fatigue and carry-over expectations. However, reaction times were reliably faster in the motor-only block than in the preceding main task blocks across experiments (all ps < 0.001), which did not support the fatigue in the motor-only block at the end of the experiment. Moreover, any residual action–sound expectation carried into the silent baseline of the motor-only block would reduce (rather than inflate) the contrast of motor–auditory versus motor-only, rendering our estimates conservative.

4.2. Preparatory P2 attenuation is context-dependent and may be facilitated by well-learned mappings

In Experiment 2, experienced pianists showed a small preparatory P2 suppression when preparing actions with well-learned key-sound contingencies. The effect was observed when the auditory probes were syllables instead of piano tones, indicating that high action-outcome precision can facilitate predictive suppression beyond instrument-specific mappings. This interpretation aligns with reports of stronger attenuation for familiar versus novel action-sound mappings (Audette and Schneider, 2023; Ross et al., 2017), and with EEG and TMS evidence that early auditory responses can be shaped by factors such as temporal predictability, consistent action-outcome timing, and the strength of action-outcome associations (e.g., Horváth et al., 2012b; Klaffehn et al., 2019; Timm et al., 2014, 2016). Our results extend forward-model

accounts by demonstrating that calibration occurs not just during execution, but also during preparation, and critically depends on the stability of well-learned sensorimotor regularities.

Conversely, when the action-sound mapping was unfamiliar or kinematically/spatially perturbed, Bayesian model comparison predominantly favored the absence of preparatory P2 attenuation. Non-pianists in Experiment 1 (arbitrary keypress→syllable) and professional pianists in Experiment 3 (piano tones triggered by a laboratory response pad) showed moderate evidence for the null relative to a passive listening baseline (Experiment 1: BF01 = 3.89; $t(21) = -0.33$, $p = 0.746$, $dz = -0.07$. Experiment 3: BF01 = 4.62; $t(23) = -0.136$, $p = 0.893$, $dz = -0.03$). A simple explanation is that the action-sound link did not match the learned instrument closely enough. Predictive mechanisms that reduce auditory responses work best when the timing, movement, touch, and spatial layout of the action match what has been learned on the instrument. Pianists do rely on learned pitch-to-position associations, although how precise this "keyboard localization memory" is outside natural playing is still debated (see Fontana et al., 2024 for evidence that mechanical cues can support spatial localization). Consistent with this idea—and given the lab response pad—Ott et al. (2013) reported no self-generated suppression and a larger N1 when button presses triggered piano tones over headphones.

The absence of P2 suppression in Experiment 3 suggests that the internal forward model is sensitive to specific action-outcome parameters (Bolt and Loehr, 2021; Timm et al., 2014; Wolpert et al., 1995; Wolpert and Miall, 1996). Two possible mechanisms could account for the null effect. First, limited generalizability. The motor plan for pressing a rigid response pad differs kinematically and somatosensorily from pressing a weighed piano key, potentially failing to engage the specialized forward model established through piano training (Lahav et al., 2007; Lee and Noppeney, 2011; Mifsud et al., 2016; Mifsud and Whitford, 2017). Second, attenuated prediction due to contextual invalidity. The arbitrary mapping between a button press and a piano sound constitutes a low-validity context in which the brain likely down-weights the precision of the predictive signal (Feldman and Friston, 2010; Han et al., 2022; see also Gordon et al., 2017). Under this account, the prediction is not necessarily absent but may be sufficiently attenuated to prevent measurable suppression of the auditory P2 response.

A caveat for cross-experiment comparisons is that "association strength" covaried with training exposure. Experiment 1 involved only minimal laboratory familiarization, whereas Experiments 2–3 included extensive real-world sensorimotor experience in pianists and a longer in-lab practice phase. Behavioral performance in all experiments was at ceiling regardless of the training amount, indicating the results of short-term plasticity were at the same level. Moreover, the preparatory P2 suppression emerged only in Experiment 2, but not in Experiment 3, even though in both experiments, participants completed the same amount of in-lab practice. This dissociation suggests that short-term laboratory training alone was insufficient to elicit the P2 effect. Instead, the observed effects are more likely to depend on long-term learned sensorimotor priors.

Acoustic and perceptual differences between stimulus sets may also contribute to the differing P2 patterns observed across experiments. The P2 component is sensitive to acoustic features and auditory experience. For instance, Shahin et al. (2005) demonstrated that P2 amplitude in musicians varies systematically with acoustic properties of musical sounds, consistent with experience-dependent encoding of stimulus features. Additionally, manipulating amplitude rise time and spectral–temporal cues in speech modulates the N1–P2 complex (Carpenter and Shahin, 2013). Because the probe sets in Experiment 2 (speech syllables) and Experiment 3 (piano tones) were not acoustically matched and necessarily differed in amplitude-envelope and spectral–temporal characteristics known to influence the N1–P2 complex. However, in the experimental design and analysis of the current study, the modulation effects were quantified by taking the differences between the responses

to the same auditory stimuli in GP and PL conditions in a given experiment. Hence, only remain the modulation effects (the interaction between preparation and probe) that are independent of the auditory stimuli. Therefore, the observed absence of the modulation effects in Experiment 3, contrasting with the suppression in Experiment 2, was most likely caused by the experimental manipulation of the unfamiliar pairing of action and its consequences.

Converging behavioral evidence comes from the probe-related RT benefit. Notably, the probe-related RT benefit (GP vs. GPns) was numerically larger in Experiment 2 with the familiar mapping (dz = –1.56) than in Experiment 3 with the unfamiliar mapping (dz = –1.06). Although this cross-experiment contrast should be interpreted cautiously, the pattern is consistent with mapping-specific optimization: well-established action–effect associations may enable performers to exploit predictive auditory cues more efficiently during motor preparation. This interpretation aligns with prior evidence that sensorimotor experience strengthens auditory–motor coupling in pianists (e.g., Drost et al., 2005a, 2005b). Together, the ERP and behavioral patterns converge on the view that stronger sensorimotor priors support more efficient preparatory use of predictive auditory information.

Note that Experiments 2–3 had fewer trials per condition by design compared to Experiment 1 (Appendix Table 1), potentially reducing the ERP signal-to-noise ratio and sensitivity to small effects. Although final trial counts were closely matched between GP and PL, conclusions regarding preparatory P2 attenuation should be interpreted with appropriate caution and warrant replication.

4.3. Precision-weighted modulation of N1 during motor preparation

Prior work on speech preparation typically reports N1 attenuation, often attributed to premotor predictions about imminent auditory feedback (Li et al., 2020; Zheng et al., 2022). By contrast, in our keypress-sound paradigm, both non-musicians and musicians showed a larger N1 to probes presented during the preparation interval than in the passive listening. We interpret this as an early boost of sensory processing: the probe categories (syllables or piano tones) are acoustically clear, whereas the action-sound link implemented with a response pad is still relatively uncertain. Under these conditions, motor-based suppression contributes little, and the system relies more on the incoming sound, yielding greater early gain (N1 enhancement; Bäß et al., 2008; Schneider and Mooney, 2018).

Because the GP condition requires motor preparation and temporal readiness for an impending go cue, the enhanced N1 could be interpreted as an attentional-like increase in early sensory gain (Fruhstorfer et al., 1970; Hillyard et al., 1973; Näätänen and Picton, 1987; Pritchard, 1981). However, our findings suggest that this modulation cannot be fully attributed to a purely tonic, non-specific elevation in vigilance or arousal (Näätänen and Picton, 1987). Instead, the effect shows structured dependence on stimulus and learning/predictability context: in Experiment 1, the N1 amplification was largest for cough and smaller for tones, a graded pattern that is difficult to reconcile with a one-size-fits-all global arousal account; moreover, under highly predictable mappings we observed a biphasic profile, characterized by early enhancement followed by later suppression, rather than a uniform gain increase for early components. Critically, the magnitude of motor preparation on each trial (CNV) predicted the magnitude of the N1 response across all three experiments. Given that CNV is widely taken to index motor preparation and anticipatory attention (Brunia and van Boxtel, 2001; Gómez et al., 2003), this robust trial-by-trial CNV–N1 coupling indicates that stronger preparation is associated with greater early auditory sensitivity—an association that is not readily explained by non-specific arousal. Accordingly, a weighted interpretation is warranted: while some contribution from non-specific attention/arousal is plausible, the observed N1 enhancement is more likely induced by a preparation-coupled, motor-driven precision modulation of auditory processing.

Why would vocal preparation yield N1 suppression (Li et al., 2020), whereas manual preparation produces N1 enhancement? We suggest this contrast stems from fundamental differences in effector-specific predictive mechanisms, namely the specificity and strength of the internal forward model and efference copy. The speech motor system has a dedicated neural circuitry for monitoring self-generated sounds, including direct efference copies that reach the auditory cortex (Eliades and Wang, 2003; Houde et al., 2002). Through development and a lifetime of speaking, individuals acquire highly overlearned, stable and one-to-one mappings between vocal motor commands and their acoustic consequences. This likely enables a precise inhibitory prediction already during preparation, reducing auditory gain and manifesting as N1 attenuation (Li et al., 2020). By contrast, a hand action like a keypress has no inherent one-to-one auditory outcome; its mapping to sound is learned and can be contextually variable. As a result, during keypress preparation, the internal forward model's auditory prediction may be less specific and less reliable. Within a predictive-coding account, lower prediction certainty should lead to higher precision-weighting (sensory gain) of incoming auditory input, which would enhance early sensory responses such as the N1. In other words, the manual efference copy or prediction might not be strong or specific enough to suppress early sensory responses, and preparatory modulation may therefore rely more on sensory gain that increases auditory precision, yielding N1 enhancement. This framework yields testable dissociations between effector and design factors: strengthening the determinism and extensive overlearning of keypress-sound mappings (e.g., a fixed one-to-one mapping trained over time) should reduce preparatory N1 enhancement and may shift it toward attenuation; conversely, increasing action-sound uncertainty during vocal preparation (e.g., probabilistic or perturbed auditory feedback) should weaken preparatory N1 suppression and could even move it toward enhancement.

In line with this view, button-press studies with tone feedback have found weaker—and sometimes reversed—self-generated attenuation under non-ecological mappings (Ott and Jäncke, 2013; Reznik et al., 2014). Although these effects were measured at action execution rather than during our preparation probes, the directional match is notable. While attention and arousal can increase N1 (Hillyard et al., 1973; Luna et al., 2023; Näätänen and Picton, 1987) and modulate self-generation (Timm et al., 2013; Schröger et al., 2015), they are unlikely to fully account for the combination of early N1 enhancement and later P2 attenuation observed here.

Importantly, a clear dissociation across experiments challenges a purely attention-based interpretation of the ERP pattern. Across experiments, the critical manipulation consistently enhanced the N1 component. Whereas the P2 component showed reliable suppression only in Experiment 2. If a unitary increase in selective attention were the sole mechanism, it could not account for why it would produce a context-invariant N1 enhancement while yielding a context-specific, directionally opposite P2 effect. This temporal and directional fractionation therefore, suggests that, although attention-related facilitation is a plausible explanation for the N1 enhancement, the selective P2 suppression in Experiment 2 likely reflects additional context-dependent processing beyond attentional allocation, such as post-perceptual evaluative or categorization-related operations engaged by that context. This dissociation motivates a precision-based account of predictive modulation.

Across all three experiments, we observed a robust and highly consistent enhancement of the N1 component under the optional mapping of keypress-inducing-sound. We interpret this early modulation as a candidate signature of generalized motor preparation that operates before response selection. Previous research has established that the N1 component is reliably sensitive to top-down attentional control and is modulated by overall state variables such as arousal (Hillyard et al., 1973; Näätänen and Picton, 1987). In the present context, the N1 enhancement likely reflects a broadly tuned increase in sensory alertness or "preparatory gain" engaged when participants anticipate the

potential need to act, before a specific mapping is selected. This interpretation is consistent with evidence that action preparation can enhance perceptual sensitivity in a gain-like manner even before movement onset (Rolfs and Carrasco, 2012).

This early, global preparatory mechanism is conceptually distinct from the later P2 suppression. Whereas the N1 reflects a generalized mobilization of sensory-motor resources under optional-mapping demands, the P2 suppression appears to reflect mapping-specific control that emerges when competing stimulus–response alternatives must be resolved, potentially through selective down-weighting or inhibition of the non-selected mappings. This interpretation converges with targeted inhibitory mechanisms described in competitor-rule suppression accounts of task switching and rule competition (Meiran, Hsieh, and Dimov, 2010, 2011). Thus, our results suggest a two-stage control hierarchy: an initial, global enhancement of processing (N1), followed by a later, mapping-specific selection/suppression process (P2).

A design-related caveat is the differences in the occurrence certainty of auditory stimuli in different conditions. Auditory probes were less certain in GP (50 % of trials) than in PL (100 % of trials). The expected stimuli in PL could lead to suppression in auditory responses (e.g., Fishman et al., 2021; Han et al., 2019; see also Auksztulewicz and Friston, 2016 for a review), and hence could result in N1 enhancements in the contrast of GP minus PL. However, several lines of evidence suggest that probe-occurrence probability alone is unlikely to fully account for the observed N1 pattern. First, previous results do not show a consistent mapping between higher occurrence probability and reduced early auditory responses, and studies manipulating action-tone contingency probability do not reliably yield the predicted N1 attenuation with increased certainty (e.g., Harrison et al., 2023; Han et al., 2022). Second, in similar delayed action-preparation paradigms with probes occurring in half of the trials, N1 enhancement was not observed. For example, Li et al. (2020) reported N1 suppression during preparation relative to passive listening despite probes occurring only on a subset of trials. These findings argue against a "lower certainty leading to a larger N1" account. Moreover, the N1 modulation was correlated with the preparatory responses. Specifically, in GP probe-present trials, single-trial CNV amplitude (indexing preparatory engagement) predicted subsequent N1 amplitude on a trial-by-trial basis across all three experiments (all $p < 0.001$). This within-condition coupling is hard to explain by the differences in occurrence probability between conditions and instead supports the interpretation that the action preparation contributes to the modulation of N1 amplitude. Taken together, while probe certainty differences may contribute to GP–PL differences to some extent, the overall pattern is less likely to be driven primarily by occurrence probability and is more consistent with modulation by preparatory engagement.

A further design consideration concerns the visual cues signaling task context. The cue–condition mappings were fixed rather than counterbalanced across participants (Experiment 1: colors; Experiments 2–3: shapes), which could introduce a minor visual confound (e.g., differences in color/shape salience or sustained visual context). The severity is partly mitigated by the use of different visual cues across experiments and by the fact that the key effects replicated across experiments, suggesting that the findings are more likely driven by preparatory processes than by specific cue properties. Nonetheless, future work would be better fully counterbalance or randomize cue–condition mappings at the participant level (and/or equalize cue salience) to eliminate this potential confound.

We also observed a clear stimulus-category gradient in preparatory N1 enhancement—largest for human vacalizations, minimal for pure tones. We take this pattern to reflect the amount of early acoustic evidence required to resolve the probes under a response pad mapping that provides low action-outcome certainty. In our stimulus set, syllables contained richer and more informative spectrotemporal cues (and sharper onsets), coughs were more variable, and pure tones offered fewer diagnostic features; this would naturally yield stronger early gain

for human vocalizations and weaker effects for tones (Beechey et al., 2022; de Lange et al., 2018; Feldman and Friston, 2010; Irsik et al., 2021). To rule out acoustic confounds, future work should orthogonalize category and onset slope and equate loudness/spectrum, and include trial-level acoustic metrics in the analysis.

4.4. Time-frequency signatures of the motor-auditory loop

Time–frequency analyses time-locked to the auditory probe onset revealed robust oscillatory modulations that complement the ERP results and provide a more direct readout of the motor–auditory loop dynamics. In the motor ROI, β-band power was lower in GP than in PL during mid-to-late post-probe periods. This pattern is consistent with prior work showing that preparatory and movement-related activity is accompanied by suppression of sensorimotor β rhythms, and that the magnitude of β suppression scales with the degree to which an action plan is specified (Pfurtscheller et al., 2003; Tzagarakis et al., 2010; van Helvert et al., 2021). GP provides a clearer requirement to execute an action (a keypress) and a stronger preparatory drive, making it easier for the motor system to shift from a waiting/maintenance mode into a preparatory mode. By contrast, PL lacks information that an imminent action is required and thus keeps the motor system in a standby state. This interpretation also aligns with theoretical accounts that link higher β activity to maintaining the current sensorimotor state ("status quo") and reduced β activity to releasing that state in preparation for updating or initiating action (Engel and Fries, 2010; Jenkinson and Brown, 2011).

In the auditory ROI, the observations in the α band diverged from the prediction of a "gating-by-inhibition" account, which would anticipate increased α power in auditory cortex during motor preparation (Jensen and Mazaheri, 2010; Foxe and Snyder, 2011). Instead, in Experiments 2 and 3 (with a weaker trend in Experiment 1), we observed stronger α desynchronization for GP than PL, suggesting that the auditory region did not show the canonical α-power increase associated with inhibitory gating, but rather appeared to maintain (or even enhance) readiness to process the probe stimulus. Although the observations appear opposite to canonical demonstrations in which α increases gate task-irrelevant sensory regions, accumulating evidence suggests that α-mediated gating is strongly context-dependent, varying with task demands, cognitive load, and the behavioral relevance of potential distractors (Fodor et al., 2020; Gutteling et al., 2022; Heinz and Johnson, 2017; van Diepen et al., 2019; Bonnefond and Jensen, 2025). Critically, in our design, the auditory probe occurred in the preparation stage, making it difficult to treat as a purely irrelevant distractor. The probe is likely integrated as a task-relevant signal, potentially useful for monitoring or calibrating the action–sound relationship, so motor preparation may benefit from keeping the auditory cortex in a more "ready" state, and hence expressed as α ERD. Taken together, the results in the α band therefore do not necessarily indicate a failure of sensory gating; rather, they suggest that under conditions of active motor preparation in a dual-task context, the motor–auditory loop may favor enhanced auditory processing over a uniform inhibitory shutdown of auditory cortex.

4.5. From gain to suppression: A precision-based account of predictive modulation

Our results—larger N1 to probes during preparation (~90–130 ms) and small P2 attenuation when sensorimotor links are strong (~170–230 ms)—are not fully explained by a purely one-way, always-suppressive forward-model view (Blakemore et al., 1998; Wolpert et al., 1995). Instead, they suggest that expectations can have opposite effects over time: an early increase in responsiveness and a later reduction for predictable inputs. A simple way to think about this is to distinguish sensory (likelihood) precision from action-outcome (prior) precision. Early on, when probe categories such as syllables or piano tones are clear but the keypress-to-sound link is uncertain, responses to the probes are larger (N1 enhancement). Later, as the action-sound mapping becomes

reliable or sounds are self-generated, responses to predicted input are reduced (P2 attenuation). This time course fits accounts that propose an early tilt toward expected events followed by a re-balancing once evidence arrives (Kok et al., 2017; Press et al., 2020; Yon and Press, 2017), and it is in line with recent behavioral/EEG work showing a shift from stronger encoding of expected events to stronger encoding of surprising ones.

On this view, corollary discharge during preparation is not always inhibitory: when the action model is uncertain or mismatched, it may mainly keep the auditory system responsive so that deviations can be detected (larger N1). As experience builds and the mapping stabilizes, processing shifts toward suppressing predictable input (smaller P2). Importantly, we take this shift to depend more on prediction confidence than on elapsed time per se (de Lange et al., 2018; Näätänen and Picton, 1987; Schneider and Mooney, 2018). Future work may systematically vary the extent of action-outcome mapping training and the stability of sensory categories to dissociate the relative contributions of these two mechanisms across processing stages.

5. Conclusion

Across three EEG experiments using a delayed keypress paradigm, we show that motor-based prediction in an optional sensorimotor mapping produces a robust preparatory modulation of auditory processing. During preparation, auditory probes reliably elicited an enhanced N1, consistent with an early gain in sensory responsiveness when action–outcome predictions are relatively imprecise. Evidence for a later attenuation of P2 was more constrained: a P2 reduction was observed in the strongest-association context (Experiment 2) but was absent in Experiments 1 and 3 and was session-dependent in the within-subject analysis, suggesting that late attenuation is conditional and less robust than early gain. At execution, self-initiated sounds nevertheless showed robust, category-general attenuation of N1/P2, consistent with forward-model mechanisms operating once timing and outcome constraints are specified.

Taken together, these findings challenge a unitary suppressive view of motor prediction. Instead, they are compatible with a precision-weighed predictive coding interpretation in which preparatory motor states can up-weigh sensory processing under low predictive precision (yielding early gain) and may down-weigh predicted input when predictions become sufficiently reliable (yielding conditional late attenuation). We therefore present this gain-to-conditional-attenuation sequence as a working hypothesis motivated by the current results, which generates testable predictions for how the modulatory functions of motor-based prediction on auditory processes change with the reliability of action-sound mapping and sensory uncertainty.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT (OpenAI) in order to edit grammar and improve readability of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Xinjing Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xiaodan Zheng:** Writing – review & editing, Visualization,

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2026.121744.

References

Alday, P.M., 2019. How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. Psychophysiology. 56 (12), e13451.

Aliu, S.O., Houde, J.F., Nagarajan, S.S., 2009. Motorinduced suppression of the auditory cortex. J. Cogn. Neurosci. 21 (4), 791–802.

Audette, N.J., Schneider, D.M., 2023. Stimulus-specific prediction error neurons in mouse auditory cortex. J. Neurosci. 43 (43), 7119–7129.

Auksztulewicz, R., Friston, K., 2016. Repetition suppression and its contextual determinants in predictive coding. Cortex 80, 125–140.

Baess, P., Widmann, A., Roye, A., Schröger, E., 2013. Selective suppression of the auditory N1 component during self-triggered tones. Clin. Neurophysiol 124 (4), 734–742.

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.

Beechey, T., 2022. Is speech intelligibility what speech intelligibility tests test? J. Acoust. Soc. Am. 152 (3), 1573–1585.

Blakemore, S.-J., Rees, G., Frith, C.D., 1998. How do we predict the consequences of our actions? A functional imaging study. Neuropsychologia 36 (6), 521–529.

Blakemore, S.J., Wolpert, D.M., Frith, C.D., 2000. Why can't you tickle yourself? Neuroreport 11 (11), R11–R16.

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

Bonnefond, M., Jensen, O., 2025. The role of alpha oscillations in resisting distraction. Trends Cogn. Sci. (Regul. Ed.) 29 (4), 368–379.

Brunia, C.H.M., van Boxtel, G.J.M., 2001. Wait and see. Int. J. Psychophysiol. 43 (1), 59–75.

Carpenter, A.L., Shahin, A.J., 2013. Development of the N1–P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. Neurosci. Lett. 544, 56–61.

Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36 (3), 181–204.

Crapse, T.B., Sommer, M.A., 2008. Corollary discharge across the animal kingdom. Nat. Rev. Neurosci. 9 (8), 587–600.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., Hari, R., 2000. Speaking modifies voiceevoked activity in the human auditory cortex. Hum. Brain Mapp. 9 (4), 183–191.

de Lange, F.P., Heilbron, M., Kok, P., 2018. How do expectations shape perception? Trends Cogn. Sci. (Regul. Ed.) 22 (9), 764–779.

Doyle, L.M., Yarrow, K., Brown, P., 2005. Lateralization of event-related beta desynchronization in the EEG during pre-cued reaction time tasks. Clin. Neurophysiol. 116 (8), 1879–1888.

Drost, U.C., Rieger, M., Brass, M., Gunter, T.C., Prinz, W., 2005a. Action-effect coupling in pianists. Psychol. Res. 69 (4), 233–241.

Drost, U.C., Rieger, M., Brass, M., Gunter, T.C., Prinz, W., 2005b. When hearing turns into playing: movement induction by auditory stimuli in pianists. Quart. J. Exp. Psychol. 58 (8), 1376–1389.

Eliades, S.J., Wang, X., 2003. Sensorimotor interaction in the primate auditory cortex during selfinitiated vocalizations. J. Neurophysiol. 89 (4), 2194–2207.

Eliades, S.J., Wang, X., 2008. Neural correlates of the Lombard effect in primate auditory cortex. J. Neurosci. 28 (31), 10737–10749.

Engel, A.K., Fries, P., 2010. Beta-band oscillations—Signalling the status quo? Curr. Opin. Neurobiol. 20 (2), 156–165.

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

Fishman, Y.I., Lee, W.W., Sussman, E., 2021. Learning to predict: neuronal signatures of auditory expectancy in human event-related potentials. Neuroimage 225, 117472.

Flinker, A., Chang, E.F., Kirsch, H.E., Barbaro, N.M., Crone, N.E., Knight, R.T., 2010. Singletrial speech suppression of auditory cortex activity in humans. J. Neurosci. 30 (44), 16643–16650.

Fodor, Z., Marosi, C., Tombor, L., Csukly, G., 2020. Salient distractors open the door of perception: alpha desynchronization marks sensory gating in a working memory task. Sci. Rep. 10, 19179.

Fontana, F., Järveläinen, H., Papetti, S., De Pra, Y., 2024. Acoustic cues of keyboard mechanics enable auditory localization of upright piano tones. J. Acoust. Soc. Am. 156 (1), 164–175.

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.M., Roach, B.J., Mathalon, D.H., 2010. Assessing corollary discharge in humans using noninvasive neurophysiological methods. Nat. Protoc. 5 (6), 1160–1168.

Foxe, J.J., Snyder, A.C., 2011. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Front. Psychol. 2, 154.

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

Frühstorfer, H., Soveri, P., Järvilehto, T., 1970. Short-term habituation of the auditory evoked response in man. Electroencephalogr. Clin. Neurophysiol. 28 (2), 153–161.

Gómez, C.M., Marco, J., Grau, C., 2003. Preparatory visuo-motor cortical network of the contingent negative variation estimated by current density. Neuroimage 20 (1), 216–224.

Gonzalez, J., Nieto, N., Brusco, P., Gravano, A., Kamienkowski, J.E., 2024. Speech-induced suppression during natural dialogues. Commun. Biol. 7, 291.

Gordon, N., Koenig-Robert, R., Tsuchiya, N., van Boxtel, J.J.A., Hohwy, J., 2017. Neural markers of predictive coding under perceptual uncertainty revealed with hierarchical frequency tagging. Elife 6, e22749.

Greenlee, J.D.W., Jackson, A.W., Chen, F., Larson, C.R., Oya, H., Kawasaki, H., Chen, H., Howard 3rd., M.A., 2011. Human auditory cortical activation during selfvocalization. PLoS. One 6 (3), e14744.

Gutteling, T.P., Sillekens, L., Lavie, N., Jensen, O., 2022. Alpha oscillations reflect suppression of distractors with increased perceptual load. Prog. Neurobiol. 214, 102285.

Han, B., Mostert, P., de Lange, F.P., 2019. Predictable tones elicit stimulus-specific suppression of evoked activity in auditory cortex. Neuroimage 200, 242–249.

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. Cogn. Neurosci. 34 (8), 1488–1499.

Harrison, P.M., Davies, C.A., Frith, U., 2023. Exploring the internal forward model: action-effect prediction and attention in sensorimotor processing. Front. Hum. Neurosci. 17, 10350834. Article.

Heinks-Maldonado, T.H., Nagarajan, S.S., Houde, J.F., 2006. Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport 17 (13), 1375–1379.

Heinz, A.J., Johnson, J.S., 2017. Load-dependent increases in delay-period alpha-band power track the gating of task-irrelevant inputs to working memory. Front. Hum. Neurosci. 11, 250.

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

Horváth, J., Maess, B., Baess, P., Tóth, A., 2012a. Action preparation modulates auditory cortex responses: evidence from delay manipulations. J. Neurosci. 32 (13), 4309–4314.

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

Houde, J.F., Jordan, M.I., 1998. Sensorimotor adaptation in speech production. Science (1979) 279 (5354), 1213–1216.

Houde, J.F., Nagarajan, S.S., Sekihara, K., Merzenich, M.M., 2002. Modulation of the auditory cortex during speech: an MEG study. J. Cogn. Neurosci. 14 (8), 1125–1138.

Hughes, G., Desantis, A., Waszak, F., 2013. Attenuation of auditory N1 results from identityspecific action-effect prediction. Eur. J. Neurosci. 37 (6), 1152–1158.

Irsik, V.C., Almanaseer, A., Johnsrude, I.S., Herrmann, B., 2021. Cortical responses to the amplitude envelopes of sounds change with age. J. Neurosci. 41 (23), 5045–5055.

Ito, M., 1991. Short-term retention of a constructed motor program. Percept. Mot. Skills. 72 (1), 339–347.

Jenkinson, N., Brown, P., 2011. New insights into the relationship between dopamine, beta oscillations and motor function. Trends. Neurosci. 34 (12), 611–618.

Jensen, O., Mazaheri, A., 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front. Hum. Neurosci. 4, 186.

Kaganovich, N., Schumaker, J., 2016. Electrophysiological correlates of individual differences in perception of audiovisual temporal asynchrony. Neuropsychologia 86.

Khalilian-Gourtani, A., Wang, R., Chen, X., Yu, L., Dugan, P., Friedman, D., Flinker, A., 2024. A corollary discharge circuit in human speech. Proc. Natl. Acad. Sci. 121 (50), e2404121121.

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.

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

Knolle, F., Schwartze, M., Schröger, E., Kotz, S.A., 2019. Auditory predictions and prediction errors in response to self-initiated vowels. Front. Neurosci. 13, 1146.

Kok, P., Mostert, P., De Lange, F.P., 2017. Prior expectations induce prestimulus sensory templates. Proc. Natl. Acad. Sci. 114 (39), 10473–10478.

Lahav, A., Saltzman, E., Schlaug, G., 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J. Neurosci. 27 (2), 308–314.

Lee, H., Noppeney, U., 2011. Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc. Natl. Acad. Sci. U.S.A. 108 (51), E1441–E1450.

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

Li, X., Chu, Q., Lu, Y., Su, Y., Tian, X., 2025. Motor-based and memory-based predictions distinctively modulate sensory processes. Neuropsychologia 218, 109242.

Lorenz, A., Mercier, M., Trébuchon, A., Bartolomei, F., Schön, D., Morillon, B., 2025. Corollary discharge signals during production are domain general: an intracerebral EEG case study with a professional musician. Cortex 186, 11–23.

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.

Luna, F.G., Aguirre, M.J., Martín-Arévalo, E., Ibáñez, A., Lupiáñez, J., Barttfeld, P., 2023. Event-related potentials associated with attentional networks evidence changes in executive and arousal vigilance. Psychophysiology 60 (8), e14272.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190.

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

Meiran, N., Hsieh, S., Chang, C.-C., 2011. Smart inhibition: electrophysiological evidence for the suppression of conflict-generating task rules during task switching. Cogn., Affect., Behav. Neurosci. 11 (3), 292–308.

Meiran, N., Hsieh, S., Dimov, E., 2010. Resolving task rule incongruence during task switching by competitor rule suppression. J. Exp. Psychol.: Learn., Memory, Cogn. 36 (4), 992–1002.

Miall, R.C., Wolpert, D.M., 1996. Forward models for physiological motor control. Neural Networks 9 (8), 1265–1279.

Mifsud, N.G., Beesley, T., Watson, T.L., Whitford, T.J., 2016. Attenuation of auditory evoked potentials for hand and eye-initiated sounds. Biol. Psychol. 120, 61–68.

Mifsud, N.G., Whitford, T.J., 2017. Sensory attenuation of self-initiated sounds maps onto habitual associations between motor action and sound. Neuropsychologia 103, 38–43.

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.

Niemi, P., Näätänen, R., 1981. Foreperiod and simple reaction time. Psychol. Bull. 89 (1), 133–162.

Niziolek, C.A., Nagarajan, S.S., Houde, J.F., 2013. What does motor efference copy represent? Evidence from speech production. J. Neurosci. 33 (41), 16110–16116.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computat. Intell. Neurosci. 2011 (1), 156869.

Ott, C.G., Jäncke, L., 2013. Processing of self-initiated speech-sounds is different in musicians. Front. Hum. Neurosci. 7, 41.

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

Pfurtscheller, G., Graimann, B., Huggins, J.E., Levine, S.P., Schuh, L.A., 2003. Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement. Clin. Neurophysiol. 114 (7), 1226–1236.

Pilling, M., 2009. Auditory event-related potentials (ERPs) in audiovisual speech perception. J. Speech, Lang., Hear. Res. 52 (4), 1073–1081.

Press, C., Kok, P., Yon, D., 2020. The perceptual prediction paradox. Trends Cogn. Sci. (Regul. Ed.) 24 (1), 13–24.

Pritchard, W.S., 1981. Psychophysiology of P300. Psychol. Bull. 89 (3), 506–540.

Rao, R.P.N., 2024. A sensory-motor theory of the neocortex. Nat. Neurosci. 27 (7), 1221–1235.

Reznik, D., Guttman, N., Buaron, B., Zion-Golumbic, E., Mukamel, R., 2021. Action-locked neural responses in auditory cortex to self-generated sounds. Cerebral Cortex 31 (12), 5560–5569.

Reznik, D., Henkin, Y., Schadel, N., Mukamel, R., 2014. Lateralized enhancement of auditory cortex activity and increased sensitivity to selfgenerated sounds. Nat. Commun. 5, 4059.

Rittershofer, K., Yon, D., Press, C., 2025. Time-resolved EEG decoding reveals a flip from enhanced expected to unexpected action outcomes. In: *Proceedings of the Computational Cognitive Neuroscience (CCN)* 2025, p. 4. Extended abstract.

Rolfs, M., Carrasco, M., 2012. Rapid simultaneous enhancement of visual sensitivity and perceived contrast during saccade preparation. J. Neurosci. 32 (40), 13744–13752.

Ross, B., Barat, M., Fujioka, T., 2017. Sound-making actions lead to immediate plastic changes of neuromagnetic evoked responses and induced β-band oscillations during perception. J. Neurosci. 37 (24), 5948–5959.

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., Schröger, E., 2013. I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when. Front. Hum. Neurosci. 7, 407.

Schneider, D.M., Mooney, R., 2015. Motor-related signals in the auditory system for listening and learning. Current Opin. Neurobiol. 33, 78–84.

Schneider, D.M., Mooney, R., 2018. How movement influences hearing. Annu Rev. Neurosci. 41, 553–574.

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.

Shahin, A., Roberts, L.E., Pantev, C., Trainor, L.J., Ross, B., 2005. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. Neuroreport 16 (16), 1781–1785.

Tandonnet, C., Burle, B., Vidal, F., Hasbroucq, T., 2003. The influence of time preparation on motor processes assessed by surface laplacian estimation. Clin. Neurophysiol. 114 (12), 2376–2384.

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

Timm, J., SanMiguel, I., Saupe, K., Schröger, E., 2013. The N1 component and its relation to preparatory processes: effects of self-initiated vs. externally triggered auditory stimuli. Psychophysiology 50 (3), 233–240.

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.

Tzagarakis, C., Ince, N.F., Leuthold, A.C., Pellizzer, G., 2010. Beta-band activity during motor planning reflects response uncertainty. J. Neurosci. 30 (34), 11270–11277.

Tzagarakis, C., West, S., Pellizzer, G., 2021. Neural encoding of the reliability of directional information during the preparation of targeted movements. Front. Neurosci. 15, 679408.

van Diepen, R.M., Foxe, J.J., Mazaheri, A., 2019. The functional role of alpha-band activity in attentional processing: the current zeitgeist and future outlook. Curr. Opin. Psychol. 29, 229–238.

van Helvert, M.J.L., Oostwoud Wijdenes, L., Geerligs, L., Medendorp, W.P., 2021. Cortical beta-band power modulates with uncertainty in effector selection during motor planning. J. Neurophysiol. 126 (6), 1891–1902.

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.

Whitford, T.J., Jack, B.N., Pearson, D., Griffiths, O., Luque, D., Harris, A.W.F., Spencer, K.M., Le Pelley, M.E., 2017. Neurophysiological evidence of efference copies to inner speech. Elife 6, e28197.

Whitford, T.J., Mathalon, D.H., Shenton, M.E., Roach, B.J., Bammer, R., Adcock, R.A., Ford, J.M., 2011. Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. Psychol. Med. 41 (5), 959–969.

Wolpert, D.M., Miall, R.C., 1996. Forward models for physiological motor control. Neural Netw. 9 (8), 1265–1279.

Wolpert, D.M., Ghahramani, Z., Jordan, M.I., 1995. An internal model for sensorimotor integration. Science (1979) 269 (5232), 1880–1882.

Yon, D., Press, C., 2017. Predicted action consequences are perceptually facilitated before cancellation. J. Exp. Psychol.: Hum. Percept. Perform. 43 (6), 1073–1083.

Zheng, X., Zhu, H., Li, S., Tian, X., 2022. The generic inhibitory function of corollary discharge in motor intention: evidence from the modulation effects of speech preparation on the late components of auditory neural responses. eNeuro 9 (6), 0309–0322.