Reviewed Preprint
v1 • May 20, 2026
Not revised

For correspondence:
haiqin.zhang@ens.psl.eu

Competing interests: No competing interests declared

Funding: See page 26

Reviewing editor: Nai Ding, Zhejiang University, China

© 2026, Zhang et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Auditory-motor surprisal reveals learning across multiple timescales during exploration and production

Haiqin Zhang[1] ✉, **Giorgia Cantisani**[2], **Shihab Shamma**[3]

[1]Laboratoire des systèmes perceptifs, CNRS, École Normale Supérieure, PSL University, Paris, France • [2]Sciences et Technologies de la Musique et du Son Lab, CNRS, IRCAM, Sorbonne Université, Paris, France • [3]Neural Systems Laboratory, University of Maryland College Park, College Park, United States

eLife Assessment

This **valuable** study builds a novel auditory-motor paradigm to investigate how the brain learns associations between movements and their auditory consequences. **Solid** evidence is provided for early ERPs (50-100 ms latency) reflecting violations of established key-pitch mappings. The writing, however, could be streamlined to better emphasize the paper's key contribution, and some statistical analyses might be improved.

https://doi.org/10.7554/eLife.111080.1.sa2

Abstract

Auditory-motor learning is critical in mastering the production of complex sounds, such as speaking and playing music. It is anchored upon internal models of interactions between actions and their sensory consequences, which are fine-tuned by minimizing the errors between the predicted and received sound. Here, we applied the concept of surprisal to a piano-playing task to probe the neural dynamics of sensorimotor learning. Specifically, during play, the key-pitch map was changed unpredictably among three map configurations: normal, inverted, and shifted-inverted. At the change boundaries, a signature of violated motor-to-auditory predictions was found in the auditory evoked responses at N100 which could not be attributed to either purely auditory surprisals or motor execution errors. This surprisal is modulated by short-term context, with greater surprise following longer periods of no map change, indicating that the brain continuously tracks short-term map contexts and rapidly adapts to them. In contrast, 30 minutes of extended goal-directed training on a single map modulated P50 amplitude only for that map, which can be explained by a slow, persistent modulation of motor predictions from the auditory signals. Hence, while auditory predictions from motor actions are rapidly and implicitly learned within short-term contexts, the complementary process of adjusting motor inferences from auditory inputs requires targeted training sustained over time. Our approach of studying auditory-motor surprisal in time-varying sequences reveals that auditory-motor learning is fast, context-sensitive, and shaped by both short- and long-term experience.

Significance statement

Understanding how the brain links motor actions with their sensory consequences is key to explaining how complex skills are acquired and how they adapt to changing environments. Prior work has shown that short-term sensory feedback supports rapid adaptation. Yet, the neural mechanisms underpinning the evolution of internal sensorimotor associations across different stages of learning remain to be elucidated. We address this challenge by extending the concept of *surprisal*, traditionally used in studies of perception, to the sensorimotor domain. Results show that surprisal responses are modulated by both short-term sensory feedback and longer-term training, suggestive of two distinct neural mechanisms underlying sensorimotor learning. These

findings advance our understanding of the neural dynamics of sensorimotor learning and inform development of technologies that interface with sensorimotor systems, such as virtual reality and brain–machine interfaces.

Introduction

Continuous interactions between the sensory and motor cortical areas are fundamental for the execution of complex actions, such as speaking, playing music, writing, and balance control, among others. Control theorists described early how sensory feedback enables rapid and accurate adjustment of motor commands to achieve the intended outcomes and cope with noise (1; 2). In this context, action execution can be seen as a closed-loop iterative process within a sensorimotor system that combines contributions from both long-term knowledge accumulated through experience and short-term sensory feedback for fine control and rapid adaptation to a changing environment during performance (3).

Three processes underlying sensorimotor learning: *exploration*, *skill acquisition* and *adaptation*

Sensorimotor interactions are fundamentally based on internal models that are flexible enough to adapt over time to reflect experience and changing circumstances. We conjecture that sensorimotor learning comprises two distinct phases involving different neural processes: an unsupervised phase of initial *exploration* of the sensorimotor space, and a supervised phase of targeted *skill acquisition* through repeated practice to control expression within that space. A third process, *adaptation*, underlies both phases, enabling the formation of motor-to-auditory associations during exploration, and the refinement of inverse auditory-to-motor ones during skill building.

During *exploration*, the sensorimotor space is freely surveyed to discover action-auditory feedback pairings, e.g., babbling for a baby or exploring the keys of a piano for the first time (4; 5; 6). In this process, an internal model emerges that links motor commands to their sensory outcomes—in our case, key presses to piano notes—with associations forming rapidly and implicitly in an unsupervised manner as we shall demonstrate. During *skill acquisition*, the brain undergoes many action-feedback iterations in a stable sensorimotor environment to master sophisticated sequences of motor actions so as to reproduce a target sensory output, e.g., a pianist learning the sequence of keys to play a specific melody, or a baby learning how to say an intended word. Such skill acquisition typically occurs in a supervised manner over an extended period of time, involving repeated, targeted training. Crucially, *adaptation* underlies both phases, with the sensorimotor system able to flexibly adapt to changes in the environment (e.g., ambient acoustics) or in the system itself (e.g., musician fatigue, mistuned instrument) by making online adjustments to the auditory-motor maps.

Here, we shall investigate the neural dynamics of learning through these different phases of auditory-motor map formation. Specifically, our work builds upon a theoretical framework, referred to as the *Mirror Network* (7), that links control-theoretic principles to neural substrates for auditory-motor tasks, where the action-feedback loops reside in complementary pathways between motor and auditory regions in the cortex. In this framework (1) the basic substrate of the exploration phase is a *forward* pathway (decoder) from the motor to the auditory regions generating *predictions* of the sounds corresponding to the actions; and (2) the substrate for the skill acquisition phase is an *inverse* pathway (encoder) from auditory to motor regions translating an intended sound into its corresponding motor commands (Figure 1). Both projections are *adaptive* and evolve as needed with each iteration by matching auditory predictions with sensory feedback (7). Hence, the resulting prediction errors yield a measure of movement accuracy that is then mapped back to the motor regions to inform how the motor plan should be adjusted. For clarity, we shall refer to the *external* movement-to-sound correspondence as the *key-pitch* map, and the neural representation of the map as the *auditory-motor* map. The key-pitch map is

determined by factors in the external environment such as the piano keyboard configuration, while the auditory-motor map exists only in the brain's sensorimotor system and is adjusted through learning.

Assuming these predictions are generated by an internal sensorimotor model, we can relate prediction errors to the information-theoretic formulation of *surprisal*, namely the information content $IC = -\log p(x \mid c)$, which quantifies the unexpectedness of an event x by a causal predictive model given its preceding context in the sequence c. In this formulation, higher surprisal corresponds to events that strongly violate the model's contextual predictions.

Disentangling surprisals in auditory perception, motor action, and auditory-motor tasks

Surprisal responses are the signature of prediction errors, and hence can be used as a neural marker for understanding how both short and long-term regularities shape sensory predictions. This concept has been largely used in auditory perception studies to reveal how predictions about upcoming sound tokens (e.g., musical notes or speech sounds) are driven both by the immediate preceding context (e.g., unfolding structure of the current melody), and by the long-term knowledge of transition probabilities between tokens as acquired through musical exposure through lifetime (8; 9).

Previous electrophysiological studies have demonstrated how auditory predictions and the sensory input itself are encoded with opposite polarity: when predictions match the sensory input, their responses cancel out, resulting in attenuated activity (10; 11). By contrast, *prediction violations* have been shown to modulate neural responses (11). Notably, continuous neural responses recorded by both invasive and non-invasive electrophysiology can be explained by time-varying surprisal for passive listening of continuous speech (12; 13), music (8; 14; 15; 16), or more generally, structured auditory sequences (17). Further, Event-Related Potentials (ERPs) time-locked to note or phoneme onsets of highly surprising tokens exhibit a higher evoked activity at N100 and P200 peaks (8; 18; 9).

Here, we extend the concept of surprisal to auditory-motor production, in which upcoming auditory inputs are predicted from motor commands in accordance with the auditory-motor associations formed during the exploration phase. As with purely perceptual auditory surprisals, we expect auditory-motor surprisals to be driven by both short-term sensory feedback and by existing knowledge acquired through sustained training, *i.e.*, skill acquisition. For example, in the short term, pianists rapidly explore and adapt their sensory expectations when playing on a keyboard tuned slightly higher or with a broken key. On the other hand, long-term musical training helps pianists build priors about the relative pitches of neighboring keys and extrapolate the expected sound of unplayed ones given a small number of exploratory keystrokes. We emphasize a key distinction between the pure *auditory surprisal* investigated by previous studies, which reflects the violation of auditory predictions based on previous sensory context, and the *auditory-motor surprisal* studied here, which reflects violations of auditory predictions given a set of motor commands and the preceding motor sequence. Thus, such auditory-motor surprisals are expected to exist for *self*-produced sounds, but not necessarily for all sounds.

Suppression of auditory responses to self-produced movements is a well-documented phenomenon (10; 19; 20). Analogous to the pure perceptual auditory surprisal case, suppression is thought to be driven by accurate internal predictions of sound given an executed movement. Violations of such internal predictions have been measured with ERPs in classical oddball paradigms introducing unexpected sensory feedback (21; 22; 23). These designs, however, confound auditory-motor with pure auditory surprisal driven by explicit target–response comparisons. Recent experiments addressed this confound by comparing a simple auditory-motor pairing that is either predictable or completely random; still, the key-pitch mapping was limited to pressing one or two buttons (23). As a result, trials are treated as independent, precluding

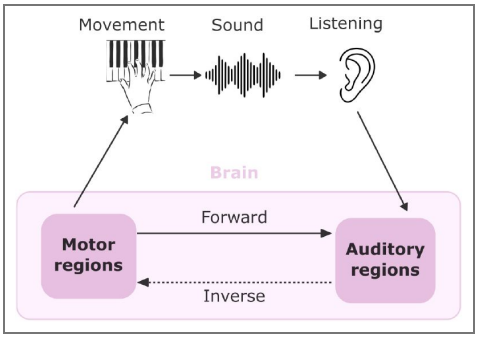


Figure 1. Schematic of the MirrorNet model proposed by Shamma et al., (7) in relation to the neural substrates of a sound production task.

The *forward* pathway (decoder) maps the motor to the auditory regions, generating *predictions* of the sounds corresponding to the actions; the *inverse* pathway (encoder) maps auditory to motor regions, translating an intended sound into its corresponding motor commands. Predictions are informed by the previous context, *i.e.*, the previous actions in the motor sequence and the previous tokens in the auditory one.

investigations of trial-to-trial learning and adaptation despite behavioral evidence that auditory-motor systems rapidly adjust to short-term perturbations (e.g., speech production with a bite block or artificial palate) (24; 25; 23).

Experimental paradigm and hypotheses

To address the above limitations, we introduce a continuous, complex auditory-motor task that dissociates sensorimotor from purely auditory surprisal, enabling us to track how neural dynamics support rapid adaptation and long-term learning across both exploration and skill acquisition. In the task—referred to as the *variable-map-playing task*—auditory-motor surprise is elicited by unpredictably changing the key-pitch mapping of a keyboard (Figure 2A). As we shall demonstrate, this design minimizes the contribution of purely auditory surprisal, ensuring that the observed electrophysiological response relates to surprisal elicited by the auditory-motor task. This is done by comparing responses to the first note played after a map change (referred to as *first* keystrokes) to those of subsequent notes (referred to as *other* keystrokes).

To assess the relative contribution of auditory and motor components to the neural signature of auditory-motor surprisal, we included an *auditory-only* condition (passive listening) and a *motor-only task* (muted playing). Neural responses to these tasks were compared with those to the variable-map-playing task, which combines auditory and motor responses. We also measured the learning of auditory-motor associations at three different time-scales. The first is the rapid learning that occurs during the exploration phase, whose effects can be observed at the keystroke-to-keystroke time scale. The second is a longer-term training (e.g., skill acquisition by targeted practice), which focused solely on the effects of a specific unfamiliar key-pitch map (Figure 2C). Finally, the third considers the effect of longer-term musical education—the participants' musical background—on the formation of the auditory-motor maps.

By examining neural markers of auditory-motor surprisal in the three different tasks, this study provides new insights into the temporal dynamics of auditory-motor learning, as well as its enhancement by short-term training and long-term expertise. We approached our investigation with the following hypotheses:

1. The *first* keystroke after a map change will violate the auditory prediction induced by the ongoing motor command (*i.e.*, forward pathway, Figure 1) and therefore elicit a stronger surprisal response than *other* (following) keystrokes (Figure 2B).
2. Since our task violates predictions by altering auditory feedback rather than introducing motor perturbations, the auditory-motor surprisal will be reflected in the auditory response as in the work of Di Liberto et al., (8).
3. Training on a specific key-pitch map will strengthen the predictions of motor commands from the sounds heard (*i.e.* inverse pathway, Figure 1), such that keystrokes within this trained map will be *less* surprising than those played within untrained ones.
4. The magnitude of surprisal responses correlates with lifetime musical expertise.

Results

Experimental sessions consisted of three main blocks (Figure 2C), each lasting about 30 minutes: pre-training, training, and post-training. Pre- and post-training blocks were identical and consisted of three 10-minute tasks: *passive listening*, *mute playing*, and *variable-map playing*. In the *variable map playing* task, participants were asked to play short, approximately 4-note melodies separated by a pause of 1-2 seconds, on a keyboard whose key-pitch map changed unpredictably every 2–10 seconds (Figure 2A). Participants were instructed to vary the melodies played (*i.e.*, to play the four keys randomly over time, both in their order and intervals). An example recording detailing the map changes, along with the auditory output, is available in the supplementary materials (Figure S1; supplementary video).

To further dissociate auditory and motor contributions to surprisal, participants also completed a passive listening and a mute playing task (Figure 2C) to generate EEG data, which were used for training modality-specific decoders, as detailed later. In addition, a 30-minute training session

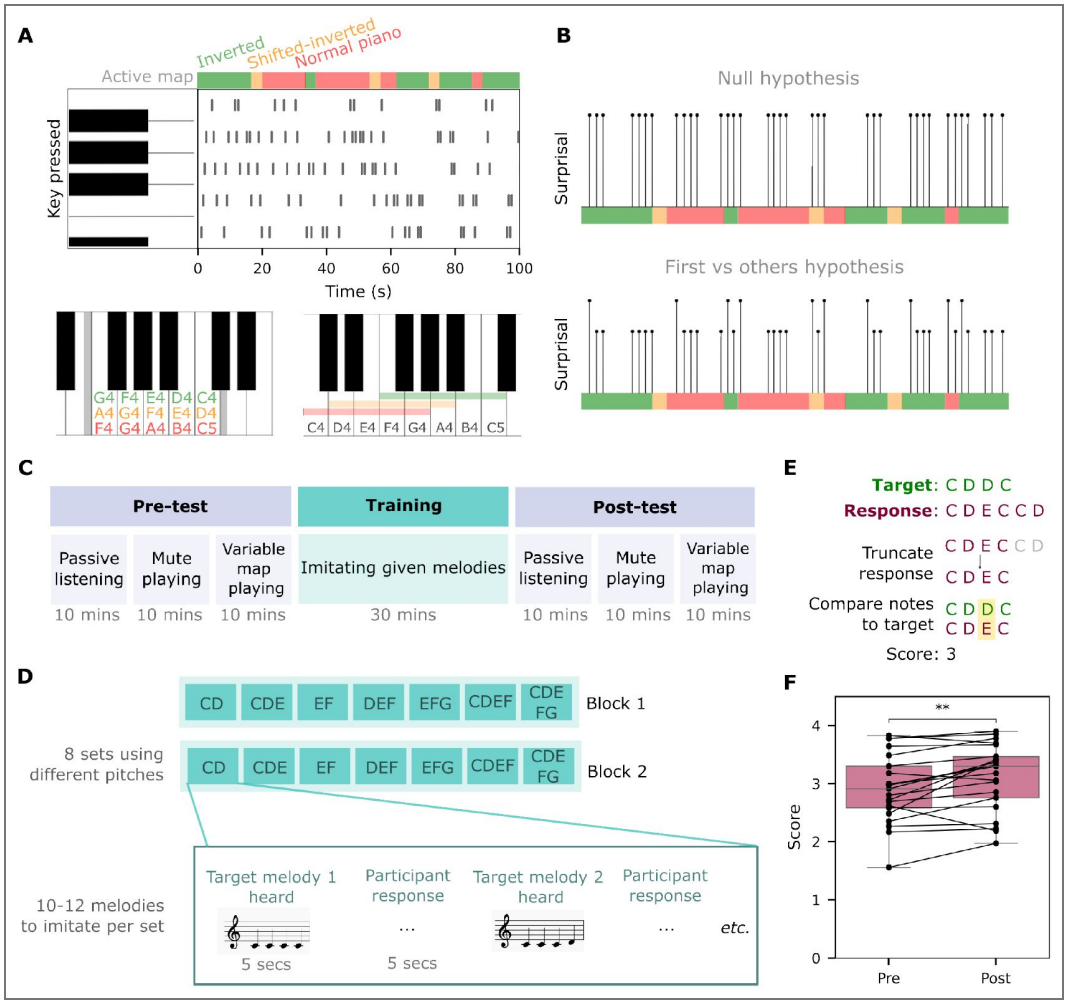


Figure 2. Experimental paradigm and hypotheses.

A. Example of keys pressed by a participant over time. Vertical lines mark individual strokes, while the colored bar shows the active key-pitch map (green: inverted, yellow: shifted-inverted, red: normal). The small keyboards show key-pitch assignments and their pitch distribution on a standard piano keyboard, for each map. **B**. Hypotheses about surprisal differences between *first* and *other* keystrokes after a map change. The null hypothesis assumes no effect of the map change; the 'first vs. others' hypothesis predicts higher surprisal for the first keystroke than for later ones. **C**. An experimental session consisted of three main blocks, each lasting approximately 30 minutes: pre-training, training, and post-training. The pre- and post-training blocks were identical, each including three 10-minute tasks: passive listening, mute playing, and variable-map playing. **D**. During training, participants imitated 4-note melodies by playing them back on the keyboard, in two blocks of increasing difficulty. **E**. Training was evaluated note-by-note against the target melody, ignoring rhythm and timing. **F**. Mean imitation scores for each block show learning over time (Wilcoxon signed-rank test, $p = 0.002$.). Lines denote individual participants.

with a fixed key-pitch map (the *inverted* map) was included to assess the effects of extended skill acquisition on surprisal responses. The training consisted of imitation trials in which participants had one attempt to play 4-note melodies immediately after hearing them once. Imitation trials were divided into two identical blocks with increasing difficulty within each block. To measure learning and engagement, each melody was scored based on how accurately it matched the target one (Figure 2E). Indeed, scores were significantly higher in the second than in the first block (Wilcoxon signed-rank test: $W = 32.0$, $p = .002$; Figure 2F).

Neural signatures of auditory-motor surprisal

We began with the hypothesis that the evoked responses to the *first* keystrokes after a map change would elicit stronger auditory-motor surprisals than those of *others* due to the change in the key-pitch map (Figure 2B). To test this, we compared amplitudes of note-onset ERPs of the two classes of keystrokes. An initial exploratory analysis (details in Section 6) identified candidate time points with a significant difference in ERP amplitude between *first* and *other* ERPs (Δ_{f-o}). We identified two time points of interest at 50 and 100 ms (P50 and N100 in Figure 3A, S2) for which we tested the robustness of the effect with a computationally-intensive bootstrap analysis at a few representative electrodes. This analysis revealed that Δ_{f-o} at N100 was significant in centro-frontal electrodes as shown in Figure 3B (only significant channels are displayed; $p < 0.05$, FDR-corrected Wilcoxon signed-rank test). Δ_{f-o} bootstrapped distributions are shown for three representative channels relative to the region of significant electrodes: Fz within the region showing the most marked difference, Cz at the edge, and Iz outside showing no difference (Figure 3C).

To distinguish contributions of purely auditory versus auditory-motor surprisal to the evoked responses, we conducted a control experiment in which a set of naïve participants passively listened to the audio recordings of the notes played by participants in the main experiment. While significant Δ_{f-o} were found at approximately 100 ms for the playing participants, listening participants only showed a significant Δ_{f-o} at 200 ms (Permutation cluster test, Figure 3D). Figure 3E illustrates the bootstrapped distributions of N100 Δ_{f-o} at the Fz electrode. Remarkably, only the playing distribution exhibited a significant difference from the null distribution ($p < 0.001$, empirical p-value). Further, playing and listening distributions were significantly different ($p < 0.001$, permutation test, Figure 3E).

To explain the finding that the only listening condition had a cluster of interest at 200 ms, we examined the statistical properties of the note sequences themselves by evaluating the sequential surprisal of each note in the played melodies using IDyOM (*Information Dynamics of Music*), a statistical model of musical structure that causally estimates the probability distribution over all possible note-pitch continuations based on previous context in the sequence (26). Interestingly, a permutation test on IDyOM surprisals revealed that notes produced by *first* keystrokes were indeed, on average, more surprising than *others* ($p < 0.001$, Figure 3F), although the overall surprisals distribution did not differ.

Context-sensitive formation of auditory-motor associations

We next tested whether Δ_{f-o} at N100 was modulated by the number of keystrokes played in the map preceding the map change. Thus, we defined a *0-keystroke* context as the case where the note immediately preceding the current *first* keystroke was itself a *first* keystroke in a previous map. A *1-keystroke* context corresponds to a case in which only one non-*first* note precedes the *first* keystroke and a *first* keystroke occurs two keystrokes prior, and so on. We hypothesized that *first* keystrokes with a longer context in the previous map would be more surprising and thus exhibit a larger N100 relative to *others* (Figure 4A). We found that Δ_{f-o} at N100 between each context category of *firsts* and the average amplitude of all *others* was significant regardless of the number of keystrokes played in the preceding map (t-test, $p = 0.005$ for *0-keystroke*, $p < 0.001$ for *1/2/3-keystroke*, Figure 4B), supporting our previous finding that *firsts* are always more surprising than *others*. Furthermore, the magnitude of this difference increased monotonically as a function

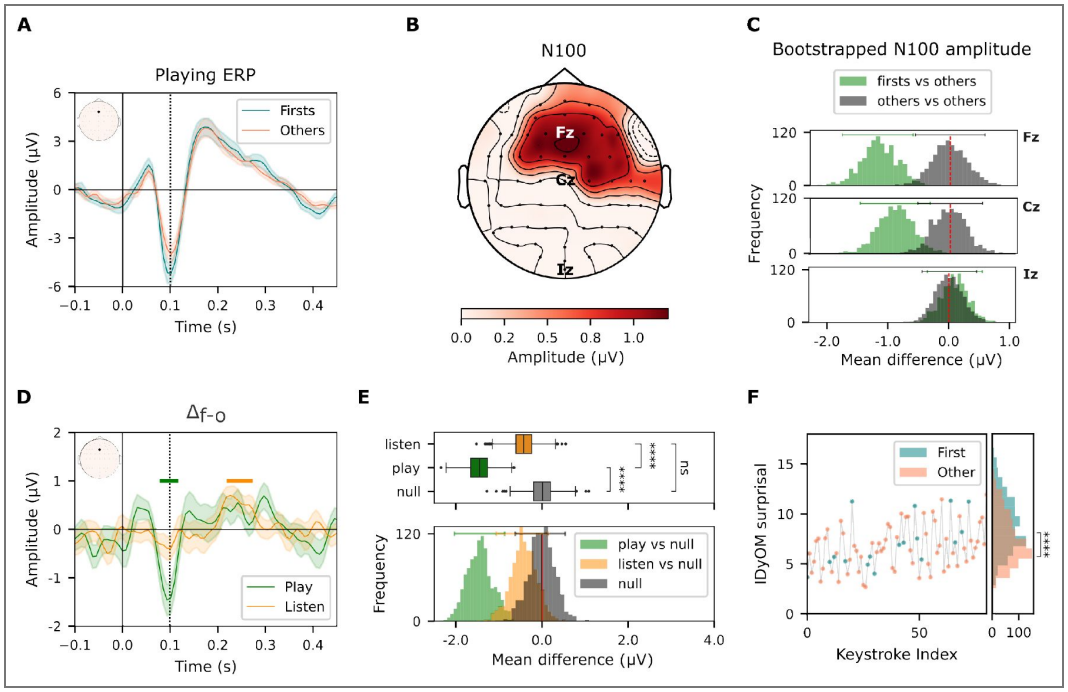


Figure 3. Neural signatures of auditory-motor surprisal

A. Grand-average ERPs of *firsts* and *others* at Fz. The dashed line marks 100 ms, where responses differ significantly **B**. Topography of N100 amplitude differences between *first* and *others* (Δ_{f-o}), masked to significant channels (FDR-corrected Wilcoxon signed-rank test). **C**. Bootstrapped Δ_{f-o} distribution (100-sample resamples, one from *firsts* and one from *others* over 1000 iterations) versus a null distribution where both samples are drawn from the pool of other keystrokes only. Horizontal bars indicate the CI95. **D**. Δ_{f-o} difference waves. Bars indicate regions of interest identified by the cluster-based permutation test ($p < 0.05$). **E**. Distribution of N100 Δ_{f-o} in the playing and listening experiments, compared with the null distribution; bars mark significant clusters from the permutation cluster test ($p < 0.05$). **F**. Surprisal of heard notes as calculated by cross-validation using IDyOM: the scatter plot shows an excerpt from one recording, the histogram shows surprisal distributions over all recordings for all *firsts* and a size-matched random sample of *others* ($p < 0.001$).

of the number of preceding keystrokes in the previous map. With a 3-keystroke context, Δ_{f-o} at N100 with was significantly greater than that of a 0-keystroke context (*t*-test, uncorrected for multiple comparisons, $p = 0.027$, Figure 4B).

In a complementary analysis, we examined the surprisal of *other* keystrokes as a function of the number of keystrokes since the last map change, with the hypothesis that it would decrease with a certain decay factor (Figure 4C). To test this hypothesis, we categorized each keystroke based on the number of preceding keystrokes since the last map change (*firsts*, by definition, have no preceding keystrokes). Again, in agreement with the results discussed in Section, *firsts* have significantly higher N100 amplitudes than all *others* (Figure 4D, *t*-test, highest *p* among all pairwise comparisons between *firsts* and others: $p = 0.00204$). However, we found no effect of the number of intervening keystrokes since key-pitch map change (*t*-test, $p > 0.05$ for all pairwise comparisons, Figure 4D).

Disentangling auditory and motor components in the playing responses

During play, auditory and motor components of the neural responses are intermingled. This is reflected in the evoked responses of the three different conditions: *playing*, which generates both components; *listening*, which produces pure auditory responses; and mute playing evoking pure motor responses (Figure 5A). The playing ERP, therefore, can be viewed as a combination of these auditory and motor components. The simplest such combination is the linear weighting of the pure components, which optimally fits the composite of the playing ERPs as illustrated in Figure S10.

We used a decoding approach to disentangle the auditory and motor components in the playing (auditory-motor) condition and assess how they are affected by training. For each participant and condition (*i.e.*, auditory and motor), we learned the optimal mapping from time-lagged EEG activity to stimulus-related features via regularized linear regression (27) (Figure 5A). The *auditory decoder* was trained to reconstruct *sound onsets* from the passive listening EEG, and the *motor decoder* to reconstruct *keystroke onsets* from *mute* playing EEG (Figure 5B). Both decoders were then applied to *variable map playing* EEG (Figure 5C-D), and the mean amplitude of their predictions at note onsets was compared in pre- and post-training (Figure 5E). Note that the decoders are temporally agnostic, being trained on a wide time window (±500 ms around onsets), so they could not rely on EEG amplitude at a single latency.

Both decoders performed above chance in reconstructing note onsets and keystrokes, with higher amplitudes at note-onset times (Figure S9E). The *auditory decoder* predicted reliably greater amplitudes at *first* than at *other* keystrokes (main effect of keystroke type: $F(1, 17) = 34.26$, $p < .001$, Greenhouse-Geisser corrected), with no main effect of training or interaction. The *motor decoder* also predicted greater amplitudes at *first* keystrokes ($F(1, 17) = 6.53$, $p = .020$, Greenhouse-Geisser corrected), again without a main effect of training nor interactions. Yet, pairwise Wilcoxon tests revealed that amplitudes at *first* keystrokes were greater than those of *others* in post-($p = 0.00769$) but not pre-training, a result that will be discussed further in the next section. Overall, these findings suggest that auditory-motor prediction violations in the playing condition are mainly reflected in the auditory, rather than motor component of the neural response.

The effects of targeted skill-acquisition training

The impact of extended training on auditory-motor surprisals was determined by comparing the *first* and *others* ERPs, as well as their difference (Δ_{f-o}) before and after training (Figure 6A, B). In both pre- and post-training, *first* keystrokes have qualitatively larger N100 than *others*. An unanticipated finding was that the P50 peak found in Δ_{f-o} before training was significantly attenuated after training (Figure 6B, C). A repeated-measures ANOVA revealed for N100 a main effect of keystroke type $(F(1, 17) = 22.92, p < .001, \eta^2_G = .067)$ and training $(F(1, 17) = 4.49, p = .049, \eta^2_G = .026)$ but no interaction $(F(1, 17) = 0.13, p = .724, \eta^2_G = .000)$. For

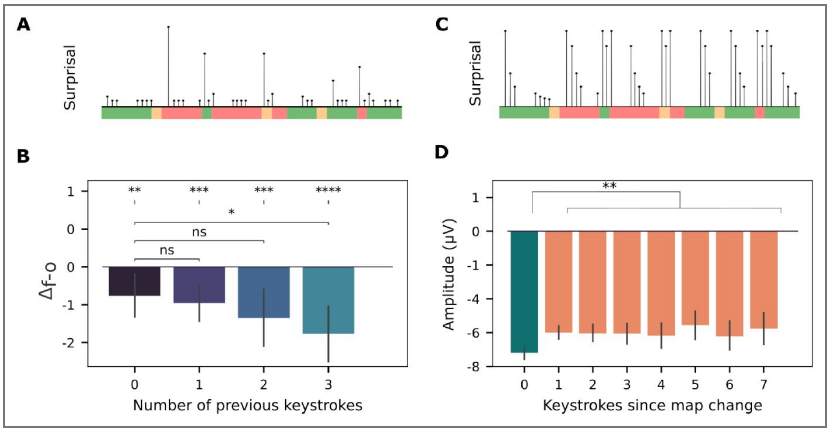


Figure 4. The influence of short-term context

A. Hypothesis: surprisal is modulated by the number of keystrokes in the preceding map. **B**. Bar plot showing mean difference in N100 amplitude between *first* and *other* keystrokes (Δ_{f-o}) as a function of the number of keystrokes in the previous map. Independent samples *t*-test for comparisons between *firsts* with different numbers of previous keystrokes (bottom stars); one-sample *t*-test for comparisons between *firsts* and others (top stars). **C**. Hypothesis: surprisal is modulated by the number of keystrokes since the *first* keystroke in the current map. **D**. Δ*f–o* sorted by the number of keystrokes since the map change. Independent samples *t*-test.

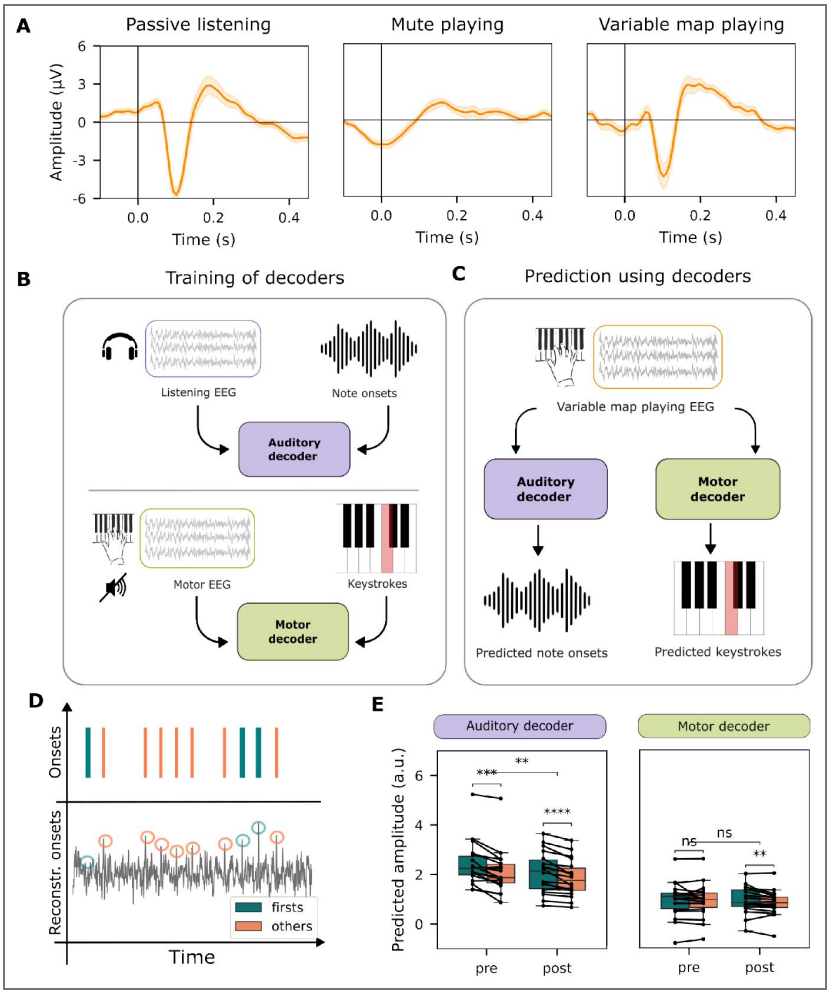


Figure 5. Disentangling auditory and motor components of playing responses

A. Grand average ERPs to note onsets at Fz for passive listening, mute playing, and variable map playing. **B**. The auditory decoder reconstructs note onsets from passive listening EEG; the motor decoder reconstructs key presses from mute playing EEG. **C**. Both decoders are then applied to the playing EEG. **D**. Example of ground truth and reconstructed onsets. **E**. Reconstruction amplitudes at *first* and *other* keystrokes are averaged and compared in pre- and post-training. Lines represent participants; bars show medians and quartiles. Wilcoxon signed-rank test, $p < 0.001$.

P50, there was no main effect of keystroke type$(F(1,17) = 0.15, p = .700, \eta^2_G = .000)$, nor training $(F(1,17) = 0.01, p = .912, \eta^2_G = .000)$. However, there was an interaction between the two $(F(1,17) = 10.45, p = .005, \eta^2_G = .046)$.

Topographic distributions were consistent with these results, revealing significant Δ_{f-o} in central-frontal electrodes in both pre- and post-training for the N100, but only before training for the P50 (FDR-corrected Wilcoxon signed-rank test, only significant electrodes are displayed, significance threshold at p = 0.05, Figure 6D).

To explore the meaning of this effect of training at P50, we compared the mean Δ_{f-o} at P50, sorted by the key-pitch map that was being entered into after a map change (for a visual summary of the sorting method, see Figure S7). We found a significant difference between pre- and post-training *only* in the inverted map (within-subjects Benjamini–Hochberg corrected *t*-test, p = 0.00281, Figure 6F). A similar analysis on N100 revealed no differences between pre- and post-training (FDR-corrected Wilcoxon signed-rank test, Figure 6F;).

To summarize, the effects of 30-minute training on the *inverted* key-pitch map resulted in a significant decrease in the Δ_{f-o} at P50. This effect stems primarily from a change in the amplitude of the motor components, which decreased relative to those of the *firsts* after training (Figure 5E). Since the motor ERPs are *negative* near 50 ms (Figure 5A), the net effect of the decrease of *other* motor component is a larger negative value in Δ*f*–*o*, resulting in the change at P50 in Figure 6C. The meaning and significance of this change is further elaborated upon in the Discussion and Supplementary (Figure S10), which includes a more detailed analysis of how the motor component affects the P50.

The effects of musical expertise

Lifelong musical experience, including formal musical training as well as more general musical sophistication of non-musicians, can be considered as another form of training that preceded our recording sessions. We hypothesized that these factors may enhance participants' ability to build auditory-motor associations and, therefore, their sensitivity to auditory-motor map violations. We use the training score of our learning task as a proxy for musical training, assuming that a higher musical background leads to higher scores in our learning task. Specifically, we measured the correlation between Δ_{f-o} at N100 and the musical ability as measured by the training score in the pre-training period (Figure 2F, Figure 6L; Pearson's *r*, $r(17)$ = 0.48, p = 0.045). The same analysis at P50, split by the active key-pitch map, did not yield any significant correlations (Figure S8).

Discussion

Learning how to map movements onto complex sounds depends on a finely tuned communication between the motor and auditory systems. To probe how short-term sensory feedback and long-term knowledge each contribute to this cross-talk, we developed the *variable map playing* paradigm. This approach adapts the idea of auditory surprisal, well studied in passive listening contexts (8; 22), and applies it to self-generated sounds in which surprisal arises from violations of learned auditory-motor associations. Taken together, our findings show that while auditory predictions in the forward pathway is modulated by recent sensory feedback, motor predictions in the inverse pathway is shaped by slowly accumulated knowledge.

The neural signature of auditory-motor surprisals

Auditory-motor surprisal is reflected in the modulations of the N100 component of the auditory response

One central hypothesis of our study was that keystrokes following a change in a key-pitch map elicit a stronger neural (surprisal) response than other keystrokes. We hypothesized that this surprise response is auditory-motor in nature (*i.e.*, due to a violation of the expected association between the motor action and resulting sound), rather than purely auditory (based on learned statistical patterns of note transitions).

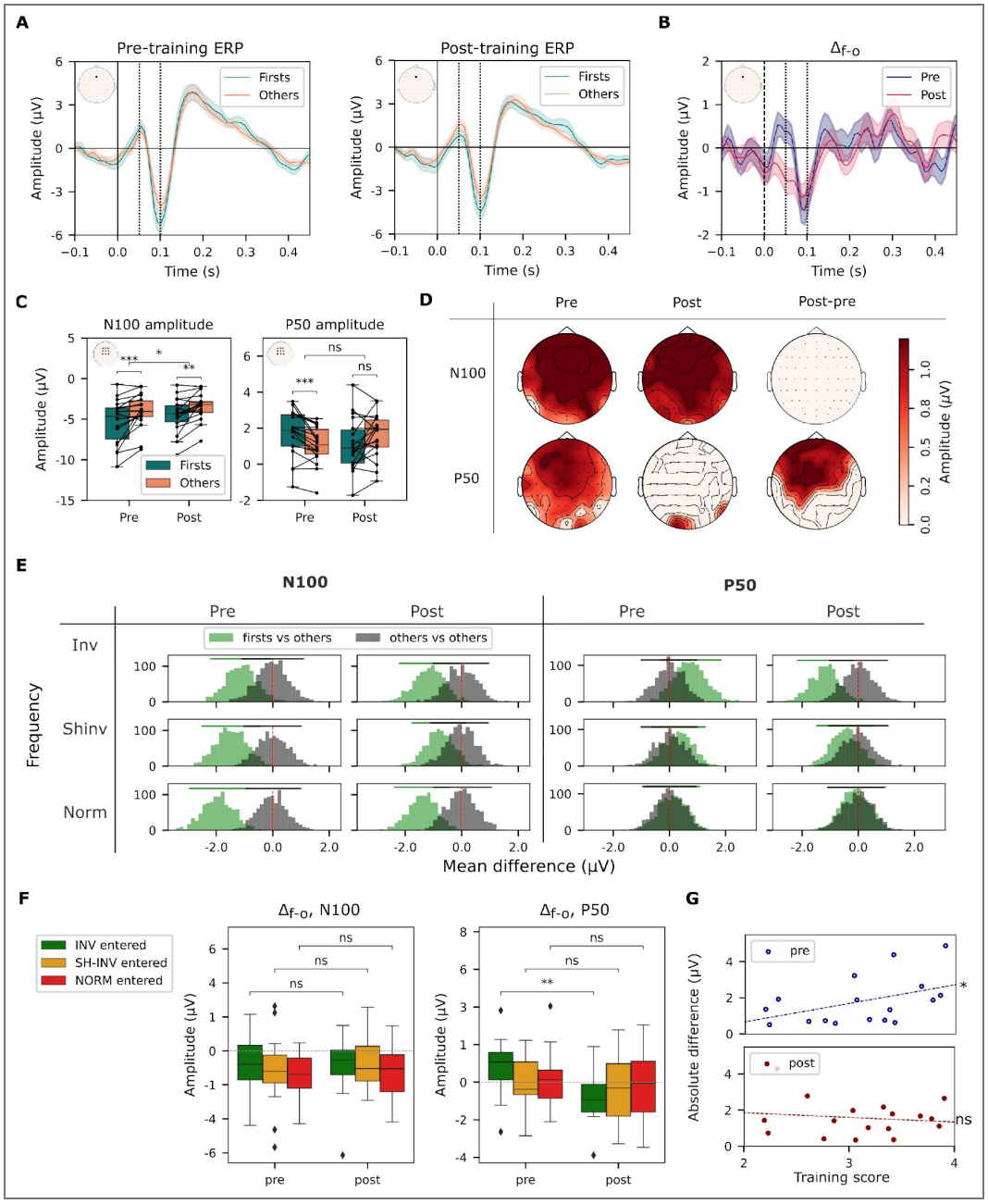


Figure 6. Effects of targeted and sustained training.

A. Grand-average ERPs of *firsts* and *others* at Fz before and after training. Mean with shading representing SEM; dashed lines mark time-points of interest (P50 and N100). **B**. Δ_{f-o} in pre- and post-training at Fz. **C**. Amplitudes of N100 (minimum over 80-120 ms), and P50 (maximum over 20-60 ms) averaged across centro-frontal channels. **D**. Topomap of Δ_{f-o} at N100 and P50 in pre and post training, as well as the difference between the two (Δf_o, post Δ_{f-o}, pre). Colours are masked to show channels with significant Δ_{f-o} (FDR-corrected Wilcoxon signed-rank test, only significant channels are displayed, significance threshold at $p = 0.05$). **E**. N100 and P50 Δ_{f-o} bootstrapped distributions as a function of training and map type against the null distribution. **F**. Subject mean N100 and P50 Δ_{f-o} as a function of training and map type ($p = 0.00281$, Wilcoxon signed-rank test). **G**. Correlation between training score and Δ_{f-o} at N100 in pre-training ($R = 0.477$, $p = 0.0451$, Pearson's correlation) and post-training (not significant). Points represent individual participants.

We found a neural signature of auditory-motor surprise in the form of a significantly larger N100 in the ERP of *first* notes after a map change compared to that of *others*. The N100 is known as a component of auditory responses to sound onsets (28; 29; 30; 31), which has been shown to be modulated during passive listening by contextual surprise within melodic sequences (referred to earlier as 'pure auditory' surprisals) (11; 8), as well as in speech (13). The modulation is explained by a model in which internal predictions of an incoming sound evoke a neural response with an opposite polarity to that of encoding the sound itself. When internal predictions match the incoming sensory input, the two components partially cancel out, resulting in an attenuated net neural response. When they do not match, the N100 amplitude is enhanced proportionally to the mismatch or surprisal (8). This process also occurs with self-generated sounds, such as speech (32) or self-triggered tones (33; 34; 35), where N100 attenuation arises from accurate motor predictions rather than prior auditory context. Here, we replicate this finding and extend it by demonstrating that when the key-pitch mapping remains stable across multiple key presses—rather than changing trial-by-trial—the brain progressively refines its auditory-motor model using information from each action (Figure3).

Effects at N100 cannot be explained by purely auditory surprisal

To distinguish between purely auditory and auditory-motor surprisal, we conducted a control experiment in which a separate group of participants listened to recordings of the original group's playing sessions. The hypothesis was that such sequences should not exhibit a structural bias between *firsts* and *others*, and therefore should not elicit any contextual surprisal modulations when only heard. If the played melodies had exhibited meaningful structure, this could have biased the results, with the first note after a map change being, on average, more surprising than the others, and hence resulting in a confounding factor. Consistent with our hypothesis, we did not find any significant difference in the N100 between *firsts* and *others* in the listening (control) group, whereas such a difference was present only in the playing group.

Interestingly, however, we found a difference between *firsts* vs *others* ERPs at P200 among the control group, a component already associated with auditory surprisal in prior work (8). This difference was absent for the variable map playing (Figure 3D and S2) likely because it was masked or modulated by the concurrent motor activity. Indeed, we have shown that the playing ERP could be approximately explained as a linear sum of the auditory and motor ERPs, obtained respectively from note onsets in the passive pure listening condition and the keystrokes in the mute playing condition (Figure S10), both exhibiting a clear P200 (Figure S4).

A computational analysis using IDyOM, a statistical model of musical surprisal (26; 36), converged on the same finding, yielding higher surprisals for *firsts* than for *others*. We chose to model surprisal using IDyOM because its surprisal predictions have been shown to closely align with behavioral and neural responses (8; 37; 38; 39; 40). Several factors can at least partially explain such contextual surprisals, including interval leaps (the largest leap possible within the same map is only 7 semitones, whereas it may reach 12 semitones at a map change; Figure 2A), and musically surprising transitions (participant played short melodies, and map changes may have rendered melodies less musically coherent). Together, results indicate a small contribution of purely auditory surprisal to the neural response, which is nevertheless insufficient to account for the stronger N100 modulation observed during active playing.

The surprisal effect cannot be explained by motor execution errors

In the MirrorNetwork framework (7), the auditory prediction generated by the motor system is projected into auditory areas through a *forward* pathway (decoder), where it is compared with the actual sensory feedback. Prediction violations elicit a surprisal signal that is projected back to motor areas through an *inverse* pathway (encoder), supporting both learning and adaptation. In this model, auditory predictions can be violated in two ways: either the motor plant misses the articulatory target (*i.e.*, a motor error), or a change in the environment alters the sound outcome (*i.e.*, an auditory prediction error). Our experiment was explicitly designed to violate auditory predictions in a sensory-motor task by inducing an erroneous *auditory feedback* while maintaining a simple motor task that minimizes motor errors (single, slow finger movements).

To determine whether the surprisal signature could be attributed to an auditory or a motor process, we trained linear decoders to reconstruct note onsets from responses to passive listening and mute playing, and evaluated their transfer performance on playing data. We found that the decoder trained on passive-listening EEG data (purely sensory) could discriminate between *first* and *other* keystrokes in the variable-map playing data both before and after training. By contrast, a decoder trained with mute playing data (purely motor) could not detect a difference before training (the post-training results shall be discussed below) (Figure 5). We interpret this as: Within the MirrorNetwork framework, these results suggest that the *forward* pathway is rapidly established during initial exploration, generating accurate auditory predictions from motor commands before motor skills are refined. Thus, early surprisal responses reflect prediction errors in the auditory domain rather than motor execution errors.

Sensorimotor interactions combine short-term sensory feedback with long-term knowledge

N100 reveals a continuous tracking of short-term sensory context for rapid adaptation

We hypothesized that short-term context primarily influences the early phase of sensorimotor learning, an *exploration* phase that is coupled to a rapid *adaptive* process, and is detectable through surprisal-related neural responses. During early exploration, predictions improve over successive key presses as movements are associated with corresponding sounds, causing surprisals to decrease steadily from one keystroke to the next. Once the auditory-motor map is acquired, any subsequent change in the map induces a new high-surprisal response that rapidly decreases, and then plateaus as auditory-motor predictions align with the appropriate new key-pitch map. Our analyses revealed that *first* keystrokes after a map change elicited significantly larger N100 amplitudes than subsequent ones. Interestingly, the N100 amplitude of subsequent keystrokes remains relatively stable (Figure 4D). This suggests that after the surprisal to the *first* note, the brain rapidly establishes accurate predictions for all other keys—either by extrapolating from the initial violation or by activating one of multiple stored auditory-motor maps. This interpretation aligns with findings from auditory feedback perturbation studies in speech (41), and language-related auditory-motor learning, where bilingual participants learn language-specific motor adjustments and apply them as a function of the language being spoken (42). Realignment of the sensorimotor maps was also shown to be possible in other modalities, such as in auditory-visual maps (43; 44).

Finally, we found that *first* keystrokes elicited a stronger N100 response as the number of keystrokes played in a previous map increased (Figure 4B). We attribute this effect to the perceived stability of the short-term sensory context, *i.e.*, as more keystrokes occur without a map change, the more stable the context. Hence, a mapping violation is more surprising in a stable context than in one characterized by frequent changes.

Prior work has theorized that the brain makes continuous predictions at multiple hierarchical levels. At a lower level, it predicts upcoming sensory events (e.g., individual note outcomes) based on the current sensorimotor mapping. At a higher level, it predicts when and how the sensorimotor mapping itself will change (45). Both prediction levels rely on sensory information, but operate on different timescales: immediate sensory feedback updates predictions about the next event, while accumulated sensory information over time updates the probability distribution governing potential map changes (46). This hierarchical predictive mechanism allows the brain to anticipate and prepare for changing consequences of motor actions.

Taken together, our results suggest that the brain continually tracks short-term context to anticipate map changes and rapidly adapt to them.

P50 as a marker of extended training on a key-pitch map

Auditory-motor surprisal responses comprise distinct auditory and motor components, which we were able to disentangle based on their different dynamics across the 30-minute targeted training. Specifically, the N100 primarily indexes auditory surprisal, while the P50 reflects motor surprisal. The N100 exhibits rapid changes during the exploratory phase of learning, whereas the P50 only shows modulation following extended training on the targeted mapping.

To begin with, although all ERPs were globally attenuated after training—likely due to the adaptation from prolonged exposure to similar, repeated stimuli (47)—, the attenuation did not affect the difference between *first* and *others* in N100, which remained stable across all key-pitch maps regardless of training (Figure 6C). This suggests that the auditory surprisal component developed rapidly during the early *exploration* phase and persisted afterwards, regardless of training. On the other hand, there was a significant effect of training on the difference between *first* and *others* in P50 (Figure 6B-D). This effect was driven by keystrokes in the key-pitch map that the participants trained on (Figure 6E-F), suggesting that the motor component of the response is significantly affected by training, with the emergence of a motor surprisal component only after (and presumably due to) the extended training.

The evidence on auditory and motor surprisal components can be further elaborated in the light of the MirrorNetwork framework described in Figure 1 (7). On the one hand, the *forward* pathway predicts a sound from a movement (*i.e.*, generating auditory expectations and therefore the auditory component of auditory-motor surprisals). As we have seen, this pathway is learned rapidly during the *exploration* phase of learning and adjusted when necessary. On the other hand, the *inverse* projection retrieves motor commands from a given sound (*i.e.*, generating motor expectations and therefore the motor component of auditory-motor surprisals). The inverse pathway is therefore learned slowly and through practice over many trial-and-error iterations. Both our hypothesis-driven analysis based on ERPs and hypothesis-free analysis using data-driven decoding provide converging evidence supporting such a model. Indeed, the auditory component of the auditory-motor surprisal response is present both before and after training, without any changes between the two phases (*i.e.*, Δ_{f-o} in N100 and in the auditory decoder predictions). The motor component, instead, emerged only *after* training and only for the trained map (*i.e.*, Δ_{f-o} in P50 and in the motor decoder predictions). This suggests that the *inverse* pathway inferring motor commands from sounds is modulated by skillful training at a timescale that is much longer than that of the *forward* pathway inferring sounds from movements.

The effects of long-term musical abilities

We expected better musical abilities to enhance the participants' ability to build reliable auditory-motor associations and, therefore, their sensitivity to map violations. Musical training has been shown to improve abilities such as motor control (48; 49), auditory perception (50; 51), and audio-motor integration (52; 53; 54; 55; 56; 57; 58). These abilities possibly contribute to musicians' ability to adapt faster to various environmental changes, such as rhythmic perturbations (59; 60; 61). Yet, the influence of long-term training on rapid adaptation in short-term, changing environments with perturbed auditory-motor associations remains unclear.

Specifically, we used melody imitation accuracy during training as a proxy for musical ability, as it is likely to be performed better by trained or musically gifted individuals (62). We found that the difference in N100 amplitude between *firsts* and *others* correlated with training score, but only before training. This suggests that musicians already possessed stronger internal predictions about the keyboard, leading to greater surprise when a key-pitch map change violated those predictions. After training, musicians may have quickly adapted their internal model to the new inverted keyboard, reducing the difference in prediction violation response relative to non-musicians.

Conclusions, limitations and future steps

Our findings reveal that auditory-motor learning operates on dual timescales: the brain rapidly and implicitly updates its predictions of sound from action within seconds, while the inverse process of refining motor commands from auditory feedback requires sustained, goal-directed

practice over extended periods. This asymmetry between forward and inverse model learning is an important step in understanding skill acquisition in music, speech, and other complex sensorimotor behaviors. However, while our simple paradigm allowed us to control for potential confounds, we acknowledge that ecological auditory-motor tasks are far more complex. For example, in piano playing, the ten fingers of the pianist are used to play on 88 keys, and there is no fixed finger-to-key assignment as was the case in our study. Future investigations should examine auditory-motor learning in more ecologically valid settings and use methods such as temporal response functions (27) that are better-suited to studying time-varying stimuli and multidimensional models. Additionally, future work could also investigate the *inverse* pathway in the motor domain, by looking into surprisal responses related to motor parameters such as finger, arm, and elbow positions. Such parameterization would allow the creation of a motion model linking sound and movement. Our work represents a step in this direction by applying temporal analysis methods to auditory-motor studies and exploiting neural signatures of auditory-motor surprisal and their modulation through training and sensory feedback.

Materials and method

Participants

Twenty-one participants (12 female, ages 19-29, mean age 24.1, 3 left-handed) were recruited from the École Normale Supérieure community. Three participants were excluded due to technical anomalies during recording. Participants had various levels of musical training, ranging from no training to over 10 years of formal study. None of the musically trained participants played piano as their primary instrument. Written informed consent was obtained from all participants, and each participant was compensated for his/her participation. The study was conducted in accordance with the Declaration of Helsinki and was approved by the CNRIPH committee. Participants were briefed on the instructions by the experimenter and had access to a printed copy of the written instructions throughout the experiment.

Experimental setup

Participants completed the experiment in a single session within a soundproof, electrically shielded booth with dim lighting. For all phases of the experiment, participants were instructed to fixate on a cross at the center of their visual field and minimize all motor activity except for the finger movements required to play the keyboard. A Novation Launchkey Mini MIDI keyboard was placed on the desk in front of the participant. Audio stimuli were presented monophonically at a sampling rate of 44.1 kHz using Sennheiser HD650 headphones at a volume comfortable for the participant. The experiment lasted approximately 3 hours, including 1 hour of setup, 1.5 hours of EEG recording, and short breaks between tasks at the participant's discretion.

Overview of tasks

The experiment consisted of three main blocks, each lasting approximately 30 minutes: pre-test, training, and post-test. The pre- and post-test blocks were identical and consisted of 3 tasks, each lasting 10 minutes: passive listening, mute playing, and variable map playing (Figure 2C). Each of these tasks is detailed below.

Variable map playing task

The participant's right hand was positioned with the experimenter's assistance on the 'playing zone' of the keyboard, comprising the keys F4, G4, A4, B4, and C5. One finger was positioned on each key. Textured velcro strips placed on the boundaries of the playing zone allowed participants to keep their fingers in the correct position throughout the experiment. Participants were asked to play 4-note sequences at approximately 60 beats per minute (bpm) separated by 2 seconds of pause, and to vary the sequences played.

The task lasted 10 minutes, during which the *key-pitch mapping* changed unpredictably every 2-10 seconds. Three mappings were possible as illustrated in Figure 2A: the *inverted* mapping (the playing zone keys, F4, G4, A4, B4, and C5, were mapped to G4, F4, E4, D4, and C4, respectively), *shifted-inverted* mapping in which pitches from the *inverted* mapping were shifted up one whole tone (to A4, G4, F4, E4, and D4, respectively), and *normal* mapping reflecting that of a normal piano (F4, G4, A4, B4, and C5). The complete assignment of mappings over the 10-minute playing session is detailed in Supplementary Figure S1. Sound synthesis and switching between mappings were automated using Ableton Live 11, a low-latency digital audio workstation for live music.

Participants performed the tasks with no visual input apart from the fixation cross to minimize eye movements. The beginning and end of each task were marked by a bell sound. The experimenter monitored the sequences played by the participant in real time from outside the experimental booth to ensure that participants followed the task instructions. The audio output was recorded and used in the control experiment described below.

Training

With the right hand placed on the keyboard in the same way as in the variable map playing task, participants completed a self-guided training session (Figure 2) designed to familiarize them with one of the three key-pitch maps used in the variable map playing task. Throughout training, *key-pitch mapping* consistently followed the *inverted* mapping.

The training session consisted of 2 identical blocks, where the same 8 sets of 10-12 melodies were played (Figure 2D). The sets of melodies were arranged in ascending difficulty based on the number of pitches used in each melody (e.g., set 1 used only C4 and D4, while set 8 used all 5 keys). Each melody consisted of four notes, played at 60 bpm. Participants were asked to imitate the melodies played to the best of their ability. There was no penalty for rhythmic errors, but participants were asked to stay within the allotted response time of 5 seconds. There was an opportunity to take a break after each set. Played melodies were recorded using the MIDI protocol during both the pre- and post-phases, as well as during training. Training lasted approximately 30 minutes and included a total of 164 melodies to imitate.

Performance on the melody imitation task during training was scored by comparing the melody played with the target one, note by note, without considering timing. If participants answered with more than four key presses, only the first four were considered. If participants answered with fewer key presses, the response was padded with non-valid events. For each note, a score of 1 was awarded if the note was correct, and zero otherwise, meaning that the maximum score was 4 (Figure 2E). We observed a significant improvement in score for all participants (Figure 2F).

Auxiliary tasks

Participants performed additional passive-listening and mute-playing tasks to generate subject-specific EEG data for training auditory and motor decoder models, as well as unimodal ERPs. The order of tasks was chosen to ensure that the listening and mute playing tasks during pre-training are not affected by participant expectations resulting from playing the keyboard (Figure 2C). Participants had the opportunity to take a break after each task.

In the passive listening task, participants listened to the corpus of melodies used in the training session. Each melody was presented once, with a 2-second pause between melodies, for a total of 11 minutes and 5 seconds of listening.

In the mute playing task, participants were asked to play 4-note sequences on a muted keyboard and to vary the sequence each time. The keyboard output (both audio and MIDI) was recorded in Ableton Live to ensure that participants played a variety of different melodies. The task lasted 10 minutes and was identical to the variable map playing task except that the participants heard no sound.

Control experiment

To control for responses potentially related to auditory surprisal, six additional participants who did not participate in the original experiment were recruited to listen to the recordings of note sequences played during the full 10-minute variable map playing task. Eight recordings were used in the experiment, and each participant listened to a semi-random combination of 4 recordings such that each recording was heard by three participants. The sample size matched the number of note events analyzed in the original experiment while accounting for inter-participant variability in recordings and individual differences in neural responses. We did not invite the original participants to listen passively to their own playing to exclude the possibility that they recalled the played melodies.

EEG acquisition and preprocessing

During the experiment, 64-channel EEG data, along with two external electrodes, were recorded at a sampling rate of 2048 Hz. The BioSemi ActiveTwo system and the accompanying software were used for EEG acquisition. In addition, 4 external electrodes (2 mastoid, 2 ocular) were used for re-referencing and eye-blink artifact removal, respectively. To ensure synchronization between EEG data and stimuli, all key presses on the MIDI keyboard, audio onsets, and trial start signals were routed to a customized analog trigger system, which was recorded as additional EEG channels.

EEG data were preprocessed and analyzed offline using MNE-Python (63) and following standard guidelines for linear modeling of neurophysiological data to auditory stimuli (64). EEG data were notch-filtered at 50 Hz and bandpass-filtered between 1 and 30 Hz using Butterworth zero-phase filters (order 3, forward and backward pass). All channels were re-referenced to the average of the two mastoid channels. Channels with a variance exceeding three times that of the surrounding ones were replaced by an estimate calculated using spherical spline interpolation (64). Finally, the data was downsampled to 128 Hz to reduce the computational load.

ERP analysis

Data were segmented using time-aligned note onsets (for motor-only tasks, the participant's headphones were muted, but the corresponding audio information was still transmitted to the EEG acquisition computer). Segments from 200 ms before to 500 ms after the note onset were isolated and averaged for analysis. Independent component analysis was performed separately for each task and participant, rejecting components closely correlated with EOG signals. ERPs were classified as *firsts* or *others* based on their relative positions to map changes (Figure 2).

To identify time points of interest for further analysis, an exploratory Wilcoxon signed-rank test without correction for multiple comparisons was performed over all time points between 0 and 500 ms at Fz, Cz, Pz, and Iz channels (Figure S2). Based on the exploratory analyses, regions around 50 ms (corresponding to the P50 peak), 100 ms (corresponding to the N100 peak), and 370 ms were selected for further analysis. Significance of differences across subjects was assessed using a one-sample Wilcoxon signed-rank test. To examine the topographic distribution of significant differences in N100 and P50, we performed Wilcoxon signed-rank tests at N100 and P50 in all channels and identified a frontal-central region of interest. The selected areas and timepoints align with regions of interest detected by the permutation cluster test over timepoints and channels using automatically selected thresholds (Figure 3). Unless otherwise indicated, all analyses of ERP amplitudes and differences between *first* and *other* keystrokes represent the average of the following centro-frontal channels: Fz, FCz, Cz, F1, FC1, C1, C2, FC2, F2.

Modeling of surprisal using IDyOM

To estimate the musical surprisal of individual notes, we applied the Information Dynamics of Music (IDyOM) model (26) using a Python implementation (36). The dataset consists of eight recordings, where each recording represents a participant's variable map playing session in the main experiment. Surprisal values were computed for each note in each of the eight recordings using leave-one-out cross-validation: for each recording, the model was trained on the other seven

recordings and then used to predict the note probabilities in the held-out recording. Both short-term (predicting probabilities based on up to 20 previous notes in the same recording) and long-term (trained on transition probabilities in all other recordings) models were used to calculate surprisal. Only pitch information was modeled; a uniform rhythmic structure was assumed, as participants did not vary the rhythm of the melodies.

Statistical inference

Comparing *firsts* versus *others* keystrokes

To address the class imbalance, *i.e.*, the fact that there are fewer *firsts* than *others* keystrokes, we employed a bootstrapping procedure at each time point of interest (50, 100, and 370 ms). For each of the $N = 1000$ bootstrap iterations, we drew $n = 100$ random samples with replacement from the set of *firsts*, and $n = 100$ from the set of *others* for each subject. We then computed the mean ERP amplitude for each class at each time point and calculated the within-subject difference between the two. The average difference across subjects was retained at each iteration, yielding a bootstrap distribution of mean differences across subjects. To generate a null distribution, the same procedure was applied using two random samples of $n = 100$ trials each drawn from the *others* keystroke class, thus controlling for differences unrelated to keystroke category. Statistical inference was based on comparing the empirical and null bootstrap distributions: we evaluated whether their 95% confidence intervals (CI95) overlapped, and whether either distribution's CI included zero. In addition to the average across centro-frontal electrodes, we applied the same bootstrapping procedure to Fz, Cz, and Iz electrodes to confirm that the observed distributions of mean differences were consistent with the spatial patterns seen in the ERP difference topographies.

Keystrokes contextual analysis

To assess whether auditory-motor surprise is modulated by the previous context, we analyzed N100 amplitudes of *first* keystrokes as a function of the number of keystrokes performed in the map preceding the map change. We defined a *0-keystroke* context as the case where the note immediately preceding the current first keystroke was itself a first keystroke. A *1-keystroke* context corresponds to a case in which one *other* keystroke (i.e., a keystroke *not* following a map change) precedes the first keystroke and a first keystroke occurs two keystrokes prior, and so on. Trials were pooled across participants, and only classes of first keystrokes with more than 300 samples were retained to ensure a sufficient signal-to-noise ratio. Due to frequent map changes, longer context sequences were increasingly rare and thus excluded. This threshold on the number of samples allowed us to analyze up to a context of 3 keystrokes.

To determine whether first keystrokes in each context category elicited larger N100 amplitudes than *other* keystrokes, we performed one-sample *t*-tests comparing the mean N100 amplitude of first keystrokes (across subjects) to that of *other* keystrokes. To assess whether the N100 response to first keystrokes varied systematically as a function of the preceding context, we conducted independent-samples *t*-tests comparing the amplitude of first keystrokes in the 0-keystroke context with those in the *1-*, *2-*, and *3-keystroke* contexts. Because this analysis was exploratory and targeted a small, progressive effect across non-independent contrasts, we did not apply multiple-comparison corrections to not obscure potentially informative trends. Instead, we report exact p-values and interpret the pattern across contexts rather than across single comparisons. Findings are thus hypothesis-generating and will be validated in future work.

Decoding analysis

Subject-specific decoders were trained separately using EEG responses to the passive listening and mute playing tasks. The *auditory decoder* was trained on EEG responses to passive listening to reconstruct note onsets, while the *motor decoder* was trained on EEG responses to mute playing to reconstruct keystrokes. Both decoders were then used to predict note onsets and keystrokes from EEG responses to the variable map playing task (Figure 5). Separate auditory and motor models were trained and tested for each participant using data exclusively from that individual, using a

cross-validation procedure. Reconstructed note onsets were then sorted according to whether they corresponded to a *first* or *other* keystroke, and averaged separately to produce a mean predicted amplitude for *firsts*, and a mean predicted amplitude for *others*.

Data and software

Statistical analyses were performed in Python using `scipy` (65), `pingouin` (66), and `statsmodels` (67). Significance stars in figures were automatically drawn using `statannotations` (68). Decoders were trained using `mTRFpy` (69).

Supplementary figures

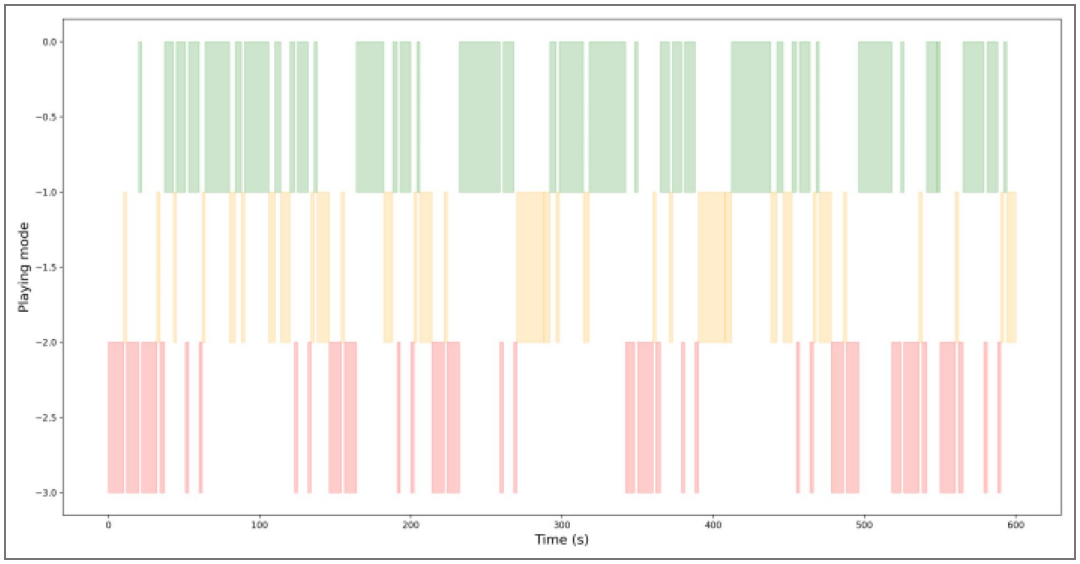


Figure S1. **Key-pitch map assignment over time during the full 10-minute task.**

Figure S2. Time points with significant differences between the amplitude of the first and others at Fz, Cz, Pz, and Iz channels.

* $p < 0.05$, Wilcoxon signed-rank test, without FDR correction.

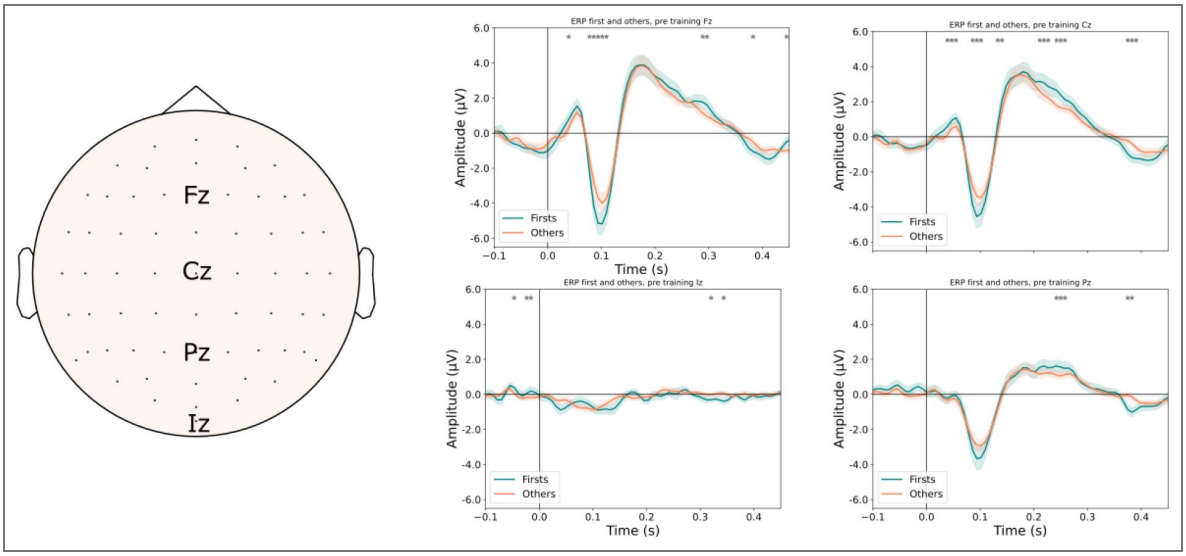


Figure S3. First keystrokes of the map versus the keystrokes immediately after the map change when entering each of the key-pitch maps, at timepoints of interest identified in Figure S2 : 50, and 370 ms.

For analysis at 100 ms, see Fig. 3.

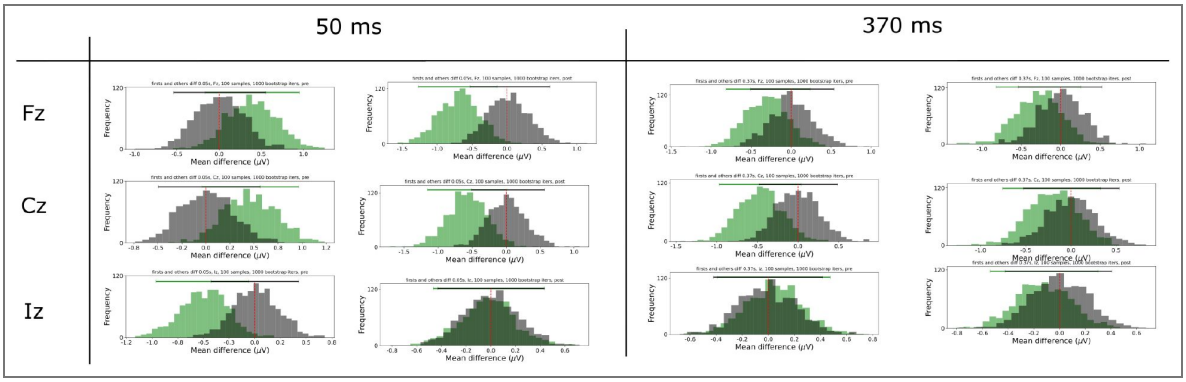


Figure S4. Grand average note onset ERPs during the passive listening, mute playing, and variable map playing tasks in all channels, in pre-training only.

Colours represent EEG channels.

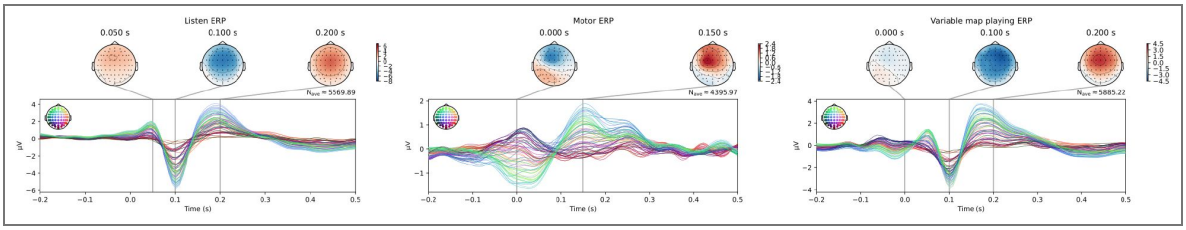


Figure S5. ERPs aligned to the note onsets during the passive listening, mute playing, and variable map playing tasks in the FCz channel, showing differences pre and post-training.

Lines represent the grand average over all subjects, shading represents SEM.

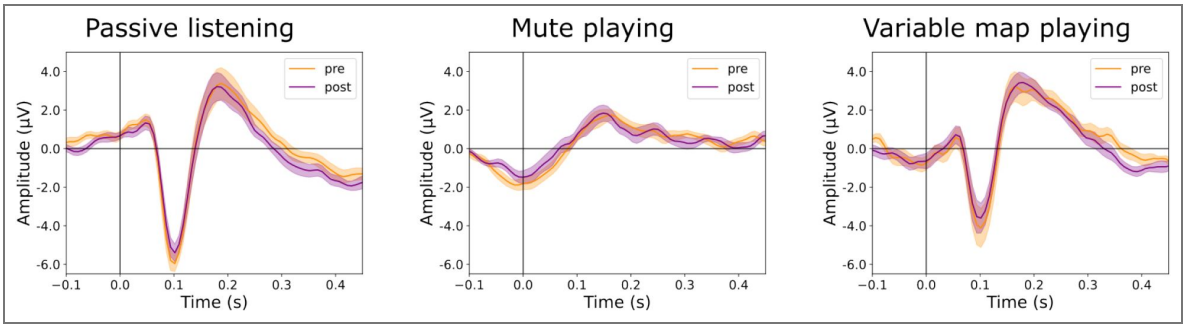


Figure S6. Note onset ERPs sorted by map, including all *firsts* and others, separated by pre and post-training.

Lines represent the grand average over all subjects, shading represents SEM.

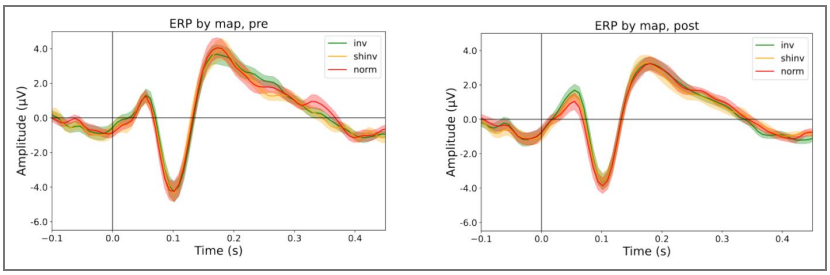


Figure S7.

A. Example of keystrokes included when analyzing entering of the maps, showing the keystrokes included when analyzing entering the INV map: INV/first and INV/other keystrokes. **B**. Difference waves (first - other ERPs) sorted by the map entered, pre- and post-training.

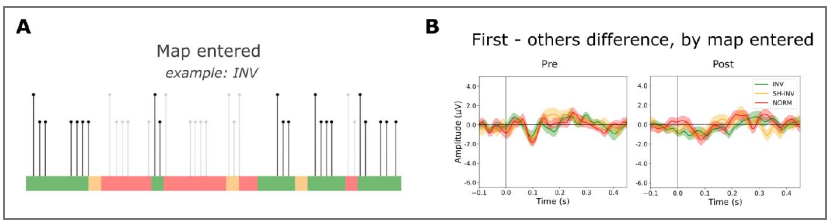


Figure S8. Correlation between training score and Δ_{f-o} at P50.

Columns show the active key-pitch map at the time of the keystroke. Rows show correlations in pre-training, post-training, and the difference between the two (post-pre). Dotted lines show linear regression. $p > 0.05$ for all panels.

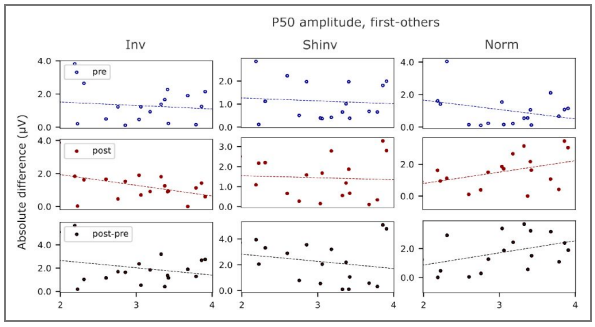
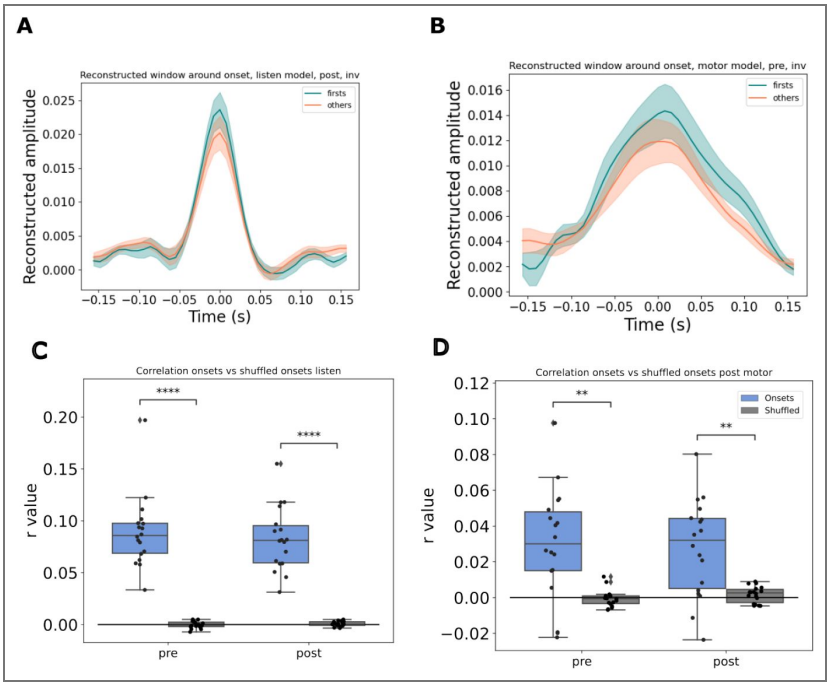


Figure S9. Reconstruction accuracy of inverse TRF decoder.

A-B. Average window around note onset times in reconstructed stimulus using listening and motor decoders, respectively. **C-D**. Correlation between ground truth (sparse vector with time of note onsets set to 1) and reconstructed stimulus when the ground truth is shuffled versus unshuffled, for listening and motor decoders, respectively. ** $p < 0.01$, **** $p < 0.0001$, Pearson's correlation.



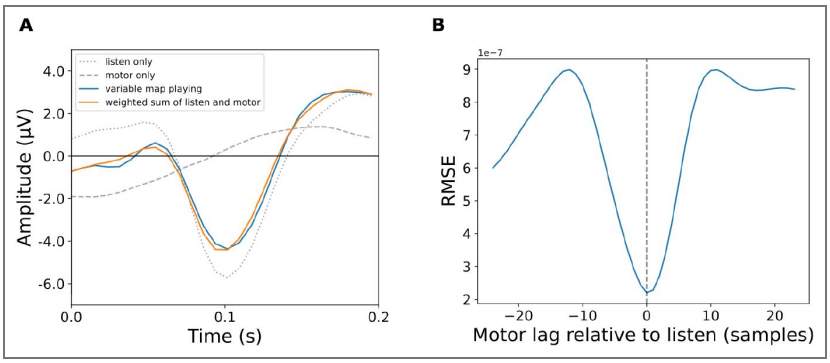


Figure S10. Comparison of weighted sum of auditory and motor ERPs with playing ERP.

A. Grand average ERPs for the auditory-only condition (locked to note onset), motor-only condition (locked to keystrokes), the variable map playing ERP, and the optimized weighted sum of auditory and motor ERPs. Weights were determined by least squares optimization. **B**. Root mean square error (RMSE) between the playing ERP and the optimized weighted sum of auditory and motor ERPs across different lags applied to the motor ERP relative to the auditory ERP. The lowest RMSE occurs at zero lag.

Data availability

The code for analyzing data and generating the figures is available on Github at *https://github.com/haiqin-zhang/AuditoryMotorPiano*. EEG data is available on Zenodo (DOI: 10.5281/zenodo.17949131).

Acknowledgements

This work was supported by an Advanced European Research Council grant (NEUME, 787836) and FrontCog grant ANR-17-EURE-0017 to SS. HZ is supported by a *Contrat Doctoral Spécifique Normalien* (CDSN) doctoral grant funded by the French Ministry of National Education, Higher Education, Research and Innovation, and administered by École Normale Supérieure – PSL. The project was based on pilot experiments and discussions at the 2023 Telluride Neuromorphic Engineering Workshop. The authors are grateful to Yves Boubenec, Giovanni Di Liberto, Claire Pelofi, and Virginie van Wassenhove for their feedback during conception and writing.

Additional files

Supplementary video Excerpt of a variable map playing session. Audio is played by the participant. Each row represents one key-pitch map (green: *inverted*, yellow: *shifted-inverted*, red: *normal*). Bars in that row represent the beginning of that map (and the end of the previous one).

Additional information

Funding

Funder	Grant reference number	Author
DOD | USN | Office of Naval Research (ONR)	MURI	Shihab Shamma
EC | European Research Council (ERC)	NEUME	Shihab Shamma
DOD | AF | AMC | AFRL | Air Force Office of Scientific Research (AFOSR)	FA9550-19-1-0408	Shihab Shamma

Author ORCID iDs

Haiqin Zhang: https://orcid.org/0009-0002-3967-8962
Giorgia Cantisani: https://orcid.org/0000-0003-3574-925X
Shihab Shamma: https://orcid.org/0000-0002-1283-6804

References

[1] **Conant Roger C.,** Ashby William Ross (1970) Every good regulator of a system must be a model of that system. *International Journal of Systems Science* **1**:89-97 https://doi.org/10.1080/00207727008920220

[2] **Wolpert D. M.,** Ghahramani Z., Jordan M. I. (1995) An internal model for sensorimotor integration. *Science* **269**:1880-1882 https://doi.org/10.1126/science.7569931 | PubMed

[3] **Körding Konrad P.,** Wolpert Daniel M. (2004) Bayesian integration in sensorimotor learning. *Nature* **427**:244-247 https://doi.org/10.1038/nature02169 | PubMed

[4] **Hickok Gregory,** Buchsbaum Bradley, Humphries Colin, Muftuler Tugan (2003) Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience* **15**:673-682 https://doi.org/10.1162/089892903322307393 | PubMed

[5] **Hickok Gregory** (2012) Computational neuroanatomy of speech production. *Nature Reviews. Neuroscience* **13**:135-145 https://doi.org/10.1038/nrn3158 | PubMed

[6] **Pa Judy,** Hickok Gregory (2008) A Parietal-Temporal Sensory-Motor Integration Area for the Human Vocal Tract: Evidence from an fMRI Study of Skilled Musicians. *Neuropsychologia* **46**:362-368 https://doi.org/10.1016/j.neuropsychologia.2007.06.024 | PubMed

[7] **Shamma Shihab,** Patel Prachi, Mukherjee Shoutik, Marion Guilhem, Khalighinejad Bahar, Han Cong, Herrero Jose, Bickel Stephan, Mehta Ashesh, Mesgarani Nima (2020) Learning Speech Production and Perception through Sensorimotor Interactions. *Cerebral Cortex Communications* **2** https://doi.org/10.1093/texcom/tgaa091 | PubMed

[8] **Di Liberto Giovanni M,** Pelofi Claire, Bianco Roberta, Patel Prachi, Mehta Ashesh D, Herrero Jose L, Alain de Cheveigné, Shamma Shihab, Mesgarani Nima (2020) Cortical encoding of melodic expectations in human temporal cortex. *eLife* **9** https://doi.org/10.7554/elife.51784 | PubMed

[9] **Koelsch Stefan,** Vuust Peter, Friston Karl (2019) Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences* **23**:63-77 https://doi.org/10.1016/j.tics.2018.10.006 | PubMed

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

[11] **Di Liberto Giovanni M.,** Marion Guilhem, Shamma Shihab A. (2021) Accurate Decoding of Imagined and Heard Melodies. *Frontiers in Neuroscience* **15**:673401 https://doi.org/10.3389/fnins.2021.673401 | PubMed

[12] **Weissbart Hugo,** Kandylaki Katerina D., Reichenbach Tobias (2020) Cortical Tracking of Surprisal during Continuous Speech Comprehension. *Journal of Cognitive Neuroscience* **32**:155-166 https://doi.org/10.1162/jocn_a_01467 | PubMed

[13] **Gillis Marlies,** Vanthornhout Jonas, Simon Jonathan Z., Francart Tom, Brodbeck Christian (2021) Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics. *The Journal of Neuroscience* **41**:10316-10329 https://doi.org/10.1523/jneurosci.0812-21.2021 | PubMed

[14] **Omigie Diana,** Pearce Marcus, Lehongre Katia, Hasboun Dominique, Navarro Vincent, Adam Claude, Samson Severine (2019) Intracranial Recordings and Computational Modeling of Music Reveal the Time Course of Prediction Error Signaling in Frontal and Temporal Cortices. *Journal of Cognitive Neuroscience* **31**:855-873 https://doi.org/10.1162/jocn_a_01388 | PubMed

[15] **Sankaran Narayan,** Leonard Matthew K., Theunissen Frederic, Chang Edward F. (2024) Encoding of melody in the human auditory cortex. *Science Advances* **10** https://doi.org/10.1126/sciadv.adk0010 | PubMed

[16] **Galeano-Otálvaro Juan-Daniel,** Martorell Jordi, Meyer Lars, Titone Lorenzo (2024) Neural encoding of melodic expectations in music across EEG frequency bands. *European Journal of Neuroscience* **60**:6734-6749 https://doi.org/10.1111/ejn.16581 | PubMed

[17] **Skerritt-Davis Benjamin,** Elhilali Mounya (2021) Computational framework for investigating predictive processing in auditory perception. *Journal of Neuroscience Methods* **360** https://doi.org/10.1016/j.jneumeth.2021.109177 | PubMed

[18] **Kern Pius,** Heilbron Micha, de Lange Floris P., Spaak Eelke (2022) Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. *eLife* **11** https://doi.org/10.7554/elife.80935 | PubMed

[19] **Wolpert D. M.,** Miall R. C. (1996) Forward Models for Physiological Motor Control. *Neural Networks: The Official Journal of the International Neural Network Society* **9**:1265-1279 https://doi.org/10.1016/s0893-6080(96)00035-4 | PubMed

[20] **Saupe Katja,** Widmann Andreas, Trujillo-Barreto Nelson J., Schröger Erich (2013) Sensorial suppression of self-generated sounds and its dependence on attention. *International Journal of Psychophysiology* **90**:300-310 https://doi.org/10.1016/j.ijpsycho.2013.09.006 | PubMed

[21] **Katahira Kentaro,** Abla Dilshat, Masuda Sayaka, Okanoya Kazuo (2008) Feedback-based error monitoring processes during musical performance: An ERP study. *Neuroscience Research* **61**:120-128 https://doi.org/10.1016/j.neures.2008.02.001 | PubMed

[22] **Lutz Kai,** Puorger Roman, Cheetham Marcus, Jancke Lutz (2013) Development of ERN together with an internal model of audio-motor associations. *Frontiers in Human Neuroscience* **7** https://doi.org/10.3389/fnhum.2013.00471 | PubMed

[23] **Tast Valentina,** Schröger Erich, Widmann Andreas (2025) Auditory N1 Suppression and Omission N1 Do Not Share a Common Underlying Mechanism. *Psychophysiology* **62** https://doi.org/10.1111/psyp.70094 | PubMed

[24] **Gay T.,** Lindblom B., Lubker J. (1981) Production of bite-block vowels: acoustic equivalence by selective compensation. *The Journal of the Acoustical Society of America* **69**:802-810 https://doi.org/10.1121/1.385591 | PubMed

[25] **McFarland D. H.,** Baum S. R., Chabot C. (1996) Speech compensation to structural modifications of the oral cavity. *The Journal of the Acoustical Society of America* **100**:1093-1104 https://doi.org/10.1121/1.416286 | PubMed

[26] **Pearce Marcus T.** (2018) Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences* **1423**:378-395 https://doi.org/10.1111/nyas.13654 | PubMed

[27] **Crosse Michael J.,** Di Liberto Giovanni M., Bednar Adam, Lalor Edmund C. (2016) The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience* **10** https://doi.org/10.3389/fnhum.2016.00604 | PubMed

[28] **Hillyard Steven A.,** Hink Robert F., Schwent Vincent L., Picton Terence W. (1973) Electrical Signs of Selective Attention in the Human Brain. *Science* **182**:177-180 https://doi.org/10.1126/science.182.4108.177 | PubMed

[29] **Näätänen Risto,** Picton Terence (1987) The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology* **24**:375-425 https://doi.org/10.1111/j.1469-8986.1987.tb00311.x | PubMed

[30] **Wolpaw J. R.,** Penry J. K. (1975) A temporal component of the auditory evoked response. *Electroencephalography and Clinical Neurophysiology* **39**:609-620 https://doi.org/10.1016/0013-4694(75)90073-5 | PubMed

[31] **Zouridakis George,** Simos Panagiotis G., Papanicolaou Andrew C. (1998) Multiple Bilaterally Asymmetric Cortical Sources Account for the Auditory N1m Component. *Brain Topography* **10**:183-189 https://doi.org/10.1023/a:1022246825461 | PubMed

[32] **Curio Gabriel,** Neuloh Georg, Numminen Jussi, Jousmäki Veikko, Hari Riitta (2000) Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping* **9**:183-191 https://doi.org/10.1002/(sici)1097-0193(200004)9:4<183::aid-hbm1>3.0.co;2-z | PubMed

[33] **Baess Pamela,** Horváth János, Jacobsen Thomas, Schröger Erich (2011) Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology* **48**:1276-1283 https://doi.org/10.1111/j.1469-8986.2011.01196.x | PubMed

[34] **Timm Jana,** SanMiguel Iria, Keil Julian, Schröger Erich, Schönwiesner Marc (2014) Motor Intention Determines Sensory Attenuation of Brain Responses to Self-initiated Sounds. *Journal of Cognitive Neuroscience* **26**:1481-1489 https://doi.org/10.1162/jocn_a_00552 | PubMed

[35] **Martikainen Mika H.,** Kaneko Ken-ichi, Hari Riitta (2005) Suppressed Responses to Self-triggered Sounds in the Human Auditory Cortex. *Cerebral Cortex* **15**:299-302 https://doi.org/10.1093/cercor/bhh131 | PubMed

[36] **Marion Guilhem,** Gao Fei, Gold Benjamin P., Di Liberto Giovanni M., Shamma Shihab (2025) IDyOMpy: A new Python-based model for the statistical analysis of musical expectations. *Journal of Neuroscience Methods* **415** https://doi.org/10.1016/j.jneumeth.2024.110347 | PubMed

[37] **Pearce M. T.** (2015) The construction and evaluation of statistical models of melodic structure in music perception and composition. City University London.

[38] **Bianco Roberta,** Ptasczynski Lena Esther, Omigie Diana (2020) Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain and Cognition* **138**:103621 https://doi.org/10.1016/j.bandc.2019.103621 | PubMed

[39] **Hansen Niels Chr,** Pearce Marcus T. (2014) Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology* **5** https://doi.org/10.3389/fpsyg.2014.01052 | PubMed

[40] **Moldwin Toviah,** Schwartz Odelia, Sussman Elyse S. (2017) Statistical learning of melodic patterns influences the brain's response to wrong notes. *Journal of Cognitive Neuroscience* **29**:2114-2122 https://doi.org/10.1162/jocn_a_01181 | PubMed

[41] **Rao Nishant,** Ostry David J. (2025) Probing sensorimotor memory through the human speech-audiomotor system. *Journal of Neurophysiology* **133**:479-489 https://doi.org/10.1152/jn.00337.2024 | PubMed

[42] **Lametti Daniel R.,** Wheeler Emma D., Palatinus Samantha, Hocine Imane, Shiller Douglas M. (2025) Language enables the acquisition of distinct sensorimotor memories for speech. *Cognition* **254**:106010 https://doi.org/10.1016/j.cognition.2024.106010 | PubMed

[43] **Knudsen E. I.,** Brainard M. S. (1991) Visual instruction of the neural map of auditory space in the developing optic tectum. *Science* **253**:85-87 https://doi.org/10.1126/science.2063209 | PubMed

[44] **Knudsen Eric I.** (2002) Instructed learning in the auditory localization pathway of the barn owl. *Nature* **417**:322-328 https://doi.org/10.1038/417322a | PubMed

[45] **Vetter P.,** Wolpert D. M. (2000) Context estimation for sensorimotor control. *Journal of Neurophysiology* **84**:1026-1034 https://doi.org/10.1152/jn.2000.84.2.1026 | PubMed

[46] **Heald James B.,** Máté Lengyel, Wolpert Daniel M. (2023) Contextual inference in learning and memory. *Trends in Cognitive Sciences* **27**:43-64 https://doi.org/10.1016/j.tics.2022.10.004 | PubMed

[47] **Rosburg T,** Haueisen J, Sauer H (2002) Habituation of the auditory evoked field component N100m and its dependence on stimulus duration. *Clinical Neurophysiology* **113**:421-428 https://doi.org/10.1016/s1388-2457(01)00727-1 | PubMed

[48] **Worschech Florian,** James Clara E., Jünemann Kristin, Sinke Christopher, Krüger Tillmann H. C., Scholz Daniel S., Kliegel Matthias, Marie Damien, Altenmüller Eckart (2023) Fine motor control improves in older adults after 1year of piano lessons: Analysis of individual development and its coupling with cognition and brain structure. *The European Journal of Neuroscience* **57**:2040-2061 https://doi.org/10.1111/ejn.16031 | PubMed

[49] **Abolghasemi Saideh,** Abolghasemi Reyhaneh, Ardalani Hossein (2024) The music effect on motor skills of healthy people, a systematic review. *Journal of Bodywork and Movement Therapies* **40**:1166-1176 https://doi.org/10.1016/j.jbmt.2024.07.005 | PubMed

[50] **Xie Xin,** Myers Emily (2015) The impact of musical training and tone language experience on talker identification. *The Journal of the Acoustical Society of America* **137**:419-432 https://doi.org/10.1121/1.4904699 | PubMed

[51] **Choi William** (2021) Musicianship Influences Language Effect on Musical Pitch Perception. *Frontiers in Psychology* **12** https://doi.org/10.3389/fpsyg.2021.712753 | PubMed

[52] **Li Qiongling,** Wang Xuetong, Wang Shaoyi, Xie Yongqi, Li Xinwei, Xie Yachao, Li Shuyu (2018) Musical training induces functional and structural auditory-motor network plasticity in young adults. *Human Brain Mapping* **39**:2098-2110 https://doi.org/10.1002/hbm.23989 | PubMed

[53] **Lahav Amir,** Saltzman Elliot, Schlaug Gottfried (2007) Action Representation of Sound: Audiomotor Recognition Network While Listening to Newly Acquired Actions. *The Journal of Neuroscience* **27**:308-314 https://doi.org/10.1523/jneurosci.4822-06.2007 | PubMed

[54] **Baumann Simon,** Koeneke Susan, Schmidt Conny F., Meyer Martin, Lutz Kai, Jancke Lutz (2007) A network for audio–motor coordination in skilled pianists and non-musicians. *Brain Research* **1161**:65-78 https://doi.org/10.1016/j.brainres.2007.05.045 | PubMed

[55] **Palomar-García María-Ángeles,** Zatorre Robert J., Ventura-Campos Noelia, Bueichekú Elisenda, Ávila César (2017) Modulation of Functional Connectivity in Auditory-Motor Networks in Musicians Compared with Nonmusicians. *Cerebral Cortex* **27**:2768-2778 https://doi.org/10.1093/cercor/bhw120 | PubMed

[56] **Jünemann Kristin,** Engels Anna, Marie Damien, Worschech Florian, Scholz Daniel S., Grouiller Frédéric, Kliegel Matthias, Van De Ville Dimitri, Krüger H. C., James Clara E., *et al.* (2023) Increased functional connectivity in the right dorsal auditory stream after a full year of piano training in healthy older adults. *Scientific Reports* **13**:19993 https://doi.org/10.1038/s41598-023-46513-1 | PubMed

[57] **Hosoda Moe,** Furuya Shinichi (2016) Shared somatosensory and motor functions in musicians. *Scientific Reports* **6**:37632 https://doi.org/10.1038/srep37632 | PubMed

[58] **Hyde Krista L.,** Lerch Jason, Norton Andrea, Forgeard Marie, Winner Ellen, Evans Alan C., Schlaug Gottfried (2009) The Effects of Musical Training on Structural Brain Development. *Annals of the New York Academy of Sciences* **1169**:182-186 https://doi.org/10.1111/j.1749-6632.2009.04852.x | PubMed

[59] **van der Steen M. C.,** Molendijk E. B.., Altenmüller E., Furuya S. (2014) Expert pianists do not listen: The expertise-dependent influence of temporal perturbation on the production of sequential movements. *Neuroscience* **269**:290-298 https://doi.org/10.1016/j.neuroscience.2014.03.058 | PubMed

[60] **Scheurich Rebecca,** Pfordresher Peter Q., Palmer Caroline (2020) Musical training enhances temporal adaptation of auditory-motor synchronization. *Experimental Brain Research* **238**:81-92 https://doi.org/10.1007/s00221-019-05692-y | PubMed

[61] **Yasuhara Masaki,** Uehara Kazumasa, Oku Takanori, Shiotani Sachiko, Nambu Isao, Furuya Shinichi (2024) Robustness and adaptability of sensorimotor skills in expert piano performance. *iScience* **27**:110400 https://doi.org/10.1016/j.isci.2024.110400 | PubMed

[62] **Mantell James T.,** Peter Q. (2013) Pfordresher. Vocal imitation of song and speech. *Cognition* **127**:177-202 https://doi.org/10.1016/j.cognition.2012.12.008 | PubMed

[63] **Gramfort Alexandre,** Luessi Martin, Larson Eric, Engemann Denis, Strohmeier Daniel, Brodbeck Christian, Goj Roman, Jas Mainak, Brooks Teon, Parkkonen Lauri, *et al.* (2013) MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* **7** https://doi.org/10.3389/fnins.2013.00267 | PubMed

[64] **Crosse Michael J.,** Zuk Nathaniel J., Di Liberto Giovanni M., Nidiffer Aaron R., Molholm Sophie, Lalor Edmund C. (2021) Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience* **15**:705621 https://doi.org/10.3389/fnins.2021.705621 | PubMed

[65] **Virtanen Pauli,** Gommers Ralf, Oliphant Travis E., Haberland Matt, Reddy Tyler, Cournapeau David, Burovski Evgeni, Peterson Pearu, Weckesser Warren, Bright Jonathan, *et al.* (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**:261-272 https://doi.org/10.1038/s41592-020-0772-5

[66] **Vallat Raphael** (2018) Pingouin: statistics in Python. *Journal of Open Source Software* **3**:1026 https://doi.org/10.21105/joss.01026

[67] **Seabold Skipper,** Perktold Josef (2010) Statsmodels: Econometric and Statistical Modeling with Python. In: Proceedings of the 9th Python in Science Conference. pp. 92-96 https://doi.org/10.25080/majora-92bf1922-011

[68] **Charlier Florian,** Weber Marc, Izak Dariusz, Harkin Emerson, Magnus Marcin, Lalli Joseph, Fresnais Louison, Chan Matt, Markov Nikolay, Amsalem Oren, *et al.* (2022) Statannotations.

[69] **Bialas Ole,** Dou Jin, Lalor Edmund C. (2023) mTRFpy: A Python package for temporal response function analysis. *Journal of Open Source Software* **8**:5657 https://doi.org/10.21105/joss.05657

Peer reviews

Reviewer #1 (Public review):

Summary:

Zhang et al. report on an ambitious study that investigates multiple aspects of the neural and behavioral underpinnings of auditory-motor surprisal in the context of an auditory-motor learning paradigm (piano keyboard). Using an intricate design comprising several sub-parts and control procedures, they report that early ERPs (50-100 ms latency) reflect violations of established key-pitch mappings.

Strengths:

This is a carefully devised and executed study. The paradigm is quite intricate and, at the same time, addresses multiple aspects of auditory-motor learning, and does so in a rigorous way.

Weaknesses:

Perhaps because of the exhaustive approach, it is sometimes difficult to follow which parts of the experimental design the results come from; there are some questions regarding appropriate statistical methods, the inclusion/treatment of musical background in participants, and the nature (latency & extent) of the identified neural components that detect auditory-motor violations.

https://doi.org/10.7554/eLife.111080.1.sa1

Reviewer #2 (Public review):

Summary:

Zhang et al. report an EEG study (n=18) of participants playing a keyboard where the correspondence between keys and pitches is varied to introduce sensory-motor mismatches (discrepancies between sensory inputs and expected sensory consequences of motor commands). They find that the auditory N100 amplitude is enhanced for the initial keystroke following a mapping switch but rapidly attenuates for subsequent keystrokes (showing rapid updating of the forward model), whereas the motor-related P50 amplitude only differentiates trained versus untrained mappings after 30 minutes of goal-directed practice (potentially showing timescales of inverse model updating). Using parallel univariate and mTRF decoding analyses, they conclude that forward models (mapping action to predicted sound) update almost instantly to track short-term context, while inverse models (mapping sound to motor commands) update slowly and require extended, targeted practice.

Strengths

(1) Methodological innovation:

The study utilizes an interesting, continuous auditory-motor paradigm that moves beyond standard trial-by-trial oddball designs, offering a more ecologically valid measure of trial-to-trial adaptation.

(2) Analytical elegance and rigor:

The combination of traditional univariate ERP analyses with multivariate temporal response function (mTRF) decoding is elegant, allowing the authors to successfully dissociate overlapping auditory and motor variance streams.

(3) The dissociation between the rapid adaptation of the N100 forward model and the slower adaptation of the P50 inverse model is interesting.

Weaknesses

(1) Confounded passive listening baseline:

The passive listening control condition lacks an orthogonal behavioural task (e.g., an occasional oddball detection task). Active playing inherently necessitates focused attention on auditory feedback to monitor performance, whereas passive playback does not. The globally weaker stimulus-evoked pattern at electrode Fz during passive listening strongly suggests that the absence of an N100 effect in this condition may simply reflect a lower state of attention, rather than isolating the absence of a motor-driven forward prediction, in particular because the pure sensory suprisal was also enhanced for "firsts" notes, so this could also lead to stronger N1, but this effect may be masked.

(2) Overclaimed theoretical novelty:

The conceptual framing leans excessively on the authors' specific "MirrorNet" framework, presenting foundational, decades-old tenets of the motor control literature (i.e., unsupervised exploration for forward models vs. supervised skill acquisition for inverse models; Wolpert, Jordan, both in the nineties) as their own novel "conjectures." This theory-heavy introduction obscures the paper's actual empirical contribution to the design and the interesting question regarding the distinct temporal adaptation scales of forward versus inverse models. I think some rewriting can improve the paper.

(3) Misplaced surprisal terminology:

In a similar vein, I find the use of the term "auditory-motor surprisal" more theoretical grandstanding than actually useful. The significance statement claims to "extend this principle from sensory processing" but in fact, the concept of sensory motor unexpectedness is again a staple of the forward motor literature. Moreover, nowhere in the paper do they actually estimate sensorimotor surprisal. While the authors compute surprisal for their auditory baseline using IDyOM, their central sensorimotor analysis relies entirely on a simple categorical mismatch (first vs. subsequent keystrokes). The phenomenon can equally be referred to by its established nomenclature-"sensorimotor mismatch" or "sensory motor unexpectedness".

(4) Incremental conceptual advance regarding the N100:

The paper frames the N100 finding as a major discovery, but as far as I know, the attenuation of the auditory N1 to self-generated sounds via accurate motor prediction-and its enhancement during sensorimotor mismatch - is one of the most heavily documented phenomena in the auditory-motor literature (e.g. Timm et al., 2013; Bendixen et al, 2012; 2013). As far as I'm concerned, the authors should clarify that the novelty lies in the novel, elegant design that provides a new way to correct for non-sensory-specific motor-induced attenuation, and characterizing the distinct adaptation timescales of forward versus inverse models -- not in demonstrating N100 modulation by sensorimotor mismatch, which is well-documented, AFAIC.

https://doi.org/10.7554/eLife.111080.1.sa0