npj | science of learning

Article

Published in partnership with The University of Queensland

https://doi.org/10.1038/s41539-025-00353-y

Note-by-note predictability modulates rhythm learning and its neural components

Check for updates

Marc Deosdad-Díez[1,2] & Josep Marco-Pallarés[1,2]

Rhythm production requires the integration of perceptual predictions and performance monitoring mechanisms to adjust actions, yet the role of auditory prediction remains underexplored. To address this, electroencephalography was recorded from 70 non-musicians as they synchronized with and reproduced rhythms containing notes of varying predictability. Participants were split into three groups, each receiving different visual cues to aid rhythm perception. Behaviorally, higher asynchrony occurred with less predictable notes. However, participants who viewed rhythms as distances between lines showed improved timing. EEG revealed that the Error Negativity component seems to reflect prediction error, increasing only when errors were clear and expected. When perceptual predictability was low, Ne response was reduced. The Error Positivity component, however, was heightened by both performance errors and unpredictable stimuli, highlighting the salience of such events. Overall, predictability plays a key role in shaping the neural and behavioral mechanisms underlying rhythm production.

Prediction plays a pivotal role in understanding human behavior. By extracting regularities from our environment and generating expectations about future events, we establish the groundwork for perception and other cognitive processes, including learning. The Active Inference Framework[1] posits that both perception and action strive to minimize prediction errors—the discrepancies between predicted and actual sensory input. This mechanism is particularly relevant in musical rhythm perception[2,3]. When auditory stimuli related to musical rhythm reach the auditory system, predictive internal models come into play. These models, shaped by statistical regularities derived from music enculturation[4], anticipate temporal patterns and relate to metrical structures—an accented isochronous pattern superimposed on auditory information—aiding our comprehension of rhythmic structure[3]. Interestingly, they are linked to motor planning activity[5], as shown by the activation of motor regions during rhythm perception[6,7].

While the impact of predictability on rhythm perception is well-established[8], its influence on rhythm production remains less explored. Indeed, rhythm production relies on a precise temporal template that must be translated into motor commands[9]. Therefore, it might be assumed that notes whose timing is better predicted will be produced more easily than those with a more unpredictable temporal appearance. Previous studies have shown that complex rhythms, particularly those presenting more syncopations (notes that deviate from metrical regularity[7,10,11]), are more difficult to learn and present higher inaccuracies than rhythms without syncopations[10]. Additionally, predictions of temporal information also integrate interval-based information, as it is possible to anticipate event occurrences by tracking transitional changes between intervals[11]. In other words, rhythm complexity affecting production could be derived from the mismatch of the sequence events with an underlying isochronous accented pattern (beat and meter) or the sequence of intervals that conform the pattern.

Importantly, predictability also plays a role in the different stages of learning to produce a rhythm, although it has never been explicitly studied. Therefore, in the initial learning stages when the temporal template is not apprehended, there is less clear the exact timing in which the response has to be produced, resulting in increased inaccuracy[10] and modulation of the component associated with error-monitoring and conflict processing[9]. Previous research has shown that this initial inaccuracy is linked to an increase in the Error Negativity (Ne), a fronto-central Event-Related Potential (ERP) peaking 50–100 ms. This ERP has been related to the commission of errors, both in cognitive control tasks such as flanker and Simon tasks[12] as well as in rhythm-learning tasks[9]. Although initial theories proposed that Ne was associated with responses to errors, current accounts propose that it captures the conflict between the activation of the correct and other concurrent potential responses[13].

[1]Department of Cognition, Development and Educational Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain. [2]Cognition and Brain Plasticity Unit, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. e-mail: marcdeosdaddiez@gmail.com; josepmarco@ub.edu

Once the rhythm template is consolidated across repetitions, errors decrease, and their associated Ne neural components diminish. Accordingly, inaccuracies are more clearly detected. This is reflected in another ERP component, Error Positivity (Pe), a centroparietal positive deflection occurring between 100 and 300 ms after an erroneous response. Different studies have linked the appearance of Pe after an erroneous response to the awareness of this error, which may elicit changes in attentional allocation, resulting in a reorienting response toward subsequent actions[14]. However, while these two components (Ne and Pe) reflect different cognitive control mechanisms critical for rhythm production, they have primarily been studied in relation to response accuracy in rhythm production tasks. However, their relationship with the predictability of the note to be produced remains unexplored.

Previous research on rhythm learning and its associated cognitive control mechanisms has primarily focused on global aspects of rhythm predictability, such as the difficulty of producing a rhythm based on the degree of syncopation[10], or how accuracy evolves with repetition[9]. However, no studies have specifically examined how the predictability of individual notes influences the accuracy of their production. To fill this gap, the present study investigates the role of individual note predictability in rhythm learning and production, as well as its impact on the underlying cognitive control processes. Given that rhythm templates derive from temporal information perception, we hypothesize that predictability significantly influences rhythm production, both during synchronization and reproduction trials in a rhythm-learning task. To investigate this, we manipulated the temporal predictability of individual notes within rhythms and analyzed them using the Information Dynamics of Music (IDyOM)[4] computational model, which captures statistical features of music. IDyOM provides the information content of each note—a measure of the temporal predictability of the note—based on the probability distribution of a predefined corpus, as well as the rhythms listened to during the experiment. In the rhythm domain, IDyOM allows for the assessment of note predictability not only in terms of syncopation, which is related to beat-based predictions, but also by integrating interval-based predictions of the sequence itself. Simultaneously, we created different learning contexts through images that provided different degrees of temporal information about the rhythms that could benefit their learning. Specifically, we created a graphical representation in which time was represented as distance and a categorical representation in which different intervals were represented as distinct geometrical figures. Additionally, a control group where images did not provide meaningful information was included.

We hypothesized that the predictability of individual notes would be a critical factor in rhythm production and its associated cognitive control mechanisms. Specifically, we predicted that the information content (IC) of individual notes, reflecting their degree of surprise regarding their timing as computed using IDyOM, would influence note accuracy. Notes with higher IC would be more difficult to produce, leading to greater asynchrony compared to notes with lower IC. Furthermore, we hypothesized that more unpredictable notes would elicit an increased magnitude of the Ne and Pe ERPs, reflecting an increase in motor conflict. Additionally, we predicted that increasing note predictability through the use of meaningful images—particularly in the graphical group—would enhance rhythm accuracy, facilitate rhythm learning, and reduce the cognitive control required for the task. This reduction in cognitive control would, in turn, be reflected in a decreased magnitude of the Ne and Pe ERPs.

Results

In this study, participants learned to produce nine different rhythms. Each trial consisted of three phases: first, the rhythm was presented without any action; next, participants tapped along with the rhythm (synchronization); finally, they tapped the rhythm without auditory feedback (repetition). This sequence was repeated ten times for each rhythm in a row. Our goal was to investigate how the perceptual predictability of individual notes influenced rhythm production, learning, and the associated cognitive control mechanisms. The predictability of each note was quantified using information content (IC) as computed by IDyOM, a measure of how expected the timing of a note was within the given context.

Additionally, participants were assigned to different groups based on the type of visual information provided, which varied in its potential to enhance note predictability: (1) the graphical group received high-enhancement visual cues (rhythmic values represented as spatial information), (2) the categorical group received low-enhancement cues (rhythmic values as shapes), and (3) the control group received meaningless visual information (random dot points). The performance during the task was assessed using the absolute asynchrony, that is, the time difference between the produced and the auditory onsets. In the behavioral analysis, the asynchrony of each note was used as a dependent measure in the LMM, with the IC (surprise) of the note, repetition of the rhythm and group as factors. In addition, cognitive control mechanisms, as reflected by the Ne and Pe of each response, were also analyzed following a similar procedure using the amplitude as dependent measure and asynchrony, IC, repetition and group as factors.

Synchronization sequences behavioral results

Performance accuracy was analyzed for sequences where synchronization was done with sounds. In particular, we hypothesized that asynchrony in this condition would increase with unpredictability of the notes and accuracy would increase with repetition. As expected, asynchrony decreased across repetitions ($F(1,68) = 131.65$, $p < 0.001$), indicating that participants learnt to perform more accurately through the task. On the other hand, Information content was associated with higher asynchrony overall ($F(1,56570) = 4482.4$, $p < 0.001$). There was also an interaction between repetition and information content. As can be seen in Fig. 1a, the obtained learning curves show different patterns across IC levels, with higher IC resulting in steeper curves. There was an interaction between repetition and IC ($F(1,56585) = 136.76$, $p < 0.001$). In other words, notes that were more predictable were performed more accurately since the beginning and did not change much throughout repetitions (Fig. 1b).

Moreover, as expected, the visual representations showed an effect ($F(2,67) = 11.32$, $p < 0.001$) on accuracy. This aide that presented intervals as distances between lines (graphical group) helped participants to have overall less asynchrony with an estimated mean of 0.08 compared to the No Information group (0.093, $z = -3.23$, $p < 0.035$) and the Categorical group (0.097, $z = 3.23$, $p < 0.001$). There was no statistically significant difference between the No Information and Categorical group ($z = -1.0$, $p = 0.58$). Additionally, this representation also reduced the effect of other variables, with an interaction between group and IC ($F(2,56570) = 64.40$, $p < 0.001$) and group and repetition ($F(2,68) = 3.81$, $p = 0.03$). For instance, repetitions for the graphical group reduced the asynchrony repetition ($\beta = -0.003$) compared to the categorical group ($\beta = -0.005$) or the no information group ($\beta = -0.005$). Only the comparison between Graphical and Categorical ($z = -2.35$, $p = 0.048$) was statistically significant, while the comparison between Graphical and No Information was marginal ($z = -2.09$, $p = 0.09$). Graphical group reduced the impact of IC in increasing asynchrony ($\beta = 0.15$) compared to No Information ($\beta = 0.2$) and Categorical group ($\beta = 0.22$). The comparison between groups showed that both Graphical compared to No Information ($z = 8.3$, $p < 0.001$) and Graphical compared to Categorical ($z = 10.64$, $p < 0.001$) were statistically significant. In other words, the visual representation reduced the learning associated by repeating the rhythms and the effect of unpredictability by boosting the accuracy in performance through congruent visual information. Additionally, a triple interaction was found between group, repetition and IC ($F(2,56595) = 3.36$, $p < 0.03$). Changes in nIC trends by repetition changed differently by group: No Information group changed by an estimate of 0.02, Categorical group by 0.02 and Graphical by 0.01. Pairwise comparison showed statistically significant change compared to Graphical Group (No-Information vs Graphical $z = 8.26$, $p < 0.001$; Categorical vs Graphical $z = 10.64$, $p < 0.001$). The conditional R^2 of the model was 0.12, and the marginal R^2 was 0.08 (Supplementary Table 1).

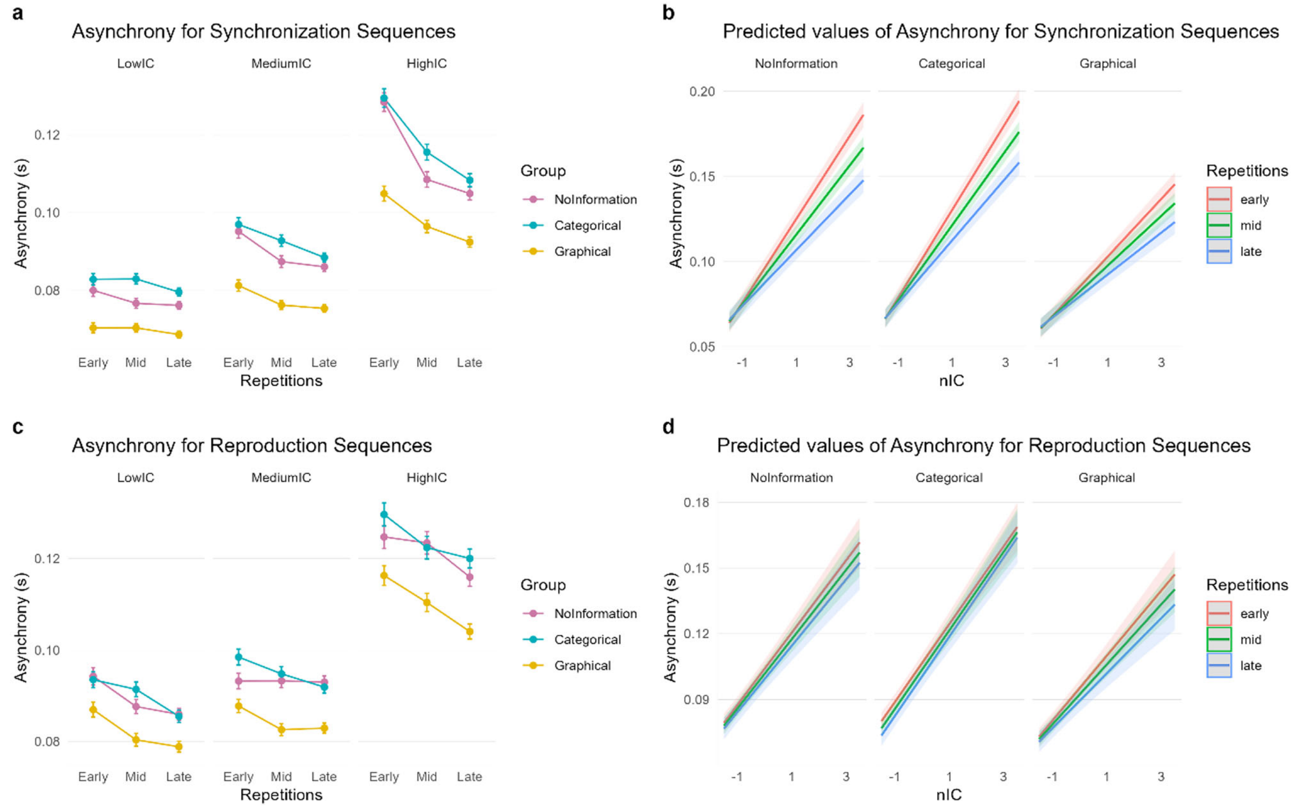

Fig. 1 | Behavioral results. **a** Asynchrony across repetitions for three levels of information content of the Synchronization trials. Asynchrony was grouped for visualization purposes in learning stages (early: 1st to 3rd repetitions, mid: 4th to 7th repetitions, late: 8th to 10th repetitions) and information content (low: lowest tertile of IC, medium: medium tertile of notes, high: highest tertile). Asynchrony decreases by repetition, increases by IC. Graphical group shows a reduction of asynchrony and a diminishing effect on repetition and IC. **b** Predicted values of the model for asynchrony of Synchronization Trials. **c** Asynchrony across repetitions for three levels of information content of Reproduction Trials. Asynchrony was grouped for visualization purposes in learning stages (early: 1st to 3rd repetitions, mid: 4th to 7th repetitions, late: 8th to 10th repetitions) and information content (low: lowest tertile of IC, medium: medium tertile of notes, high: highest tertile). Asynchrony decreases by repetition, increases by IC. Graphical group shows a reduction of asynchrony and a diminishing effect on repetition and IC. **d** Predicted values of the model for asynchrony of Reproduction Trials.

ERP results of synchronization sequences

Visual inspection of the response-related evoked potentials associated with the tapping responses of the participants showed a series of components compatible with rhythm-learning literature (Fig. 2a).

Increased activity in the first time window compatible with an Ne component was found peaking around 50 ms after the response. We expected an increase for this component's magnitude with asynchrony due to the relationship between the Ne and error, and a reduction of this component through repetition. In addition, we expected an increase in the Ne with IC due to its potential relationship with conflict. Finally, we also expected a reduction of Ne in the graphical group due to an increase in the predictability of the rhythms in this condition.

Contrary to what we expected, the mixed-effect linear model (Supplementary Table 3, Fig. 2b) predicted a reduction in the negativity for this component with IC values ($\beta_{\mathrm{NoInfomation}} = 0.24$, $\beta_{\mathrm{Categorical}} = 0.22$, $\beta_{\mathrm{Graphical}} = 0.13$, $F(1,78) = 12.03$, $p < 0.001$) and asynchrony ($\beta_{\mathrm{NoInfomation}} = -0.22$, $\beta_{\mathrm{Categorical}} = 0.08$, $\beta_{\mathrm{Graphical}} = -0.19$, $F(1,53424) = 4.1$, $p = 0.04$). There was also an interaction between repetition and asynchrony ($F(1,55267) = 15.62$, $p < 0.001$) and IC and Asynchrony ($F(1,52373) = 4.37$, $p = 0.4$). The trend of asynchrony at first repetition ($\beta = 0.24$) and the last ($\beta = -0.43$) one changed ($t = 4.08$, $p < 0.001$). This suggests that in later repetitions, asynchrony decreases the component as sign of error. The trends of asynchrony at low IC (-0.01) compared to high IC (-0.41) changes ($t = 1.95$, $p = 0.5$). This suggests an increased magnitude for high IC and high asynchrony.

Finally, there was also an interaction between group x repetition x asynchrony ($F(2,52981) = 4.14$, $p = 0.02$), although pairwise comparisons are not statistically significant. Additionally, repetition x IC x Asynchrony ($F(1,54059) = 7.24$, $p < 0.01$). The conditional R^2 of the model was 0.02 (marginal $R^2 = 0.002$).

Then we analyzed the Pe component at the Cz electrode in the 90–150 ms range (Fig. 2a). As this component has been linked with error awareness, we expected to increase with asynchrony and information content, and to be reduced in the graphical group. The model (Supplementary Table 4, Fig. 2c) showed an increase in the positivity for higher values of information content ($\beta_{\mathrm{NoInfomation}} = 0.72$, $\beta_{\mathrm{Categorical}} = 0.47$, $\beta_{\mathrm{Graphical}} = 0.69$, $F(1,55046) = 218.42$, $p < 0.001$) and asynchrony ($\beta_{\mathrm{NoInfomation}} = 0.50$, $\beta_{\mathrm{Categorical}} = 0.91$, $\beta_{\mathrm{Graphical}} = -0.58$, $F(1,75) = 52.86$, $p < 0.001$). Additionally, there was a significant negative interaction between IC and asynchrony($F(1,44601) = 12.69$, $p < 0.001$) and repetition and IC ($F(1,55253) = 6.40$, $p = 0.01$). Components magnitude increased by asynchrony in lower IC notes (0.82) than high IC notes (0.10) ($t = 3.56$, $p < 0.001$). Additionally, Pe magnitude increased less at early repetitions more by IC (0.45) rather than in later repetitions (0.80) ($t = -2.54$, $p = 0.01$). Also, there was an interaction between group x repetition x asynchrony ($F(2,53960) = 6.22$, $p < 0.01$). $\beta_{\mathrm{NoInfomation}} = 0.49$, $\beta_{\mathrm{Categorical}} = 0.90$, $\beta_{\mathrm{Graphical}} = 0.58$, although pairwise comparisons do not reach statistical significance. The total explanatory power of the model was $R^2 = 0.03$ (conditional), and the fixed effects power (marginal R^2) was 0.01.

Behavioral results of reproduction sequences

Reproduction sequences were produced without auditory information, just after each synchronous trial. Participants repeated the previously listened-to and tapped stimuli with only the visual aids on the screen. The asynchrony

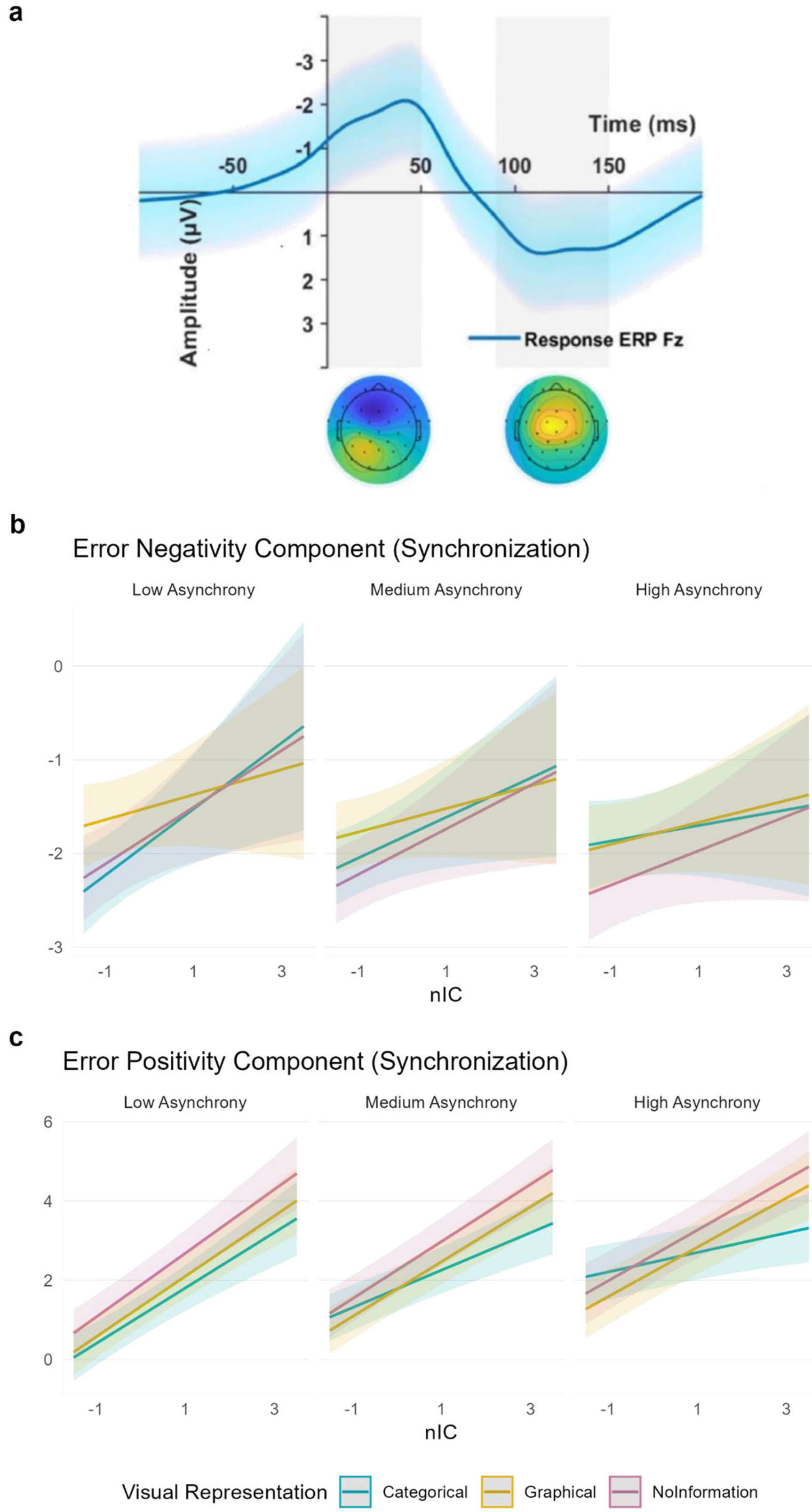

Fig. 2 | Error negativity and error positivity for synchronization trials. **a** Grand average of the response-locked evoked potentials for synchronization trials. In gray, the time windows used for the note-by-note analysis with their topographic maps. The first window corresponds to the Error Negativity (frontocentral negativity between 0 and 50 ms), and the second window corresponds to the Error Positivity (central positivity between 90 and 150 ms). **b** Predicted values of the Error Negativity for Information content and degree of asynchrony corresponding to its response. Information Content reduces the component, while asynchrony increases it for high information content notes. **c** Predicted values of the Error Positivity for Information content and degree of asynchrony corresponding to its response. Information content increases the components' magnitude, while asynchrony increases the component only for low IC notes.

across learning stages showed a similar pattern to the synchronization sequences (Fig.1c).

Asynchrony for reproduction sequences was analyzed with a mixed-effect linear model (Fig. 1d, conditional $R^2 = 0.07$, marginal R^2 of 0.05). Linear mixed-effect model (Supplementary Table 2) showed that asynchrony decreased with repetition ($\beta = 0.003$, $F(1,65) = 36.56$, $p < 0.001$) and increased with information content ($F(1,67) = 587.5$, $p < 0.001$).

A group effect was found ($F(1,67) = 7.91$, $p < 0.001$). The graphical group presented an overall lower mean asynchrony of 0.09, compared to the mean asynchrony of 0.10 of Categorical and No Information. Comparisons between No Information and Graphical ($z = 3.01$, $p < 0.01$) and Categorical and Graphical ($z = 3.7$, $p < 0.001$) showed statistical differences. Moreover, there was an interaction between group and IC ($F(2,67) = 3.52$, $p = 0.03$). Pairwise comparison of the trend slopes ($\beta_{NoInfomation} = 0.16$,

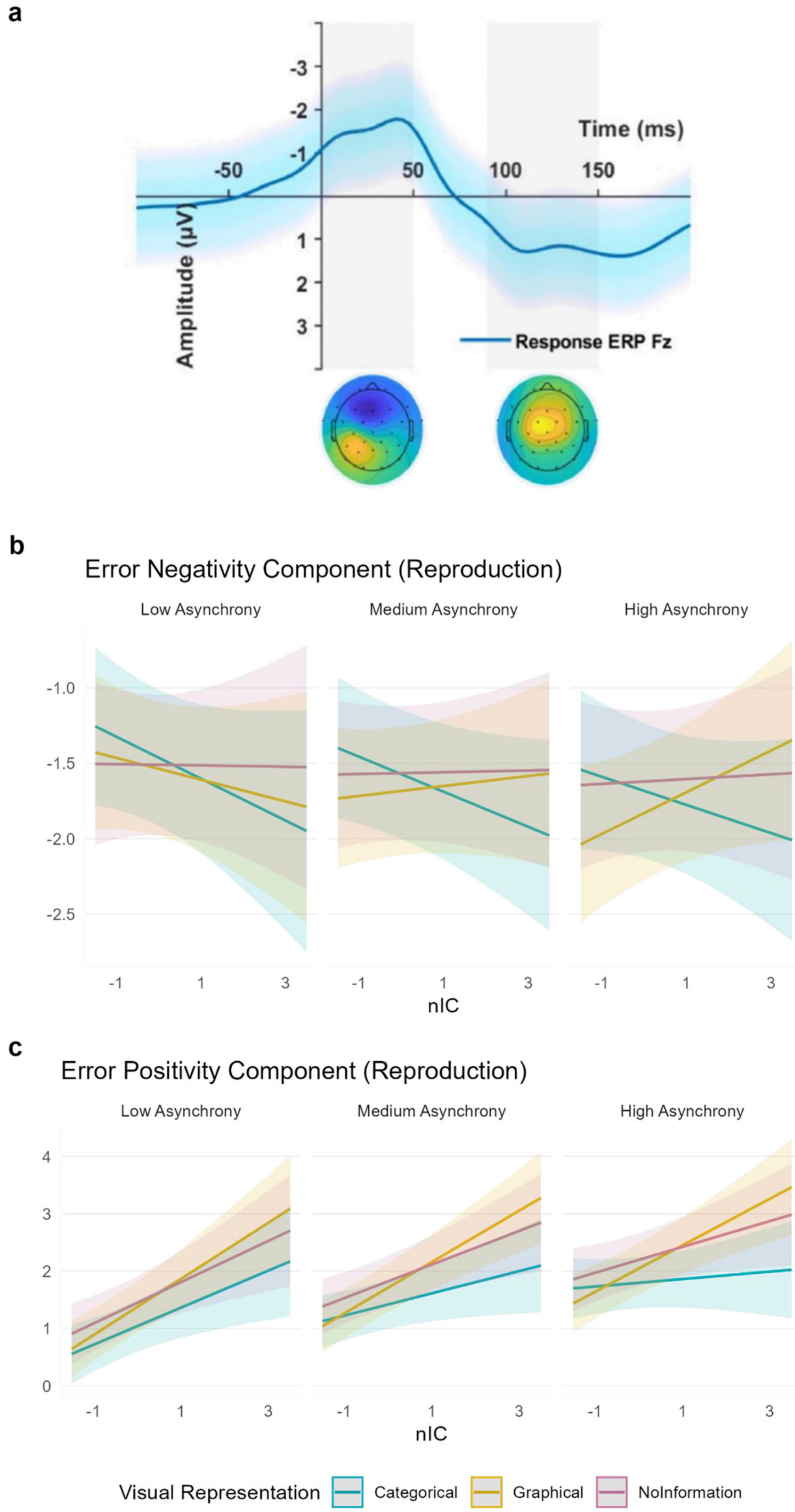

Fig. 3 | Error negativity and error positivity for reproduction trials. a Grand average of the response-locked evoked potentials for reproduction trials. In gray, the time windows used for the note-by-note analysis with their topographic maps. The first window corresponds to the Error Negativity, and the second window corresponds to the Error Positivity. **b** Predicted values of the Error Negativity for Information content and degree of asynchrony corresponding to its response. No variables showed a statistically significant modulation of the component's magnitude. **c** Predicted values of the Error Positivity for Information content and degree of asynchrony corresponding to its response. Error Positivity magnitude increases for information content. Asynchrony increases the component magnitude for low IC notes.

$\beta_{Categorical} = 0.18$, $\beta_{Graphical} = 0.14$) was only statistically significant between Categorical and Graphical groups ($z = 2.68$, $p < 0.2$). This shows that in this case, the visual representation compared to the categorical representation reduced the effect of the note unpredictability in asynchrony. No significant effects were found for the no-information group compared to the categorial group.

ERP results of reproduction sequences

In the EEG domain, the Ne component (Supplementary Table 5, Fig.3b, conditional $R^2 = 0.02$, marginal R^2 of 0.002) was modulated by asynchrony ($F(1,51298) = 7.05$, $p < 0.01$) and repetition ($F(1,51274) = 4.30$, $p = 0.04$). No statistically significant differences between groups were found regarding asynchrony ($\beta_{NoInfomation} = -0.07$, $\beta_{Categorical} = -0.14$, $\beta_{Graphical} = -0.19$)

and repetition ($\beta_{NoInfomation} = 0.11$, $\beta_{Categorical} = 0.12$, $\beta_{Graphical} = 0.01$). On the other hand, the Pe component (Supplementary Table 6, Fig.3c) was increased with repetition ($\beta_{NoInfomation} = 0.21$, $\beta_{Categorical} = 0.18$, $\beta_{Graphical} = -0.05$, $F(1,51256) = 7.09$, $p < 0.01$), information content ($\beta_{NoInfomation} = 0.29$, $\beta_{Categorical} = 0.18$, $\beta_{Graphical} = 0.45$, $F(1,77) = 36.07$, $p < 0.001$) and asynchrony ($\beta_{NoInfomation} = -0.49$, $\beta_{Categorical} = 0.5$, $\beta_{Graphical} = 0.44$, $F(1,51310) = 84.25$, $p < 0.001$), conditional R^2 of the model = 0.02 (marginal $R^2 = 0.004$), as expected. Additionally, there was an interaction between group and repetition ($F(2,51256) = 5.06$, $p < 0.01$). Only statistically significant differences between groups were found for repetition between graphical group and the other two groups (Graphical vs No Information $z = -2.73$, $p < 0.02$; Categorical vs Graphical $z = 2.42$, $p < 0.04$). This indicates that for graphical group, the nIC does not increase the magnitude compared to the other groups. There was also an interaction between asynchrony and information content ($F(1,50830) = 5.95$, $p < 0.05$). Asynchrony estimates at low IC ($\beta = 0.58$) to high IC changes ($\beta = 0.13$), indicating that Pe magnitude does not increase much for asynchrony of higher IC notes ($t = 2.40$, $p < 0.2$).

Discussion

Learning to produce a rhythm is a complex skill that depends on the development of robust motor templates. The timing predictability of each note is crucial, yet research on its influence on rhythm production is limited. Our study explored this by examining the information content of each note and improving precision through visual cues provided to participants. We found that note predictability was associated with performance accuracy, with more surprising notes being produced with less precision (higher asynchrony) across all learning stages. Moreover, providing visual timing information resulted in a marked reduction of asynchrony. Notably, the Ne component was reduced in notes with higher IC complexity only when rhythms were produced with their corresponding sound. Conversely, the Pe component escalated with note complexity and asynchrony. These findings underscore the pivotal role of the predictability of individual notes in learning new rhythms.

The first main finding of the study is the significant effect of predictability in rhythm production, with unpredictable notes showing less synchronization accuracy. These notes, however, showed the most significant improvements with practice, especially in synchronization tasks. Previous studies have described reduced performance in complex rhythms, particularly those presenting more syncopation. However, our study linked differences in performance to specific notes and its associated surprise, which it was not only related to syncopation but to other temporal aspects captured by IDyOM model, such as inter-onset intervals and metric accent of the notes. This general measure, IC, which reflects the surprise of individual rhythms, directly influences the difficulty of accurately tapping a note. Therefore, rhythm's internal template for each note hinges upon the note's information content, with greater difficulty in processing notes that are less predictable, suggesting that they initially demand more from performance-monitoring systems, a demand that decreases as learning progresses. As the internal rhythm template refines, it allows for more accurate timing of note production, especially for challenging notes. An important question is how the IC of a note might influence taps preceding its onset. This could be puzzling if participants were encountering the rhythm for the first time. However, in our experiment, participants had already heard the rhythm before tapping, even in the first trial. Our findings suggest that IC captures the difficulty of tapping a previously heard note, highlighting its role in shaping rhythmic performance. Crucially, the manipulation of the internal template by providing explicit information on the temporal representation of the notes in the graphical group yielded, on one hand, a global reduction of the asynchrony and, on the other, a reduction in the effect of learning, supporting the effectiveness of this procedure to reduce the uncertainty in rhythm templates and enhancing performance. However, since we did not include a group without any visual information, we cannot determine whether the graphical procedure improved accuracy compared to the absence of a visual aid. Future studies incorporating such a control group could help assess the effectiveness of this intervention.

The relationship between the predictability of the individual notes and performance-monitoring mechanisms is supported by the modulation of the Ne component. The main result is the modulation of the Ne with the IC. Unexpectedly, Ne amplitude during synchronization trials was inversely related to IC, being smaller for notes with higher IC. In addition, Ne was also enhanced in more complex notes when the asynchronies were high. Therefore, Ne was not related solely to erroneous responses, as previously suggested in performance-monitoring[15] and rhythm production[9] tasks, but associated with the predictability or the notes to be produced. Therefore, we propose that Ne is signaling a discrepancy between expected and actual response modulated by the predictability of the note, which would be indexed by the IC. We propose that notes with low IC would generate highly precise prediction errors that would be further translated to higher levels in the hierarchy to induce changes in the generative models and improve internal templates. Therefore, the results of the current study show that modulation of Ne is related to those notes that are more likely to be used to update the model. Ne is higher for notes with low IC, independently of their asynchrony, as prediction errors in those notes are very precise and, therefore, according to the model, likely to be used to update internal models. In contrast, more unpredictable notes at a perceptual level (that is, those presenting higher IC) yield to imprecise prediction errors, which would be virtually ignored and not sent to higher levels of the hierarchy unless the errors are very large, in which case they can be clearly detected and will be useful for updating the model. This also corresponds to the modulation of Ne, with lower amplitude for low asynchronies compared to high asynchronies in notes with high IC.

Interestingly, these results nicely fit within the active inference framework[2]. In the music domain, this framework proposes that prediction errors in the incoming notes are associated with precision. Precise prediction errors would be translated to higher levels in the hierarchy to induce changes in the generative models, while unprecise ones would be discarded. In our study, the precision of the prediction errors would likely inversely escalate with IC, with prediction errors of low IC notes being more precise than the high IC ones.

Current results are at odds with previous results on rhythm-learning production[9] in which responses with higher asynchrony (labeled as errors in that study) showed an increase in Ne in the early stages of learning. This finding is not replicated in our current study in the no-information group in the synchronization trials, which is the one that only relies on auditory information. Our findings show an increased negativity overall for responses that were more asynchronous without the reduction in latter repetitions. This discrepancy could be related to the fact that in this study, the IC of the sounds was not considered, with the possibility of mixing the different factors affecting the Ne. In addition, this study divided the trials into correct and erroneous responses, while we have not categorized the responses, but used asynchrony and repetition as factors in the linear mixed-effects model. Our results suggest that the Ne is not evaluating responses only on the basis of accuracy, but also on the relationship to the surprise of the notes to be produced. Therefore, the categorization of the trials could overlook the modulation of the Ne.

Additionally, the modulations of Ne observed during synchronization trials by IC disappear in reproduction trials, making the role of the Ne in the reproduction trials similar to previous literature[9]. In contrast, in these trials, this component is modulated by repetition and asynchrony, suggesting that a comparison between the expected and actual response is still taking place. Di Gregorio and colleagues[16] found that when the explicit expected response was absent, but an error was detected through other contextual cues, the Ne disappeared while the Pe persisted. In this context, the representation of the expected response—specifically, the timing of the perceived note—is maintained in memory rather than actively perceived. This distinction could result in a less precise expected response, weakening the comparison process and removing the effect of IC on the reproduction trials.

In contrast, the modulation of Pe follows a different pattern than Ne. It was expected that the Pe would increase its amplitude along repetitions and asynchrony. Our findings aligned with these expectations. These results fit with previous literature suggesting that once the rhythm template is learnt, error awareness increases[9]. In addition, responses that produced notes that were more unpredictable also presented higher amplitudes in this component. The Pe has been proposed to index the accumulation of evidence that an error has occurred and its conscious awareness[17], integrating information of different sources[18]. This awareness leads to an orienting response to unprovable events in the environment that will require a behavioral adjustment[18]. According to our results, the Pe would increase for those notes that arouse higher saliency, either because they are surprising (higher IC), the response was not precise (higher asynchrony), or a combination of both (e.g., low complexity but high asynchrony). Therefore, Pe could signal saliency of the response rather than awareness of error per se.

Predictability modulating the Ne and Pe may indicate effects at different stages of the cognitive control processing. Stimulus to goal-appropriate action mapping takes place in the pre-Supplementary Motor Area (pre-SMA)/Supplementary Motor Area (SMA), where predicted action forward models are transferred to be selected in the basal ganglia and effected in the motor cortex[5]. In addition, in the basal ganglia, a copy of the motor command is generated and transferred back to pre-SMA/SMA, where an evaluation is made to update the action models[12]. Our results support the idea that complexity in the auditory level weights the comparison to the copy of the motor command. This is reflected by the weighted prediction error indexed by the Ne. However, the Pe might be signaling the recruitment of attentional resources in the feedback controllers to adjust the behavior-integrating saliency aspects of the given response. Additionally, visual information about the timing of the notes in the graphical group reduces asynchrony without presenting a main effect on Ne and Pe. This behavioral effect and the modulation of asynchrony, repetition and IC on the Ne and Pe components by visual information suggest that explicitly representing the temporal structure of the rhythm might influence other stages of the performance control processing. Specifically, visual information could be affecting the feedback controllers by modifying attentional resources required to update control forward models or by enhancing the action-selection mechanisms by which the appropriate motor command will be chosen. Future research is needed to understand how integrating visual information in auditory-motor tasks affects performance monitoring processing.

Methods

Participants and experimental design

Seventy-two subjects (M = 22.24, SD = 4.15, 52 women) were recruited for this study. All participants reported having less than 3 years of musical training and were not currently learning to play any instrument. Participants also reported no psychological or neuropsychiatric disorders or motor dysfunction of the superior limbs. All participants signed a written informed consent, were paid €10 per hour and were randomly allocated to one of the three groups corresponding to each visual representation. All procedures were approved by the Institutional Review Board of the University of Barcelona (IRB0003099), in accordance with the Declaration of Helsinki. Two participants were excluded due to data loss, with a final sample of 70 participants (24 in categorical group, 24 in graphical group and 22 in no information group).

The experimental paradigm consisted of an adaptation of the Rhythm Synchronization Learning Task[9,10]. The participants had to first listen to an auditory rhythm. Subsequently, they listened to the same rhythm and tapped simultaneously using a response pad (RB-840 Cedrus Response Pad). Finally, they were required to reproduce the rhythm without sound using the same pad. Each of the three sequences was preceded by two beeps separated by 0.5 s each, which indicated the beginning of the rhythm (Fig. 4a). These three sequences were repeated ten times for each of the nine rhythms in a block design. After the 2nd, 4th, 7th, and 10th reproduction trials, participants were asked to rate on a Likert scale how much they liked the rhythm that they were listening to and tapping.

While participants had to listen to, synchronize with, or reproduce a rhythmic sequence, an image was presented on the screen. Therefore, visual information appeared in all three sequences, whereas auditory information only appeared in the listening and synchronization conditions.

After the Rhythm Synchronization Learning Task, participants were asked to rate the stimuli after listening to each in the lab using a rating scale like the one used in the pre-test phase. Participants of the categorical and graphical groups also had a test phase in which they had to reproduce three new rhythms using only visual information, without previously hearing them. This test phase was presented to assess whether the participants were able to decode visual representations of the rhythms. This data is not analyzed in the present manuscript.

The duration of the experiments was approximately 1 h. The rhythms were presented in a block design with a fix order (see Table 1 for the characteristics of the rhythms). Every time the rhythm changed (block of 10 repetitions), there was a pause for participants to rest. Before starting the task, the participants practiced with an easy rhythm. In this practice, they were presented with two repetitions of a sequence of three parts. Once the practice was finished, the task began.

The task was built using PsychoPy software (version 3.0.5) in Python programming language[19,20] on a PC, and the sound was presented using stereo headphones at a comfortable intensity level adapted for each participant.

Stimuli

The rhythms were adapted from previous experiments[7,9,10] and consisted of 11 snare drum notes of 200 ms generated using the online software of music edition Bandlab. Each rhythm had the following rhythmic category intervals: five eighth notes (IOI = 250 ms), three quarter notes (500 ms), one dotted quarter note (750 ms), one half note (1000 ms) and one dotted half note (1500 ms). These 11 intervals were randomly scrambled to generate a pool of 300 different rhythms varying in complexity.

The 9 selected rhythms were chosen according to their information content (Table 1), a predictability measure given by the Information Dynamics of Music model[4]. Information content (IC) gives a quantitative measure of the unpredictability of an event based on extracting the statistical properties of a corpus of music by which the model is trained. In this study, the model was trained in extracting the probabilities of onsets, inter-onset intervals and metric accent of the notes of 185 chorale melodies by Bach[21]. The result was a note-by-note measure of the rhythm, in which higher values indicated more unpredictability. The IDyOM model tries to simulate the process of musical enculturation and statistical learning of musical features of humans, and IC measure is related to surprise and prediction error produced by the analyzed event[4].

In our study, we used three specific viewpoints of the IDyOM model to analyze temporal structure and its predictability. Specifically, onset, metric accent and Inter-onset interval were used. Probabilities of each viewpoint were extracted from the corpus and the actual stimuli to understand their structure. Onset viewpoint represented the timing of a note event, analyzing its time of onset. Inter-Onset Interval (IOI) viewpoints consider the spacing between notes, analyzing how different intervals can be followed from one to another. Lastly, metrical accenting analyzed the strength of the note given a metrical framework, assigning stronger accenting to events that occur in the downbeat or relevant metrical positions.

We opted for analyzing note-by-note predictability instead of static overall complexity measure because musical rhythms are perceived sequentially. This approach allows a fine-grained analysis of the dynamic interplay between expectation and occurrence that can be critical in performance.

Meaningful visual representations (Fig. 4b) of the rhythms were provided for two of the three groups. The first one consists of a categorical transcription of musical notation used in Western music (categorical group). A geometrical figure was associated with each of the five intervals: a

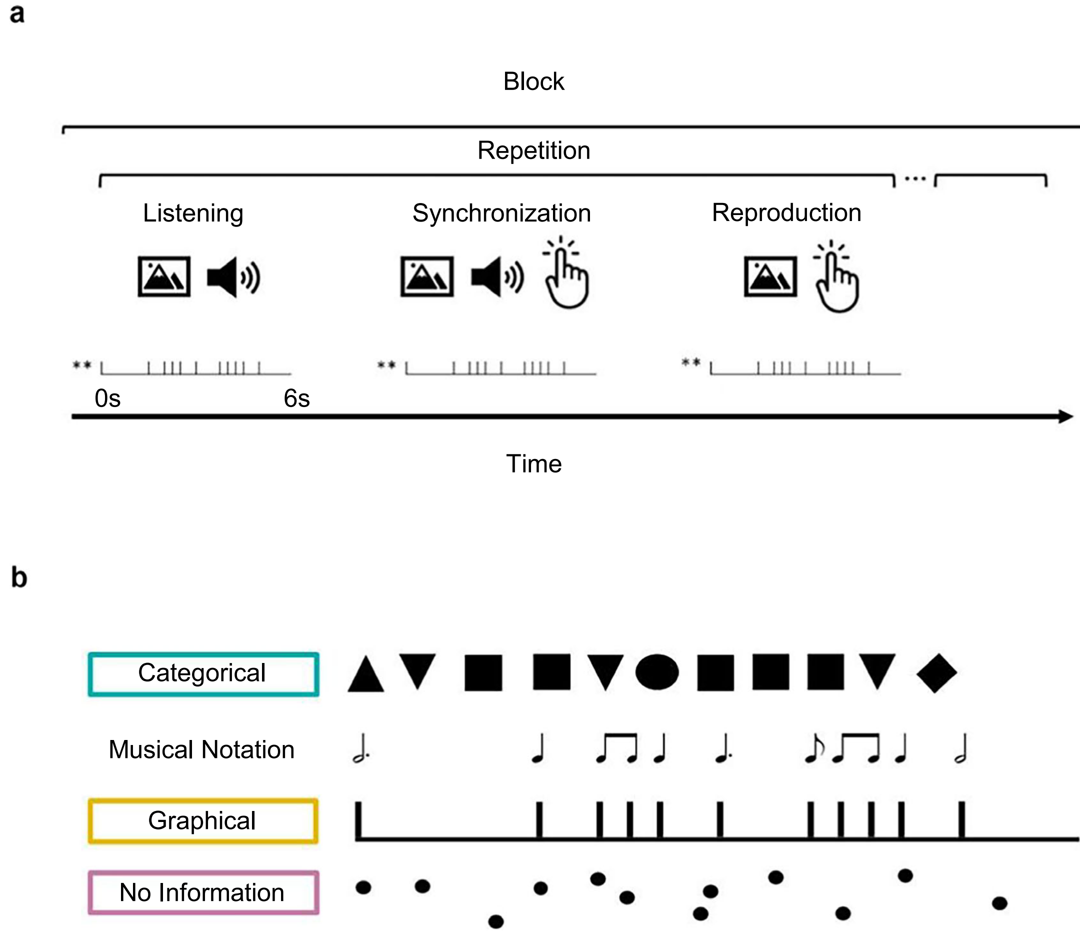

Fig. 4 | Paradigm and visual representations. a Experimental paradigm: rhythms were presented in sequences of 3 trials. Listening Trials required participants to listen to and watch the visual representation. Synchronization Trials required to synchronize pressing a button to the auditory rhythm with the help of the visual presentation. Reproduction Trials required participants to reproduce the rhythm without sound, helped only by the representation. This sequence was repeated 10 times for each rhythm. **b** Types of visual representation and their correspondence to Western musical notations. Categorical: different geometric figures representing the different intervals. Graphical: distance represents the intervals between notes. No Information: no correspondence with auditory rhythm.

Table 1 | IC for the rhythms used in the experiment

Rhythm	Mean	SD	Min	Max
1	2.28	2.48	0.11	7.96
2	2.25	1.98	0.16	5.71
3	2.29	2.17	0.18	6.92
4	2.69	2.39	0.18	6.95
5	3.62	3.74	0.16	12.29
6	4.30	3.26	0.19	10.95
7	3.98	3.45	0.12	10.74
8	4.02	2.16	0.16	7.86
9	2.72	2.51	0.15	6.85

Information content of the 9 rhythms used in the experiment: mean information content, standard deviation, minimum and maximum value of the IC of the notes of the rhythm.

square was associated with an eighth note, an inverted triangle represented a quarter note, a circle represented a dotted quarter note, a rhomb represented a half-note, and an upward-pointed triangle represented a dotted half-note. The figures are evenly spaced. The second group of participants (graphical group) received a graphical representation consisting of marking each sound with a vertical line, where the distance between lines represented the time between the sounds (e.g., twice the distance represented twice the time between the sounds). We hypothesized that in this group, participants would extract the temporal information more easily and that it would be more useful in the performance of the rhythm. Finally, the third group (no information group) received a series of dots on the screen that had no relationship with the rhythm they were listening to. Participants were randomly assigned to one of three groups before starting the experiment.

Behavioral analysis

To quantify the performance of the participants, we used the asynchrony computed as the absolute difference between the produced and the real auditory onsets in the case of the synchronization phase or the onset where the note would appear in the case of reproduction. A window was selected with half of the duration of the interval before the onset of the note and half of the duration of the interval after the onset of the note. All the responses given during this window were selected, and their asynchrony was calculated[9]. Only those responses with minimum asynchrony for each window were computed as actual responses and considered for behavioral and EEG analysis. Only rhythms between 8 and 14 taps were analyzed.

Note-by-note asynchrony was computed separately for the trials that required synchronization to the actual sounds and those that required the participant to reproduce them. A linear mixed-effect model was built to predict asynchrony as a function of the predictability of each note in the rhythm (IC), the repetition in which this sound was produced, and the visual aid that was concurrent with learning and their interactions. Repetitions and

ICs were normalized to reduce fitting problems due to the different scales of the variables.

Linear mixed-effect models were selected due to the repeated nature of the sequences being learned and the repetition effect. Also, some produced rhythms that had fewer notes than sounds were listened to, so missing values were not avoided in this experimental design.

For all models, the maximum random effect structure was selected with random slopes for each variable and reduced using rePCA function[22] until the overfitting of the model was solved. The starting random structure of each model was participants and groups as random intercepts, while rhythm, IC and Repetition were accounted for as random slopes. We run an ANOVA to assess the statistical significance of the variables of the models and report the means and trends of the significant variables' slopes.

EEG analysis

EEG signals were recorded from an active electrode cap with 29 electrodes mounted at a standard position (Fp1/2, Fz, F7/8, F3/4, Fc1/2, Fc5/6, Fcz, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T5/6, PO1/2, Oz) with the FCz electrode as reference. Eye movement was monitored using an electrode at the infraorbital ridge of the right eye. The electrode impedances were kept to less than 10 kΩ during the entire experiment. The signal was digitalized at a rate of 250 Hz and filtered online between 0.01 and 70 Hz. All the electrodes were re-referenced offline to the mastoids. While performing the rhythms, participants were instructed to look at the visual representations of the rhythms. Between each sequence, they had 2 s to blink before continuing the task.

The selected epochs were locked to the motor response from −1000 to 1000 ms. The baseline was corrected from −200 to −50ms according to previous studies[9]. Trials exceeding 100 μV were rejected offline. The EEG signal was band-filtered between 0.1 and 30 Hz. Visual inspection and previous literature analysis of similar responses showed the following evoked activity: a negativity between 0 and 50 ms around the frontal region, and a positivity between 90 and 150 ms at central regions[9]. These components corresponded to error negativity and error positivity waveforms, respectively, and have been described in the context of a skill acquisition process, such as rhythm learning. In this situation, learning to synchronize with or reproduce different rhythms does not produce an error or correct response but a continuity from less to more accurate responses. Regarding that, changes in amplitude in single-trial EEG signals locked to the motor responses were analyzed without a classification between correct or erroneous responses and were related to their asynchrony value, repetition, the information content of the note to produce and the visual representations used during the learning.

A linear mixed-effect model was built to predict the mean activity of each time window, with each response as a function of asynchrony, normalized Information Content, repetitions, and visual representations as main effects (all variables were standardized). The random structure consisted of participant and visual representation group as random intercepts, while rhythm, asynchrony, IC and repetitions were introduced as random slopes.

Data availability

The data of this study are available on request.

Code availability

The code used for this study is available on request.

Received: 6 June 2025; Accepted: 11 August 2025;
Published online: 21 August 2025

References

1. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. Active inference: a process theory. *Neural Comput.* **29**, 1–49 (2017).
2. Koelsch, S., Vuust, P. & Friston, K. Predictive processes and the peculiar case of music. *Trends Cogn. Sci.* **23**, 63–77 (2019).
3. Vuust, P. & Witek, M. A. G. Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Front. Psychol.* **5**, 1111 (2014).
4. Pearce, M. T. Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Ann. N. Y. Acad. Sci.* **1423**, 378–395 (2018).
5. Proksch, S., Comstock, D. C., Médé, B., Pabst, A. & Balasubramaniam, R. Motor and predictive processes in auditory beat and rhythm perception. *Front. Hum. Neurosci.* **14**, 578546 (2020).
6. Kung, S.-J., Chen, J. L., Zatorre, R. J. & Penhune, V. B. Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* **25**, 401–420 (2013).
7. Chen, J. L., Penhune, V. B. & Zatorre, R. J. Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* **18**, 2844–2854 (2008).
8. Kasdan, A. V. et al. Identifying a brain network for musical rhythm: a functional neuroimaging meta-analysis and systematic review. *Neurosci. Biobehav. Rev.* https://doi.org/10.1016/j.neubiorev.2022.104588 (2022).
9. Padrão, G., Penhune, V., de Diego-Balaguer, R., Marco-Pallares, J. & Rodriguez-Fornells, A. ERP evidence of adaptive changes in error processing and attentional control during rhythm synchronization learning. *Neuroimage* **100**, 460–470 (2014).
10. Chen, J. L., Penhune, V. B. & Zatorre, R. J. Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* **20**, 226–239 (2008).
11. Teki, S., Grube, M., Kumar, S. & Griffiths, T. D. Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* **31**, 3805–3812 (2011).
12. Fu, Z., Sajad, A., Errington, S. P., Schall, J. D. & Rutishauser, U. Neurophysiological mechanisms of error monitoring in human and non-human primates. *Nat. Rev. Neurosci.* **24**, 153–172 (2023).
13. Vidal, F., Burle, B. & Hasbroucq, T. On the comparison between the Nc/CRN and the Ne/ERN. *Front. Hum. Neurosci.* https://doi.org/10.3389/fnhum.2021.788167 (2022).
14. Orr, J. M. & Carrasco, M. The role of the error positivity in the conscious perception of errors. *J. Neurosci.* **31**, 5891–5892 (2011).
15. Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E. & Donchin, E. A neural system for error detection and compensation. *Psychol. Sci.* **4**, 385–390 (1993).
16. Di Gregorio, F., Maier, M. E. & Steinhauser, M. Errors can elicit an error positivity in the absence of an error negativity: evidence for independent systems of human error monitoring. *Neuroimage* **172**, 427–436 (2018).
17. Steinhauser, M. & Yeung, N. Decision processes in human performance monitoring. *J. Neurosci.* **30**, 15643 (2010).
18. Ullsperger, M., Harsay, H. A., Wessel, J. R. & Ridderinkhof, K. R. Conscious perception of errors and its relation to the anterior insula. *Brain Struct. Funct.* **214**, 629–643 (2010).
19. Peirce, J. W. Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* **2**, 10 (2009).
20. Peirce, J. W. PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
21. Riemenschneider, A. *371 Harmonized Chorales and 69 Chorale Melodies with Figured Bass* (G. Schirmer, 1941).
22. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

Acknowledgements

The present project has been funded by the European Regional Development Fund (ERDF), the Spanish Ministry of Science and Innovation to J.M.-P. (PID2021-126477NB-I00), ICREA Academia program 2018 to J.M.-P., Ministerio de Universidades de España to MDD (FPU18/05977) and the Government of Catalonia (2021 SGR 00352).

Author contributions

M.D. contributed to the design, data collection and analysis, results interpretation and writing the manuscript. J.M. contributed to the design, results interpretations and writing of the manuscript. The final version of the manuscript was approved by both authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41539-025-00353-y.

Correspondence and requests for materials should be addressed to Marc Deosdad-Díez or Josep Marco-Pallarés.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.