Cognition 268 (2026) 106345

Contents lists available at ScienceDirect

Cognition

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

Abstract representations underlie rhythm perception and production: Evidence from a probabilistic model of temporal structure

Fleur L. Bouwer [a,b,c,*,1], Atser Damsma [b,d,e], Thomas M. Kaplan [f], Mohsen Ghorashi Sarvestani [c], Marcus T. Pearce [f,g]

[a] Cognitive Psychology Unit, Institute of Psychology & Leiden Institute for Brain and Cognition, Leiden University, the Netherlands
[b] Amsterdam Brain and Cognition, University of Amsterdam, the Netherlands
[c] Department of Psychology, University of Amsterdam, the Netherlands
[d] Music Cognition Group, Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands
[e] Conservatorium van Amsterdam, Amsterdam University of the Arts, the Netherlands
[f] School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom
[g] Department of Clinical Medicine, Aarhus University, Denmark

ARTICLE INFO

Keywords:
Rhythm
Temporal expectations
Probabilistic modelling
Pattern
Complexity
Tapping
Music

ABSTRACT

Rhythm, such as in music, contains structure in the form of rhythmic patterns: the more or less predictable successions of longer and shorter intervals (i.e., the “morse code” of the rhythm). Listeners can use rhythmic patterns to predict the timing of sounds and guide their perception and action. It is still unclear how rhythmic patterns are represented in the human mind. Here, we used a probabilistic model of auditory expectations to simulate the perception and production of rhythmic patterns. We modelled expectations in rhythmic sequences at three different levels of abstraction: as the predictability of absolute inter-onset intervals (IOI), ratios between successive intervals (ratio), and the direction of change of successive intervals (contour). Subsequently, we selected rhythms that varied maximally in their modelled predictability across the three levels of abstraction for three behavioral tasks: a target detection task in which the rhythm was not task-relevant (implicit task), a complexity rating task (explicit task), and a tapping task (motor task). We found that both ratio and contour affected behavioral responses across all tasks, with the largest effects in the explicit rating task. IOI only affected responses for the explicit and motor tasks, where the rhythm was task-relevant, and to a greater extent when an imprecise, categorical representation of IOI was assumed. These findings suggest that humans rely mostly on imprecise representations of rhythmic patterns, but may flexibly adapt their representation based on task demands.

1. Introduction

To efficiently process our dynamic environment, we use temporal structure in sensory input to guide the timing of perception and action. Temporal structure is especially prominent in rhythmic signals like language and music, and the ability to process it has consequences not just for perception (Nobre & van Ede, 2018), and movement (Damm et al., 2019), but also for the pleasurable and social responses associated with music (Fiveash et al., 2023; Savage et al., 2020). Rhythmic signals (e.g., sequences of events in time) can contain several types of structure that may allow accurate predictions for the timing of events (Bouwer et al., 2021; Nobre & van Ede, 2018). Temporal structure in rhythm can be (quasi)periodic in nature (Doelling et al., 2022), such as a musical beat, and may also be hierarchically structured (Bouwer et al., 2021), such as meter in music (e.g., a march or a waltz). Although it is less commonly investigated, structure can also occur in the surface pattern of the rhythm: the succession of longer and shorter intervals that make up the rhythmic pattern (e.g., the “morse code” of the rhythm) can be predictable, for example through pattern repetition (Bouwer et al., 2020). This rhythmic patterning is the focus of the current work.

* Corresponding author at: Cognitive Psychology Unit, Institute of Psychology & Leiden Institute for Brain and Cognition, Faculty of Social and Behavioural Sciences, Leiden University, Postbus 9555, 2300 RB Leiden, the Netherlands.
E-mail address: f.l.bouwer@fsw.leidenuniv.nl (F.L. Bouwer).

[1] FLB is supported by a VENI grant from the Dutch Research Council NWO (VI.Veni.201G.066). AD is supported by an Amsterdam Brain and Cognition (ABC) Project Grant awarded to Henkjan Honing and Pilou Bazin.

https://doi.org/10.1016/j.cognition.2025.106345
Received 4 April 2025; Received in revised form 12 August 2025; Accepted 9 October 2025

One approach to studying how humans process rhythmic patterns is to examine the motor production of patterns. Serial reaction time (SRT) tasks, in which a motor sequence is implicitly learned, have shown that the presence of predictable temporal intervals improves motor response times (Brandon et al., 2012; Heideman et al., 2018a, 2018b; Nobre & van Ede, 2018; O'Reilly et al., 2008; Schultz et al., 2013; Shin & Ivry, 2002), and that humans can implicitly learn rhythmic patterns of at least eight temporal intervals, even in the absence of other structure like a beat (Schultz et al., 2013). Similar learning of rhythmic patterns has been shown in explicit tapping synchronization studies, even when patterns do not have a clear beat (Bulger et al., 2022; Dean et al., 2021; Matthews et al., 2016; Repp, 2005; Repp & Su, 2013). While in SRT and synchronization tasks, participants respond and synchronize to concurrent stimuli, another approach has been to train participants to perform a motor sequence without a pacing signal (Hardy & Buonomano, 2016). Such paradigms have demonstrated learning to motorically reproduce rhythmic patterns, even when patterns did not contain a metrical beat, and when the temporal structure of a pattern was not accompanied by structure in the identity of stimuli (e.g., ordinal structure) (Cameron & Grahn, 2014; Kornysheva et al., 2013; Kornysheva & Diedrichsen, 2014; Laje et al., 2011; Ullén & Bengtsson, 2003).

The presence of a rhythmic pattern does not just aid motor performance but also perception, as evidenced by increased target detection in the presence of predictable temporal patterns in auditory sequences (Bouwer et al., 2020; Morillon et al., 2016). Humans can also reliably judge whether two rhythmic patterns are the same (Grahn & Brett, 2007), and whether the continuation of a pattern fits with the learned temporal structure (Bouwer et al., 2023). Evidence from EEG research suggests that rhythmic patterns can be learned not just explicitly, but also implicitly, when they are not task-relevant (Bouwer et al., 2020). Thus, to sum up, the temporal structure in rhythmic patterns aids motor and perceptual performance, and is readily learned by humans.

Despite similar behavioral benefits, the mechanisms underlying processing of predictable patterns and regular beats in rhythm have been suggested to be distinct (Aharoni et al., 2024; Bouwer et al., 2023; Breska & Deouell, 2017; Teki et al., 2011). While beat-based processing has been explained as a process of entrainment – the phase alignment of some internal process in a listener with an external rhythmic signal (Doelling et al., 2022; Obleser & Kayser, 2019), pattern-based processing can be explained by the broader framework of predictive coding and predictive processing (Friston, 2005; Koelsch et al., 2019; Vuust et al., 2022). According to the general predictive processing framework, the brain continuously tries to predict incoming sensory information based on past experience, and adjusts its predictions based on prediction error: the discrepancy between incoming information and the predictions made by the brain (Koelsch et al., 2019). Temporal expectations in rhythm in general have been explained within this framework (Vuust & Witek, 2014). To model rhythmic pattern processing in particular, fitting within a predictive processing framework, several computational models have used a probabilistic approach (Cannon, 2021; Kaplan et al., 2022; Milne et al., 2023; Sadakata et al., 2006; Sauvé & Pearce, 2019; Senn, 2023; van der Weij et al., 2017).

These models, while all using some probabilistic method for predicting the timing of the next event, make different assumptions about the underlying cognitive processes involved in perceiving and producing rhythmic patterns. Several models require some other type of structure to be predefined, which is then used to predict rhythmic patterns. For example, models have used a predefined beat, meter, or other template to predict rhythmic patterns fitting such existing structure (Cannon, 2021; Senn, 2023). While it is indeed likely that processing different structural aspects of rhythm (e.g., pattern, beat, meter) is related and interactive, such models have two disadvantages. First, it is unclear in these models how the listener arrives at the perceived beat or template that is used to predict the pattern, with some models using a limited set of manually configured templates, or templates derived from music corpora (Kaplan et al., 2022). Second, these models do not shed light on the processing of rhythmic patterns without a beat or metrical structure, while humans can perceive and reproduce such patterns (Cameron & Grahn, 2014; Schultz et al., 2013).

A possible advantage of models that derive rhythmic patterns from a predefined meter (Senn, 2023) or cyclical template (Cannon, 2021; Cannon & Kaplan, 2024) is that such models represent rhythmic patterns as a (hierarchical) whole. Hence, timing is not computed linearly, on an event-by-event basis, but rather, can be represented as time since the onset of a pattern (Laje et al., 2011). Such a representation of rhythmic patterns allows for events to be skipped, with expectations carried forward to the next event. This account is supported by experiments in the visual domain, in which participants were able to use the temporal structure of a spatiotemporal pattern even when it was interspersed with distractor events (Boettcher et al., 2021). However, a hierarchical representation considering rhythmic patterns as a whole is not consistent with results from a behavioral experiment in the auditory domain, which found that ratings for goodness of fit for pattern-based rhythms reflected only one interval after the end of a rhythmic pattern whereas ratings for beat-based rhythms continued beyond this (Bouwer et al., 2023). This finding suggests distinct mechanisms for beat-based and pattern-based processing of rhythm, the latter involving a linear, non-hierarchical representation, in which a rhythmic pattern is learnt as a succession of temporal intervals.

Several models have indeed simulated rhythmic pattern processing as a linear sequence learning process. In some of these models (Milne et al., 2023; Sauvé & Pearce, 2019), rhythmic patterns are represented as a sequence of absolute inter-onset intervals (*IOIs*). Probabilistic information about absolute IOIs has been shown to contribute to behavior in the form of complexity ratings (Sauvé & Pearce, 2019) and tapping accuracy (Milne et al., 2023). Related to such models of IOI predictability is a model that represents rhythm in an absolute manner, but with less precision: as a succession of categorical "short" and "long" intervals (Ravignani et al., 2018). While it was suggested that this model is neurally plausible, empirical evidence is lacking.

A representation of rhythmic patterns based on absolute IOIs or intervals is difficult to reconcile with temporal scaling: the phenomenon that listeners perceive two rhythms as being the same, despite being played at different speeds (Hardy & Buonomano, 2016). To account for this, some models have represented rhythmic patterns as a succession of *ratios* between IOIs, a relative rather than absolute representation (Gold et al., 2019; Kaplan et al., 2022; Ravignani et al., 2018), and consistent with the notion that some neurons in the primate brain encode time relative to the mean rate of a sequence (Meirhaeghe et al., 2021). At an even more abstract level, human listeners may represent rhythmic patterns in a relative and imprecise manner as rhythmic *contour*, only indexing the direction of change of the interval length (e.g., "same", "shorter", or "longer"; (Schmuckler & Moranis, 2023)).

To summarize, several different probabilistic models have been used to examine rhythmic pattern processing, but modelling of rhythmic pattern processing has generally been conducted in the context of metrical structure (Cannon, 2021; Senn, 2023; van der Weij et al., 2017), other rhythmic features (Bulger et al., 2022; Dean et al., 2021; Milne et al., 2023), and other musical features, like pitch (Gold et al., 2019; Sauvé & Pearce, 2019), making it harder to understand rhythmic pattern processing in isolation. In addition, it is still unclear at which level of representation humans process rhythmic patterns, with different studies using different representations of rhythmic patterns, as a sequence of absolute intervals, ratios between intervals, or directions of change between intervals (longer, shorter, same). On a final note, previous modelling research has simulated a large variety of tasks, ranging from explicit ratings of complexity (Sauvé & Pearce, 2019; Senn, 2023) to tapping along to sequences (Bulger et al., 2022; Dean et al., 2021; Milne et al., 2023), to unpaced reproduction of rhythmic patterns (Kaplan et al., 2022). Different task demands in perceptual and motor tasks may prioritize different cognitive processes in temporal cognition, further complicating comparison between studies (Shalev et al., 2019).

Given the above, here, we examine at what level of representation humans process rhythmic patterns, using stimuli without metrical structure or pitch information, assessed across both perceptual and motor tasks. Motivated by modelling studies investigating human pitch processing (Pearce, 2018) and previous research investigating rhythmic patterns (Gold et al., 2019; Ravignani et al., 2018; Sauvé & Pearce, 2019; Schmuckler & Moranis, 2023), we simulated processing of rhythmic patterns at three different levels of representation, with increasing degrees of abstraction: as sequences of absolute *IOIs* (akin to learning successions of absolute pitch, e.g., "this interval was 800 ms", similar to Sauvé and Pearce (2019)), as sequences of relative *ratios* between IOIs (akin to learning pitch intervals, e.g., "this interval was exactly twice as long as the previous interval", similar to Gold et al. (2019) and Ravignani et al. (2018)), and as sequences denoting *contour* (direction of change) of successive intervals (akin to learning pitch contour, e.g., "this interval was longer than the previous interval", similar to Schmuckler and Moranis (2023)). IOI here is a precise and absolute representation of the pattern, while ratio and contour are relative representations, with ratio being a more precise representation than contour. To prevent any bias in constructing the stimuli, we first generated a large space of all rhythmic patterns matching constraints on length, number of events and temporal resolution, regardless of temporal structure. Subsequently, we computed the predictability of all rhythmic patterns at each of the three levels of abstraction (*IOI, ratio, contour*) using a variable order Markov model, based on IDyOM, an information-theoretic model of auditory expectation (Pearce, 2005, 2018). Using the modelled predictability, we selected a set of stimuli for the experiment that maximally differentiated between the three levels of abstraction.

We then examined behavioral responses to these stimuli in three tasks. First, participants completed an implicit target detection task, in which we measured their speeded response to infrequent softer sounds embedded in the rhythmic patterns (Bouwer et al., 2020). Second, participants gave complexity ratings for all patterns, in an explicit perception task (Sauvé & Pearce, 2019; Senn, 2023). Finally, all participants performed a tapping task, tapping along to the rhythmic patterns (Bulger et al., 2022). Based on previous research, we expected that the modelled predictability for *IOI* (Sauvé & Pearce, 2019) and *ratio* (Gold et al., 2019) would affect task performance, with better target detection, lower complexity ratings, and higher tapping accuracy for more predictable patterns. We also explored whether the reliance on a type of representation differed between tasks, whether modelling IOI in an imprecise manner, akin to contour, would improve the model predictions, and, finally, whether the same model features used to predict performance for whole rhythmic patterns could also predict performance in real time, as the rhythm unfolds.

2. Materials and methods

2.1. Participants

45 participants (30 women), aged between 18 and 70 years (M = 24, SD = 9) completed the full experiment, one additional participant dropped out halfway through the session. None of the participants reported a history of neurological or hearing disorders. All participants provided written consent prior to the study and received either monetary reimbursement or study credit for their participation. The study was approved by the Ethics Review Board of the Faculty of Social and Behavioral Sciences of the University of Amsterdam. After exclusion of 5 participants with inadequate responses (see section 2.5.3), the remaining 40 participants (13 male, 27 female) were on average 23.6 years old (SD = 8.9). On average, the participants scored 25.2 (SD = 12.5) on the Gold-MSI musical training subscale (Müllensiefen et al., 2014), ranging from 7 (the minimum score on the scale) to 49 (the maximum score on the scale). This is comparable to the norm scores.

In the absence of directly comparable studies providing effect sizes for a full power analysis, we looked at previous studies with similar designs to estimate our sample size. One study used a probabilistic model of pitch perception, combined with one rhythm feature, to predict musical preferences (Gold et al., 2019). While no effect sizes are reported, responses were predicted well by the model with data from 43 participants, and 44 trials per participant. Another study used a probabilistic model including pitch and harmony features, along with one rhythm feature, to successfully predict behavioral data from 28 participants, with 48 trials each. Hence, with 40 participants, and 56 rhythmic sequences, we expect to have sufficient power to detect whether the model can be used to predict rhythmic behavior.

2.2. Stimuli

2.2.1. Creating a rhythm space

To create unbiased rhythms agnostic of metrical structure and predictability, we first generated all possible rhythmic patterns within three constraints: 1) Rhythms were 4 s long, which allows for pattern repetition without inducing a beat, as the repetition rate of 0.25 Hz is outside the range of human beat perception (London, 2012), and is short enough to allow (implicit) learning of the pattern (Schultz et al., 2013); 2) Rhythms consisted of 8 sound events, for an average event rate of 2 Hz, consistent with the average event rate of music (Ding et al., 2017), previous studies examining processing of rhythmic patterns (Cameron & Grahn, 2014; Chen et al., 2008; Grahn & Schuit, 2012), and preferred tapping rate for humans (Drake et al., 2000; Honing & Bouwer, 2019; London, 2012); 3) Inter-onset intervals (IOIs) were between 200 and 800 ms, with a 50 ms spacing (so 200, 250, 300, 350 etc.). The lower limit was chosen to account for restrictions in how fast people can tap, which has a lower limit of 150–200 ms between taps (Repp, 2005). The upper limit of 800 ms was chosen to avoid very uneven rhythms in terms of event density. Within these three constraints (4 s patterns with 8 events and IOIs between 200 and 800 ms), a total of 30,162,301 rhythmic patterns could be generated. Since we repeated the rhythmic patterns in streams during the experiments, rhythmic patterns with the same sequence of intervals starting at a different position would be perceptually identical. Therefore, next, we pruned the large set by removing circularly identical rhythms. We always kept the pattern with the longest interval last, given a suggestion that these are perceived as most stable (Jacoby et al., 2021). This resulted in a set of 3,770,475 rhythms. Subsequently, the final selection of stimuli used for the behavioral experiments was achieved using the modelled predictability of each rhythm.

2.2.2. Three levels of representation: IOI, ratio, and contour

As input for the probabilistic model, we used three representations of each rhythmic pattern (see Fig. 1A): *IOI, Ratio,* and *Contour*. For absolute inter-onset intervals (IOIs), there were 13 possible values. For ratios between successive IOIs, initially, the number of possible values was >90. Many rhythms contained ratios occurring exactly once, leading to most rhythms being at ceiling in terms of unpredictability as computed by the model. Also, many ratios were based on large integers (>10) but irreducible to smaller integer ratios (e.g., a ratio of 650/450 would be equivalent to a 13/9 ratio in integers; similarly, ratios of 16/15, 16/13, 13/11 etc. occurred in the dataset), while perceptually, listeners prefer small integer-ratio rhythms (Jacoby et al., 2021; Jacoby & McDermott, 2017), and may perceive large-integer ratios in rhythm as belonging to the nearest small integer-ratio category (Desain & Honing, 2003; Ravignani et al., 2018). Therefore, we rounded each ratio to the nearest small-integer ratio including integers 1 to 4 (e.g., 13/9 was rounded to 3/2), leaving 11 possible ratio values.

The third and most abstract rhythmic representation, contour, is comparable to melodic contour (i.e., whether the pitch of a melody goes up or down or stays the same). In rhythm, contour denotes whether the duration of an interval is longer, shorter, or the same as the preceding interval, and contour therefore only takes one of three values

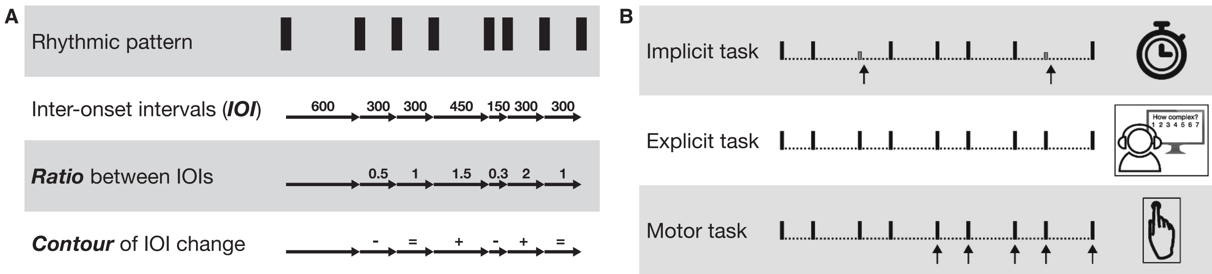

Fig. 1. Schematic overview of model features and behavioral tasks. A) Information content and entropy were computed for three features (viewpoints): Inter-onset interval (*IOI*), its derivative *ratio*, and as a further abstraction *contour*. Ratios were rounded to integer ratios based on integers 1 to 4. Contour was derived from ratio, with all ratios >1 having a contour value of "longer", ratios equal to 1 having a contour value of "equal", and ratios <1 having a contour value of "shorter". B) Participants performed three behavioral tasks. In the implicit task, they were asked to detect infrequent softer targets in streams of rhythmic patterns, and both hit rates and reaction times were recorded. In the explicit task, after listening to a rhythmic pattern, participants indicated how complex they thought the pattern was. In the motor task, participants were asked to tap along to every tone in the pattern, starting after two repetitions of each pattern.

(Schmuckler & Moranis, 2023). However, similar to how we quantized raw ratios to account for perceptual categorization, we adjusted the contour representation to account for perceptual acuity. In duration discrimination tasks, single intervals are, on average, perceived as different if their lengths differ more than 20 to 25 % (Dalla Bella et al., 2017). In sequences, the threshold for detecting length differences is around 10 % of a repeating interval (Dalla Bella et al., 2017). To account for this, here, we derived contour from quantized ratio. That is, any ratio < 1 was coded as "shorter" for contour, ratios >1 were coded as "longer", and ratios of 1 as "equal". Given how we computed the ratio representation, raw ratios between 0.875 and 1.167 were thus coded as equal for contour, akin to a difference in interval length of <12 % being assumed to be perceptually indetectable, in line with previous research (Dalla Bella et al., 2017).

2.2.3. *Modelling predictability as a probabilistic process using a PPM model*

We used a probabilistic model based on the IDyOM (Information Dynamics of Music) model, which uses the Prediction by Partial Matching (PPM) algorithm to generate probabilistic predictions (Pearce, 2005, 2018). PPM is a statistical learning model that progresses through symbolic sequences (e.g., "ABACADABRA"), to generate, at each event, a predictive distribution of the most likely symbol to be encountered as the next event. PPM works with symbolic data, so each event will be of a given type that is drawn from a finite alphabet of possible symbols (e.g., a set of "A", "B", "C", "D", and "R"). The predictive distribution generated by PPM is based on several n-gram models of different order (or context length, n) that are combined, where the n-gram represents a subset of adjacent symbols (e.g., "AB" is a 2-gram, and "CADA" is a 4-gram). For each n-gram model, the probability distribution of the next event is computed from all occurrences of the n-gram (e.g., after observing "ABACADABRA", for an order-2 model upon encountering "AB" again, the probability of the next symbol being "A" is 0.5, the probability of the next symbol being "R" is also 0.5, and the probability of the other symbols 0). PPM combines different n-gram models to arrive at a single probability distribution at each event with all possible continuations having a non-zero probability of occurrence. Combining models of different order allows the model to benefit from both the structural specificity of high-order models and the statistical robustness of low-order models. There are multiple ways of combining the different n-gram models into one distribution. Here, we used interpolated smoothing with escape method C, as in the original implementation of IDyOM (Pearce, 2005). Since we did not aim to examine enculturation, nor long-term learning, we did not include prior expectations in the model, rather, the model started afresh for each rhythmic sequence. The two information-theoretic properties used to quantify the predictability of the rhythms were information content (IC) and entropy (Shannon, 1948). Information content is the negative logarithm, base 2, of the probability of the timing of a sound event, given the conditional probability distribution at that point in time. IC values thus reflect the unpredictability of an event, with higher values being associated with more surprising events. Entropy of the probability distribution at a given point in time reflects the uncertainty of the prediction prior to the event occurring.

2.2.4. *Quantifying predictability of the rhythmic patterns*

To quantify predictability of each rhythmic pattern, we computed IC and entropy using the PPM algorithm for each event in each rhythm and for each representation (i.e., IOI, ratio, and contour) separately (see Fig. 1). We used a model of order 4, that is, for each event, a probability distribution was calculated based on smoothing the distributions from models that considered n-grams of 1, 2, 3, 4 and 5 symbols. An order of 4 was found to be appropriate for musical applications of PPM (Harrison et al., 2020), and it is known that people can learn temporal sequences of at least this length (Bouwer et al., 2020; Schultz et al., 2013).

We ran the PPM model on two repetitions of each rhythm (3,770,475 in total) concatenated to itself, followed by a final event to demark the end of the second pattern (e.g., 17 events, to demark 16 temporal intervals). However, both the model and humans need about one cycle to detect the predictability of a rhythmic pattern. After two events, there will be only one interval, so the model will not yet be possible to compute a representation of the derived representations of ratio or contour. Only at event three will the first ratio or contour be defined by the model. But, at that point the prediction will be a uniform distribution since no learning has yet taken place. Also, behaviorally, listeners typically need around one cycle to start synchronizing to short rhythmic patterns (Jacoby et al., 2024). Hence, we discarded the IC and entropy values for the first iteration of each rhythm, and only used IC and entropy values for the second iteration (this entails values associated with events 10 to 17, as an interval is only apparent at the second event marking it). Given that the explicit rating task was based on responses to entire rhythms rather than individual events, we averaged the IC and entropy values for each rhythm, following previous research (Gold et al., 2019; Sauvé & Pearce, 2019). Averaged values were used for both stimulus selection and the main analyses. As PPM is ideally suited to also simulate expectations in real time – on an event-by-event basis, we performed exploratory analyses examining real-time changes in predictability (see below).

2.2.5. *Stimulus selection*

Fig. 2 shows the steps taken in selecting the stimuli for the experiment, as well as the average computed predictability for all rhythms on all features (*IOI, ratio, contour*) using IC as an information-theoretic measure of predictability. For completeness, Supplementary Fig. 1 shows the same statistics for both IC and entropy. After removal of circular duplicates, we first removed rhythms with large variation in predictability values (e.g., IC values) between events. For example, a

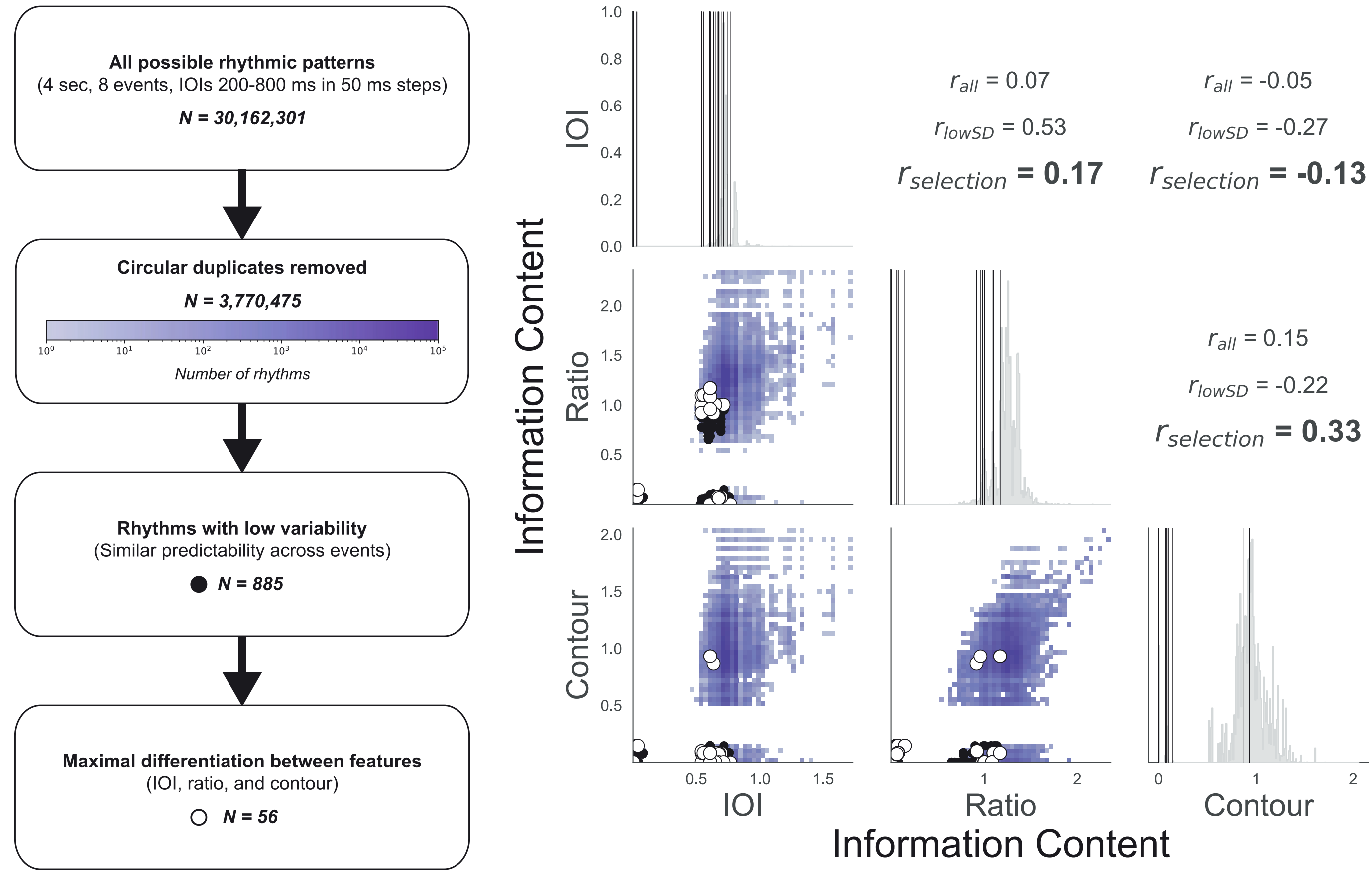

Fig. 2. Computed information content for all features and all rhythms and schematic depiction of the selection process. The information content for the full set of rhythms is shown in purple shades. Information content is averaged over all 8 events in each rhythmic pattern (using the second iteration, see text). Note that the coloring is on a logscale. From the full set of rhythms (~30 million), we first removed circular duplicates. From the remaining rhythms (~3.7 million), only rhythms with limited variability were considered for the experiment. These rhythms ($N = 885$) are shown as black dots. For all these rhythms the standard deviation of the IC and entropy values across the 8 events was below a threshold of one SD below the mean of the whole rhythm set. Finally, a set of 56 rhythms was selected for the experiment to maximally differentiate between the three features of interest. These rhythms are depicted as white dots. For all three sets (full set, low SD set, selected set), the Spearman rank correlations between the different features are given in the upper half of the figure. Distributions for IC values on the diagonal are for the full set of rhythms; the selected rhythms for the experiment are overlaid as vertical black lines to indicate where in the distribution they fall. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

rhythm may be largely very predictable, but with one deviation (e.g., IOIs of 450-450-450-450-450-450-450-850 would yield very low IC values for the first 7 events, and an extremely high IC value for the last event). For such a rhythm, the average IC and entropy values would in practice be meaningless, as the rhythm as a whole is not predictable or unpredictable, but rather, separate events are either very predictable or very unpredictable. Hence, we only considered rhythms with a standard deviation of IC values lower than the threshold of one standard deviation below the mean (of the distribution of standard deviations). This left 885 rhythms to consider (see Fig. 2, black dots), with a low standard deviation on all features. As is apparent from Fig. 2, the rhythms with the greatest unpredictability (e.g., the highest surprisal or IC values) were eliminated during this step of the stimulus selection.

Next, we selected rhythms for the experiment that had a maximally different modelled predictability for the three different features (see Fig. 2). We treated the modelled predictability as a six-dimensional space, with the position of each rhythm defined by six values (e.g., IC values for IOI, ratio, and contour, and entropy values for IOI, ratio and contour). All values were normalized between 0 and 1. We subsequently selected rhythms that had the smallest Euclidian distance to points in space that represented the extremes in terms of predictability (e.g., [0 0 0 0 0 0] would represent the most predictable rhythm on all features, while [0 1 0 0 1 0] would be the point in space where a rhythm was unpredictable based on ratio, but not the other features). For simplicity, we harmonized the two information-theoretic measures by picking rhythms that were predictable or unpredictable based on both (e.g., we selected rhythms close to [0 1 0 0 1 0], but not [0 1 0 0 0 0]). Thus, in our final sample of rhythms, entropy and IC are highly correlated, as we did not aim to differentiate between these two measures of uncertainty and surprise. In total, we selected 56 rhythms (Fig. 2, white dots) closest to eight points in the space (7 per category), representing combinations of high and low predictability for all three features (i.e., the stimulus categories were high/low predictability crossed with IOI/ratio/contour). Fig. 3 shows an example of one rhythm for each of the eight categories, with the modelled IC values in real time.

Note that some regions of the six-dimensional space were not inhabited (see Fig. 2). For example, it is difficult to find rhythms that are highly predictable in terms of their IOI, but unpredictable in terms of their contour. This is not surprising given that contour is derived from IOI. We did not adjust for this, and simply selected rhythms closest to the preferred point in the space. For some categories of rhythms, like those that are highly unpredictable on all features, this meant that the selected rhythms were close to the ideal point in space, while for others, like the aforementioned rhythms with unpredictable contour but predictable IOI, rhythms were further removed from the ideal point in space. As we are not using a factorial design but the actual predictability values in a mixed-effects regression model, we do not regard this as problematic. All included rhythms can be found in Supplementary Table 1.

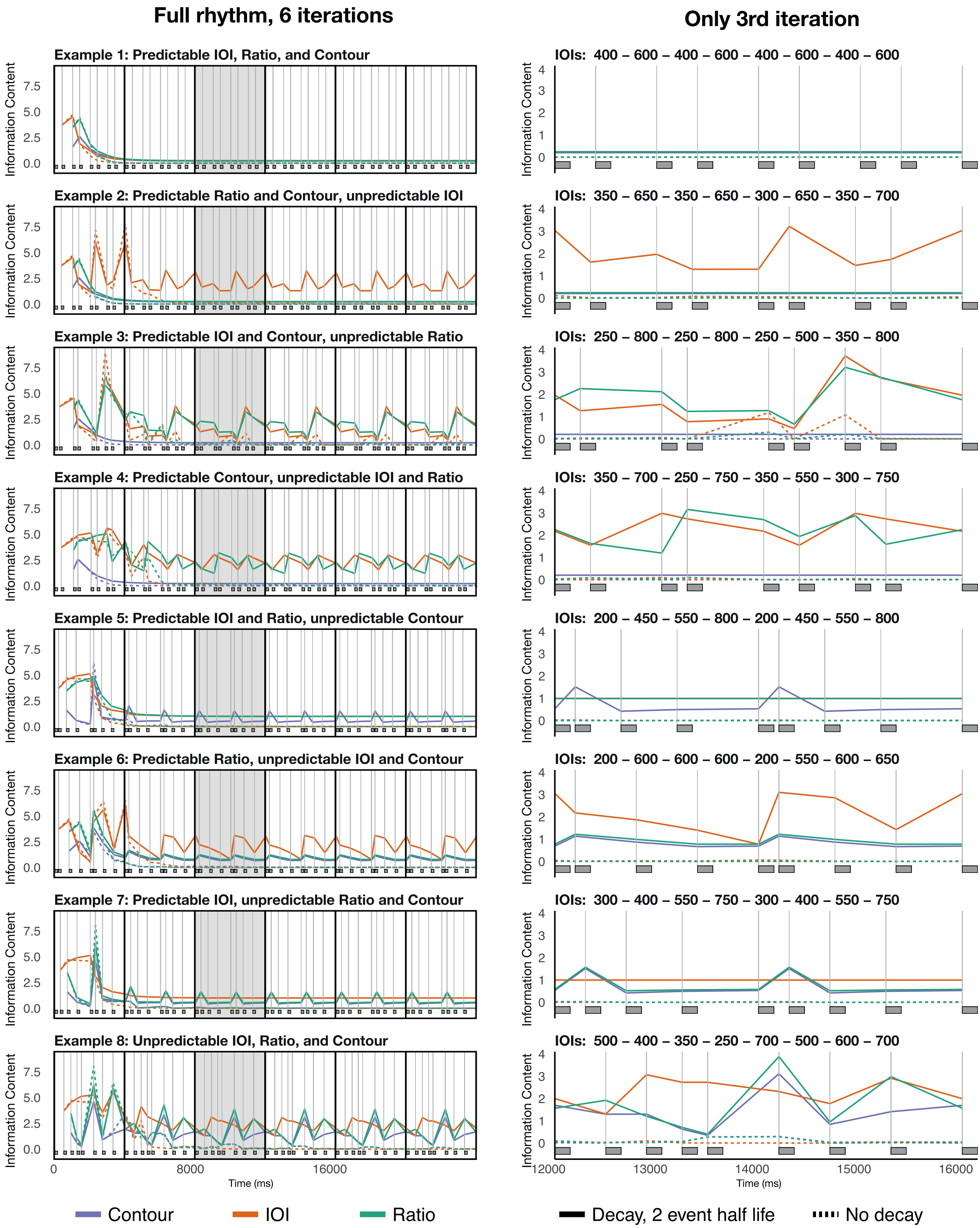

Fig. 3. Examples of rhythms selected for the experiment. Rhythms were selected to conform as much as possible to the eight categories derived from crossing high/low predictability with IOI/ratio/contour. Due to the features being related to each other, the high and low points are relative, but not absolute (see under Stimulus selection). Left: modelled IC for a full 6 iterations of the rhythms. Right: only the 3rd iteration, equivalent to the greyed area on the left.

2.2.6. Sounded rhythmic patterns

For all tasks, rhythms were concatenated into streams of multiple repetitions of the same four-second pattern. Sounds were 440 Hz sine tones, with 5 ms rise and fall time. Each stream ended with a final tone to demark the final interval.

2.3. Tasks

2.3.1. Implicit perception

To measure the implicit perception of rhythmic structure, we used a deviant detection task, as has previously been used both in behavioral (Bouwer et al., 2020; Bouwer & Honing, 2015) and EEG research (Bouwer et al., 2016; Bouwer et al., 2020; Bouwer et al., 2024; Geiser et al., 2009; Vuust et al., 2009). Participants were presented with a stream of 10 concatenated repeating rhythmic patterns, and were asked to detect occasional softer tones, which were softer than the standard sounds by 6 dB. The assumption here is that if participants can predict sound timing, they should detect the deviants faster and better, as was found previously (Bouwer et al., 2020). For each rhythm, each event was presented as a deviant once, amounting to 8 deviants per rhythm. The placement of the deviants was randomized, with two constraints. First, deviants could only occur between the fourth event in iteration 2 and the fourth event in iteration 9, to make sure that a deviant could only occur after the participant had already encountered the whole rhythmic pattern once. Second, to allow participants to make responses, deviants occurred with a minimum of 3 standard tones and a minimum of 2500 ms in between. Additional deviants were added to the beginning of all sequences (e.g., somewhere in the first iteration), to avoid expectations for a deviant building up towards the beginning of the second iteration, and to the end of half the sequences (iteration 10), to avoid expectations for the number of targets. Responses for these deviants were discarded, leaving one target for each event in each rhythm. Each rhythm was presented once, for a total of 56 trials. Rhythms appeared in random order, in blocks of 8 trials, with a one second inter-trial interval and longer breaks in between blocks. Participants responded by pressing the space bar as fast as possible upon hearing a target.

2.3.2. Explicit perception

In the explicit perception task, participants listened to a stream of 4 concatenated rhythmic patterns. After being presented with the rhythm, they were asked to answer three questions, which appeared on the screen in random order: “How complex do you think this rhythm was?”, “How much did this rhythm make you want to move?”, and “How much did you like listening to this rhythm”. The latter two questions were for exploratory analyses into groove and liking (Witek et al., 2014) that are not reported here. For the purposes of the current study, we focus on the question about complexity, similar to previous research using IDyOM as a model to link predictability to complexity (Sauvé & Pearce, 2019). Participants answered the questions on a 7-point Likert scale, which for the complexity measure ranged from “extremely simple” to “extremely complex”. As for the implicit task, the rhythms were all presented exactly once, in 7 blocks of 8 trials.

2.3.3. Motor task

In addition to the two perceptual tasks, participants completed a motor tapping task. Here, they were presented with streams of 6 repetitions of each rhythm. After 2 repetitions, the word “TAP” appeared in the screen, indicating that we expected them to start tapping along to the rhythmic pattern. If they felt comfortable at an earlier time, they were allowed to start tapping earlier. Participants were specifically instructed to tap the rhythmic pattern itself, and not the beat. Moreover, we instructed them to try and follow the pattern as closely as possible, so if the pattern was slightly irregular, their tapping should follow suit. As for the perceptual tasks, each rhythm was presented exactly once, in 7 blocks of 8 trials.

2.4. Procedure

After providing informed consent, participants first proceeded with the implicit task, to avoid participants taking their experiences from the other tasks into the implicit measure. The order of the subsequent explicit and motor tasks was randomized over participants. For each task, participants were given a few practice trials to familiarize them with the procedure. After finishing all tasks, participants were asked to fill out the musical training subscale from the Gold-MSI (Müllensiefen et al., 2014), to assess their musical expertise. Participants were allowed breaks between tasks. An entire experimental session lasted for about 2 h.

Participants were tested individually in a dedicated lab at the University of Amsterdam. Sounds were presented at a comfortable hearing level with headphones, using custom scripts in Psychopy (Peirce et al., 2019). For the tapping task, we used the setup as described in Experiment 2 in Anglada-tort et al. (2021), with a microphone on the table connected to one input channel of a Solid State Logic SSL 2 USB soundcard, one output channel looped back into a second input channel, and one output channel connected to headphones for the participants. Thus, we recorded simultaneously the taps and the sound into one audio file for further analysis, for extremely precise temporal resolution.

2.5. Data preprocessing

2.5.1. Implicit task

For the implicit task, any response within 150 to 1500 ms after a target tone was considered a hit. For the initial mixed-effects model, two dependent variables, *reaction times* and *hit rates* were averaged across events in each rhythm. Note that first, for hit rates, the possible values for average hit rates were limited to integers divided by 8 – the number of targets (e.g., 0, 0.125, 0.25 etc.). Second, participants may have failed to respond to events with high unpredictability, biasing reaction time data for predictable events. This may have diminished the overall power of the reaction time analysis by eliminating data selectively for highly unpredictable events, making the data more homogeneous. To address this issue, we examined the responses in real time, taking the unpredictability of each single event into account (see below).

2.5.2. Motor task

To extract taps from the audio signal obtained from the motor task, we searched the audio signal for amplitude peaks. Participants varied in the strength of their tapping, with large differences in SNR of the tapping to the background noise not only between, but also within participants, sometimes even within a trial. To account for this, we used a peak detection algorithm with a dynamically adjusted threshold. The threshold was adjusted to arrive at exactly 24 taps over the last 3 iterations of the rhythm. If no threshold could be found that yielded 24 taps, the trial was considered incorrect, yielding a per rhythm score for *tapping correctness*. We then assigned each tap to the corresponding sound and computed the asynchrony in ms (e.g., the time between the tap and the sound).

From the raw asynchronies, we computed our main dependent variable of interest: the *tapping variability*, which was quantified as the standard deviation of all 24 asynchronies in one trial. This measure accommodates systematically consistent early (or late) tapping by a participant, since the standard deviation of asynchronies in a pattern that is perfectly replicated, but consistently tapped early (or late) would be (close to) 0, while a pattern that is reproduced with an average asynchrony of 0, but inconsistently, with some taps preceding and some taps following the sounds, would have a higher standard deviation of asynchronies. Note that this measure of tapping asynchronies did not account for differences in absolute IOI, which according to the Weber fraction, should lead to deviations from the tapped interval that are proportional to the absolute IOI (e.g., for a 200 ms interval, a deviation in tapping of 50 ms would be equivalent to a deviation of 200 ms from an 800 ms interval). We chose to use a measure based on raw asynchronies – not fractions, as an analysis of the tapped intervals showed that the Weber fraction did not hold in our data (see Supplementary Fig. 2).

On some trials, subjects did tap 24 times but added extra taps to one part of the trial while omitting taps in another part of the trial (e.g., the correct number of taps, but not correctly assigned to the tones). To account for these trials, we classified trials with >2 taps with an asynchrony in the 95th percentile or higher, or with a tapping variability that exceeded the 90th percentile as incorrect. We checked the outcomes of these procedures visually and corrected a few trials in which participants exhibited a very large negative asynchrony, but tapped the pattern correctly (e.g., given a very high raw asynchrony, these trials were initially classified as incorrect, but given that the pattern was tapped correctly – albeit early – these trials were subsequently manually corrected to be included in the analysis).

For the analysis in real time (see below), we computed tapping deviation as the deviation in ms from the intended interval. Hence, tapping deviation was computed by taking the absolute difference between the tapped interval (e.g., the time of the current tap minus the time of the previous tap) and the intended interval (e.g., the IOI preceding the current tap). Like for tapping variability, this measure accounted for participants having a stable large asynchrony (i.e., a consistent phase offset), whilst tapping the pattern correctly, as the interval instead of the raw asynchrony was used.

2.5.3. Participant exclusion

We excluded participants based on three heuristics: First, participants ($N = 3$) with a hit rate in the implicit task of <50 % were excluded. Second, we excluded participants ($N = 2$) with on average > 3 false alarms on each trial in the implicit task. Finally, we excluded participants (N = 3) with <50 % correctly tapped trials. As some participants were excluded based on multiple measures, the total number of excluded participants was 5, leaving 40 complete datasets for the analysis.

2.6. Statistical analyses

2.6.1. Main mixed-effect regression analyses

For all three tasks we modelled the average responses for each rhythm using mixed-effects models, with the dependent variables *reaction time*, *hit rate* (implicit task), *complexity rating* (explicit task), *tapping correctness* and *tapping variability* (motor task). For each model, we included three independent variables of interest: IC based on *IOI*, *Ratio*, and *Contour*. We report all the models based on the IC values, as these are a direct index of predictability. In addition, we included a fixed factor *Evenness,* computed as the standard deviation of the IOIs in each rhythm. At the level of the whole rhythm, *Evenness* is an index of the variability of the rhythm, and directly related to how much intervals deviate from the mean of 500 ms. The deviation from the mean interval has previously been used to model human rhythmic behavior in a probabilistic way, but at the level of single intervals (Doelling et al., 2023), as opposed to at the level of patterns, as in the current work. We included *Evenness* in our models to account for such single interval predictions, and compared the full models, including the IC predictors, to a baseline model with only *Evenness*, to test whether our sequence learning predictors explained additional variance in behavior. Models additionally included random intercepts for each participant. Random slopes were omitted, as random slope models yielded singular fits. All models were run in *R* (R Development Core Team, 2008). Table 1 shows the correlations between all fixed factors in the models (note that here, as the models were linear models, we used Pearson correlations, while Fig. 3, accounting for the non-normality of the data, shows Spearman correlations). To check whether the correlations posed a problem for the mixed-effects models, we computed the variance inflation factors (VIFs), using the *check_collinearity()* function from the *performance* (Lüdecke et al., 2021) and *easystats* (Lüdecke et al., 2022) packages. For all variables, the VIFs were < 2 indicating no issues with collinearity for the mixed-effects models.

Table 1
Correlations between fixed factors in the main linear mixed models, using the mean modelled IC values for each rhythm as a proxy for predictability.

	IOI	Ratio	Contour
Ratio	0.488		
Contour	0.090	0.311	
Evenness	−0.161	0.232	0.020

Note: Pearson's r is shown.

To account for the nature of the responses, we used different types of models for each task and response type. For reaction times and tapping variability, we used a mixed linear model, as per the *lmer()* function in the *lme4* package (Bates et al., 2015). For complexity ratings, as these were on a Likert scale, we used an ordinal mixed model as per the *clmm()* function in the ordinal package (Christensen, 2019). Finally, for the hit rates and tapping correctness, we used a generalized linear mixed model with a binomial link as per the *glmm()* function from the lme4 package. Statistical significance of the factors in the models was assessed using the *Anova()* and *Anova.clmm()* functions from the *car* (Fox & Weisberg, 2019) and *RVAideMemoire* (Hervé, 2018) packages respectively. All models were compared to the baseline models using likelihood ratio tests as per the inbuilt R function *anova()*.

2.6.2. Control mixed-effects regression analyses

To test the robustness of our results, we ran three control analyses. First, we ran all models again using entropy values as predictors. Because of our stimulus selection method, IC and entropy values were highly related in our stimuli (see Supplementary Fig. 1). Thus, we did not include both information theoretic measures (e.g., IC and entropy) into each model, but we did check whether the results would remain the same when using entropy instead of IC values as predictors. Second, as we were interested in rhythmic patterns, but not necessarily metrical structure, we completely disregard metricality of the patterns throughout this paper. However, especially the easiest patterns (e.g., the isochronous pattern and patterns with repeating alternating intervals, patterns 1–7 in Supplementary Table 1) could be interpreted in a metrical framework. To check that our results were not driven by the highly predictable, but also possibly metrical, patterns, we repeated the analyses while removing the patterns that were selected based on having lowest unpredictability on all features (e.g., low IC values on IOI, ratio, as well as contour). Finally, as can be seen in Fig. 2, the 56 patterns selected for the experiment did not cover the whole space of possible values for each feature of interest (i.e., the distribution of IC values is somewhat bimodal for all features). To account for the fact that this may hamper the linear interpretation of the models, we repeated the analyses using a dummy coding for each fixed factor, with two categories: high and low IC. For all features, we used a value of 0.4 as a cutoff between the categories, as this fell between the peaks of the bimodal distributions (see Fig. 2). All models and results from the control analyses are reported in the Supplementary Information.

2.6.3. Comparison across tasks

We also aimed to test whether for different task demands, listeners rely on different features, which requires comparing the effects of the three predictors across the different behavioral tasks. Due to the nature of the responses, the coefficients from the separate models cannot be directly compared, as they are either beta weights, for the linear mixed models, or log odds ratios, for the ordinal and logistic models. To be able to compare the tasks directly, we therefore normalized the continuous data for the explicit task (complexity ratings), for the implicit task (RTs) and for the tapping task (tapping variability). Subsequently, we ran a linear mixed-effects model including these data, with the modelled IC values for *IOI, Ratio,* and *Contour* as fixed factors, as well as a factor *Task* (explicit, implicit, or tapping) and the interactions between *Task* and the other three fixed factors. In addition, we included a random intercept for participant and a fixed factor *Evenness*. As for the main regression analyses, we used the *lmer()* function in the *lme4* package (Bates et al., 2015) combined with the *Anova()* function from the *car* package (Fox & Weisberg, 2019). Significant interactions were pursued using the *emtrend()* function to compare slopes from the *emmeans* package (Lenth, 2019).

2.6.4. Categorizing IOI values

We based the *IOI* feature on the absolute IOIs, as was done in previous research (Sauvé & Pearce, 2019). Given that we constrained the rhythm space to use only 11 possible IOIs, this automatically yielded a somewhat discontinuous, categorical space of IOIs. This approach is compatible with the idea that humans perceive temporal intervals in a categorical way (Ravignani et al., 2018). While we based the stimulus selection on all 11 IOI values for the main analyses, which replicates previous research, we also explored whether IOI, like ratio, may be represented in a more imprecise way. Therefore, we compared the main regression model (see 2.6.1, with IOI based on 11 categories of possible intervals) with a model in which the IOI feature was computed based on only 5 categories of intervals. Given that no previous research to our knowledge has tested how listeners perceptually categorize IOIs in rhythmic patterns, we simply divided IOIs in 5 categories based on the categories having roughly equal numbers of IOIs, and the distribution of the IOIs retaining its original shape. Hence, IOIs of 200 and 250 ms were recoded as “short”, IOIs of 300, 350, and 400 ms were recoded as “medium short”, 450, 500, and 550 ms as “medium”, 600, 650, and 700 ms as “medium long” and 750 and 800 ms as “long” (see Supplementary Fig. 2). The original model (i.e., with predictability based on raw IOI values) was compared to the imprecise model (i.e., with predictability based on 5 categories of IOI values) using AIC, given the non-nested nature of these models. We used the heuristic that two models with an AIC difference < 2 are considered equivalent (Burnham & Anderson, 2004). VIF values as computed using the *performance* package (see 2.6.1) showed no collinearity issues (all VIFs <1.6).

2.6.5. Exploring predictability in real time

To assess whether the predictors were not just useful to characterize an entire rhythm as predictable or unpredictable, but also carried meaningful information in real time, we conducted additional exploratory analyses modelling responses as the rhythm unfolds. For these analyses, we included the dependent variables that were collected on an event-by-event basis: hits (either hit or miss for each target), reaction times to each correctly identified target, and tapping deviation from the intended interval. For tapping data, all taps starting with the second tap of iteration 3 of the rhythm were included, provided the trial was deemed a correctly tapped trial (see section 2.5.2). As the main fixed factors of interest in these analyses, we used the three predictors from the PPM model: *IOI, Ratio,* and *Contour*.

To account for memory constraints and the fact that humans do not always act like ideal observers when learning sequences (Bianco et al., 2020; Harrison et al., 2020), for each predictor, values in real time were recomputed using the *PPM-decay* package in *R*, using interpolated smoothing and escape method A, as implemented in the package (Harrison et al., 2020). Using the *PPM-decay* package, we included a decay parameter in the model that decreased the weighting of the learned n-grams as a function of time elapsed since the observation of that n-gram. Previous research has found that for melody, humans typically take into account 5–10 previous events in making predictions about the next event (Kern et al., 2022), and it has been suggested that time intervals are represented as discrete items in working memory (Fan & Yotsumoto, 2018). Hence, we included decay kernels with a half-life parameter of between 1 and 10 events, to search for the best fitting model.

We included three additional fixed factors in the event-by-event mixed-effect models to account for low-level acoustic processing. First, we included the absolute IOI of the interval preceding the response (*Previous IOI*). This accounted for differences in responses based on the length of the interval (i.e., though the Weber fraction did not hold, see Supplementary Fig. 2, we did observe differences in responses based on absolute IOI), for example due to masking of the subsequent stimulus, or motor constraints. Second, we included a fixed factor coding for the deviation from the average IOI (*Deviation from mean IOI*, e.g., an interval of 200 ms was a deviation of 300 ms from the average IOI, and an interval of 800 ms had the same deviation). This accounted for the effects of learning the average interval, similar to the *Evenness* parameter used at the level of a whole rhythm and related to an existing Bayesian model of processing of quasi-isochronous sequences (Doelling et al., 2023). Finally, we added a fixed factor for the order of appearance of the sound, as we expected that participants may get progressively better at tapping, and possibly worse at deviant detection, due to lapses in attention, throughout the trial (*Sound number*).

All fixed factors (*IOI, Ratio, Contour, Previous IOI, Deviation from mean IOI, Sound number*) were scaled to have a minimum of 0 and a maximum of 1. As for the models of whole rhythms, random slopes were omitted, as some of the models yielded a singular fit with random slopes added. Models included random intercepts for participant as well as random intercepts for each rhythm, accounting for variation in difficulty for each rhythm not accounted for by the PPM model. Models including the predictors of interest with varying decay parameters (10 models in total) were compared to a baseline model with only the three low-level predictors (*Previous IOI, Deviation from mean IOI, Sound number*) and to each other, to assess the best fitting decay parameter.

Finally, we compared the models with real-time predictors as well as the baseline model to models with predictors indexing rhythmic predictability at the level of the whole rhythm. For these models, the mean values for each rhythm as computed for the main regression analyses (see section 2.6.1), representing the average, overall predictability of each rhythm, were assigned to all events in each rhythm. Thus, computationally, these models were identical to the main regression analyses at the rhythm level, but ran at the level of single events, which allowed for the inclusion of the single event acoustic control predictors, and for direct comparison with the models run in real time. Hence, this allowed for assessing whether the rhythmic predictability did not just affect the overall performance, but also performance fluctuations within a rhythm, in real time.

To summarize, for each dependent variable (hits, reaction times, tapping deviations), we compared a total of 12 models: a baseline model with only low-level predictors, a model with the average rhythmic predictability as quantified by the mean IC values for *IOI, Ratio,* and *Contour* across events, and 10 models with varying levels of decay and predictability values for *IOI, Ratio,* and *Contour* in real time. Given the non-nested nature of these models, we compared models using AIC, using the heuristic that two models with an AIC difference < 2 are considered equivalent (Burnham & Anderson, 2004). All model comparisons, as well as collinearity checks, were performed using the *performance* (Lüdecke et al., 2021) and *easystats* (Lüdecke et al., 2022) packages. For reference, Table 2 shows the correlations between the

Table 2
Correlations between fixed factors in the exploratory linear mixed models, using modelled IC values for each sound separately as a proxy for predictability.

	IOI	Ratio	Contour	Previous IOI	Deviation from mean IOI
Ratio	0.530				
Contour	0.155	0.321			
Previous IOI	0.035	0.054	−0.054		
Deviation from mean IOI	0.026	0.225	0.157	0.027	
Sound number	−0.010	0.000	−0.011	−0.063	−0.068

Note: Pearson’s r is shown. Data from the model with a decay parameter with a half life of 2 intervals is shown.

fixed factors for the model with a decay kernel with a half-life parameter of two intervals. For all models used in the comparisons, VIF was below 4.3, showing no issues with collinearity. Fig. 3 shows modelled IC values for a selection of the rhythms used, both without decay, and with a decay kernel with a half-life parameter of two intervals.

3. Results

3.1. Rhythmic predictability affects behavior across tasks

The relationship between all dependent variables and the features of interest included in the models can be seen in Fig. 4 (for the same graph depicting the relationship between modelled entropy and performance, see Supplementary Fig. 3; for graphs depicting the raw data, see Supplementary Figs. 4 and 5). Generally, greater predictability (i.e., lower IC values) is associated with better performance across all tasks (i.e., higher hit rate, lower RT, lower complexity rating, higher number of correctly tapped trials, lower tapping variability). To understand the contribution of each separate feature to task performance, regression analyses are reported including all three features as fixed factors, and a factor *Evenness* to account for low-level rhythmic variation and single interval predictions (see section 2.6.1).

For all five dependent variables, the addition of the rhythmic features (*IOI, Ratio, and Contour*) significantly improved the model fit compared to the baseline model with only an *Evenness* feature (all χ^2s > 23.2, all *ps* < 0.001). Table 3 shows the results from all main regression analyses, and Fig. 5 shows a summary of the coefficients found in each model for each of the fixed factors. As can be seen in both the table and the figure, the *Contour* feature significantly affected responses across all three tasks, with significant effects for hit rates, complexity ratings, correctly tapped trials, and tapping variability (all *ps* < 0.001). Only for reaction times, the effect of *Contour* did not reach significance (p = 0.145). Similarly, *Ratio* significantly affected reaction times, complexity ratings, correctly tapped trials, and tapping variability (all *ps* < 0.008),

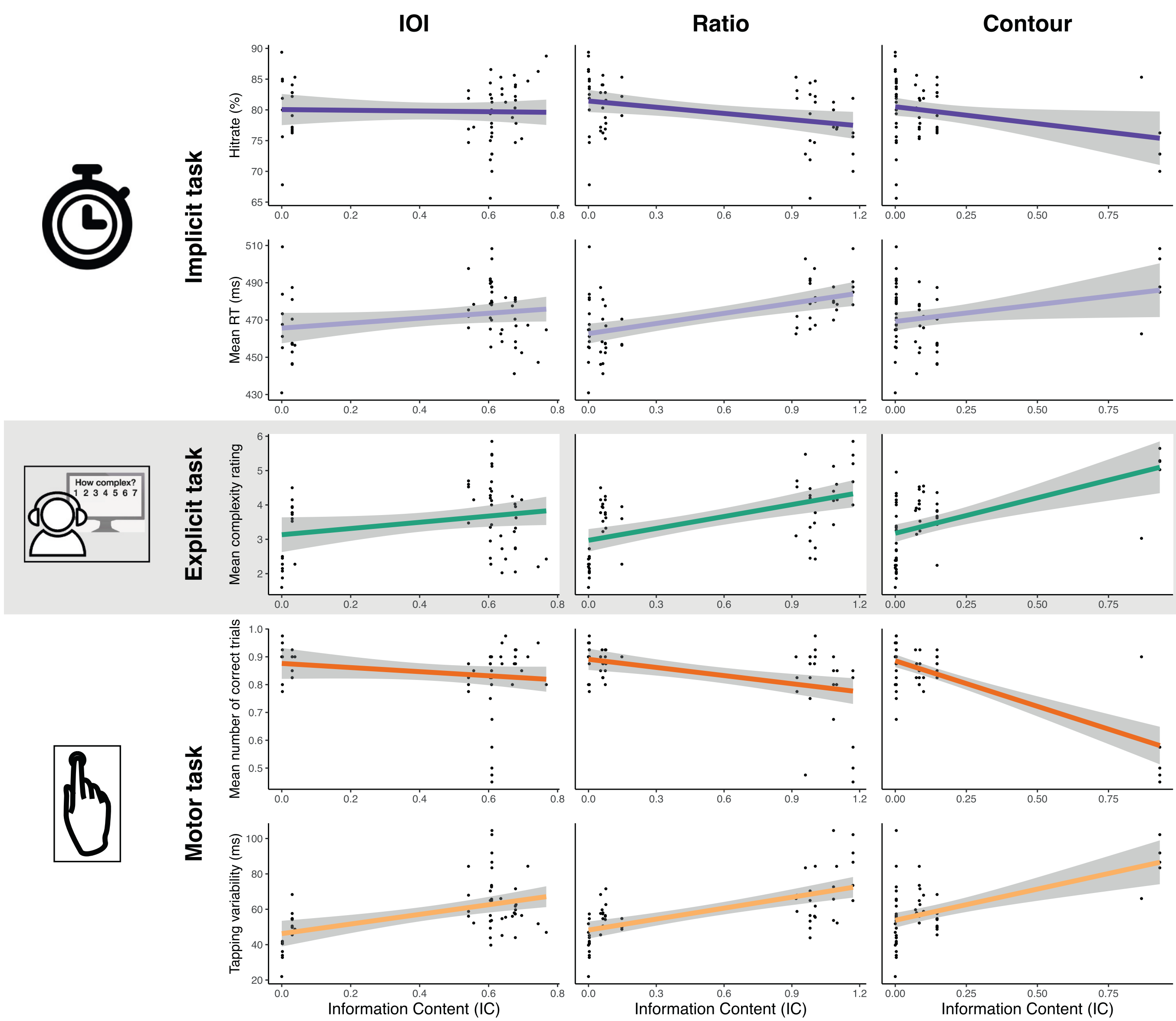

Fig. 4. Higher modelled predictability is associated with better performance across all tasks. The relationship between all five dependent variables and the three features of interest (*IOI, Ratio,* and *Contour*) is shown. Here, Information Content is used as an index of modelled predictability. Data are averaged for each rhythm over participants to account for between-subject variability for visibility purposes. Each dot in these graphs thus represents the average response for one rhythm. The data for the entropy predictors, as well as the raw data, can be found in the Supplementary Information. Note that in these graphs, the raw data are shown, not the fitted values of the statistical analyses (i.e., the linear mixed models). Hence, the graphs are not corrected for the collinearity between the three predictors.

Table 3
Modelled information content for Ratio and Contour but not IOI significantly affects both rhythm perception and production.

Task	Variable	Feature	Fixed effect coefficient	LR	p	
		IOI	−0.140	4.242	0.039	*
		Ratio	−0.012	0.038	0.846	
		Contour	−0.313	18.027	<0.001	**
	Hit rate	Evenness	−0.292	168.817	<0.001	**
		IOI	4.872	0.845	0.358	
	Reaction time	Ratio	13.462	7.493	0.006	**
		Contour	9.171	2.119	0.145	
Implicit		Evenness	7.044	15.612	<0.001	**
		IOI	0.398	10.208	0.001	**
		Ratio	0.817	50.623	<0.001	**
	Complexity rating	Contour	2.115	188.002	<0.001	**
Explicit		Evenness	0.403	86.277	<0.001	**
		IOI	−0.003	<0.001	0.989	
		Ratio	−0.531	6.961	0.008	**
		Contour	−1.748	72.513	<0.001	**
	Correct trials	Evenness	<0.001	<0.001	0.997	
		IOI	13.179	90.209	<0.001	**
		Ratio	12.614	95.165	<0.001	**
	Tapping variability	Contour	26.253	180.251	<0.001	**
Tapping		Evenness	2.105	20.614	<0.001	**

but not hit rates ($p = 0.846$). Results for *IOI* were somewhat more variable: while *IOI* clearly affected complexity ratings and tapping variability (both $ps < 0.001$), the effect on hit rates ($p = 0.039$) was less strong, and the effects on reaction times and correctly tapped trials were nonsignificant (both $ps > 0.35$). *Evenness* affected all dependent variables, except for the correctly tapped trials (all $ps < 0.001$; correctly tapped trials $p = 0.997$). In summary, the modelled predictability for all three features predicted behavior across all three tasks to some extent, with the largest effects for *Contour*, followed by *Ratio*, then *IOI*.

3.2. Control regression analyses confirm the main analyses

We tested whether our results were robust by running three additional analyses, based on (1) entropy as a predictability measure, (2) avoiding potential metricality by removing the most predictable patterns, and (3) a categorical treatment of the fixed factors (see section 2.6.2). The results from the analyses using entropy instead of IC values as fixed factors can be found in Supplementary Table 2. The results from the analysis without the easiest, and possibly metrical, rhythms can be found in Supplementary Table 3. The results from the analyses with dummy coded fixed factors (e.g., categorical, with either high or low predictability) can be found in Supplementary Table 4. The results, while numerically necessarily somewhat less strong for the latter two analyses (because these are based on less data), do not change in these control analyses: for all three tasks, the Ratio and Contour features exert the largest effects on responses, while the IOI feature has only weak and more variable effects. Hence, we directly replicate our main analyses using entropy instead of IC, using only a subset of rhythms, and using a different model setup, highlighting the robustness of the results.

Results from the mixed-effects models. The type of models that were run are different depending on the nature of the dependent variable, with hit rate and correctly tapped trials modelled with a binomial link model, complexity rating modelled with an ordinal mixed model, and reaction times, tapping variability and relative pattern deviation modelled with a normal linear mixed model. Because of the differences in nature between the models and the dependent variables, the coefficients from the models cannot be directly compared, as these are beta weight in the case of the linear mixed models, and log odds ratios in the case of the logistic and ordinal models. LR = Likelihood Ratio. $^*p < 0.05$, $^{**}p < 0.01$.

3.3. Larger effects of predictability in the explicit than in the implicit and motor tasks

To directly compare the effects of the different predictors across tasks, we ran a mixed-effects regression with all continuous task outcomes combined (see section 2.6.3). In line with the results of the separate regression analyses, the analysis across different tasks yielded main effects of all predictors (*IOI, Ratio, Contour, Evenness,* and *Task,* with all $ps < 0.001$). More importantly, we found interactions between *Task* and *Ratio* ($\chi^2 = 20.6$, $p < 0.001$), between *Task* and *Contour* ($\chi^2 = 83.9$, $p < 0.001$), and between *Task* and *Evenness* ($\chi^2 = 33.5$, $p < 0.001$), though the interaction between *Task* and *IOI* did not reach significance ($\chi^2 = 2.78$, $p = 0.25$). The follow-up analysis of the simple slopes showed that the effect of *Ratio* was larger in the explicit task (estimated trend 0.406) than in the implicit task (estimated trend 0.133, test of the difference in slopes $p < 0.001$) and the tapping task (estimated trend 0.130, test of the difference in slopes $p < 0.001$). Similarly, the estimated trend for *Contour* was larger in the explicit task (0.870) than in the implicit (0.090) and tapping (0.257) tasks (both $ps < 0.001$), and the estimated trend for *Evenness* was larger in the explicit task (0.165) than in the implicit (0.070) and tapping (0.019) tasks (both $ps < 0.001$). The difference between the implicit and tapping tasks did not reach significance for any of the features (all $ps > 0.122$). In summary, these results suggest that participants were more affected by the predictability of the rhythm in the explicit task than in the implicit and motor tasks.

3.4. IOI representations are imprecise

To explore whether IOIs may be represented in a less precise manner (Ravignani et al., 2018) than assumed previously (Sauvé & Pearce, 2019), we compared the main regression models with models in which IOI predictability was based on only 5 categories of IOIs (e.g., ranging from “short” to “long”). For reaction times and correctly tapped trials, there was no difference between the models (in both cases, the difference in AIC was <1.84). For hit rates, the main regression model performed better, but the difference was small (ΔAIC = 3.1) and the IOI predictor did not reach significance in the more imprecise model ($\chi^2 = 1.143$, $p = 0.285$). However, both for the explicit task and the tapping variability, the imprecise model performed markedly better than the original model (explicit task: ΔAIC = 134.1; tapping variability: ΔAIC = 102.3). In both models, all predictors affected behavior significantly, and the effects of the imprecise IOI feature were markedly larger than in the original model with predictability based on raw IOI values, though still smaller than the coefficients for Contour. Table 4 shows the outcomes of the imprecise models for IOI.

Results from the mixed-effects models using imprecise representations for IOI, with 5 categories of interval lengths. Only the models for the explicit task and the tapping task are shown, as these models outperformed the original regression models as reported in Table 3. Because of the differences in nature between the models and the dependent variables, the coefficients from the models cannot be directly compared, as these are beta weight in the case of the linear mixed models, and log odds ratios in the case of the logistic and ordinal models. LR = Likelihood Ratio. $^*p < 0.05$, $^{**}p < 0.01$.

3.5. Predictability in real time predicts tapping better than average predictability

Finally, we aimed to assess whether information content explains perception at the level of single events, rather than simply indexing the overall predictability of entire rhythms. Thus, we compared models with event-level predictors (i.e., real-time modelling of predictability) with models using rhythm-level predictors (i.e., using the same average

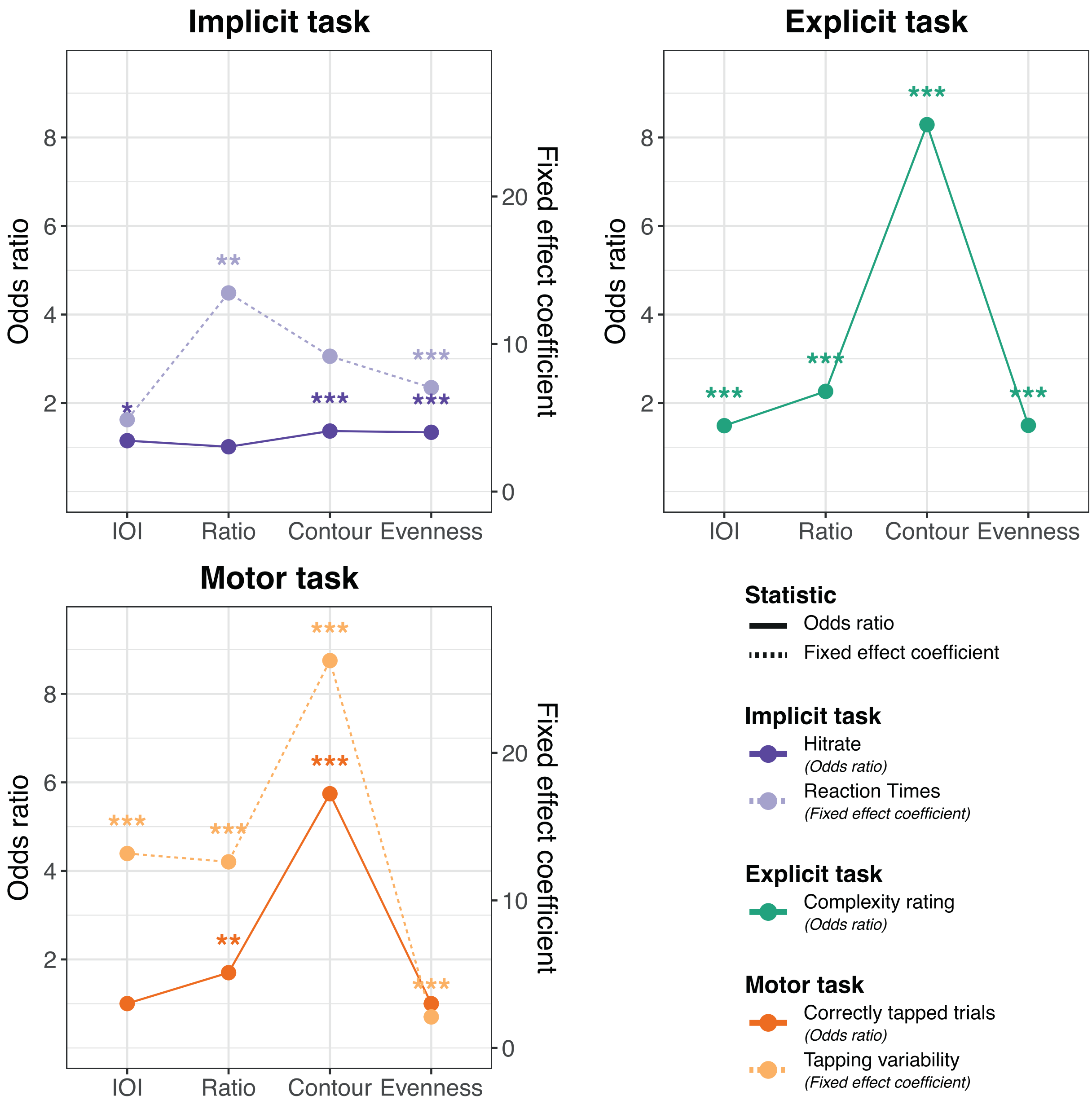

Fig. 5. Contour and Ratio exert the biggest effects on rhythmic behavior across all tasks. The graph depicts the fixed effect coefficients for the reaction time and tapping variability models, and the odds ratios for the hit rate, complexity rating, and tapping correctness models. Values for hit rates and correctly tapped trials were inverted, so for every task, a positive value indicates an effect in the expected direction (better performance for more predictable rhythm). Across all three tasks, the coefficients were largest for the Contour and Ratio features, and smaller and more variable for IOI. Note that the Odds Ratios and Fixed Effect Coefficients cannot be directly compared; they are depicted in this graph together for purposes of illustration. For a direct comparison between tasks, see under *Differences across tasks*. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 4
Information content for IOI yields larger effects on rhythm perception and production with imprecise representations.

Task	Variable	Feature	Fixed effect coefficient	LR	*p*	
		IOI (5 categories)	1.646	144.295	0.001	**
		Ratio	0.337	13.081	<0.001	**
		Contour	2.085	157.679	<0.001	**
Explicit	Complexity rating	Evenness	−0.418	101.618	<0.001	**
		IOI (5 categories)	21.058	200.325	<0.001	**
		Ratio	9.174	78.06	<0.001	**
		Contour	23.052	127.655	<0.001	**
Tapping	Tapping variability	Evenness	−1.203	7.981	0.005	**

Table 5
Model parameters for the best performing model for tapping (with a decay kernel with a half life of 2 intervals) and the models with average IC values for each rhythm for hits and reaction times.

Task	Variable	Feature	Fixed effect coefficient	LR	p	
Implicit	Hit	Average IOI	0.014	0.021	0.885	
		Average Ratio	0.134	2.474	0.116	
		Average Contour	0.267	5.731	0.017	*
		Previous IOI	−0.451	43.41	<0.001	**
		Deviation from mean IOI	0.788	95.057	<0.001	**
		Sound number	0.704	106.45	<0.001	**
	Reaction time	Average IOI	2.59	0.243	0.622	
		Average Ratio	17.07	9.048	0.003	**
		Average Contour	8.444	1.083	0.298	
		Previous IOI	24.393	49.056	<0.001	**
		Deviation from mean IOI	25.572	48.794	<0.001	**
		Sound number	21.468	45.876	<0.001	**
Motor	Tapping deviation	Real-time IOI	16.88	121.311	<0.001	**
		Real-time Ratio	18.359	74.243	<0.001	**
		Real-time Contour	38.271	204.455	<0.001	**
		Previous IOI	33.187	1506.564	<0.001	**
		Deviation from mean IOI	−21.714	374.204	<0.001	**
		Sound number	−0.389	0.261	0.609	

Note: Results from the mixed-effects models. For hits, the results are coded with higher values indicating higher chance of a miss (e.g., worse performance). LR = Likelihood Ratio. $^{*}p < 0.05$, $^{**}p < 0.01$.

predictability values as in the main analyses, but run at the level of single events, see section 2.6.4). For the implicit task, none of the real-time models outperformed the rhythm-level model, while the rhythm-level model did outperform the baseline model (i.e., only acoustic parameters, see Supplementary Table 5). Table 5 shows the rhythm-level models for the implicit task. Reaction times were longer and miss rates higher for higher modelled *Ratio* and *Contour* IC respectively (e.g., worse performance for more unpredictable rhythms). Not surprisingly, given that they are based on the same predictability values, these results corroborate the main analysis (Table 3).

For tapping, the best performing model, shown in Table 5, was the event-level model with a decay half-life parameter of 2 events. This model outperformed the rhythm-level model (ΔAIC >713 points), as well as all other event-level models (ΔAIC >49 points), and the baseline model (ΔAIC >749 points, see Supplementary Table 5). For the tapping task, performance decreased (e.g., larger tapping deviations) with increasing modelled information content of the *IOI*, *Ratio*, and *Contour* of the preceding interval, in line with the main analysis (Table 3). The model comparison suggests that for the tapping task, the PPM model indeed performed best when used at the level of single events.

Note that the acoustic control parameters also significantly affected responses, with longer reaction times and higher tapping deviations (e.g., decreased performance) for events following longer intervals (e.g., higher *Previous IOI*). Similarly, reaction times were longer and hit rates lower following intervals that were further from the mean interval (e.g., higher *Deviation from mean IOI*), and following intervals that came late in the sequence (e.g., higher *Sound number*). Contrary to these findings, events were less likely to be a hit when following shorter intervals (e.g., lower *Previous IOI*), possibly because reaction times are only based on hits, and therefore, may give a somewhat biased result. Also, tapping deviations were smaller (i.e., better performance) for intervals that were far from the mean interval. This may have been due to the Weber fraction partly affecting results (see also Supplementary Fig. 3).

4. Discussion

In the current research, we aimed to understand representations of rhythmic patterns in auditory perception and production. We differentiated between three levels of representation, with increasing degrees of abstraction: the absolute durations in a pattern (*IOI*), the succession of integer ratios (*Ratio*), and the direction of change (*Contour*, whether an interval is longer, shorter, or equal to the previous interval). Across three different behavioral tasks, we tested whether these levels of representation affected responses, using a computational model to quantify the predictability of rhythmic patterns for each of the representations.

We found that the modelled predictability of the contour of a rhythm, as well as the predictability of the succession of ratios affected behavior across an implicit, an explicit, and a motor task. The modelled predictability of the succession of absolute IOIs did affect responses, but to a lesser extent, and not across all tasks, with only weak or nonsignificant results in the implicit task. The effects of predictability based on absolute intervals was larger when the representation of the intervals was less precise (e.g., when IOIs were grouped into only five categories) – but only in the explicit and motor tasks. The effects of contour and ratio predictability were larger in the explicit rating task than in the implicit task and the motor task. Finally, the predictability of IOI, ratios, and contour predicted responses for the motor task best when computed in real time, on an event-by-event basis. For the implicit task, the event-by-event predictions did not affect behavior to the same extent: here, responses were equally well predicted by the overall predictability of the rhythm.

Notably, while previous research has used absolute IOIs (Milne et al., 2023; Sauvé & Pearce, 2019) or ratios between intervals (Gold et al., 2019; Kaplan et al., 2022; Ravignani et al., 2018) as the representation for rhythmic patterns, here we show that listeners likely represent patterns at an even more abstract and imprecise level: the level of contour. While contour was already suggested to play a role in rhythmic processing some decades ago by music theorists (Marvin, 1991), only recently has empirical research shown that human listeners indeed form duration contours in memory (Schmuckler & Moranis, 2023). Here, we extend these findings by showing that firstly, contour is used by listeners to guide and optimize perception and action when listening to rhythmic sequences, and secondly, contour is used more readily than more precise representations, like ratio and sequences of IOI. An imprecise representation of rhythm based on contour is in line with a recent study in which goodness of fit ratings for events following rhythmic patterns were higher in a broad window around an expected timepoint (Bouwer et al., 2023).

However, we found evidence that more temporally precise information, such as sequences of ratios, and to a lesser extent, absolute sequences of IOIs, were also used to represent and predict rhythm in perceptual and motor tasks. This can be explained in two ways. First, it is possible that the human brain represents rhythm at multiple levels of abstraction, and the different levels are engaged flexibly, depending on the task, the stimulus, and the individual listener. In the present study, the predictability of the absolute intervals (be it IOI or a categorized version of IOI) did not contribute to performance in an implicit task, but did contribute to performance in an explicit rating task and a tapping task. While rhythmic patterns can be learned even when they are not task-relevant (Bouwer et al., 2020), it is possible they are represented in a different way depending on the goal the representation needs to serve, be it perception, or action (Shalev et al., 2019), and implicitly, or explicitly (Droit-Volet & Coull, 2016). Possibly, absolute representations of rhythmic patterns are mostly engaged for explicit tasks, rather than when rhythm has an implicit effect on task performance. Thus, while, predictability in general may assert a bigger effect in explicit tasks (as is apparent from a larger effect for both *Contour* and *Ratio* in the explicit task than in the other tasks), the level of abstraction of the representation may differ depending on task demands. Similarly, different stimuli may emphasize different representations (e.g., IOI, Ratio, or Contour), and listeners may be sensitive to these differences, flexibly using the representation that allows for the most accurate predictions, or even refining their representation with learning and repeated exposure to the same pattern. Different listeners may also be prone to use different levels of representation. Musical training has been shown to lead to more precise perception (Banai et al., 2012) and production (Cicchini et al., 2012) of temporal intervals. Hence, experience-dependent plasticity may lead to the use of more precise representations of intervals and rhythmic patterns.

Foreshadowing a possible neural implementation of this, previous research has repeatedly implicated the cerebellum in the processing of single, absolute durations (Breska & Ivry, 2016; Teki et al., 2011), and the hippocampus in processing of absolute durations in a sequence (Lee et al., 2020). Relative timing, be it beat based, or processing of time relative to a template sequence that may be aperiodic, has been associated with the supplementary motor area (SMA) and basal ganglia (Cannon & Patel, 2021). One possibility is that representations based on absolute intervals (i.e., IOIs) are thus processed in a network including the cerebellum and medial temporal lobes, while more abstract representations (i.e., ratios and contour) rely on SMA and basal ganglia. This hypothesis does raise the question how beat-based timing, which similarly is thought to rely on SMA and basal ganglia (Grahn & Brett, 2007), relates to processing of relative durations in a sequence, as investigated here, and whether the two types of timing may share overlapping neural representations and processes (Pretto & James, 2015) – something for future studies to focus on.

A second possible explanation for the results is that while we modelled three different types of representation, the brain represents rhythmic patterns only at one level, which falls somewhere in between the more and less abstract representations we used in the current study. The aim of the current study was not to create an optimal model of rhythmic predictions, but rather, to disentangle the effect of different representations using an existing model. As such, the approach is not best suited to test the idea that listeners possess a single flexible representation of rhythmic patterns, which may rely on, for example, ratios between intervals at a slightly different level of abstraction (e.g., with a different rounding parameter than the integers between 1 and 4 that we used here). A model that does not use a discrete alphabet, like the PPM model we used here, could be used to estimate an optimal parameter for the imprecision with which humans represent rhythm on a continuous scale. Recent flexible Bayesian implementations of PPM-like models are promising in this regard (Cannon, 2021; Cannon & Kaplan, 2024; Kaplan et al., 2022, 2023).

Modelling in a continuous manner may also be beneficial in examining the categorical representation of IOI in the PPM model. First, the models with a smaller number of IOI categories clearly outperformed the models with all IOIs represented as a separate category (at least for the explicit and motor tasks), in line with the idea that rhythmic sequences are represented with only a limited number of categorical intervals (Ravignani et al., 2018). However, whereas it is well established that listeners interpret successive intervals as having integer ratio relationships (Jacoby & McDermott, 2017), there is no established threshold for what intervals would be interpreted as "long" or "short" for sequences of absolute IOIs, making the cutoff values we used rather arbitrary. Second, the fixed alphabet as used in the current modelling approach does not allow the model to represent time as a continuum, which means the model does not "know" that 350 ms is closer to 400 ms than to 800 ms: it just "knows" that these intervals are different. In future work, a modelling approach with intervals represented in a continuous manner may aid in establishing how listeners categorize absolute intervals in rhythmic sequences, and could account for the relationships between intervals.

Interestingly, the task context not only affected which representation influenced responses, but also whether responses were affected more by a pattern's overall predictability or by real-time, event-by-event predictability. In the implicit task, rhythmic predictability at the level of the whole rhythm affected responses, and adding parameters for predictability in real time did not improve the model, while in the tapping task, the best performing model was one with real-time predictors of rhythmic predictability. This again indicates differences between the implicit and tapping tasks in the way the representation of rhythm affects behavioral performance. Not only may the representation in the former be more imprecise, behavior in the implicit task may also be more dependent on average predictability, rather than real-time fluctuations in predictability. It is possible that listeners are not always primed to use rhythm in an implicit way, but rather, only do so when a rhythm is sufficiently predictable as a whole. However, another explanation is related to potential differences in power between the two tasks: with the tapping task providing datapoints for every event, and the implicit task only probing a maximum of eight events in each sequence, the failure of the real-time model in the implicit task may have been due to low power.

The real-time model for tapping showed that the optimal model incorporated a strong exponential decay parameter, with a half life of 2 intervals, which indicates faster decay than found previously for pitch sequences (Bianco et al., 2020; Kern et al., 2022). In the rhythmic sequences used here, there is silence between events, hence the memory representation for the sequence exists at a longer timescale than for the pitch sequences used in previous experiments (Bianco et al., 2020; Harrison et al., 2020), with pitched events every 50 ms. Nonetheless, the decay value that fitted the tapping data best was much faster than that previously found for pitch, both in time and number of events. While the 2-interval half life in our study corresponds to a decay parameter of 1 s (based on the average IOI of 500 ms), previous work in pitch obtained a half-life parameter of more than 3 s (Bianco et al., 2020), or around 7–8 events (Kern et al., 2022). These results may suggest that processing rhythmic sequences depends more on recent information than processing pitch sequences. One explanation for this is that the representation for temporal intervals may necessarily contain a higher information rate, as it codes not just when events appear, but also the content (e.g., the identity, or "what") of the event itself (Fan & Yotsumoto, 2018). With higher information load, memory decay may be faster (Harrison et al., 2020). Another possibility is that the short, repeated rhythms used here allowed for accurate predictions based only on the local context: Rhythmic patterns containing regularities over longer timescales might require listeners to consider contextual events at longer timescales when generating temporal predictions.

In the current study, we aimed to focus exclusively on rhythmic pattern processing. As such, we decided to disregard influences of regularity in the form of beat and meter. Beat and metrical structure have been shown to not only be driving forces behind rhythmic behavior (Honing & Bouwer, 2019; Snyder et al., 2024), but also, to interact with pattern processing, leading to improved discrimination of rhythmic patterns in the presence of a beat (Grahn & Brett, 2007). The patterns we used were not designed to contain metrical regularity: most patterns lacked durations related to each other at small integer ratios, and did not contain regular events at a rate consistent with the preferred beat rate for humans – as is typically found in metrical rhythm (Grahn & Brett, 2007). However, we cannot completely rule out that unintended metrical structure did affect the responses. Given the design choices we made, we cannot know whether listeners perceived any metricality in the stimuli and, if so, its characteristics in terms of tactus rate and any dynamic shifts of rate or phase within the patterns. Accounting for metricality in the current design would hence be non-trivial and post-hoc, given that the stimuli were selected considering only predictability, but not metricality. However, when excluding patterns that were most likely to contain metrical structure, the results remained the same (see Supplementary Table 3), supporting their validity in explaining pattern-based, but not beat-based processing. Future studies may be designed accounting for both predictability and metricality to examine their relationship and unique contributions to rhythm processing.

One open question in this regard is whether an abstract perceptual representation of rhythmic patterns, such as contour, ratio, or categories of IOIs, can account for any rhythmic behavior that has been ascribed to beat perception. Whether the mechanisms for beat and pattern processing are similar or separate remains an open discussion (Aharoni et al., 2024; Bouwer et al., 2020; Bouwer et al., 2023; Morillon et al., 2016; Nobre & van Ede, 2018; Rimmele et al., 2018; Solli et al., 2025). Creating more sophisticated models of rhythm that represent patterns in more abstract ways may show that such representations can account for both beat-based and pattern-based timing perception without assuming different underlying mechanisms, similar to a model in which beat is represented as a pattern amenable to probabilistic learning (Cannon, 2021). Indeed, it is possible to represent and learn metrical expectations within a probabilistic predictive model such as IDyOM (van der Weij et al., 2017).

Similar to the influence of beat and meter, we cannot exclude the influence of previous exposure to rhythm on responses. Statistical learning of patterns in music has been shown to be susceptible to cultural influences and long term learning (Kaplan et al., 2022; Pearce, 2018), as has the categorization of rhythmic patterns (Jacoby et al., 2021; Jacoby & McDermott, 2017). Although the stimuli resulting from model-based selection did not resemble metrical rhythmic patterns common in Western music, in future work, the present approach could be used to simulate the influence of cultural exposure on rhythmic pattern perception.

Finally, in our design, we did not consider possible non-linear influences of rhythmic complexity. It is well-established that the pleasurable urge to move to music (also known as groove) has a non-linear, inverted U-shaped relationship with rhythmic complexity, with groove being strongest for medium complexity rhythms (Etani et al., 2024; Spiech et al., 2022; Spiech et al., 2024; Witek et al., 2014). Moreover, in one study, ratings of beat presence were shown to be following a binary pattern, with the beat being perceived as present or not, but without a linear relationship (Bouwer et al., 2018). Given the design and selection of rhythms in our study – with a largely binary distribution between rhythms with relatively high and low predictability on each feature – our data do not allow for testing these potential non-linear relationships. Similarly, we cannot claim that the linearity is present over the whole distribution of rhythms. Our rhythms were somewhat unevenly distributed over the possible IC values in the rhythm space (e.g., for contour, many rhythms had IC values between 0 and 0.2, with only a few with much higher IC values). Given the somewhat binary distribution of IC in the selected rhythms, the data cannot shed light on the precise psychophysical relationship between IC and complexity perception.

5. Conclusion

In the current work, we examined cognitive representation of rhythmic patterns in auditory perception. Using stimuli selected from a large universe of rhythmic patterns by a model of auditory expectation, we aimed to study processing of rhythmic patterns in isolation, unaffected by metrical structure and unbiased by cultural experience. Across three tasks we found that temporal predictions in rhythmic patterns relied mostly on abstract and imprecise representations. Moreover, we found that explicit task demands increased reliance on rhythmic predictability for task performance in general, and that reliance on absolute representations of rhythmic patterns was particularly apparent under explicit and motor task demands. Overall, this implies that rhythm is represented in an abstract manner, reflecting interval categories and sequences of interval ratios and contour rather than absolute timing, with the representation flexibly adapted based on task demands. Given the ubiquitousness of rhythmic patterns in the environment, especially in language and music, such flexibility in the human perceptual system may strengthen performance on particular tasks by emphasizing rhythmic representations at the most useful or predictive level of abstraction, be it for listening to music, reciting poetry, or dancing to a rhythm.

CRediT authorship contribution statement

Fleur L. Bouwer: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Atser Damsma:** Writing – review & editing, Methodology, Investigation. **Thomas M. Kaplan:** Writing – review & editing, Software, Methodology, Investigation. **Mohsen Ghorashi Sarvestani:** Writing – review & editing, Project administration, Investigation. **Marcus T. Pearce:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106345.

Data availability

All scripts for stimulus design, modelling, and running the experiments, as well as the raw data, and analysis scripts can be found through OSF (https://osf.io/kzg2e/overview?view_only=34af7d2af4854bffbaa247d8513bb14b).

References

Aharoni, M., Breska, A., Müller, M. M., & Schröger, E. (2024). Mechanisms of sustained perceptual entrainment after stimulus offset. *European Journal of Neuroscience, 59*(5), 1047–1060. https://doi.org/10.1111/ejn.16032

Anglada-tort, M., Harrison, P. M. C., & Jacoby, N. (2021). REPP : A robust cross-platform solution for online sensorimotor synchronization experiments. *bioRxiv*, 1–34.

Banai, K., Fisher, S., & Ganot, R. (2012). The effects of context and musical training on auditory temporal-interval discrimination. *Hearing Research, 284*(1–2), 59–66. https://doi.org/10.1016/j.heares.2011.12.002

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Bianco, R., Harrison, P. M., Hu, M., Bolger, C., Picken, S., Pearce, M. T., & Chait, M. (2020). Long-term implicit memory for sequential auditory patterns in humans. *eLife, 9*, Article e56073. https://doi.org/10.7554/eLife.56073

Boettcher, S. E. P., Shalev, N., Wolfe, J. M., & Nobre, A. C. (2021). Right place, right time: Spatiotemporal predictions guide attention in dynamic visual search. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0000901

Bouwer, F. L., Burgoyne, J. A., Odijk, D., Honing, H., & Grahn, J. A. (2018). What makes a rhythm complex? The influence of musical training and accent type on beat perception. *PLoS One, 13*(1), Article e0190322. https://doi.org/10.1371/journal.pone.0190322

Bouwer, F. L., Fahrenfort, J. J., Millard, S. K., Kloosterman, N. A., & Slagter, H. A. (2023). A silent disco: Differential effects of beat-based and pattern-based temporal expectations on persistent entrainment of low-frequency neural oscillations. *Journal of Cognitive Neuroscience, 35*(6), 990–1020. https://doi.org/10.1162/jocn_a_01985

Bouwer, F. L., Háden, G. P., & Honing, H. (2024). Probing beat perception with event-related potentials (ERPs) in human adults, newborns, and nonhuman Primates. In H. Merchant, & V. de Lafuente (Eds.), *Neurobiology of interval timing* (pp. 227–256). Springer International Publishing. https://doi.org/10.1007/978-3-031-60183-5_13.

Bouwer, F. L., & Honing, H. (2015). Temporal attending and prediction influence the perception of metrical rhythm: Evidence from reaction times and ERPs. *Frontiers in Psychology, 6*(July), 1094. https://doi.org/10.3389/fpsyg.2015.01094

Bouwer, F. L., Honing, H., & Slagter, H. A. (2020). Beat-based and memory-based temporal expectations in rhythm: Similar perceptual effects, different underlying mechanisms. *Journal of Cognitive Neuroscience, 32*(7), 1221–1241. https://doi.org/10.1162/jocn_a_01529

Bouwer, F. L., Nityananda, V., Rouse, A. A., & ten Cate, C. (2021). Rhythmic abilities in humans and non-human animals: A review and recommendations from a methodological perspective. *Philosophical Transactions of the Royal Society B, 376*, 20200335. https://doi.org/10.1098/rstb.2020.0335

Bouwer, F. L., Werner, C. M., Knetemann, M., & Honing, H. (2016). Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia, 85*, 80–90. https://doi.org/10.1016/j.neuropsychologia.2016.02.018

Brandon, M., Terry, J., Stevens, C. J., & Tillmann, B. (2012). Incidental learning of temporal structures conforming to a metrical framework. *Frontiers in Psychology, 3* (August), 294. https://doi.org/10.3389/fpsyg.2012.00294

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biology, 15*(2), Article e2001665. https://doi.org/10.1371/journal.pbio.2001665

Breska, A., & Ivry, R. B. (2016). Taxonomies of timing: Where does the cerebellum fit in? *Current Opinion in Behavioral Sciences, 8*, 282–288. https://doi.org/10.1016/J.COBEHA.2016.02.034

Bulger, D., Milne, A. J., & Dean, R. T. (2022). A point-process model of tapping along to difficult rhythms. *Journal of Mathematical Psychology, 111*, Article 102724. https://doi.org/10.1016/j.jmp.2022.102724

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Cameron, D. J., & Grahn, J. A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Frontiers in Human Neuroscience, 8* (December), 1003. https://doi.org/10.3389/fnhum.2014.01003

Cannon, J. (2021). Expectancy-based rhythmic entrainment as continuous Bayesian inference. *PLoS Computational Biology, 17*(6), Article e1009025. https://doi.org/10.1371/journal.pcbi.1009025

Cannon, J., & Kaplan, T. (2024). Inferred representations behave like oscillators in dynamic Bayesian models of beat perception. *Journal of Mathematical Psychology, 122*, Article 102869. https://doi.org/10.1016/j.jmp.2024.102869

Cannon, J., & Patel, A. D. (2021). How beat perception co-opts motor neurophysiology. *Trends in Cognitive Sciences, 25*(2), 137–150. https://doi.org/10.1016/j.tics.2020.11.002

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience, 20*(2), 226–239. https://doi.org/10.1162/jocn.2008.20018

Christensen, R. H. B. (2019). Ordinal—Regression models for ordinal data. *R package version, 2019*, 4–25. http://www.cran.r-project.org/package=ordinal/.

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *The Journal of Neuroscience, 32* (3), 1056–1060. https://doi.org/10.1523/jneurosci.3411-11.2012

Dalla Bella, S., Farrugia, N., Benoit, C.-E., Bégel, V., Verga, L., Harding, E., & Kotz, S. A. (2017). BAASTA: Battery for the assessment of auditory sensorimotor and timing abilities. *Behavior Research Methods, 49*(3), 1128–1145. https://doi.org/10.3758/s13428-016-0773-6

Damm, L., Varoqui, D., De Cock, V. C., Bella, S. D., & Bardy, B. (2019). Why do we move to the beat? A multi-scale approach, from physical principles to brain dynamics. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/J.NEUBIOREV.2019.12.024

Dean, R. T., Bulger, D., & Milne, A. J. (2021). Onthe roles of complexity and symmetry in cued tapping of well-formed complex rhythms:Descriptive statistics and time series analysis. *Music Perception, 39*(2), 202–225. https://doi.org/10.1525/MP.2021.39.2.202

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*(3), 341–365. https://doi.org/10.1068/p3370

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews, 81*, 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Doelling, K. B., Arnal, L. H., & Assaneo, M. F. (2023). Adaptive oscillators support Bayesian prediction in temporal processing. *PLoS Computational Biology, 19*(11), Article e1011669. https://doi.org/10.1371/journal.pcbi.1011669

Doelling, K. B., Herbst, S. K., Arnal, L. H., & van Wassenhove, V. (2022). Psychological and neuroscientific foundations of rhythms and timing [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/hkgdr

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition, 77*(3), 251–288. https://doi.org/10.1016/S0010-0277(00)00106-2

Droit-Volet, S., & Coull, J. T. (2016). Distinct developmental trajectories for explicit and implicit timing. *Journal of Experimental Child Psychology, 150*, 141–154. https://doi.org/10.1016/j.jecp.2016.05.010

Etani, T., Miura, A., Kawase, S., Fujii, S., Keller, P. E., Vuust, P., & Kudo, K. (2024). A review of psychological and neuroscientific research on musical groove. *Neuroscience & Biobehavioral Reviews, 158*, Article 105522. https://doi.org/10.1016/j.neubiorev.2023.105522

Fan, Z., & Yotsumoto, Y. (2018). Multiple time intervals of visual events are represented as discrete items in working memory. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01340

Fiveash, A., Ferreri, L., Bouwer, F. L., Kösem, A., Moghimi, S., Ravignani, A., … Tillmann, B. (2023). Can rhythm-mediated reward boost learning, memory, and social connection? Perspectives for future research. *Neuroscience & Biobehavioral Reviews, 105153*. https://doi.org/10.1016/j.neubiorev.2023.105153

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression*. Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Geiser, E., Ziegler, E., Jancke, L., Meyer, M., Jäncke, L., Meyer, M., Jancke, L., Meyer, M., Jäncke, L., & Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex, 45*(1), 93–102. https://doi.org/10.1016/j.cortex.2007.09.010

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: A reward for learning? *The Journal of Neuroscience, 39*(47), 9397–9409. https://doi.org/10.1523/JNEUROSCI.0428-19.2019

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience, 19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Grahn, J. A., & Schuit, D. (2012). Individual differences in rhythmic ability: Behavioral and neuroimaging investigations. *Psychomusicology: Music, Mind, and Brain, 22*(2), 105–121. https://doi.org/10.1037/a0031188

Hardy, N. F., & Buonomano, D. V. (2016). Neurocomputational models of interval and pattern timing. *Current Opinion in Behavioral Sciences, 8*, 250–257. https://doi.org/10.1016/j.cobeha.2016.01.012

Harrison, P. M. C., Bianco, R., Chait, M., & Pearce, M. T. (2020). PPM-decay: A computational model of auditory prediction with memory decay. *PLoS Computational Biology, 16*(11), Article e1008304. https://doi.org/10.1371/journal.pcbi.1008304

Heideman, S. G., van Ede, F., & Nobre, A. C. (2018a). Early Behavioural facilitation by temporal expectations in complex visual-motor sequences. *Neuroscience, 389*, 74–84. https://doi.org/10.1016/J.NEUROSCIENCE.2018.05.014

Heideman, S. G., van Ede, F., & Nobre, A. C. (2018b). Temporal alignment of anticipatory motor cortical beta lateralisation in hidden visual-motor sequences. *European Journal of Neuroscience, 48*(8), 2684–2695. https://doi.org/10.1111/ejn.13700

Hervé, M. (2018). RVAideMemoire: Testing and plotting procedures for biostatistics. *R package version, 0*(9–69), 3 (R).

Honing, H., & Bouwer, F. L. (2019). Rhythm. In P. J. Rentfrow, & D. Levitin (Eds.), *Foundations in music psychology: Theory and research* (pp. 33–69). MIT Press.

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology, 27*(3), 359–370. https://doi.org/10.1016/j.cub.2016.12.031

Jacoby, N., Polak, R., Grahn, J., Cameron, D. J., Lee, K. M., Godoy, R., … Mcdermott, J. (2021). Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/b879v

Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Godoy, R., … McDermott, J. H. (2024). Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries. *Nature Human Behaviour, 8*(5), 846–877. https://doi.org/10.1038/s41562-023-01800-9

Kaplan, T., Cannon, J., Jamone, L., & Pearce, M. (2022). Modeling enculturated bias in entrainment to rhythmic patterns. *PLoS Computational Biology, 18*(9), Article e1010579. https://doi.org/10.1371/journal.pcbi.1010579

Kaplan, T., Jamone, L., & Pearce, M. (2023). Probabilistic modelling of microtiming perception. *Cognition, 239*, Article 105532. https://doi.org/10.1016/j.cognition.2023.105532

Kern, P., Heilbron, M., de Lange, F. P., & Spaak, E. (2022). Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. *eLife, 11*, Article e80935. https://doi.org/10.7554/eLife.80935

Koelsch, S., Vuust, P., & Friston, K. J. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences, 23*(1), 63–77. https://doi.org/10.1016/J.TICS.2018.10.006

Kornysheva, K., & Diedrichsen, J. (2014). Human premotor areas parse sequences into their spatial and temporal features. *eLife, 3*, Article e03043. https://doi.org/10.7554/eLife.03043

Kornysheva, K., Sierk, A., & Diedrichsen, J. (2013). Interaction of temporal and ordinal representations in movement sequences. *Journal of Neurophysiology, 109*(5), 1416–1424. https://doi.org/10.1152/jn.00509.2012

Laje, R., Cheng, K., & Buonomano, D. V. (2011). Learning of temporal motor patterns: An analysis of continuous versus reset timing. *Frontiers in Integrative Neuroscience, 5*, 61. https://doi.org/10.3389/fnint.2011.00061

Lee, A. C. H., Thavabalasingam, S., Alushaj, D., Çavdaroğlu, B., & Ito, R. (2020). The hippocampus contributes to temporal duration memory in the context of event sequences: A cross-species perspective. *Neuropsychologia, 137*, Article 107300. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2019.107300

Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version, 1*, 4. https://cran.r-project.org/package=emmeans.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Wiernik, B. M., Bacher, E., Thériault, R., & Makowski, D. (2022). Easystats: Framework for easy statistical modeling, visualization, and reporting. *CRAN*. https://doi.org/10.32614/CRAN.package.easystats

Marvin, E. W. (1991). The perception of rhythm in non-tonal music: Rhythmic contours in the music of Edgard Varèse. *Music Theory Spectrum, 13*(1), 61–78. https://doi.org/10.2307/745974

Matthews, T. E., Thibodeau, J. N. L., Gunther, B. P., & Penhune, V. B. (2016). The impact of instrument-specific musical training on rhythm perception and production. *Frontiers in Psychology, 7*(February), 69. https://doi.org/10.3389/fpsyg.2016.00069

Meirhaeghe, N., Sohn, H., & Jazayeri, M. (2021). A precise and adaptive neural mechanism for predictive temporal processing in the frontal cortex. *Neuron, 109*(18), 2995–3011.e5. https://doi.org/10.1016/j.neuron.2021.08.025

Milne, A. J., Dean, R. T., & Bulger, D. (2023). The effects of rhythmic structure on tapping accuracy. *Attention, Perception, & Psychophysics, 85*(8), 2673–2699. https://doi.org/10.3758/s13414-023-02778-2

Morillon, B., Schroeder, C. E., Wyart, V., & Arnal, L. H. (2016). Temporal prediction in lieu of periodic stimulation. *The Journal of Neuroscience, 36*(8), 2342–2347. https://doi.org/10.1523/JNEUROSCI.0836-15.2016

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One, 9*(2), Article e89642. https://doi.org/10.1371/journal.pone.0089642

Nobre, A. C., & van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience, 19*, 34–48. https://doi.org/10.1038/nrn.2017.141

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences, 23*(11), 913–926. https://doi.org/10.1016/J.TICS.2019.08.004

O'Reilly, J. X., McCarthy, K. J., Capizzi, M., & Nobre, A. C. (2008). Acquisition of the temporal and ordinal structure of movement sequences in incidental learning. *Journal of Neurophysiology, 99*, 2731–2735. https://doi.org/10.1152/jn.01141.2007

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition.* City University.

Pearce, M. 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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pretto, M. D., & James, C. E. (2015). Principles of parsimony: fMRI correlates of beat-based versus duration-based sensorimotor synchronization. *Psychomusicology: Music, Mind, and Brain, 25*(4), 380–391. https://doi.org/10.1037/pmu0000122

R Development Core Team. (2008). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. www.R-project.org.

Ravignani, A., Thompson, B., Lumaca, M., & Grube, M. (2018). Why do durations in musical rhythms conform to small integer ratios? *Frontiers in Computational Neuroscience, 12*(November), 1–9. https://doi.org/10.3389/fncom.2018.00086

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*(6), 969–992. https://doi.org/10.3758/BF03206433

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin and Review, 20*(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences, 22*(10), 870–882. https://doi.org/10.1016/J.TICS.2018.08.003

Sadakata, M., Desain, P., & Honing, H. (2006). The Bayesian way to relate rhythm perception and production. *Music Perception: An Interdisciplinary Journal, 23*(3), 269–288. https://doi.org/10.1525/mp.2006.23.3.269

Sauvé, S. A., & Pearce, M. T. (2019). Information-theoretic modeling of perceived musical complexity. *Music Perception: An Interdisciplinary Journal, 37*(2), 165–178. https://doi.org/10.1525/mp.2019.37.2.165

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2020). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences, 44*, Article e59. https://doi.org/10.1017/S0140525X20000333

Schmuckler, M. A., & Moranis, R. (2023). Rhythm contour drives musical memory. *Attention, Perception, & Psychophysics, 85*(7), 2502–2514. https://doi.org/10.3758/s13414-023-02700-w

Schultz, B. G., Stevens, C. J., Keller, P. E., & Tillmann, B. (2013). The implicit learning of metrical and nonmetrical temporal patterns. *The Quarterly Journal of Experimental Psychology, 66*(2), 360–380. https://doi.org/10.1080/17470218.2012.712146

Senn, O. (2023). A predictive coding approach to modelling the perceived complexity of popular music drum patterns. *Heliyon, 9*(4), Article e15199. https://doi.org/10.1016/j.heliyon.2023.e15199

Shalev, N., Nobre, A. C., & van Ede, F. (2019). Time for what? Breaking down temporal anticipation. *Trends in Neurosciences, 42*(6), 373–374. https://doi.org/10.1016/J.TINS.2019.03.002

Shannon, C. E. (1948). *A Mathematical Theory of Communication. 55*.

Shin, J. C., & Ivry, R. B. (2002). Concurrent learning of temporal and spatial sequences. *Journal of Experimental Psychology, 28*(3), 445–457. https://doi.org/10.1037/0278-7393.28.3.445

Snyder, J. S., Gordon, R. L., & Hannon, E. E. (2024). Theoretical and empirical advances in understanding musical rhythm, beat and metre. *Nature Reviews Psychology, 1–14*. https://doi.org/10.1038/s44159-024-00315-y

Solli, S., Danielsen, A., Leske, S., Blenkmann, A. O., Doelling, K. B., Solbakk, A.-K., & Endestad, T. (2025). Rhythm-based temporal expectations: Unique contributions of predictability and periodicity. *Journal of Cognitive Neuroscience, 37*(3), 555–581. https://doi.org/10.1162/jocn_a_02261

Spiech, C., Danielsen, A., Laeng, B., & Endestad, T. (2024). Oscillatory attention in groove. *Cortex, 174*, 137–148. https://doi.org/10.1016/j.cortex.2024.02.013

Spiech, C., Sioros, G., Endestad, T., Danielsen, A., & Laeng, B. (2022). Pupil drift rate indexes groove ratings. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-15763-w. Article 1.

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *The Journal of Neuroscience, 31*(10), 3805–3812. https://doi.org/10.1523/JNEUROSCI.5561-10.2011

Ullén, F., & Bengtsson, S. L. (2003). Independent processing of the temporal and ordinal structure of movement sequences. *Journal of Neurophysiology, 90*(6), 3725–3735. https://doi.org/10.1152/jn.00458.2003

Vuust, P., Heggli, O. A., Friston, K. J., & Kringelbach, M. L. (2022). Music in the brain. *Nature Reviews Neuroscience, 23*(5), 287–305. https://doi.org/10.1038/s41583-022-00578-5

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music—Brain responses to rhythmic incongruity. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 45*(1), 80–92. https://doi.org/10.1016/j.cortex.2008.05.014

Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology, 5*(October), 1111. https://doi.org/10.3389/fpsyg.2014.01111

van der Weij, B., Pearce, M. T., & Honing, H. (2017). A probabilistic model of meter perception: Simulating enculturation. *Frontiers in Psychology, 8*(May), 824. https://doi.org/10.3389/fpsyg.2017.00824

Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS One, 9*(4), Article e94446. https://doi.org/10.1371/journal.pone.0094446