Influence of rhythmic contexts on perception: No behavioral and eye-tracker evidence for rhythmic entrainment

Rafael Román-Caballero[*,a,b], Elisa Martín-Arévalo[a], Paulina del Carmen Martín-Sánchez[a], Juan Lupiáñez[a], and Mariagrazia Capizzi[a]

[a] Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain; Department of Experimental Psychology, University of Granada, Granada, Spain

[b] Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Canada; McMaster Institute for Music and the Mind, McMaster University, Hamilton, Canada

Abstract

Keywords:
rhythm, entrainment, Dynamic Attending Theory, foreperiod, pupillometry

Entrainment theories propose that attention inherently oscillates between moments of attentional enhancement and disengagement. Consequently, perceptual and response benefits have been reported in tasks with a rhythmic structure. In the present study, we report two preregistered auditory experiments attempting to replicate previous supporting behavioral evidence of entrainment theories. In addition, we incorporated eye-tracker measures. Both Experiment 1 (duration discrimination task) and Experiment 2 (pitch discrimination task) showed no phase-specific benefit of rhythmic sequences compared to arrhythmic ones. Importantly, a tonic larger pupil size for arrhythmic conditions was observed irrespective of target phase, suggesting higher processing demands or arousal state imposed by a sustained uncertain context. Overall, the present results call into question whether the perceptual benefits predicted by entrainment theories are generalizable across all experimental designs and paradigms. On the contrary, our findings join a large group of studies that have failed to replicate the foundational results of attentional entrainment.

Highlights

- Entrainment theories propose that attentional fluctuations can be synchronized with external rhythms.
- Behavioral and eye-tracker results from two auditory rhythmic tasks failed to find an entrainment effect.
- The present findings call into question the generalizability of the perceptual benefits of rhythm across paradigms.
- A complex interplay between temporal attention effects other than rhythmic entrainment might explain the present results.
- Greater tonic pupil dilation in arrhythmic blocks suggests higher processing demands and arousal in unpredictable contexts.

1. Introduction

Being able to selectively attend to the moment in time when a relevant event is expected to occur is essential for many cognitive processes, such as perception and action selection (Capizzi & Correa, 2018; Nobre & van Ede, 2018). Different temporal contexts in our environment help to allocate attentional resources in time. In the simplest case, temporal attention is shaped by the passage of time itself, assuming an expected event must occur if it has not occurred yet (i.e., hazard function; Janssen & Shadlen, 2005; Visalli et al., 2023). In the laboratory, this temporal context can be simulated by introducing different intervals, or foreperiods, between a warning signal and a target that always appears (i.e., without catch trials). The longer the foreperiod, the higher the expectation that the target will appear, resulting in faster and/or more accurate responses when it does. This phenomenon, where performance is better at longer compared to shorter foreperiods, is known as the foreperiod effect (Coull, 2009; Niemi & Näätänen, 1981).

In dynamic environments, temporal attention is also influenced by the rhythmic structure naturally inherent in many events, such as music and speech. Entrainment theories, like the Dynamic Attending Theory (DAT; Large & Jones, 1999; Jones, 1976), leverage this idea by proposing that temporal attention functions as an oscillatory system that alternates between moments of heightened attention and moments of attentional disengagement. The synchronization between internal oscillations and external rhythms optimizes the processing of those events occurring in-phase with a preceding regular rhythm compared to those occurring out-of-phase. Supporting the predictions of DAT, a pioneering study by Jones and colleagues (2002) found that accuracy in a pitch comparison task was higher when the comparison tone was presented in-phase with a preceding rhythm rather than out-of-phase (early or late), resulting in an inverted U-shaped

Email address: rrarroca@ugr.es (Rafael Román-Caballero)
URL: https://rrarroca.com

Rafael Román-Caballero https://orcid.org/0000-0003-0943-6217
Elisa Martín-Arévalo http://orcid.org/0000-0002-4546-6440
Juan Lupiáñez https://orcid.org/0000-0001-6157-9894
Mariagrazia Capizzi https://orcid.org/0000-0001-7880-7320

performance pattern (the so-called rhythmic entrainment). Despite the popularity of DAT, more evidence increasingly challenges the predicted behavioral outcomes. For example, Bauer and colleagues (2015) undertook both conceptual and exact replications of the Jones et al. (2002) paradigm, but were unable to replicate the inverted U-shaped pattern in task accuracy. Moreover, when considering the level of musical expertise among participants, although musicians performed better overall compared to non-musicians, they still did not exhibit the inverted U-shaped pattern.

Notably, some studies that failed to replicate the inverted U-shaped pattern found that this could be masked by the foreperiod effect (Jones, 2015; Sanabria et al., 2011). Specifically, the certainty of target appearance counteracted the effects of rhythms by speeding up responses even to targets occurring later than expected, thereby eliminating the difference between in-phase and out-of-phase late targets. Additionally, it has been shown that the foreperiod effect can be modulated by rhythmic contexts, becoming more pronounced in a rhythmic condition (e.g., Martin et al., 2005). Overall, these findings suggest a complex interplay between different sources of temporal attention that can critically influence each other and could partly account for the failure to observe rhythmic entrainment as predicted by DAT.

Finally, complementing traditional behavioral measures, recent research has shown that temporal expectation can be monitored using eye-tracker measures. In two visual sustained attention tasks, Shalev and Nobre (2022) found better participants' discrimination in rhythmic blocks (isochronous sequence with a fixed interval of 3500 ms), compared to blocks where stimuli appeared after a variable interval (2500–4500 ms). In addition, rhythmic regularity reduced pupil size across the two tasks, suggesting differences in processing demands and an arousal adjustment according to temporal certainty. Another previous study using a visual vigilance task (Dankner et al., 2017) found that saccades were inhibited prior to the target onset at predictable compared with less predictable moments, using fixed (2000 ms) and variable (1000, 1500, 2000, or 2500 ms) intervals, respectively. Reduced eye motion for a short interval prior to the target may prevent transient visual distortions and enhance acuity (Dankner et al., 2017). Thus, pupil dilation and saccades are both sensitive indices of temporal expectation and context uncertainty that could help to disentangle the processes occurring from trial onset and response execution. High uncertainty, such as that experienced in arrhythmic, but not rhythmic, contexts might favor a continuous state of high tonic arousal and a slight transient increase leading to larger pupil size (Shalev & Nobre, 2022), as well as a reduced prestimulus saccadic inhibition given the unpredictable onset. Interestingly, both pupil size and saccade rate have also exhibited a foreperiod-like pattern, with a transient increase of the pupil and a decrease in saccade rate as a function of time (Abeles et al., 2020; Shalev & Nobre, 2022).

However, the existence of a single cue–target interval for the fixed inter-onset interval (IOI) condition in the available studies makes it difficult to differentiate between rhythmic entrainment and the temporal expectation for a fully predictable moment, as reflected by the foreperiod effect. Based on the above-mentioned behavioral and eye-tracker evidence, our study aimed to further test the assumptions of DAT through two different experiments and task manipulations. Although these experiments were conducted at different times, as indicated by their preregistrations (https://osf.io/njty6/), they are presented together here due to their shared goal and complementary nature.

In the first experiment, we used a continuous rhythmic task with no intertrial interval to emphasize reliance on rhythm as a temporal cue. The task was modeled after McAuley and Fromboluti (2014), who employed an auditory oddball paradigm where participants judged whether an oddball tone lasted longer or shorter than a series of standard tones. The rhythm context was either regular or irregular, and the oddball could appear in-phase or out-of-phase with respect to the preceding rhythm. The decision to use a temporal discrimination task was based on previous evidence indicating that tasks requiring temporal judgments benefit more from precise temporal attention than tasks involving judgments in other domains (e.g., pitch, timbre, or loudness discrimination; Prince & Sopp, 2019). In line with the original study (McAuley & Fromboluti, 2014), we expected less distortion in duration judgments for oddball tones when they aligned with the entrained rhythm as compared to when they occurred earlier or later. Moreover, in Experiment 1, we measured participants' musical abilities and spontaneous motor tempo, motivated by recent studies showing that rhythmic entrainment can vary based on individual differences in these skills (Bauer et al., 2015; Snapiri at al., 2023). For example, Snapiri et al. (2023) explored the relationship between discrimination performance in a visual rhythmic task and motor tempo, assessed through a spontaneous tapping task. They found that participants with slower spontaneous tempi benefited more from rhythmic manipulation than those with faster tempi, suggesting that individual differences in spontaneous motor tempo may influence the effectiveness of rhythmic benefits. Therefore, the inclusion of musical tests and the spontaneous tapping task was explorative.

In the second experiment, we used a pitch comparison task where participants estimated on a trial-by-trial basis if a comparison tone was lower or higher in pitch compared to a series of standard tones, which could be presented either regularly or irregularly. Here, we also expected to replicate the inverted U-shaped performance pattern in task accuracy, as predicted by DAT and seminal studies (e.g., Jones et al., 2002). The foreperiod between the last standard tone and the target was also manipulated to explore the relationship between entrainment and foreperiod effects. Additionally, in both experiments, we capitalized on eye-tracker measures to better characterize online cognitive processes and shed more light on the physiological processes associated with rhythmic entrainment. Following previous studies of pupillometry/saccade rate and rhythmic tasks (Dankner et al., 2017; Shalev & Nobre, 2022), we predicted a priori a phase-specific increase in pupil dilation (Experiments 1 and 2) and a reduction in saccade rate (Experiment 2) before target onset, as an index of entrainment with moments in phase with the rhythm. To preview the results, although we observed clear rhythmicity and foreperiod effects, we did not replicate the main predictions of DAT across both experiments, thus aligning with previous studies that call for a critical reappraisal of entrainment attention theories.

2. Experiment 1

2.1. Method

2.1.1. Participants

Two power analyses with different analytic methods were conducted to ensure enough statistical power. Considering the significant main effect of target onset observed in Experiment 1 by McAuley and Fromboluti (2014; η_p^2 = .26/Cohen's f = 0.59), we conducted a power analysis using the *Superpower* R package (Lakens & Caldwell, 2021) with an α of .05. With these parameters, a sample size of 17 participants is needed for a power of .82. Moreover, considering the observed difference between in-phase and late targets (Cohen's d_z = 0.47; the early–in-phase difference was slightly higher, d_z = 0.51), a sample size of 30 participants is needed for a power of .81 in a one-sided paired t test (α = .05). Therefore, we decided to collect data from a final sample of 36 participants to compensate for possible data rejection due to poor general performance or technical problems.

A final sample of 36[1] students from the University of Granada participated in the study (**Table 1**). All of them signed an informed consent form before the experiment and received €15 for their participation. They self-reported normal or corrected-to-normal vision and normal hearing and were naïve to the purpose of the experiment. The study was conducted at the Mind, Brain and Behavior Research Center (CIMCYC) of the University of Granada during the Spring of 2023, following the ethical guidelines laid down by the University of Granada (protocol number: 2488/CEIH/2021) and in accordance with the ethical standards of the 1964 Declaration of Helsinki (last update: Brazil, 2013).

Unlike McAuley and Fromboluti's study (2014), we detected outlying participants with poor performance (i.e., mean accuracy, and reaction time, RT, for RT analyses) in terms of the standard deviation from the mean (> 2), studentized residuals (> 2), and Cook's D_i (> 4/n)[2].

2.1.2. Stimuli and apparatus

Following McAuley and Fromboluti (2014), we used a 350-ms 440-Hz sine tone as standard and a 350-ms 880-Hz sine tone as the oddball target. For both stimuli, 44.1 kHz was the sampling frequency. All the sounds were created in Audacity, with a 10 ms rise and fall times. Throughout the entire task, a fixation cross was presented in the center of the screen as a reference point to keep the gaze stable on the screen and avoid artifacts in the eye-tracker measurement. Stimulus presentation and response collection were controlled by E-Prime 2.0 software running on a standard Pentium 4 PC and a 17-inch widescreen monitor with a 1,280 × 1024-pixel resolution and 60 Hz refresh rate. Temporal accuracy of event onsets was checked by recording the audio of the trials with an external device. For the presentation of the auditory stimuli, participants wore headphones (Philips SHP2000). A video-based eye tracker (EyeLink 1000, SR Research, Ontario, Canada) with a chin rest was used to measure pupil diameter as well as to monitor eye movements and blinks at 1000 Hz. The recorded data was saved to a different computer.

Musical skills were assessed with the Scale test from the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003), and the Out-of-key and Off-beat tests from the Online Test of Amusia (OTA; Peretz et al., 2008). The three tests comprised 30 melodies specifically composed for the MBEA to ensure being unfamiliar to participants. The melodies are in major key and played at a tempo of 120 beats/min. Spontaneous motor tempo (SMT) was measured with the spontaneous finger-tapping task, which was conducted using E-Prime 2.0 software (Psychology Software Tools, Inc.). Stimuli for the SMT consisted of a fixation cross presented in the center of the screen.

2.1.3. Procedure

First, participants carried out two separate blocks of the spontaneous finger-tapping task, which consisted of pressing the spacebar of the computer keyboard with the index finger of their dominant hand at their preferred rate for 60 s since the fixation cross appeared on the center of the screen. Then, they completed the Scale test from MBEA and the Out-of-key and Off-beat tests from the OTA, followed by the other two blocks of the spontaneous finger-tapping task (i.e., four finger-tapping blocks in total to ensure a reliable measure of SMT). Afterward, participants performed an oddball task similar to that used by McAuley and Fromboluti (2014), with some minor differences detailed below. As in the original study, participants listened to a sequence of standard tones in which an oddball tone was embedded. Their task was to indicate whether the duration of the oddball tone was shorter or longer than the standard tone by pressing the "z" or "m" keys on a computer keyboard (key–response mapping was counterbalanced across participants). The oddball varied in duration compared to the standard stimuli (i.e., 350 ms) by 350 ms ± 5%, 10%, 15%, or 20%. Unlike McAuley and Fromboluti (2014), we eliminated the +0% condition (where the oddball matched the standard duration) to ensure that all oddballs differed from the standard tones. This change was made because the task always required a shorter/longer comparison rather than a same/different comparison, thus forcing participants to guess in the 0% condition, rather than correctly categorize targets.

Table 1. Demographic information of the sample in Experiment 1.

Age	23.8 years, *SD* = 5.6, range 18–42
Sex	18 women and 18 men
Handedness	33 right-handed and 3 ambidextrous
Nationality	31 Spanish, 2 Italian, 1 Colombian, 1 Cuban, and 1 Peruvian
Musicianship	10 with long-term musical training and 26 without

[1] Data from four more participants were collected but not included in the sample due to technical problems during the task or dropping out of the task.
[2] Note that the thresholds selected for the *SD* and residual criteria deviated slightly from those preregistered to make detection somewhat less conservative. However, the main results were not affected by this change or even by the inclusion of outliers in the analyses.

Two types of blocks were administered based on the IOI spacing the standard stimuli: fixed blocks, with a constant IOI of 700 ms between tones, and variable blocks, with IOIs uniformly varying between 400 and 1000 ms, also with an average of 700 ms. Whereas fixed and variable blocks were manipulated in two separate experiments in the original study (i.e., Experiments 1 and 3, respectively), we manipulated the type of block within participants. In both block types, the oddball could appear in the 5th, 6th, 7th, 8th, or 9th position in the sequence, whereas the original study only used four positions (5th, 6th, 7th, and 8th). We added a 9th position to have an odd number of positions for the analysis (see the Design and Data analysis section). The IOI preceding the oddball could be early, in phase, or late relative to the standard tones (i.e., 469, 700, and 931 ms, respectively; see **Figure 1**). Therefore, the combination of block type (fixed vs. variable), oddball position (5th, 6th, 7th, 8th, and 9th), and oddball onset (early, in-phase, and late) resulted in a 2 × 5 × 3 within-participants design.

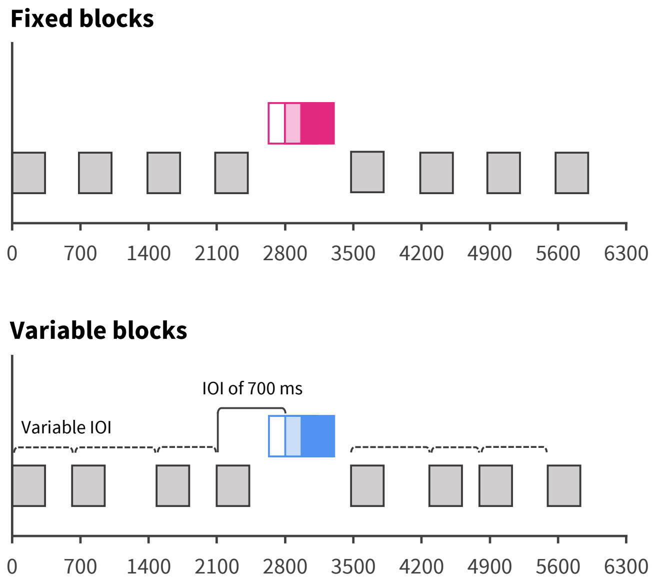

Figure 1. Graphical representation of the duration discrimination task in Experiment 1. Gray squares represent the sequence of standard tones, while colored squares represent oddball tones. For illustration purposes, only the eight standard tones in one trial are shown, although the actual sequence was continuous without an intertrial interval (which leads to the perception that the last standard tones are connected to the first standards of the next trial). In both fixed and variable blocks, the oddball appeared either 700 ms after the preceding standard tone (i.e., in-phase), or early or late (i.e., out-of-phase). The darker the color of the oddball tone, the later its onset. Additionally, oddball tones could occur in the 5th (as shown in the figure), 6th, 7th, 8th, or 9th position. IOI stands for inter-onset interval.

Participants carried out fixed and variable blocks in a counterbalanced order (i.e., half of the participants started with two fixed blocks (the IOIs between standard tones were 700 ms) followed by two variable blocks (with a variable IOI between 400 and 1000 ms). For eye-tracker measures, the dominant eye of the participants was determined with the hole-in-the-card test. After measuring the participants' visual dominance, the experimental task was delivered, starting with three blocks of 20 practice trials, followed by two fixed and two variable experimental blocks as explained above. A short break with a fixed interval of 15 s between the 1st and the 2nd blocks, and between the 3rd and the 4th blocks was allowed. After the 2nd block (just before the start of the second block type) and the 4th block, participants were asked to rate the difficulty of the two previous blocks (from 1, extremely easy, to 9, extremely difficult). The experiment comprised a total of 480 experimental trials equally divided between the fixed and variable blocks. Finally, participants filled in an online demographic and musical questionnaire. The average duration of the experiment was 90 minutes.

The questionnaire, conducted through the LimeSurvey platform, included general sociodemographic questions (i.e., age, sex, handedness, nationality) and questions on the general state of health (e.g., presence of vision and hearing problems, medical and psychiatric illnesses, use of medications, alcohol, and other drugs that might affect performance), and musical experience such as years of musical training, instruments played, hours of daily practice, and hours of listening.

2.1.4. Preregistered hypotheses

Taking Experiments 1 and 3 by McAuley and Fromboluti (2014) as reference, we expected to observe higher temporal accuracy (measured as Distortion Duration Factor/DDF close to 1) for targets presented in phase compared to targets presented earlier and later only in rhythmic blocks (*preregistered hypothesis 1*).

Moreover, we decided to extend the analyses of the reference study to other classic behavioral measures in which temporal and rhythmic context effects have been observed (e.g., Chang et al., 2019; Jones et al., 2002). We expected to observe a general effect of target position in RTs and percentage of correct responses, with faster and more accurate responses for later positions (i.e., *foreperiod effect*; Niemi & Näätänen, 1981). We also predicted significant differences between fixed and variable blocks, with overall faster and more accurate responses for fixed as compared to variable blocks (*general block effect*), and, as predicted by DAT, a local improvement for in-phase targets in fixed blocks (i.e., properly, *rhythmic entrainment effect*; formalized here as a block-by-target onset interaction): lower RTs and higher accuracy for targets appearing at the moment predicted by the rhythmic sequence (*preregistered hypothesis 2*). Variable blocks were anticipated to show only a foreperiod-like pattern.

According to recent eye-tracker evidence (Shalev & Nobre, 2022), pupil size was expected to increase prior to the target and decrease after it, resembling a pupillometric foreperiod effect. We also expected a general effect of type of block, with overall smaller pupil size in fixed blocks as compared to variable blocks, suggesting reduced attentional demands in blocks with regular sequences. Finally, we anticipated more flexible dynamics of pupil size in fixed blocks (i.e., steeper reduction at the beginning of the trial and a steeper increase just before the target onset; *preregistered hypothesis 3*).

Musical skills and SMT might predict the effect of rhythmic entrainment (i.e., the difference between in-phase vs. out-of-phase in fixed blocks). Therefore, we expected a higher entrainment effect in participants with higher musical skills and with spontaneous tempo closer to the entrained frequency

(1.43 Hz; *preregistered hypothesis 4*), as can be derived from DAT and the *preferred period hypothesis* (McAuley et al., 2006). Another possibility is the result observed by Snapiri et al. (2023), in which individuals with slower spontaneous tempi exhibited greater differences between in-phase and out-of-phase targets, independently of the task tempo.

2.1.5. Design and data analysis

The design of the present study is a within-participant factorial design with 2 (type of block: fixed vs. variable blocks) × 5 (target position: 5th, 6th, 7th, 8th, and 9th position) × 3 (target onset: early/IOI of 469, in phase/IOI of 700, and late/IOI of 931 ms) factors. In-phase onset was made the reference level for the categorical variable *target onset*. On the other hand, the target position was mean-centered to make the intercept meaningful[3].

Analyses of hypothesis 1

Like the reference study, the proportion of "longer" responses for each of the eight target durations (350 ms ± 5, 10, 15, 20%) in each target onset condition and each type of block served to estimate PSE and temporal discrimination (i.e., just noticeable difference; JND). First, "longer" response probabilities were transformed into *z* scores[4] and used to estimate PSE by fitting a regression model with *z* scores as the dependent variable and target duration as a covariate (i.e., PSE = $-\beta_0/\beta_1$, where β_0 represents the model intercept and β_1 the covariate coefficient; Macmillan & Creelman, 2005). PSE measures the duration for which a participant makes shorter and longer responses 50% of the time (i.e., the duration at which the target is perceived to be the same as the standard). PSE was used to compute a duration distortion factor (DDF) by calculating the ratio of the actual duration of the standard (i.e., 350 ms) to PSE. Following McAuley and Fromboluti (2014), we expected a DDF closer to 1 (PSE is equal to the duration of the standard with no systematic distortion) for targets presented in-phase in fixed blocks, while a DDF below 1 and above 1 for earlier and later onsets, respectively (i.e., underestimation and overestimation), suggesting that the perceived target duration is more accurate when the temporal onset of the oddball occurs in-phase relative to the entrained rhythm. By contrast, we expected that variable blocks should elicit similar DDF across all target onsets, supporting that irregularly timed (arrhythmic) sequences weaken the onset timing effects.

On the other hand, JND, representing the perceptual discrimination (or the slope of the psychometric curve), was estimated by transforming the proportion of "longer" responses for each of the eight target durations to z scores. As in McAuley and Fromboluti (2014), a linear regression with z scores as the dependent variable and target duration as a covariate was conducted. JND was estimated as the ratio of the z score for the 75th percentile between the covariate coefficient in the regression model (i.e., $0.6745/\beta 1$). Following McAuley and Fromboluti (2014), JND was expressed in relative terms as a percentage of the standard duration (dividing the obtained JND by the standard duration and multiplying by 100). In contrast to DAT, McAuley and Fromboluti observed no modulation of JND, such that we a priori hypothesized to replicate a similar null finding.

Following the analytical approach of McAuley and Fromboluti (2014), for *Hypothesis 1*, two linear mixed-effects models were conducted with the R packages *lme4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017), and *afex* (Singmann et al., 2016), one with DDF and another with JND as the dependent variable, target onset, block type as fixed factors, and their interaction as fixed factors, and participants as a random factor[5]. Cohen's f^2 was used as an estimate of effect size.

To test the robustness of our findings, we additionally conducted (non-preregistered) alternative analyses. First, we conducted the exact analysis as in the original work, a two-way repeated-measures ANOVA with type of block and target onset as factors, and DDF or JND as a dependent measure in each case. For this type of analysis, η_p^2 was used as an estimate of effect size. Second, we used the Bayesian homologous of the previous two-way repeated-measures ANOVA with JASP (0.19.0; JASP Team, 2024) to characterize null results obtained with frequentist analyses (i.e., if there is evidence in favor of the null effect, Bayes factor giving the evidence for H1 over H0 or $BF_{10} < 0.33$, or inconclusive evidence, $0.33 \leq BF_{10} < 3$). Third, we fitted a logistic non-linear mixed-effects model with *nlme* R package (Pinheiro et al., 2023), including type of block and target onset as fixed effects, and intercepts for participants for the two curve parameters (i.e., midpoint and slope) as random effects.

Analyses of hypothesis 2

For *Hypothesis 2*, a linear and a binomial generalized linear mixed-effects model on single-trial data were respectively conducted for correct RT and accuracy. Block type, target onset, target position, and all their interactions were included as fixed factors, and participants as a random factor. In RT analysis, correct responses slower or faster than 2 *SD*s of the sample's mean were not included (i.e., outlier trials).

Analyses of hypothesis 3

Regarding pupillometry measures, blink intervals were identified following Hershman et al.'s procedure (2018) and removed from the analyses. Additional strange values were identified by removing groups of samples from trials with extreme values (> 2 *SD*s of block mean) and outlying individual samples within trials (> 2 *SD*s of the trial mean). Data were smoothed using a Hanning window of 50 ms and missing values below an interval of 333 consecutive samples were interpolated using a cubic spline interpolation. Afterward, raw pupil size was converted to *z* scores, by calculating the mean and standard deviation for each participant. We examined the differences between fixed and variable blocks and their temporal evolution in pretarget

[3] Without mean-centering target position, the intercept of the model would correspond to the position 0, which is meaningless. Instead, by mean-centering, the 0 in the transformed position variable corresponds to the 7th position, in the middle between 5 and 9.

[4] For response proportions of 0 and 1, we respectively used $1/(2n)$ and $1 - 1/(2n)$, where n is the number of observations.

[5] Other structures of random effects led to similar conclusions. For the sake of consistency across analyses and given some models did not reach convergence, we selected participant's intercept as a random effect in all the analyses.

samples by conducting two mixed-effects models, one with the 1st half of the trial samples and another with the 2nd half, with block type (fixed vs. variable) and time as fixed covariates, and participants as a random factor. This division would capture the two preregistered temporal trends prior to the target: a decrease in pupil size at the beginning of the trial (and after the previous target), and an increase in pupil dilation prior to the target. Note that the interval from the beginning of the trial to the target onset varied from trial to trial (from 2569 to 5831 ms). In addition, we compared moment-to-moment differences between the two blocks (i.e., millisecond-by-millisecond) using a cluster-based permutation test with 1,000 permutations and a one-tailed t test for each moment). The analysis was conducted using the R package permutes (Voeten, 2023), which is designed to perform cluster-based permutation tests for mixed-effects models. Rather than examining each time point individually, this method identifies clusters of consecutive time points where the t-statistic of the mixed-effects model exceeds a specified threshold. These clusters are then compared to a null distribution generated from the permuted data. By focusing on cluster-level statistics, the approach helps to control for two critical issues in time-series data: temporal autocorrelation and multiple comparisons across time points. Second, in a (non-preregistered) target-locked analysis, we examined the differences between blocks and between in-phase and out-of-phase onsets (–100 to 0 ms relative to the target onset). A mixed-effects model with block and target onset as fixed variables was conducted in this temporal window. Finally, pupil size across the whole experiment was compared between both blocks also with a linear mixed-effects model.

To test in a complementary way that the differences between blocks were not fully accounted for by objective and subjective measures of task difficulty, we ran a (non-preregistered) linear mixed-effects model with participants' RT and accuracy and reported difficulty as predictors in addition to the type of block. This model was conducted with pretarget pupil-size samples as the dependent variable.

Analyses of hypothesis 4

To assess the influence of spontaneous motor preference and musical skills over rhythmic entrainment, we conducted exploratory (non-preregistered) bivariate correlations. To test whether deviance from the entrained frequency or simply SMT was linked to the rhythmic entrainment (indexed as the difference between in-phase and late-onset behavioral values in fixed blocks), we used mean inter-tap interval (ITI) for each participant and its absolute deviance from the fixed IOI (i.e., 700 ms). We selected the difference between in-phase and late-onset moments because its direction clearly indicates if the behavioral pattern was consistent with a foreperiod-like pattern (i.e., higher accuracy and lower RT in late onset, so positive and negative difference values, respectively) or with DAT (i.e., the reverse is predicted under DAT, a negative difference for accuracy and a positive difference for RTs). Conversely, the patterns of differences between early and in-phase targets predicted by the foreperiod-like pattern and DAT are the same (i.e., slower responses and lower accuracy for early onset). In addition, the sum of MBEA and OTA test scores was used as an index of musical skills. Importantly, this series of analyses had a secondary role in our study and so our power analysis did not contemplate them. Thus, we expected that our correlational analyses were underpowered and therefore had an exploratory value.

2.2. Results

Within our sample of 36 participants, three outliers with extremely low accuracy were identified ($M \approx .60$ vs. whole-sample $M = .78$) and two extra participants based on their disparate RTs ($M > 1000$ ms vs. whole-sample $M = 800$ ms). While the former group of outliers was removed from all the analyses, participants with extreme RTs were only excluded from the RT analysis. Regarding ITI, two participants were not included in the correlation analyses with the two measures of SMT (mean ITI and absolute deviance from 700 ms) given their extreme ITI mean values (> 1800 ms; vs. $M = 677$ ms, $SD = 296$). Finally, two participants were excluded from eye tracker analyses due to an excess of samples identified as blinks and missing values (40% and 99%; M of the included sample = 9.5%, $SD = 6.2$).

2.2.1. Perceived difficulty

The perceived difficulty laid slightly below the midpoint of the full range of the scale [$M = 3.1$, $SD = 0.7$; scale; contrast against 4.5: $t(32) = -11.98$, $p < .001$, $d_z = -2.08$]. Interestingly, there was no difference in the perceived difficulty between fixed and variable blocks, $t(32) = -0.47$, $p = .640$, $d_z = -0.08$, and perceived difficulty did not significantly correlate with JND, $r(31) = .21$, $p = .237$, DDF, $r(31) = -.11$, $p = .545$, overall accuracy, $r(31) = -.20$, $p = .276$, and RT, $r(29) = .20$, $p = .282$.

2.2.2. Duration distortion factor and just noticeable difference

In contrast to McAuley and Fromboluti (2014), the preregistered mixed-effects model with DDF showed no general effect of target onset, $F(2, 160) = 0.71$, $p = .493$, $f^2 < 0.01$ (early: 1.01; in-phase: 1.02; late: 1.02), and an overall overestimation bias (intercept of the model = 1.02; test against 1: $t(45.69) = 2.10$, $p = .041$; **Figure 2**). The overall effect of the type of block was also not significant, $F(1, 160) = 0.57$, $p = .450$, $f^2 < 0.01$ (fixed: 1.01 vs. variable: 1.02). However, there was a numerical, non-significant trend in the type of block-by-target onset interaction, $F(2, 160) = 2.35$, $p = .099$, $f^2 < 0.01$. Two separate mixed-effects models for each type of block suggest that the interaction was driven by lower DDF at late target onset compared to in-phase targets in the variable blocks, $\beta = -0.02$, $p = .034$ (estimated values: 1.01 vs. 1.03, respectively), whereas early onset did not differ in DDF, $\beta = -0.006$, $p = .518$ (1.02 vs. 1.03). Only DDF for late targets did not significantly differ from the maximum accuracy value (DDF = 1; tests against 1: early, $t(32) = 2.14$, $p = .040$, $d_z = 0.37$; in-phase, $t(32) = 2.63$, $p = .013$, $d_z = 0.46$; late, $t(32) = 0.60$, $p = .552$, $d_z = 0.10$). On the other hand, DDF showed no modulation with target onset in fixed blocks, $F(2, 64) = 1.35$, $p = .266$, $f^2 < 0.01$ (intercept of the model = 1.02; test against 1: $t(51.82) = 1.37$, $p = .176$; all one-sample t-tests with ps > .050).

These findings were confirmed using the original analysis (i.e., two-way repeated-measures ANOVA), a two-way repeated-measures Bayesian ANOVA, and a logistic non-linear mixed-effects model (**Table 2**; **Figure 3**). The logistic model showed again a general pattern of overestimation: estimated PSE = 333.39 ms/DDF = 1.05 [test against 350 ms: $t(15369) = -4.16$, $p < .001$].

As in McAuley and Fromboluti (2014), the three analytical models yielded no overall effect or interaction concerning JND. Therefore, rhythmic context did not impact duration discrimination, with a non-significant type of block-by-target onset interaction, $F(2, 160) = 0.53$, $p = .592$, $f^2 < 0.01$.

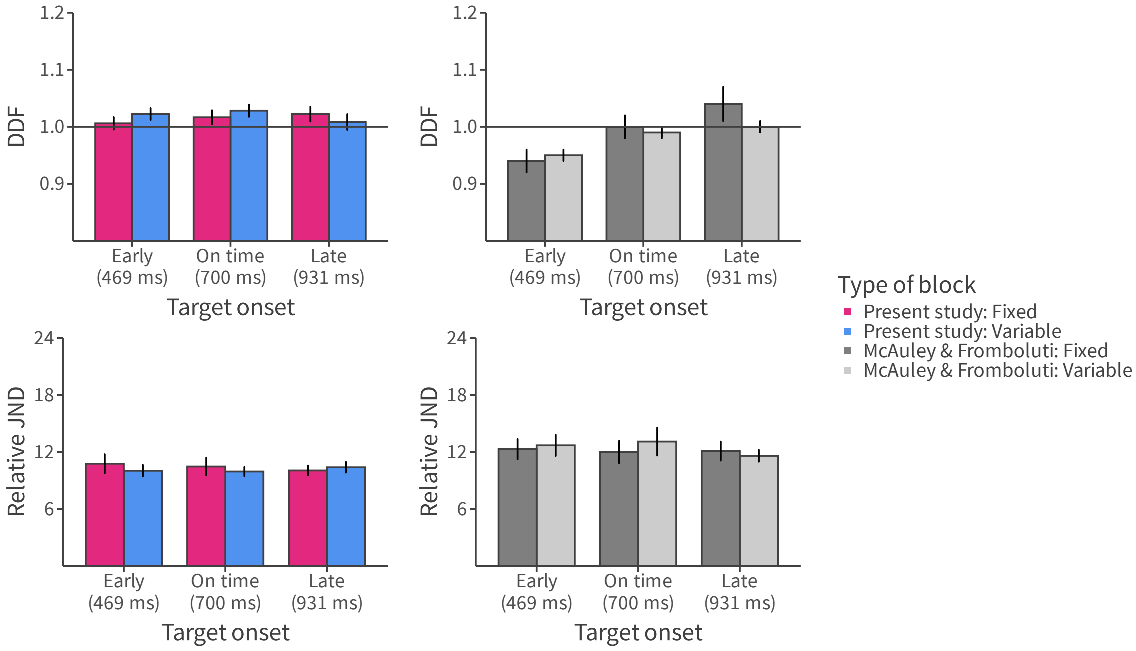

Figure 2. Means of temporal accuracy (Duration Distortion Factor, DDF) and discrimination (Just Noticeable Difference, JND) in Experiment 1 compared to the means of the original study by McAuley and Fromboluti (2014). Error bars represent standard error of the means.

Table 2. Results of the three models proposed for DDF and JND analyses in Experiment 1.

DDF/midpoint	*Linear mixed-effects model*	
	Type of block	$F(1, 160) = 0.57$, $p = .450$, $f^2 < 0.01$
	Target onset	$F(2, 160) = 0.71$, $p = .493$, $f^2 < 0.01$
	Type of block × target onset	$F(2, 160) = 2.35$, $p = .099$, $f^2 < 0.01$†
	Two-way repeated-measures ANOVA (BF_{10} from the Bayesian homologous)	
	Type of block	$F(1, 32) = 0.34$, $p = .567$, $\eta_p^2 = .01$; $BF_{10} = 0.21$
	Target onset	$F(2, 64) = 0.61$, $p = .544$, $\eta_p^2 = .02$; $BF_{10} = 0.10$
	Type of block × target onset	$F(2, 64) = 4.83$, $p = .011$*, $\eta_p^2 = .13$; $BF_{10} = 0.63$
	Logistic non-linear mixed-effects model	
	Type of block	$F(1, 15369) = 2.10$, $p = .147$, $f^2 < 0.01$
	Target onset	$F(2, 15369) = 7.31$, $p < .001$, $f^2 < 0.01$*
	Type of block × target onset	$F(2, 15369) = 3.04$, $p = .048$, $f^2 < 0.01$*
JND/slope	*Linear mixed-effects model*	
	Type of block	$F(1, 160) = 0.47$, $p = .493$, $f^2 < 0.01$
	Target onset	$F(2, 160) = 0.08$, $p = .927$, $f^2 < 0.01$
	Type of block × target onset	$F(2, 160) = 0.53$, $p = .592$, $f^2 < 0.01$
	Two-way repeated-measures ANOVA (BF_{10} from the Bayesian homologous)	
	Type of block	$F(1, 32) = 0.17$, $p = .682$, $\eta_p^2 < .01$; $BF_{10} = 0.19$
	Target onset	$F(2, 64) = 0.13$, $p = .876$, $\eta_p^2 < .01$; $BF_{10} = 0.06$
	Type of block × target onset	$F(2, 64) = 0.96$, $p = .390$, $\eta_p^2 = .03$; $BF_{10} = 0.15$
	Logistic non-linear mixed-effects model	
	Type of block	$F(1, 15369) = 1.74$, $p = .188$, $f^2 < 0.01$
	Target onset	$F(2, 15369) = 0.11$, $p = .900$, $f^2 < 0.01$
	Type of block × target onset	$F(2, 15369) = 1.65$, $p = .192$, $f^2 < 0.01$

Note: * $p < .05$; † $.05 \leq p < .10$.

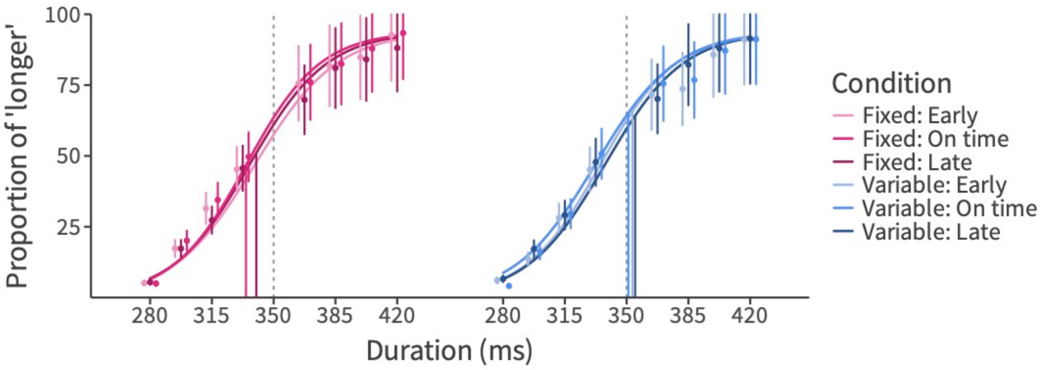

Figure 3. Estimated psychometric curves as a function of type of block and target onset in Experiment 1. Vertical lines represent PSE for each condition, while the dotted line is the actual duration of standards (i.e., 350 ms). Points and error bars represent across-participants mean and standard error for each duration value.

2.2.3. Reaction time, accuracy, and misses

No other behavioral measure showed a type of block-by-target onset interaction (**Table 3**; **Figure 4**). While the overall reaction time (RT) and percentage of correct responses were respectively 803 ms and 79%, no foreperiod effect (i.e., overall effect of target position) or rhythmic entrainment was observed for these dependent variables. Only a type of block effect appeared for RT, $F(1, 10968.43) = 9.32$, $p = .002$, $f^2 < 0.01$, with slightly faster responses during fixed blocks (fixed: 801 ms vs. variable: 806 ms). On the other hand, the mixed-effects model with misses predicted a reduction in the proportion of misses with later target positions consistent with a foreperiod pattern, $\chi^2(1) = 26.54$, $p < .001$, $f^2 = 0.08$ (5th position: 3.4% vs. 9th position: 2.0%). Moreover, the proportion of misses was lower for targets with an early onset, $\chi^2(2) = 52.64$, $p < .001$, $f^2 = 0.06$ (early: 1.1%; in-phase: 2.6%; late: 2.9%), and this pattern was more pronounced for early and in-phase targets than late onsets (early: $\beta = -0.32$; in-phase: $\beta = -0.34$; late: $\beta = -0.30$). Nevertheless, no rhythmic entrainment was found, $\chi^2(2) = 3.20$, $p = .202$, $f^2 < 0.01$.

2.2.4. Pupillometry

Eye-tracker measures showed that pupil size was overall larger for variable blocks than for fixed blocks, both at the trial level [pupil size previous to the target, 1st half: $F(1, 173566) = 3097.29$, $p < .001$, $f^2 = 0.02$; 2nd half: $F(1, 173566) = 2442.72$, $p < .001$, $f^2 = 0.04$; **Figure 5A**] and across the whole task [$F(1, 151138) = 818.93$, $p < .001$, $f^2 < 0.01$; **Figure 5B**]. Before the target, pupil size decreased along the trial [1st half: $F(1, 173566) = 4128.09$, $p < .001$, $f^2 = 0.02$; 2nd half: $F(1, 173566) = 89.57$, $p < .001$, $f^2 = 0.02$], and this happened more pronouncedly in the fixed blocks [1st half: $F(1, 173566) = 21.94$, $p < .001$, $f^2 < 0.01$; 2nd half: $F(1, 173566) = 5.97$, $p = .015$, $f^2 = 0.02$]. This trend appeared in place of a pattern of pupil increase, which would have been consistent with a foreperiod preparation pattern. On a target-locked analysis, there was no difference between in-phase and out-of-phase target, $F(1, 12490) = 0.02$, $p = .889$, $f^2 < 0.01$, but pupil size was again smaller for fixed blocks, $F(1, 12490) = 739.97$, $p < .001$, $f^2 = 0.06$, independently of the target onset (i.e., no block-by-onset interaction), $F(1, 12490) = 0.04$, $p = .849$, $f^2 < 0.01$ (**Figure 5C**). Interestingly, the difference between blocks in the pretarget range (i.e., samples previous to the target onset) was significant even after adding participants' RT and accuracy, and reported difficulty to the model as predictors, $F(1, 324718) = 14830.54$, $p < .001$, $f^2 = 0.07$, which suggests that the effect cannot be ruled out by performance on the task and perceived difficulty.

2.2.5. Correlations with spontaneous motor tempo and musical skills

Among the correlations conducted to test the relationship between rhythmic entrainment and individual differences in motor preference and musical skills, two **correlations** were statistically significant: ITI–modulation on accuracy, $r(29) = -.36$, $p = .046$, and musical skills–modulation on RT, $r(28) = .42$, $p = .021$ (**Table 4**; **Figure 6**). In accordance with previous literature (Snapiri et al., 2023), participants with slower tempo preferences showed greater rhythmic entrainment in their accuracy, rather than displaying a pattern that reflected a relationship between their tempo preference and its proximity to the entrained frequency, $r(29) = -.12$, $p = .507$. In addition, participants with higher perceptual musical skills showed a more pronounced entrainment pattern than participants with lower musical scores in RT measures.

Table 3. Results of the linear (RT) and binomial generalized linear (accuracy and misses) mixed-effects models in Experiment 1.

Effect	RT	Accuracy	Misses
Type of block	$F(1, 10968.43) = 9.32$, $p = .002$, $f^2 < 0.01$*	$\chi^2(1) = 0.05$, $p = .826$, $f^2 < 0.01$	$\chi^2(1) = 0.21$, $p = .646$, $f^2 < 0.01$
Target onset	$F(2, 10968.20) = 0.90$, $p = .405$, $f^2 < 0.01$	$\chi^2(2) = 0.14$, $p = .931$, $f^2 < 0.01$	$\chi^2(2) = 52.64$, $p < .001$, $f^2 < 0.01$*
Target position	$F(1, 10968.20) = 0.30$, $p = .582$, $f^2 < 0.01$	$\chi^2(1) = 1.70$, $p = .192$, $f^2 < 0.01$	$\chi^2(1) = 26.54$, $p < .001$, $f^2 < 0.01$*
Block × onset	$F(2, 10968.14) = 0.43$, $p = .649$, $f^2 < 0.01$	$\chi^2(2) = 0.64$, $p = .726$, $f^2 < 0.01$	$\chi^2(2) = 3.20$, $p = .202$, $f^2 < 0.01$
Block × position	$F(1, 10968.11) = 1.58$, $p = .209$, $f^2 < 0.01$	$\chi^2(1) = 0.03$, $p = .866$, $f^2 < 0.01$	$\chi^2(1) = 0.67$, $p = .415$, $f^2 < 0.01$
Onset × position	$F(2, 10968.13) = 0.41$, $p = .662$, $f^2 < 0.01$	$\chi^2(2) = 0.46$, $p = .795$, $f^2 < 0.01$	$\chi^2(2) = 5.03$, $p = .081$, $f^2 < 0.01$†
Block × onset × position	$F(2, 10968.15) = 0.48$, $p = .621$, $f^2 < 0.01$	$\chi^2(2) = 0.72$, $p = .698$, $f^2 < 0.01$	$\chi^2(2) = 2.35$, $p = .309$, $f^2 < 0.01$

Note: * $p < .05$; † $.05 \leq p < .10$.

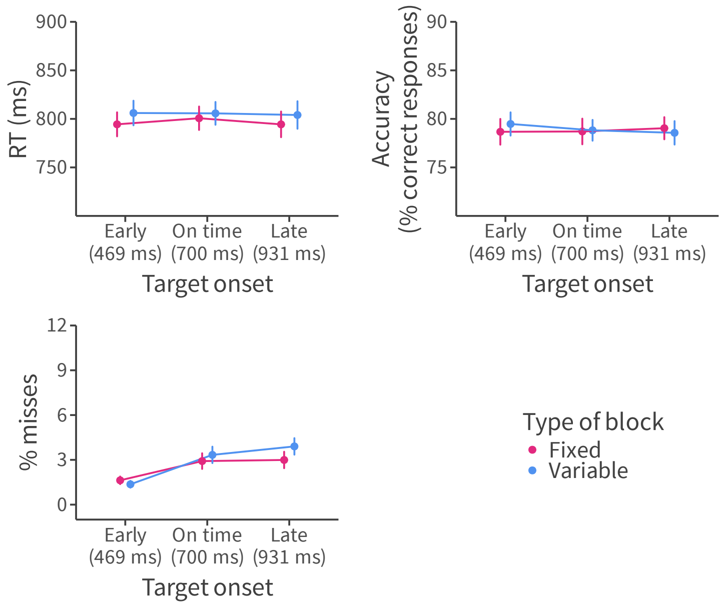

Figure 4. Means of reaction time, accuracy, and percentage of misses as a function of type of block and target onset in Experiment 1. Error bars represent standard error of the means.

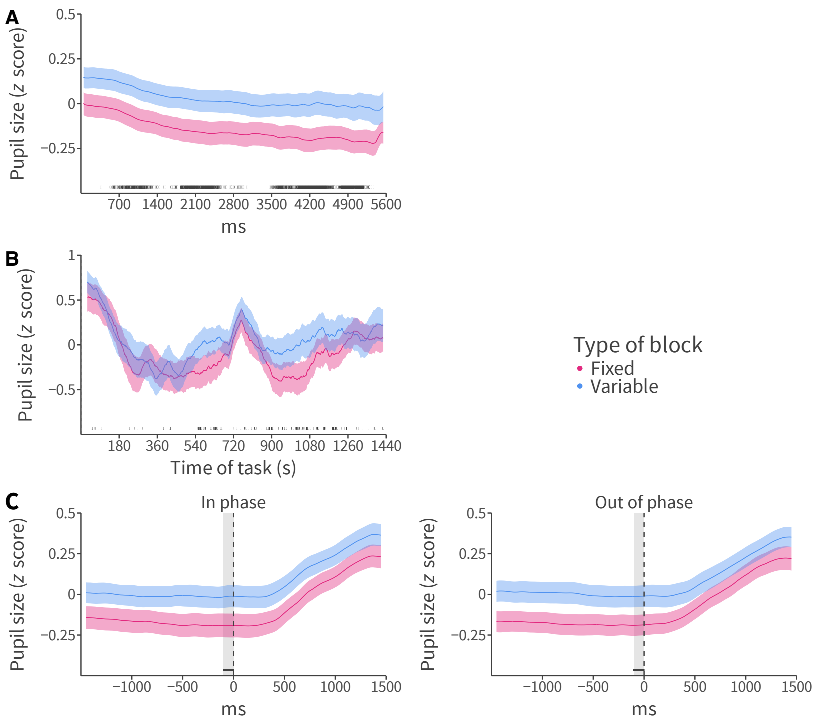

Figure 5. Pupil size (**A**) in temporal samples previous to target onset, (**B**) across the whole task, and (**C**) in target-locked samples (1500 ms before and after target onset) for both in-phase and out-of-phase target in Experiment 1. Black points at the bottom of the plots depict moments in which the difference between both conditions had a $p < .05$ (in a one-sided test). Vertical grey band in panel C indicates the range (from −100 to 0 ms) in which the differences between blocks were tested. For illustrative purpose only, values were smoothed using a sliding window of 100 ms in panels A and C, and a sliding window of 10 s (~ two trials) in panel B. The positive middle peak in panel B represents the beginning of the second block of task, which occurred after a short break, suggesting recovery from fatigue.

2.3. Discussion

Contrary to the predictions of DAT and our preregistered hypotheses following McAuley and Fromboluti (2014), the results did not support a behavioral pattern of rhythmic entrainment. The main prediction of DAT would be an inverted U-shaped pattern in just noticeable difference (JND; see Figure 2 in Henry & Herrmann, 2014). However, unnoticed during the preregistration phase and by the reference study, temporal discrimination (JND) was similar for oddball tones presented in phase and out of phase with a stream of isochronous standard tones. Therefore, the findings in the present experiment as well as the results of McAuely and Fromboluti (2014) contradict the hypothesized attention-driven perceptual enhancement by DAT. Similarly, other behavioral measures that were not explored in the reference study (i.e., RT, overall percentage of correct responses, and misses) did not show benefits for in-phase targets, suggesting the absence of attentional entrainment and a comparable level of response preparation and perceptual discrimination for stimuli appearing in phase.

On the other hand, McAuley and Fromboluti (2014) observed higher temporal accuracy (DDF close to 1) for oddballs presented in phase, as opposed to targets with an early and late onset, whose durations were underestimated and overestimated, respectively. Although this finding was interpreted as an index of attentional entrainment, an alternative explanation is possible considering the new data in our study. It is plausible that the inclusion of an intertrial interval in McAuley and Fromboluti's task (not explicitly reported in the manuscript)[6] made clear the beginning and end of each trial, thereby producing an increase in the level of attention across the possible temporal target positions (i.e., foreperiod effect on DDF). This might explain why McAuley and Fromboluti (2014) observed that later positions yielded longer perceived durations (see right panel of **Figure 2**). A foreperiod-like pattern in DDF could be transferred locally to each position and explain the progressive increase in perceived duration in both rhythmic and arrhythmic blocks, without the need to invoke an attentional entrainment account.

Table 4. Bivariate correlations between behavioral outcomes in the duration discrimination task and individual differences in inter-tap interval and musical skills.

Rhythmic modulation	Absolute deviance	ITI	Total musical score
Accuracy	$r(29) = -.12, p = .507$	$r(29) = -.36, p = .046$*	$r(29) = -.02, p = .908$
RT	$r(28) = .26, p = .159$	$r(28) = .26, p = .167$	$r(28) = .42, p = .021$*
DDF	$r(29) = .28, p = .134$	$r(29) = -.19, p = .317$	$r(29) = -.32, p = .072$†
JND	$r(29) = -.03, p = .857$	$r(29) = .24, p = .185$	$r(29) = .03, p = .851$

Note: * $p < .05$; † $.05 \leq p < .10$.

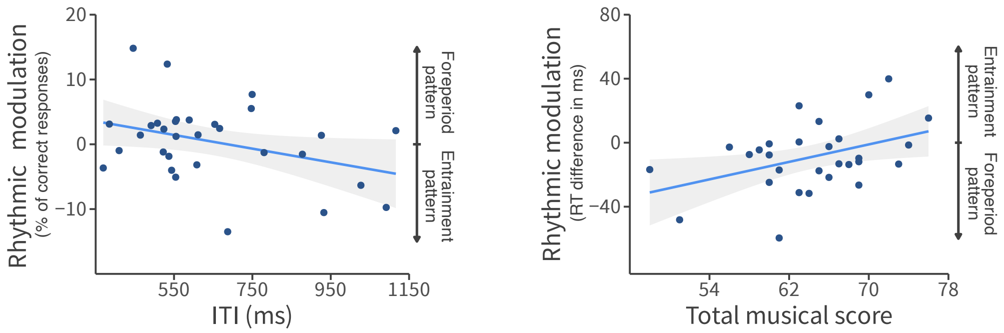

Figure 6. Significant correlations between rhythmic entrainment modulation and individual differences in ITI (modulation on accuracy) and musical skills (on RT). Shaded area represents the 95% confidence interval. Note that y axis shows the score difference between the late and in-phase onsets (i.e., late onset minus in-phase onset).

In the present study, we removed the intertrial interval, an unintentional deviation from the original design that may be also responsible for the absence of foreperiod effects in behavioral outcomes that usually present it (i.e., RT, accuracy) and the reason why our DDF results differed from the original study. In addition, the absence of an increase in pupil size prior to the target also indicates that our fully continuous design prevented a foreperiod effect in pupil dilation, as was predicted in our preregistered hypotheses based on previous literature (Shalev & Nobre, 2022).

Another contrasting piece of evidence against the original predictions of DAT comes from the correlation between rhythmic entrainment (i.e., the difference between in-phase and late-onset oddballs) on accuracy and individual differences in SMT. According to the preferred period hypothesis (McAuley et al., 2006), the perceptual benefits of performing a rhythmic task should be maximal when the rhythm is close to the participant's rhythmic motor preference. The difference between the participant's SMT and task rhythm was captured in our absolute deviance measure (i.e., difference in absolute value of the SMT and the fixed IOI, 700 ms), but no behavioral outcome correlated with that measure. Conversely, participants with slower SMT showed larger rhythm benefits relative to faster tempi (i.e., correlation between ITI and accuracy), a result that replicates the findings of a previous study (Snapiri et al., 2023). However, there is no clear explanation for the fact that individuals with slower SMT benefit more from the presence of a rhythmic context. Further research is needed to fully understand this intriguing correlation pattern.

On the other hand, individuals with higher musical skills seemed to make more use of rhythmic structures and showed faster responses for targets in phase. This suggests that the effects predicted by DAT might be specific to subsamples of individuals with musical training and higher natural musical skills, rather than being a phenomenon generalizable to all populations as thought. Indeed, long-term musical training has been related to differences in attention (Román-Caballero

[6] Please note that our experimental design was based on the original work by McAuley and Fromboluti (2014). Initially, we interpreted their lack of explicit mention of an intertrial interval as a sign of a continuous design. However, after conducting the experiment, we reconsidered it and recognized the possibility that they might have used an intertrial interval larger than IOI (700 ms). In light of this, we have discussed our results with this alternative design to account for the differential outcomes.

et al., 2021). Nevertheless, previous studies have failed to observe an association between musical training and a higher rhythmic entrainment effect (Bauer et al., 2015). It is however important to note that the current results come from a large set of exploratory correlations uncorrected for multiple testing and whose sample size was not motivated by these correlations. Future studies with larger samples may shed light on the robustness of these findings and allow drawing firmer conclusions.

The most prominent finding in Experiment 1 was a greater pupil dilation in the fixed blocks compared to variable blocks in a tonically manner and irrespective of target phase (both for trials with targets in phase and out of phase). This pattern of results was also replicated in Experiment 2 and will therefore be discussed in detail in the **General Discussion**. Of note, this pupillometric effect aligns with the observed behavioral results and points to the absence of rhythmic entrainment also at the pupillometry level.

In sum, Experiment 1 partially replicated the results in McAuley and Fromboluti (2014), providing several converging evidence for the absence of a rhythmic entrainment effect in this experimental paradigm. While we replicated the lack of rhythmic modulation in temporal discrimination (JND) and other behavioral outcomes (DDF, accuracy, and misses), smaller tonic pupil dilation in fixed blocks independent of target phase provides a new source of evidence (eye-tracker measures in this case) in which rhythmic entrainment was also not found. However, these results might be specific to our task as it poses some characteristics that might be responsible for the lack of entrainment effect. First, studies that support DAT typically use tasks with a clear separation between trials, which makes it easier to detect the beginning of the trial (Jones et al., 2002). Although a trial-by-trial design could favor the emergence of a foreperiod effect with results opposite to the prediction of DAT (i.e., improved behavioral outcomes for targets appearing late compared to in-phase), it is relevant to test the robustness of our findings, especially those of pupillometry, with a task that allows perceiving the separation between trials. Second, the type of task might also be influential. Other empirical studies consistent with DAT's predictions have used pitch discrimination paradigms in which target tones should be identified as higher or lower in pitch in comparison to a stream of standard tones. Indeed, Jones et al. (2002) observed in a series of experiments a clear perceptual benefit for tones expected in phase with the rhythm, whereas performance in an arrhythmic condition showed a pattern congruent with a foreperiod effect. These results suggest that rhythmic contexts in pitch discrimination tasks might modulate attention to override hazard rate expectancy and enhance attention for in-phase moments. Another experiment conducted in our laboratory addressed these questions (here referred to as Experiment 2). In addition to pupillometry, we included another eye-tracker measure, saccade rates, which have been used in previous studies investigating the effects of temporal expectation on discrimination tasks (Abeles et al., 2020; Dankner et al., 2017) and could provide valuable insights into the present research question.

3. Experiment 2

3.1. Method

3.1.1. Participants and design

We selected the three-way interaction [type of block (2) × foreperiod (4) × phase (2); for details, see below] to estimate the sample size in the present experiment. In a previous pilot study with 11 participants, we found an effect size of Cohen's f = 0.34. Using the *Superpower* R package (Lakens & Caldwell, 2021), at least 43 participants would be necessary for a small-to-medium effect size of $f = 0.30$ (similar to that observed in the pilot) with an alpha of .05 and a power of .80. Accordingly, we chose a sample size of 45 participants to compensate for potential missing data as a consequence of technical issues or misunderstandings of the instructions. Outlier detection was based on performance (i.e., mean reaction times and accuracy) identified as poor in terms of meeting all the following indices: standard deviation from the mean of the sample (> 2), studentized deleted residuals (> 2), and Cook's Di (> 4/n)[7].

A new sample of forty-five students from the University of Granada participated in the experiment (**Table 5**). They were naïve to the purpose of the experiment and received €15 for their participation. The study was conducted at the Mind, Brain and Behavior Research Center of the University of Granada during the Summer of 2021.

To reach a reliable discrimination threshold for each participant, the 1-up/4-down procedure finished after the occurrence of ten peaks or valleys in each block.

After the estimation of the discrimination threshold, participants carried out the pitch discrimination task. The instructions appeared on the screen and were explained to the participant by the experimenter, who also answered any questions about the procedure. In each trial, the participants heard five, six, or seven standard tones with an IOI of 500 ms or a random value between 250 and 750 ms, depending on the block (fixed vs. variable block; the sequence length was on average the same in both block types; **Figure 7**). After the sequence of standard tones, a comparison tone appeared with an IOI randomly selected from 250, 500, 750, 1000, 1250, 1500, 1750, 2000, and 2250 ms. The participants were required to discriminate as fast and accurately as possible whether the comparison tone had a higher or lower pitch than the standard tones, pressing "z" or "m" (key assignment was counterbalanced across participants). A 1500-ms interval

Table 5. Demographic information of the sample of Experiment 2.

Age	22.9 years, *SD* = 3.2, range 18–31
Sex	34 women and 11 men
Handedness	41 right-handed and 4 left-handed
Musicianship	14 with long-term musical training and 31 without

[7] As in Experiment 1, the thresholds selected for the *SD* and residual criteria deviated slightly from those preregistered to make detection somewhat less conservative (with no substantial consequences).

followed the target onset. The discrimination task began with two practice blocks of eight trials each. Feedback for incorrect responses was provided only during practice blocks. Then, participants completed one block of 252 trials with fixed IOIs of 500 ms for the sequence of standard tones or with variable IOIs between 250 and 750 ms (the order was counterbalanced across participants). As there are nine possible IOIs between the last standard tone and the target, and two pitches, there were 14 trials for each type (with the same IOI and the same pitch). Subsequently, participants were asked about the difficulty of the last task block. After 1 or 2 minutes of rest, they carried out a second task block, with variable IOIs if the first block was with fixed IOIs or with fixed IOIs if the previous block was with variable IOIs. Finally, the participants filled the question about the difficulty. The experimental task compromised 28 trials for each experimental condition.

3.1.4. Preregistered hypotheses

We expected to observe a classic foreperiod effect in both RT and accuracy. That is, the later the target onset, the faster the response and the better the discrimination (higher accuracy) of the participants. Moreover, according to DAT and, particularly, Jones et al.' study (2002), the presence of a rhythmic context before the target would enhance the perceptual discrimination and the preparation for the critical moment and would modulate the foreperiod effect. The presentation of several isochronous standard tones would lead to better performance at moments in phase with the entrained rhythm (500, 1000, 1500, and 2000 ms after the last standard tone; *preregistered hypothesis 1*).

Regarding eye-tracker measures, we also expected a modulation of the rhythmic context over the foreperiod effect. Specifically, we hypothesized a decrease of the pupil size at the beginning of the sequence of standard tones along with an increase before the target onset, that would be more marked in the rhythmic block (*preregistered hypothesis 2*). Furthermore, we expected to observe a higher saccade rate before the onset of in-phase targets in variable blocks than in fixed blocks (*preregistered hypothesis 3*).

3.1.5. Design and data analysis

The experiment has a three-factor repeated-measures design. The first factor was type of block, with two levels, fixed vs. variable. The second factor was position, with four levels: 500–750, 1000–1250, 1500–1750, and 2000–2250 ms. Finally, the third variable was phase, with two levels: in-phase targets (which included 500, 1000, 1500, and 2000 ms) and out-of-phase positions (750, 1250, 1750, and 2250 ms). We expected a higher number of misses in the cue–target interval of 250 ms, and, subsequently, a reduced proportion of trials to be analyzed (as it was the case; see **Supplementary Figure 1**). Therefore, we excluded this condition from the main analyses and got four levels of in-phase condition as for out-of-phase. Target position was mean-centered.

Analyses of hypothesis 1

As preregistered, we conducted repeated-measures ANOVAs with mean correct RTs and accuracy. For consistency with Experiment 1, we additionally conducted two (non-preregistered) mixed-effects models with the R packages *lme4*, *lmerTest*, and *afex*, one with RT (linear) and another with accuracy (binomial generalized linear). In RT analysis, correct responses slower or faster than 2 *SD*s of the mean of participants were not included.

Analyses of hypothesis 2

The analyses with pupillometry measures were identical to those of Experiment 1.

Analyses of hypothesis 3

For saccades detection, we used a similar procedure to Dankner et al. (2017). Saccades were identified by eye moments that exceeded a threshold of 6 standard deviations from the median velocity (Engbert & Kliegl, 2003) during 7 ms and we imposed a minimum interval of 50 ms between one detected saccade and the next. In addition, we excluded

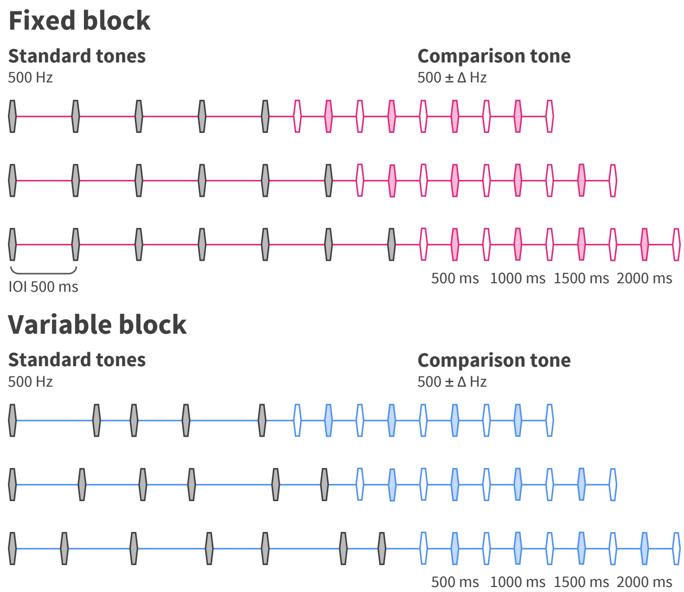

Figure 7. Graphical representation of the pitch discrimination task in Experiment 2. Grey figures represent sequences of standard tones. Color-filled figures are in-phase target tones while white figures with borders in color are out-of-phase comparison tones.

saccades with extremely large displacements (> 2 *SD*s)[8] and trials in which a blink overlaps with the target interval (–100 to 0 ms relative to the target onset). Horizontal and vertical displacement data were smoothed using a 50 ms Hanning window and missing values were interpolated using cubic spline interpolation. The dependent variable is the saccade rate (i.e., sum of the number of saccade onsets for each time point, divided by the number of analyzed trials and multiplied by the sampling rate) at –100 to 0 ms relative to the target onset. For consistency with pupil size analyses, moment-to-moment differences between both blocks were compared using a cluster-based permutation test, as well as mixed-effects models with block type (fixed vs. variable) as a fixed covariate and participants as a random factor. For pretarget analysis, the model also included time as a covariate.

3.2. Results

Within our sample of 45 participants, one participant did not complete the experiment, three outliers with extremely low accuracy were identified ($M < .77$ vs. whole-sample $M = .95$), and one participant was excluded based on their disparate RTs ($M > 800$ ms vs. whole-sample $M = 656$ ms). While the former group of outliers was removed from all the analyses, the participant with extreme RTs was only excluded in the RT analysis. Data from one participant in the POMS scales were missing. Finally, two participants were excluded from eye-tracker analyses due to technical problems during the recording, and three participants because of excess of samples identified as blinks and missing values (93, 84, and 87%; M of the included sample = 12.9% of both blink samples and missing values, $SD = 11.5$).

3.2.1. Discrimination threshold, perceived difficulty, fatigue, and vigor

The average discrimination threshold was 12.8 Hz ($SD = 7.9$; range 1–32 Hz) and it did not correlate with either perceived difficulty, $r(39) = -.01$, $p = .935$, accuracy, $r(39) = .19$, $p = .468$, and RT, $r(39) = -.20$, $p = .220$. On the other hand, the perceived difficulty laid at the midpoint of the full range of the scale [$M = 4.5$, $SD = 1.8$; scale from 1 (extremely easy) to 9 (extremely difficult)]. Interestingly, there was no difference in the perceived difficulty between types of blocks, $t(40) = -1.48$, $p = .146$, $d_z = -0.23$, but perceived difficulty significantly correlated with overall accuracy, $r(39) = -.54$, $p < .001$, and RT, $r(39) = .43$, $p = .005$. The POMS fatigue and vigor scales were associated with each other, $r(38) = .47$, $p = .002$, indicating higher vigor score in participants with lower fatigue, but they did not correlate with overall performance [fatigue: accuracy, $r(38) = .24$, $p = .128$, and RT, $r(38) = .06$, $p = .691$; vigor: accuracy, $r(38) = .28$, $p = .084$, and RT, $r(38) = .07$, $p = .688$].

3.2.2. Reaction time and accuracy

Similar to Experiment 1, the crucial type of block-by-phase interaction did not emerge in any behavioral measure (**Table 6**; **Figure 8**)[9]. In contrast, a clear foreperiod effect appeared in RT with this experimental paradigm, with faster responses for later than earlier targets, $F(3, 117) = 51.49$, $p < .001$, $\eta_p^2 = .57$. A similar pattern arose within phase factor: out-of-phase targets triggered faster responses than in-phase targets, $F(1, 39) = 23.90$, $p < .001$, $\eta_p^2 = .38$, as all out-of-phase tones appeared later in the trial timeline. Therefore, opposite to an entrainment pattern, participants showed a foreperiod effect at the local level (i.e., within each target position). That local difference between out-of-phase and in-phase targets was progressively reduced across target positions, being more pronounced at earlier positions [phase-by-position interaction: $F(3, 117) = 6.29$, $p < .001$, $\eta_p^2 = .14$], coherent with the foreperiod effect reaching asymptotic levels. Interestingly, a general advantage for fixed versus variable block arose in RT, $F(1, 39) = 6.75$, $p = .013$, $\eta_p^2 = .15$. Linear mixed-effects model with RT showed similar results, with the only exception that a type of block-by-target position interaction became significant, $F(1, 16068.05) = 6.43$, $p = .011$, $f^2 < 0.01$, suggesting that the difference between blocks was larger at earlier target positions. No significant results were found with accuracy in either analytical model.

3.2.3. Pupillometry

Again, eye-tracker measures showed that pupil size was overall larger for variable blocks than for fixed blocks at the trial level [pupil size previous to the target, 1st half: $F(1, 188889) = 912.69$, $p < .001$, $f^2 < 0.01$; 2nd half: $F(1, 188959) = 684.75$, $p < .001$, $f^2 < 0.01$; **Figure 9A**] and across the whole task, [1st half: $F(1, 59977) = 1.03$, $p = .310$, $f^2 < 0.01$; 2nd half: $F(1, 76131) = 87.17$, $p < .001$, $f^2 < 0.01$; **Figure 9B**]. At the beginning of the trial, pupil size decreased, $F(1, 188889) = 1726.73$, $p < .001$, $f^2 < 0.01$, followed by a subsequent increase, $F(1, 188959) = 472.75$, $p < .001$, $f^2 < 0.01$. While the decrease in pupil diameter suggests a state of lower preparation at the beginning of the trial, the subsequent increase is consistent with a foreperiod preparation pattern. This pattern was more pronounced in fixed block [initial decrease: $F(1, 188889) = 62.50$, $p < .001$, $f^2 < 0.01$; before-target-onset increase, $F(1, 188959) = 61.79$, $p < .001$, $f^2 < 0.01$], which suggests a more flexible dynamic of pupil size in rhythmic contexts. In a target-locked analysis, there was no difference between in-phase and out-of-phase target, $F(1, 14505) = 0.78$, $p = .378$, $f^2 < 0.01$, but pupil size was again smaller for fixed blocks, $F(1, 14505) = 128.02$, $p < .001$, $f^2 < 0.01$ (**Figure 9C**). This time, the block-by-onset interaction was significant, $F(1, 14505) = 12.31$, $p < .001$, $f^2 < 0.01$, although in the opposite direction to that predicted by DAT, with a smaller between-blocks difference for in-phase targets. Despite the interaction, pupil size was smaller for fixed blocks in both in-phase targets, $F(1, 7235) = 30.71$, $p < .001$, $f^2 < 0.01$, and out-of-phase targets, $F(1, 7235) = 111.53$, $p < .001$, $f^2 < 0.01$. Again, pupil size was significantly larger in pretarget samples even after adding participants' RT and accuracy and reported difficulty to the model, $F(1, 367489) = 1638.77$, $p < .001$, $f^2 < 0.01$.

[8] Saccades larger than 3° is a conservative threshold in our paradigm.

[9] The type of block-by-phase interaction was not statistically significant even when a linear detrending procedure was applied to RT to remove the foreperiod trend, $F(1, 16068.04) = 0.18$, $p = .674$, $f^2 < 0.01$.

Table 6. Results of the three-way repeated measures ANOVAs and mixed-effects models with reaction time and accuracy measures in Experiment 2.

Three-way repeated measures ANOVA		
Effect	**RT**	**Accuracy**
Type of block	$F(1, 39) = 6.75$, $p = .013$, $\eta_p^2 = .15$; $BF_{10} = 37377.67$*	$F(1, 40) = 0.29$, $p = .591$, $\eta_p^2 < .01$; $BF_{10} = 0.12$
Phase	$F(1, 39) = 23.90$, $p < .001$, $\eta_p^2 = .38$; $BF_{10} = 41.52$*	$F(1, 40) < 0.01$, $p = .973$, $\eta_p^2 < .01$; $BF_{10} = 0.09$
Target position	$F(3, 117) = 51.49$, $p < .001$, $\eta_p^2 = .57$; $BF_{10} = 4.7 \times 10^{25}$*	$F(3, 120) = 1.11$, $p = .349$, $\eta_p^2 = .03$; $BF_{10} = 0.02$
Block × phase	$F(1, 39) = 1.42$, $p = .241$, $\eta_p^2 = .04$; $BF_{10} = 0.18$	$F(1, 40) < 0.01$, $p = .948$, $\eta_p^2 < .01$; $BF_{10} = 0.11$
Block × position	$F(3, 117) = 2.48$, $p = .064$, $\eta_p^2 = .06$; $BF_{10} = 0.10$	$F(3, 120) = 0.13$, $p = .941$, $\eta_p^2 < .01$; $BF_{10} = 0.02$
Phase × position	$F(3, 117) = 6.29$, $p < .001$, $\eta_p^2 = .14$; $BF_{10} = 1.79$*	$F(3, 120) = 0.85$, $p = .471$, $\eta_p^2 = .02$; $BF_{10} = 0.04$
Block × phase × position	$F(3, 117) = 1.22$, $p = .306$, $\eta_p^2 = .03$; $BF_{10} = 0.19$	$F(3, 120) = 1.04$, $p = .378$, $\eta_p^2 = .03$; $BF_{10} = 0.09$
Mixed-effects model		
Effect	**RT**	**Accuracy**
Type of block	$F(1, 16068.19) = 63.42$, $p < .001$, $f^2 < 0.01$*	$\chi^2(1) = 0.87$, $p = .351$, $f^2 < 0.01$
Phase	$F(1, 16068.04) = 26.43$, $p < .001$, $f^2 < 0.01$*	$\chi^2(1) = 0.03$, $p = .855$, $f^2 < 0.01$
Target position	$F(1, 16068.05) = 263.80$, $p < .001$, $f^2 = 0.01$*	$\chi^2(1) = 1.99$, $p = .158$, $f^2 < 0.01$
Block × phase	$F(1, 16068.04) = 1.22$, $p = .270$, $f^2 < 0.01$	$\chi^2(1) = 0.01$, $p = .929$, $f^2 < 0.01$
Block × position	$F(1, 16068.05) = 6.43$, $p = .011$, $f^2 < 0.01$*	$\chi^2(1) = 0.09$, $p = .769$, $f^2 < 0.01$
Phase × position	$F(1, 16068.05) = 11.54$, $p < .001$, $f^2 < 0.01$*	$\chi^2(1) = 1.85$, $p = .174$, $f^2 < 0.01$
Block × phase × position	$F(1, 16068.03) = 0.22$, $p = .643$, $f^2 < 0.01$	$\chi^2(1) = 0.31$, $p = .578$, $f^2 < 0.01$

Note: * $p < .05$; † $.05 \le p < .10$.

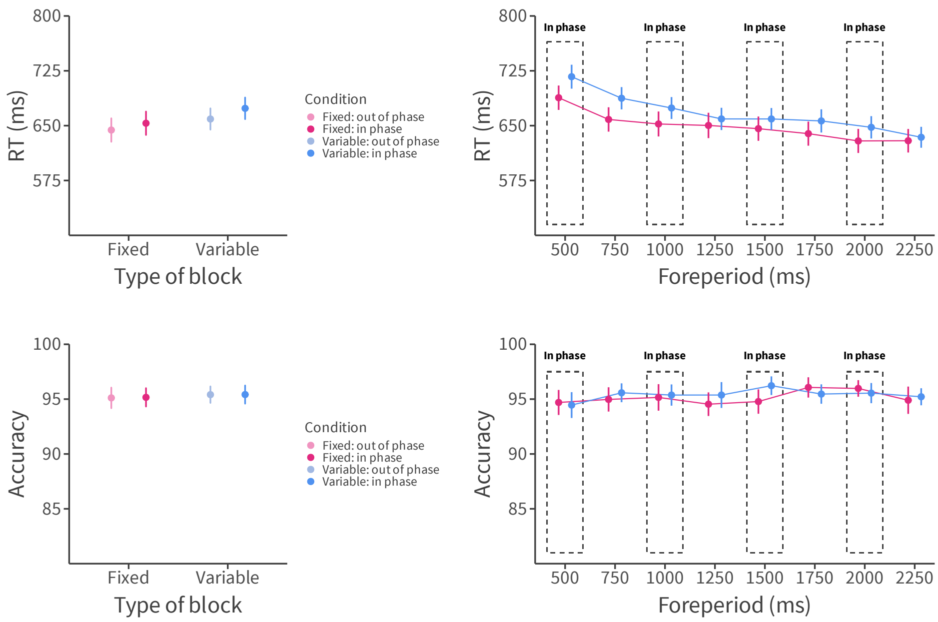

Figure 8. Means of reaction time (RT; upper row) and correct responses (lower row) as a function of type of block and target phase (left), and type of block and target position (right) in Experiment 2. Error bars represent standard error of the means.

3.2.4. Saccade rate

In the comparison interval (i.e., 100 ms before the target onset), the saccade rate was similar between fixed and variable blocks, $F(1, 105) = 1.81$, $p = .182$, $f^2 < 0.01$; and between in-phase and out-of-phase targets, $F(1, 105) = 1.32$, $p = .252$, $f^2 < 0.01$ (**Figure 10**). The type of block-by-phase interaction was also not significant, $F(1, 105) = 0.56$, $p = .454$, $f^2 < 0.01$. However, when a larger pretarget period was analyzed (i.e., 1500 ms before the target onset), the fixed block showed overall larger saccade rates, $F(1, 106305) = 966.72$, $p < .001$, $f^2 < 0.01$. In addition, the saccade rate decreased progressively before the target onset, $F(1, 106305) = 1217.93$, $p < .001$, $f^2 < 0.01$; a trend that was more pronounced for fixed block, $F(1, 106305) = 6.48$, $p = .011$, $f^2 < 0.01$.

3.3. Discussion

Once again, the findings from Experiment 2 did not support the preregistered hypotheses, which were drawn according to DAT and previous empirical studies showing a phase modulation (e.g., Jones et al., 2002). In this experiment, we observed that participants reported again a similar level of perceived difficulty for both types of blocks and no evidence of rhythmic entrainment. However, a clear foreperiod effect appeared this time due to the discontinuous design of our task (i.e., trials were separated by an interval of silence). Consistently, response preparation was slower for in-phase targets than out-of-phase. However, contrary to the

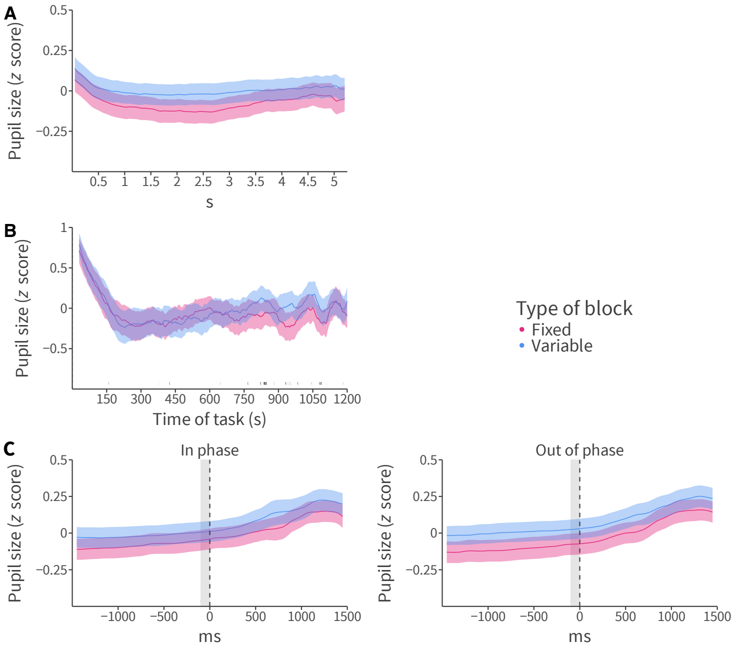

Figure 9. Pupil size (**A**) in temporal samples previous to target onset, (**B**) across the whole task, and (**C**) in target-locked samples (1500 ms before and after target onset) for both in-phase and out-of-phase target in Experiment 2. Black points at the bottom of the plots depict moments in which the difference between both conditions had a $p < .05$ (in a one-sided test). For illustrative purposes only, values were smoothed using a sliding window of 100 ms in panels A and C, and a sliding window of 10 s (~ two trials) in panel B.

predictions of DAT, this was entirely explained by the fact that phase always overlapped in Experiment 2 with later moments within each temporal position (that is, the first possible position for an in-phase target, 500 ms, was earlier than its out-of-phase counterpart, 750 ms). The analysis of pupil dilation and saccade rate provided similar results: while pupil diameter progressively increased during the pretarget period, saccade rate decreased (foreperiod effect). Moreover, there was a tonic difference between blocks, with fixed blocks showing a smaller pupil size and higher saccade rate, which indicates that unpredictable temporal contexts increase processing demands and arousal. Finally, task performance and reported difficulty did not explain these between-block differences, and the attentional and processing demands induced by arrhythmic contexts could not be fully accounted for by objective and subjective measures of task difficulty.

Interestingly, we observed a larger overall difference in RTs between blocks than in Experiment 1 (fixed block: 648 ms; vs. variable: 666 ms; compared to the 5-ms difference in Experiment 2) that was independent of the target onset. This general benefit might also be related to the discontinuous design of the trials. From the perspective of temporal expectancies in the present task, after a certain number of trials, participants might learn that the target had to appear after an interval of 2000-3000 ms embedded with standard tones during which they were not required to respond. That *non-demand period* coincides with a decrease in pupil diameter, suggesting a moment of lower preparation. Once this period had elapsed, an increase in pupil size was observed consistent with a progressive increment in preparation. Therefore, a fundamental aspect of the performance of the task is to determine *when* the target appears, and to do this it is important to predict *how long* the *non-demand period* lasts.

Time intervals filled with stimuli are perceived as longer than the same empty intervals according to the classic *filled-duration illusion* (Hall & Jastrow, 1886; see for a review Wearden & Odgen, 2021), and the same applies to intervals filled with isochronous tones in comparison to anisochronous fillers (Horr & Di Luca, 2015). If that were the case, sequences with fixed IOI would produce an overestimation of their duration and delay preparation for the *response-demanding period*, which would eventually produce slower RTs in the fixed IOI block. However, most of the studies investigating the filled-duration illusion used short intervals (≤ 1 s). In our paradigm, the time intervals fell in the supra-second range, for which the evidence shows that the filled-duration illusion is limited (especially above 3 s; Ihle & Wilsoncroft, 1983). Moreover, the illusion in the supra-second range could be in part driven by a general tendency to underestimate long intervals (Vierordt's Law; Lejeune & Wearden, 2009), and filled intervals would be estimated more accurately relative to the actual duration (and thus, still longer durations than unfilled intervals). Following that logic, sequences with variable IOI would produce an estimate of the non-demand period that is less accurate, which would subsequently affect response preparation and produce slower RTs in the variable IOI block. Alternatively, humans show better time estimation and better motor tuning mechanisms for short time intervals (Hasbroucq et al., 1997). Here, we speculate that the general benefit in response preparation with rhythmic contexts might also be a consequence of the fact that isochronous sequences might help subdivide pretarget periods. Thus, filling the *non-demand period* with an isochronous sequence could lead to a perception of supra-second intervals subdivided into multiple intervals of equal and shorter length, in the sub-second range (500 ms in Experiment 2), which participants estimate better and would lead to more time-accurate preparation.

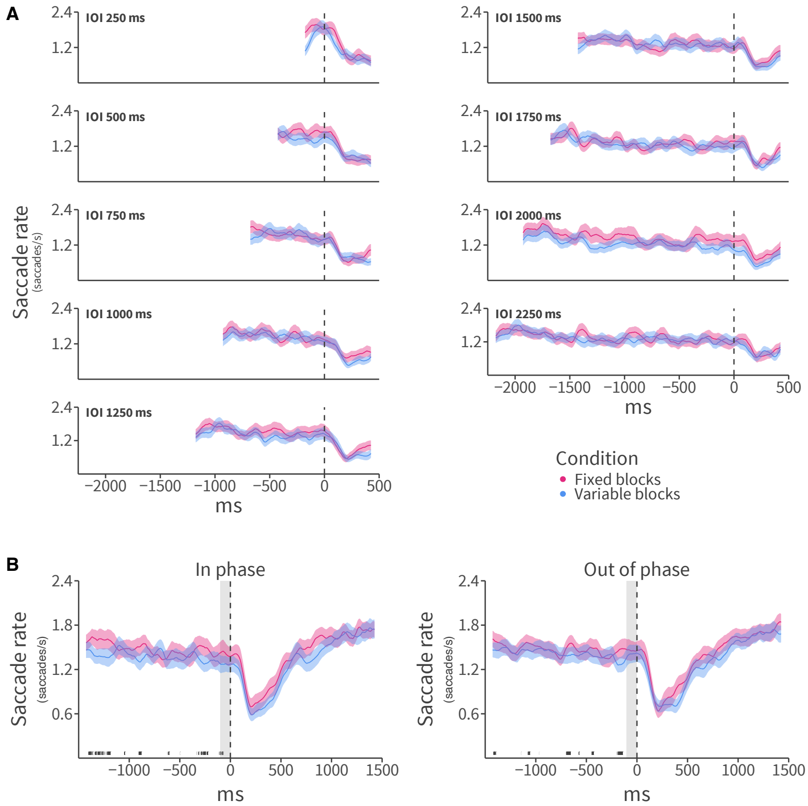

Figure 10. Saccade rate (**A**) in temporal samples previous to target onset and (**B**) target-locked samples (1500 ms before and after target onset) in Experiment 2. Dashed lines indicate target onset while gray bars denote the relevant pretarget interval for analyses. Black points at the bottom of the plots depict moments prior to the target in which the difference between both conditions had a $p < .05$ (in a one-sided test). For illustrative purposes only, values were smoothed using a sliding window of 100 ms.

4. General Discussion

The two experiments reported here showed no evidence of rhythmic entrainment in both behavioral and eye-tracker measures. The lack of rhythmic facilitation arose regardless of the type of task in each experiment (duration and pitch discrimination task) and the manipulation of the feeling of a continuous or a discontinuous task (removing vs. including an intertrial interval), among other differences. Arguably, the specific conditions of our experiments might prevent the finding of an entrainment effect. For example, the out-of-phase moments selected in Experiment 2 (but not in Experiment 1) were all at antiphase, which might also elicit an attentional enhancement because of their harmonic relationship and the possibility that participants could perceive an additional beat at the faster subdivision rate (Bouwer et al., 2021). However, multiple studies have also failed to observe a pattern of performance consistent with DAT, regardless of the perceptual modality of the rhythm and target (both auditory: Bauer et al., 2015; Lin et al., 2022; Pomper et al., 2023; both tactile: Jones, 2019; Pomper, 2023; both visual: Pomper et al., 2023; audio-visual: Pomper et al., 2023; Schirmer et al., 2021), the type of task (detection: Jones, 2019; Pomper, 2023; Schirmer et al., 2021; pitch discrimination: Bauer et al., 2015; Lin et al., 2022; visuospatial discrimination: Pomper et al., 2023; tactile discrimination: Jones, 2019), or the type of rhythm (isochronous sequences: Bauer et al., 2015; Lin et al., 2022; Pomper, 2023; Pomper et al., 2023; complex metrical sequences: Schirmer et al., 2021). As in Experiment 2, two of those studies (Bauer et al., 2015; Lin et al., 2022) were conceptual replications of the classic report by Jones et al. (2002), and even the exact replication included in the experimental series in Bauer et al. did not observe phase-dependent facilitation. In addition, some of the studies that have served to support DAT often do not include conditions in which the target is presented at out-of-phase times (e.g., Chang et al., 2019; Cravo et al., 2013; Lange, 2009; Rohenkohl et al., 2012) or out-of-phase targets correspond with earlier time points within the trial than in-phase targets (Bouwer et al., 2020). Taken together, all these findings suggest that there is a complex interplay of temporal expectancy phenomena that are elicited in rhythmic tasks and that could explain differences between rhythmic and arrhythmic conditions without the need to invoke an attentional entrainment account (e.g., foreperiod effect, filled-duration illusion, cue validity, catch trials, etc.). For example, faster responses for in-phase targets in paradigms in which out-of-phase targets appeared on average earlier (Pomper et al., 2023; Bouwer et al., 2020) is aligned with a variable foreperiod effect (Niemi & Näätänen, 1981). Also, a quadratic RT pattern is expected for tasks that include catch trials because of the integration of two different types of uncertainty, a discrete type of uncertainty about whether an event happens (modulated by catch trials; Jones et al., 2017) and a continuous type about when it happens (modulated by foreperiod; Grabenhorst et al., 2021). Therefore, the U-shaped pattern of Jones et al. (2017) could also be the product of the temporal expectation resulting from the presence of trials in which the target did not appear.

The findings of the two current experiments as well as those of other studies with inconsistent results suggest that some of the foundational evidence of DAT may not be as robust as previously thought. To disentangle the specific conditions under which the behavioral predictions of DAT are fulfilled, experimental designs should include both rhythmic and arrhythmic conditions, as well as in-phase and out-of-phase targets. The mere presence of faster responses or higher accuracy in fixed IOI conditions does not necessarily correspond with a proof of rhythmic entrainment as other temporal features might underlie such a difference (e.g., more accurate estimation of supra-second intervals filled with isochronous sequences; Lejeune & Wearden, 2009). At the same time, phase modulations in paradigms that only employ rhythmic conditions are at the risk of measuring different types of temporal expectation than attentional entrainment (e.g., variable foreperiod effect or foreperiod effect with catch trials). The complex landscape of studies and inconsistent results point out that the predictions of DAT are less generalizable across paradigms than originally thought. To unravel the specific conditions under which attentional entrainment emerges, it is necessary that further studies directly test the theory including all experimental conditions in their paradigms.

In addition, the two experiments in this study showed a consistent pattern of greater pupil dilation in arrhythmic blocks compared to rhythmic blocks in a tonically manner. Crucially, this pattern was independent of the target phase (both for trials with targets in phase and out of phase), indicating an effect of a different nature than rhythmic entrainment. Previous studies have shown increases in pupil diameter in response to ambiguous perceptual stimuli (Brunyé & Gardony, 2017), unexpected targets following a cue predicting a different event (Becker et al., 2024; Richter & Lange, 2019), and deviant stimuli embedded in predictable contexts (Bianco et al., 2020; Liao et al., 2016; Marois et al., 2018; Quirins et al., 2018). Deviants and target events in those tasks represent forms of '*unexpected uncertainty*' (i.e., changes in the environment that strongly violate top-down expectations set by the experimental task up to that moment; Yu & Dayan, 2005). The elicited transient changes in pupil dilation would reflect the phasic activity of noradrenergic projections from the locus coeruleus-norepinephrine system (Reimer et al., 2016) in parallel to the level at which stimuli violate the regularities of their context (Becker et al., 2024; Maoris et al., 2018). These norepinephrine-induced pupil changes are thought to indicate higher arousal and greater saliency-driven allocation of attentional resources to the deviant processing (Alink & Blank, 2021), suppressing top-down information in favor of bottom-up signals and promoting learning (Yu & Dayan, 2005), or difficulty in processing the unexpected sensory stimulus (Becker et al., 2024). Other oculomotor measures are also sensitive to this type of unexpected events, with fewer and longer fixations, and slower and shorter saccades (Brunyé & Gardony, 2017). In contrast, the experiments in our study showed a general pupillary difference that was associated with the regularity of the context in which the target appeared, but irrespective of its onset phase. Thus, the variable IOI blocks created a context of constant unreliability in the presentation of the stimuli that introduced a different type of uncertainty, an '*expected certainty*' (Yu & Dayan, 2005). Sustained activation of acetylcholine is hypothesized to play a major role in computing *expected uncertainty* and signaling to ignore internal models and top-down expectations to favor the processing of bottom-up information. Previous studies with predictable sequences have observed a similar smaller pupil diameter in response to regular relative to random patterns, regardless of whether the stimuli (context and target) were auditory (Milne et al., 2021) or visual (Shalev & Nobre, 2022), and regardless of whether the sequences were rhythmically predictive (constant IOI as in the present study; Shalev & Nobre, 2022) or derived from a predictable order of tones (Milne et al., 2021). Irregular contexts might hinder processing and increase overall processing demands in the task (Milne et al., 2021), something that is consistent with participants showing a general overestimation of the target duration in variable blocks in Experiment 1 (i.e., DDF above 1) and slightly slower responses in both experiments (Experiment 1: 5 ms slower; Experiment 2: 18 ms slower). According to the scalar expectancy theory (Gibbon, 1977), states of higher arousal, such as the one that arrhythmic contexts might induce, would subjectively expand the perceived duration of the target due to a greater accumulation of pulses in the internal clock during the same time interval (i.e., target duration). However, the sustainment of higher attention would be to achieve similar levels of task performance in blocks with variable IOI (Henry & Herrmann, 2014), without leading to changes in discrimination ability. Thus, the differences in pupil size between blocks persisted after regressing performance on the task and perceived difficulty in our two experiments. That suggests that larger pupil size in variable blocks might be the result of higher processing demands in arrhythmic blocks but that is not translated to observable differences in behavioral outcomes or both objective and subjective task difficulty.

Therefore, the present study adds evidence that rhythmic contexts modulate attention. Previous studies have proved the ability of rhythm as an orienting cue (Sanabria et al., 2011; Triviño et al., 2011) and lengthening the perceived duration of the whole sequence (Lejeune & Wearden, 2009). Here, we showed that arrhythmic sequences, usually employed as comparison conditions for the impact of metrical or isochronous sequences, generate a context of *expected uncertainty* that increases arousal and processing demands, which predominantly affect response speed and perceived duration but not perceptual discrimination. This modulation seems to be domain-general and leads to changes in pupil size and oculomotor responses independently that the tasks in our two experiments were purely auditory and ocular changes have been proposed to favor visual acuity (Abeles et al., 2020). A difference between our tasks with previous eye-tracker studies investigating temporal expectation is that in our design target appeared in a wide range of possible temporal moments (vs. only one constant interval; Abeles et al., 2020; Dankner et al., 2017; Shalev & Nobre, 2022). While the report by Shalev and Nobre (2022) included both rhythmic and arrhythmic conditions in which the target was presented, the design did not include targets in phase and out of phase making it impossible to compare the level of attention for moments aligned with the rhythm and moments misaligned. The study by Shalev and Nobre (2022), along with the findings from our two experiments, point to a tonic and general effect of rhythmic contexts that is independent of the phase of the target onset. Thus, overall smaller pupil dilation might indicate higher processing demands in arrhythmic blocks, which have little impact on perceptual discrimination and task performance.

5. Conclusions

In line with several recent reports, the present study failed to observe findings in support of DAT. Regardless of the type of task, or the presence or absence of a foreperiod effect, the perceptual discriminability of the participants, preparation, and oculomotor response were similar for events occurring in-phase as for those out-of-phase with the rhythm of the task. Therefore, the specific parameters and the sample characteristics that lead to an attentional entrainment remain unclear. Future studies should use designs that test the predictions of DAT appropriately (i.e., comparison between rhythmic and arrhythmic conditions, and in-phase vs. out-of-phase targets) and avoid confusion with other effects of temporal expectation, such as the foreperiod effect.

Data Availability

All data and R script for the analyses are fully available in https://osf.io/njty6/. The preregister of this study is available at https://osf.io/kqx4w and https://osf.io/zbyje.

References

Abeles, D., Amit, R., Tal-Perry, N., Carrasco, M., & Yuval-Greenberg, S. (2020). Oculomotor inhibition precedes temporally expected auditory targets. *Nature Communications, 11*(1), 3524. https://doi.org/10.1038/s41467-020-17158-9

Alink, A., & Blank, H. (2021). Can expectation suppression be explained by reduced attention to predictable stimuli? *NeuroImage, 231*, 117824. https://doi.org/10.1016/j.neuroimage.2021.117824

Andrade, E., Arce, C., Torrado, J., Garrido, J., De Francisco, C., & Arce, I. (2010). Factor structure and invariance of the POMS mood state questionnaire in Spanish. *The Spanish Journal of Psychology, 13*(1), 444–452. https://doi.org/10.1017/S1138741600003991

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, A. K. R., Jaeger, M., Thorne, J. D., Bendixen, A., & Debener, S. (2015). The auditory dynamic attending theory revisited: A closer look at the pitch comparison task. *Brain Research*, 1626, 198–210. https://doi.org/10.1016/j.brainres.2015.04.032

Becker, J., Viertler, M., Korn, C. W., & Blank, H. (2024). The pupil dilation response as an indicator of visual cue uncertainty and auditory outcome surprise. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.16306

Bianco, R., Ptasczynski, L. E., & Omigie, D. (2020). Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain and Cognition, 138*, 103621. https://doi.org/10.1016/j.bandc.2019.103621

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

Brunyé, T. T., & Gardony, A. L. (2017). Eye tracking measures of uncertainty during perceptual decision making. *International Journal of Psychophysiology, 120*, 60–68. https://doi.org/10.1016/j.ijpsycho.2017.07.008

Capizzi, M., & Correa, A. (2018). Measuring temporal preparation. In A. Vatakis, F. Balci, M. Di Luca, & Á. Correa (Eds.). *Timing and time perception: Procedures, measures, and applications*. Ed. Brill, Leiden Boston, pp. 216-232.

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2019). Rhythmicity facilitates pitch discrimination: Differential roles of low and high-frequency neural oscillations. *NeuroImage*, 198, 31–43. https://doi.org/10.1016/j.neuroimage.2019.05.007

Coull, J. T. (2009). Neural substrates of mounting temporal expectation. *PLOS Biology*, 7(8), e1000166. https://doi.org/10.1371/journal.pbio.1000166

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience, 33*(9), 4002–4010. https://doi.org/10.1523/JNEUROSCI.4675-12.2013

Dankner, Y., Shalev, L., Carrasco, M., & Yuval-Greenberg, S. (2017). Prestimulus inhibition of saccades in adults with and without attention-deficit/hyperactivity disorder as an index of temporal expectations. *Psychological Science, 28*(7), 835–850. https://doi.org/10.1177/0956797617694863

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84* (3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Grabenhorst, M., Maloney, L. T., Poeppel, D., & Michalareas, G. (2021). Two sources of uncertainty independently modulate temporal expectancy. *Proceedings of the National Academy of Sciences, 118*(16), e2019342118. https://doi.org/10.1073/pnas.2019342118

Grove, J. R., & Prapavessis, H. (1992). Preliminary evidence for the reliability and validity of an abbreviated profile of mood states. *International Journal of Sport Psychology, 23*(2), 93–109.

Hall, G. S., & Jastrow, G. (1886). Studies of rhythm. *Mind, 11*, 55–62.

Hasbroucq, T., Kaneko, H., Akamatsu, M., & Possamaï, C. A. (1997). Preparatory inhibition of cortico-spinal excitability: a transcranial magnetic stimulation study in man. *Cognitive Brain Research, 5*(3), 185–192. https://doi.org/10.1016/S0926-6410(96)00069-9

Henry, M. J., & Herrmann, B. (2014). Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time Perception*, 2(1), 62–86. https://doi.org/10.1163/22134468-00002011

Hershman, R., Henik, A., & Cohen, N. (2018). A novel blink detection method based on pupillometry noise. *Behavior Research Methods, 50*, 107–114. https://doi.org/10.3758/s13428-017-1008-1

Horr, N. K., & Di Luca, M. (2015). Taking a long look at isochrony: perceived duration increases with temporal, but not stimulus regularity. *Attention, Perception, & Psychophysics, 77*, 592–602. https://doi.org/10.3758/s13414-014-0787-z

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature Neuroscience*, 8, 234–241. https://doi.org/10.1038/nn1386

JASP Team (2024). JASP (Version 0.19.0)[Computer software].

Jones, A. (2019). Temporal expectancies and rhythmic cueing in touch: The influence of spatial attention. *Cognition, 182*, 140–150. https://doi.org/10.1016/j.cognition.2018.09.011

Jones, A., Hsu, Y. F., Granjon, L., & Waszak, F. (2017). Temporal expectancies driven by self-and externally generated rhythms. *NeuroImage, 156*, 352–362. https://doi.org/10.1016/j.neuroimage.2017.05.042

Jones, M. R., (1976) Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323–355. https://doi.org/10.1037/0033-295X.83.5.323

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, 13(4), 313–319. https://doi.org/10.1111/1467-9280.00458

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakens, D., & Caldwell, A. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, 4(1). https://doi.org/10.1177/2515245920951503

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition, 69*(1), 127–137. https://doi.org/10.1016/j.bandc.2008.06.004

Large, E. W., & Jones, M. R., (1999) The dynamics of attending: how people track time-varying events. *Psychological Review*, 106, 119–159. (doi:10.1037/0033- 295X.106.1.119)
Lejeune, H., & Wearden, J. H. (2009). Vierordt's The Experimental Study of the Time Sense (1868) and its legacy. *European Journal of Cognitive Psychology, 21*(6), 941–960. https://doi.org/10.1080/09541440802453006
Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America, 49*(2B), 467–477. https://doi.org/10.1121/1.1912375
Liao, H. I., Yoneya, M., Kidani, S., Kashino, M., & Furukawa, S. (2016). Human pupillary dilation response to deviant auditory stimuli: Effects of stimulus properties and voluntary attention. *Frontiers in Neuroscience, 10*, 154761. https://doi.org/10.3389/fnins.2016.00043
Lin, W. M., Oetringer, D. A., Bakker-Marshall, I., Emmerzaal, J., Wilsch, A., ElShafei, H. A., ... & Haegens, S. (2022). No behavioural evidence for rhythmic facilitation of perceptual discrimination. *European Journal of Neuroscience*, 55(11–12), 3352–3364. https://doi.org/10.1111/ejn.15208
Marois, A., Labonté, K., Parent, M., & Vachon, F. (2018). Eyes have ears: Indexing the orienting response to sound using pupillometry. *International Journal of Psychophysiology, 123*, 152–162. https://doi.org/10.1016/j.ijpsycho.2017.09.016
Martin, T., Egly, R., Houck, J. M., Bish, J. P., Barrera, B. D., Lee, D. C., & Tesche, C. D. (2005). Chronometric evidence for entrained attention. *Perception & Psychophysics*, 67, 168–184. https://doi.org/10.3758/BF03195020
McAuley, J. D., & Fromboluti, E. K. (2014). Attentional entrainment and perceived event duration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130401. https://doi.org/10.1098/rstb.2013.0401
McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General, 135*(3), 348–367. https://doi.org/10.1037/0096-3445.135.3.348
Milne, A. E., Zhao, S., Tampakaki, C., Bury, G., & Chait, M. (2021). Sustained pupil responses are modulated by predictability of auditory sequences. *Journal of Neuroscience, 41*(28), 6116–6127. https://doi.org/10.1523/JNEUROSCI.2879-20.2021
Niemi P, Näätänen R. (1981) Foreperiod and simple reaction time. *Psychological Bulletin*, 89, 133– 162. https://doi.org/10. 1037/0033-2909.89.1.133
Nobre, A. C., & van Ede, F. (2018). Anticipated moments: temporal structure in attention. Nature reviews. Neuroscience, 19(1), 34–48. https://doi.org/10.1038/nrn.2017.141
Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999*(1), 58–75. https://doi.org/10.1196/annals.1284.006
Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Babnon, B., Trimmer, C. G., ... Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception*, 25, 331–343. https://doi.org/10.1525/mp.2008.25.4.331
Pinheiro, J., Bates, D., & R Core Team (2023). nlme: Linear and nonlinear mixed effects models. r package version 3.1-164, https://CRAN.R-project.org/package=nlme
Pomper, U. (2023). No evidence for tactile entrainment of attention. *Frontiers in Psychology, 14*, 1168428. https://doi.org/10.3389/fpsyg.2023.1168428
Pomper, U., Szaszkó, B., Pfister, S., & Ansorge, U. (2023). Cross-modal attentional effects of rhythmic sensory stimulation. *Attention, Perception, & Psychophysics, 85*(3), 863–878. https://doi.org/10.3758/s13414-022-02611-2
Quirins, M., Marois, C., Valente, M., Seassau, M., Weiss, N., El Karoui, I., ... & Naccache, L. (2018). Conscious processing of auditory regularities induces a pupil dilation. *Scientific Reports, 8*(1), 14819. https://doi.org/10.1038/s41598-018-33202-7
Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications, 7*(1), 13289. https://doi.org/10.1038/ncomms13289
Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *eLife, 8*, e47869. https://doi.org/10.7554/eLife.47869
Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *Journal of Neuroscience, 32*(24), 8424–8428. https://doi.org/10.1523/JNEUROSCI.0804-12.2012
Román-Caballero, R., Martín-Arévalo, E., & Lupiáñez, J. (2021). Attentional networks functioning and vigilance in expert musicians and non-musicians. *Psychological Research, 85*, 1121–1135. https://doi.org/10.1007/s00426-020-01323-2
Sanabria, D., Capizzi, M., & Correa, Á. (2011). Rhythms that speed you up. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 236 –244. https://doi.org/10.1037/a0019956
Schirmer, A., Wijaya, M., Chiu, M. H., Maess, B., & Gunter, T. C. (2021). Musical rhythm effects on visual attention are non-rhythmical: evidence against metrical entrainment. *Social Cognitive and Affective Neuroscience, 16*(1–2), 58–71. https://doi.org/10.1093/scan/nsaa077
Shalev, N., & Nobre, A. C. (2022). Eyes wide open: Regulation of arousal by temporal expectations. *Cognition, 224*, 105062. https://doi.org/10.1016/j.cognition.2022.105062
Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. (2024). afex: Analysis of factorial experiments. R package version 1.3-1, https://CRAN.R-project.org/package=afex
Snapiri, L., Kaplan, Y., Shalev, N., & Landau, A. N. (2023). Rhythmic modulation of visual discrimination is linked to individuals' spontaneous motor tempo. *European Journal of Neuroscience, 57*(4), 646–656. https://doi.org/10.1111/ejn.15898
Triviño, M., Arnedo, M., Lupiáñez, J., Chirivella, J., & Correa, Á. (2011). Rhythms can overcome temporal orienting deficit after right frontal damage. *Neuropsychologia, 49*(14), 3917–3930. https://doi.org/10.1016/j.neuropsychologia.2011.10.009 Visalli, A., Capizzi, M., Ambrosini, E., Kopp, B., & Vallesi, A. (2023). P3-like signatures of temporal predictions: a computational EEG study. *Experimental Brain Research*, *241*(7), 1919–1930. https://doi.org/10.1007/s00221-023-06656-z
Voeten, C. C. (2023). permutes: Permutation tests for time series data. R package version 2.8. https://CRAN.R-project.org/package=permutes
Wearden, J. H., & Ogden, R. S. (2021). Filled-duration illusions. *Timing & Time Perception, 10*(2), 97–121. https://doi.org/10.1163/22134468-bja10040
Yu, J. A., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026