A silent disco: Persistent entrainment of low-frequency neural oscillations underlies beat-based, but not pattern-based temporal expectations

Short title: Persistent effects of beat- and pattern-based temporal expectations

Fleur L. Bouwer[1,2,3*], Johannes J. Fahrenfort[1,3], Samantha K. Millard[3,4,5], Niels A. Kloosterman[6,7], & Heleen A. Slagter[1,3]

[1] Department of Experimental and Applied Psychology, Institute Brain and Behavior Amsterdam (IBBA), Vrije Universiteit Amsterdam
[2] Institute for Logic, Language and Computation (ILLC), University of Amsterdam
[3] Department of Psychology, Amsterdam Brain and Cognition (ABC), University of Amsterdam
[4] Centre for Pain IMPACT, Neuroscience Research Australia (NeuRA), Sydney, New South Wales, Australia
[5] Faculty of Medicine, University of New South Wales (UNSW), Sydney, New South Wales, Australia
[6] Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany
[7] Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

[*] Correspondence concerning this article should be addressed to Dr. Fleur Bouwer, Department of Psychology, University of Amsterdam (bouwer@uva.nl)

FLB is supported by an ABC Talent Grant awarded by Amsterdam Brain and Cognition, a Veni grant awarded by the Dutch Research Council NWO (VI.Veni.201G.066), and the ERC starting grant awarded to HAS. HAS is supported by a European Research Council (ERC) starting grant (679399).

Abstract

The brain uses temporal structure in the environment, like rhythm in music and speech, to predict the timing of events, thereby optimizing their processing and perception. Temporal expectations can be grounded in different aspects of the input structure, such as a regular beat or a predictable pattern. One influential account posits that a generic mechanism underlies beat-based and pattern-based expectations, namely entrainment of low frequency neural oscillations to rhythmic input, while other accounts assume different underlying neural mechanisms. Here, we addressed this outstanding issue by examining EEG activity and behavioral responses during silent periods following rhythmic auditory sequences. We measured responses outlasting the rhythms both to avoid confounding the EEG analyses with evoked responses, and to directly test whether beat-based and pattern-based expectations persist beyond stimulation, as predicted by entrainment theories. To properly disentangle beat-based and pattern-based expectations, which often occur simultaneously, we used non-isochronous rhythms with a beat, a predictable pattern, or random timing. In Experiment 1 (N = 32), beat-based expectations affected behavioral ratings of probe events for two beat-cycles after the end of the rhythm, while the effects of pattern-based expectations subsided almost instantly. In Experiment 2 (N = 27), using EEG, we found enhanced spectral power at the beat frequency for beat-based sequences both during listening and the silence, but for pattern-based sequences, enhanced power at a pattern-specific frequency was only present during listening, not silence. Moreover, we found a Contingent Negative Variation (CNV) following pattern-based, but not beat-based sequences. Finally, we show how multivariate pattern decoding and multi scale entropy – measures sensitive to non-oscillatory components of the signal – can be used to probe temporal expectations. Together, our results suggest that different mechanisms implement temporal expectations, depending on the input structure, with climbing activity reflecting pattern-based, and persistent low frequency oscillations reflecting beat-based expectations specifically.

Introduction

Predicting the timing of incoming events optimizes processing in our dynamic environment (Nobre & van Ede, 2018), as it allows the brain to increase sensitivity to events at predicted times (Auksztulewicz, Myers, Schnupp, & Nobre, 2019), without the need for constant vigilance (Breska & Deouell, 2017; Rimmele, Morillon, Poeppel, & Arnal, 2018; Schroeder & Lakatos, 2009b, 2009a). Entrainment models (Large & Jones, 1999) provide a mechanistic explanation for temporal expectations, by assuming that the phase and period of low-frequency neural oscillations synchronize to external rhythmic stimulation, causing optimal neural excitability at expected times (Haegens & Zion Golumbic, 2018; Henry & Herrmann, 2014; Schroeder & Lakatos, 2009a). In line with this, behavioral performance is improved for events in phase with an external rhythm (Bouwer & Honing, 2015; Jones, Moynihan, MacKenzie, & Puente, 2002; Large & Jones, 1999), behavioral responses depend on the phase of delta oscillations (Arnal, Doelling, & Poeppel, 2014; Cravo, Rohenkohl, Wyart, & Nobre, 2013; Henry, Herrmann, & Obleser, 2014; Henry & Obleser, 2012), and low frequency oscillations phase lock to rhythmic input (Doelling, Assaneo, Bevilacqua, Pesaran, & Poeppel, 2019; Nozaradan, Peretz, Missal, & Mouraux, 2011; Stefanics et al., 2010).

Entrainment has mainly been studied in the context of periodic ("beat-based") sensory input, but temporal expectations can also be based on memory for absolute intervals (Breska & Deouell, 2017; Breska & Ivry, 2016; Morillon, Schroeder, Wyart, & Arnal, 2016; Teki, Grube, Kumar, & Griffiths, 2011), either in isolation ("cue-based"), or as part of a rhythmic pattern of intervals ("pattern-based", see (Nobre & van Ede, 2018)). Rhythmic patterns may be especially important in speech and non-Western music, which is not necessarily periodic. Expectations based on predictable patterns in aperiodic sequences afford similar behavioral benefits as expectations based on a beat (Bouwer, Honing, & Slagter, 2020; Heideman, van Ede, & Nobre, 2018; O'Reilly, McCarthy, Capizzi, & Nobre, 2008), but pose a possible challenge for

entrainment models, which are arguably better suited to explain temporal expectations for periodic input (Breska & Deouell, 2017; Rimmele et al., 2018). Some have suggested that entrainment models can account for pattern-based expectations by assuming multiple coupled oscillators at different frequencies and with different phases (Tichko & Large, 2019), or by assuming flexible top-down phase resets at expected moments, though this would entail some top-down mechanism, making observed entrainment the consequence, rather than the cause of expectations (Meyer, Sun, & Martin, 2019; Obleser & Kayser, 2019; Rimmele et al., 2018).

Alternatively, however, pattern-based and beat-based expectations could arise from dissociable neural mechanisms. For cue-based expectations, tentative evidence for a different underlying mechanism comes from a series of studies looking at the contingent negative variation (CNV), an event-related potential (ERP) component that peaks at expected moments (Praamstra, Kourtis, Kwok, & Oostenveld, 2006). The CNV resolved faster for beat-based than cue-based expectations (Breska & Deouell, 2017), and cerebellar patients showed selective impairments in forming cue-based, but not beat-based expectations (Breska & Ivry, 2018, 2020). However, in these studies, the intended beat-based sequences were isochronous. Isochronous sequences can, in addition to a beat, elicit temporal expectations through learning the repeated, identical interval (Breska & Ivry, 2016; Keele, Nicoletti, Ivry, & Pokorny, 1989). Thus, the differences in responses may here be explained by more precise cue-based or pattern-based expectations in the isochronous, beat-based condition. In our own recent work, we found no difference in the effects of beat-based and pattern-based expectations on early auditory ERP responses, suggestive of similar modulation of sensory processing (Bouwer et al., 2020), but we observed suppression of sensory processing of unexpected events in beat-based rhythms, even when these events were fully predictable based on their pattern, suggesting different underlying mechanisms (Bouwer et al., 2020). However, to our knowledge to date, no study has directly probed and contrasted the neural mechanisms underlying beat-based and pattern-

based expectations, rendering it unclear whether they are subserved by shared or separate neural dynamics.

In the current study, we directly examined the role of entrainment in beat-based and pattern-based expectations, using non-isochronous rhythms designed to properly disentangle these two types of expectations. When studying entrainment, an important challenge has been to differentiate between real entrainment ("in the narrow sense", see Obleser & Kayser, 2019) and regular evoked potentials, or similar phase locked responses that resemble entrainment with common analysis techniques (Zoefel, ten Oever, & Sack, 2018), and that may not differentiate between beat-based and memory-based expectations (Breska & Deouell, 2017). Crucially, to sidestep these issues, here we examined responses in a silent window after cessation of the rhythmic input, directly testing the prediction of entrainment models that entrainment should outlast sensory stimulation (Haegens & Zion Golumbic, 2018; Obleser & Kayser, 2019; Pesnot Lerousseau, Trébuchon, Morillon, & Schön, 2021; Zoefel et al., 2018).

Behaviorally, persistent entrainment has been shown for auditory rhythm (Hickok, Farahbod, & Saberi, 2015; Jones et al., 2002), though this effect is not always found (Bauer, Jaeger, Thorne, Bendixen, & Debener, 2015; Lin et al., 2021), possibly due to heterogeneity in the population and effects of musical training (Assaneo et al., 2019; Cameron & Grahn, 2014; Sun, Michalareas, & Poeppel, 2021). At a neural level, several studies reported persistent entrainment in the visual (de Graaf et al., 2013; Mathewson et al., 2012), and auditory (Pesnot Lerousseau et al., 2021; van Bree, Sohoglu, Davis, & Zoefel, 2021; Wilsch, Mercier, Obleser, Schroeder, & Haegens, 2020) domain. However, in all these studies, isochronous stimuli were used, making it unclear whether the expectations probed were based on a beat, or were formed based on the single repeating interval (i.e., cue- or pattern-based). Moreover, persistent entrainment was not specific to the frequency of the input (Wilsch et al., 2020), or only occurred in the gamma (Pesnot Lerousseau et al., 2021), or alpha ranges (de Graaf et al., 2013;

Mathewson et al., 2012), while humans have a preference for forming temporal expectations at slower rates (Ding et al., 2017; Merchant, Grahn, Trainor, Rohrmeier, & Fitch, 2015), as naturally present in speech and music (i.e., the delta and theta range). Thus, not only is evidence for whether entrainment can account for pattern-based expectations lacking, evidence for persistent entrainment in response to beat-based rhythms remains elusive as well.

In the current study, participants listened to non-isochronous auditory sequences with either a regular beat (eliciting beat-based expectations), a predictable pattern (eliciting pattern-based expectations), or random timing (no expectations). Each sequence was followed by a silent period. In Experiment 1, we asked participants to rate how well probe tones, presented at various time points during the silent period, fitted the preceding rhythm. Consistent with entrainment underlying beat-based, but not pattern-based expectations, we show that the effects of beat-based expectations on behavior outlasted the effects of pattern-based expectations – a finding we replicate in the second experiment. The behavioral effects of expectations correlated with musical experience, for both beat-based and pattern-based expectations. In Experiment 2, we recorded EEG activity both during presentation of the rhythms and during the silence. We found enhanced power at the beat frequency for beat-based sequences during listening, and, importantly, during the silence, while for pattern-based sequences, we observed enhanced power at frequencies related to the rhythms only during listening, but not during the silence. We also find unique neural markers associated with pattern-based expectations: while the pattern-based sequences were followed by a CNV, related to the expectation of the next sound, for the beat-based sequences, the CNV was absent.

Finally, we explored two methods to index temporal expectations that do not rely on the EEG signal being a static oscillation. As recently argued, the oscillatory dynamics underlying such expectations may be subject to changes in power and frequency, depending on coupling between sound and brain, and on the properties of the neural dynamics themselves (Doelling &

Florencia Assaneo, 2021). Also, the presence of non-sinusoidal recurring activity may not be captured by traditional analyses relying on Fourier transforms (Donoghue, Schaworonkow, & Voytek, 2021). Therefore, we here explore indexing temporal expectations in the silence using multi scale entropy – a measure of signal irregularity (Kosciessa, Kloosterman, & Garrett, 2020) – and multivariate pattern decoding. We found higher entropy in the silence following pattern-based than random sequences, but we could not rule out that these entropy differences could be explained by differences in overall neural variability (i.e., the signal standard deviation). Finally, we could classify above chance, from the raw EEG signal in the silence, whether participants had previously listened to a beat-based, pattern-based, or random sequence. Together, these findings support the notion that different neural mechanisms underlie beat-based and pattern-based expectations, and provide several tools to explore temporal expectations in a time resolved way.

Results

Behavioral effects of beat-based expectations outlast those of pattern-based expectations

Participants listened to three types of sequences (Figure 1). The sequences either contained 1) a regular, periodic beat every 600 ms, with varying surface structure in the pattern (beat-based sequences), 2) a repeating – and therefore predictable – surface pattern without an underlying periodicity (pattern-based sequences), or 3) neither repeating patterns, nor a regular beat (random sequences). Each sequence was followed by a silent period of at least 2100 ms, equaling at least 3 beat cycles or a full cycle of the repeating pattern. To assess the persistence of temporal expectations behaviorally, a probe tone could be presented in each trial at different temporal positions across the silent period, and participants provided ratings for how well probe tones fitted the preceding rhythm (see Figure 1 and Materials and Methods for more details). In Experiment 1 (32 healthy adult participants), no EEG was recorded, and 6 positions in the silence were probed. In Experiment 2 (27 healthy adult participants), probe tones appeared at only 3 positions, and on only 25% of the trials, to allow for the remaining 75% of the trials to be silent for use in the EEG analyses.

The effects of Position (600, 780, 885, 990, 1200, and 1485 ms; only the latter three in Experiment 2) and Condition (beat-based, pattern-based, and random) were examined with a mixed-effects ordinal regression model. To explore the possible effects of musical expertise, we added the Gold-MSI musical training subscale (see Materials and Methods) as a covariate to the model. Since the model with and without Musical Training yielded highly similar results, we will here only report the full model, including this covariate. Figure 2A shows the average ratings for each condition and probe position from Experiment 1. Visual inspection of this figure suggests that beat-based expectations were associated with higher fitness ratings for sounds at expected times than unexpected times for two beat cycles in the silence window (at 600 and 1200 ms), while the effects of pattern-based expectations appeared to subside after the

first expected time point in the silence window (780 ms). This was confirmed by our statistical analyses. The ordinal regression showed main effects of Condition ($\chi^2(2) = 300.72$, $p < 0.001$), Position ($\chi^2(5) = 1067.05$, $p < 0.001$), as well as Musical Training ($\chi^2(1) = 5.16$, $p = 0.023$). However, crucially, these main effects were accompanied by a very large two-way interaction between Condition and Position ($\chi^2(10) = 2478.98$, $p < 0.001$), showing that the effects of beat-based and pattern-based expectations on fitness ratings differed, depending on the position of the probe. We found additional smaller interactions between Position and Musical Training ($\chi^2(5) = 67.26$, $p < 0.001$), Condition and Musical Training ($\chi^2(2) = 6.01$, $p = 0.05$), and, interestingly, Condition, Position, and Musical Training ($\chi^2(10) = 204.88$, $p < 0.001$). Following the interactions, tests of simple main effects showed that the effect of Position was significant in all conditions (all *p*s < 0.001), and the effect of Condition was significant for all probe positions (all *p*s < 0.001). The main effect of Position in the random condition showed that even after sequences in which no specific temporal structure was present, ratings depended on the position of the probe. This likely was due to recency effects. To account for these effects when comparing the ratings at different positions in the beat-based and pattern-based conditions, we subtracted the ratings in the random condition at each position from the ratings in the other conditions. The baseline corrected model (Figure 2B) showed similar interactions between Position and Condition ($\chi^2(5) = 2147.55$, $p < 0.001$), and between Position, Condition, and Musical Training ($\chi^2(5) = 154.46$, $p < 0.001$). The latter indicated that the correlation between Musical Training and the rating score depended on both the Position and the Condition of the probe tone.

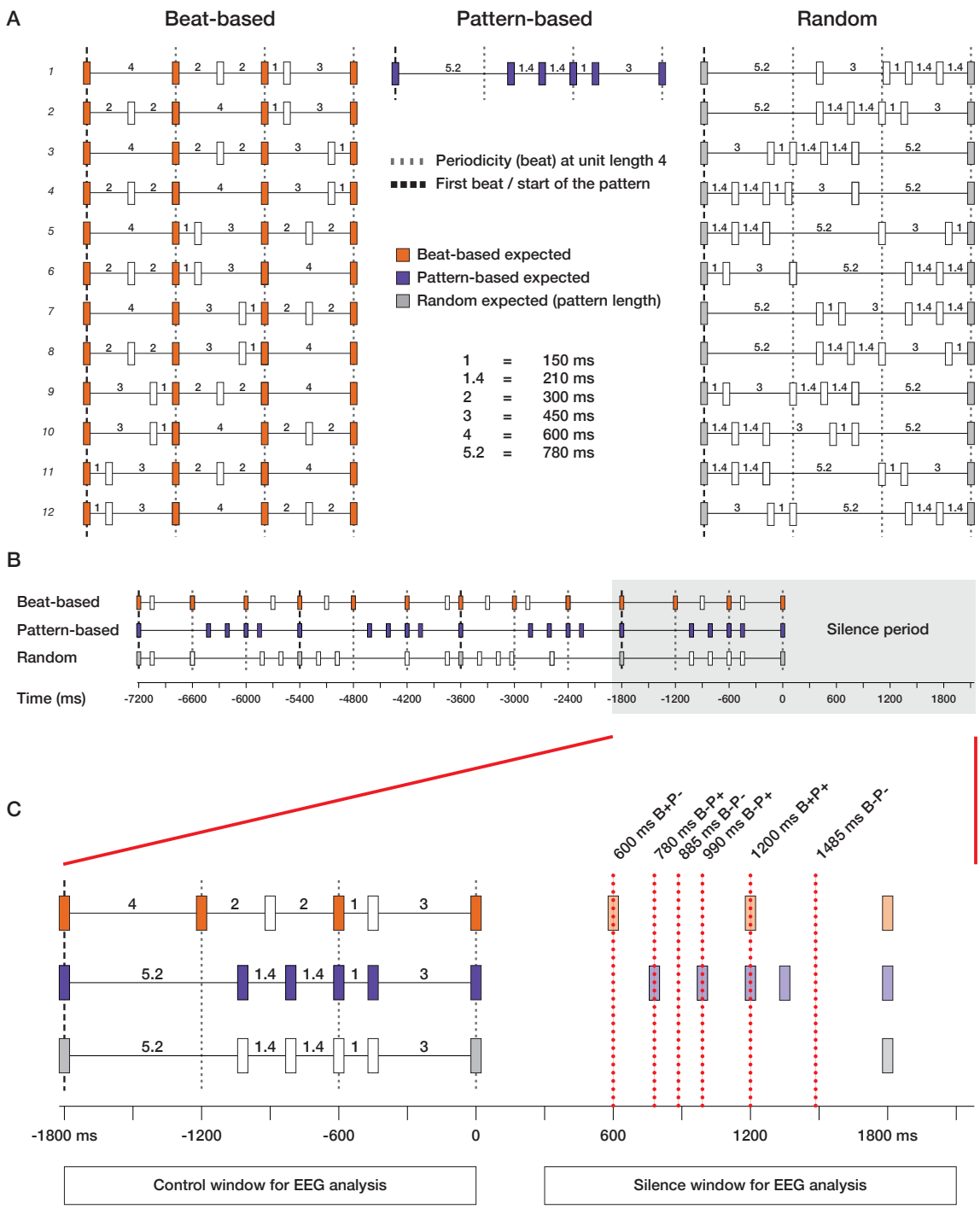

Figure 1. Schematic overview of the rhythmic stimuli used and the task. A) Twelve patterns of five temporal intervals with integer ratio durations and an event at each 600 ms period were created to form beat-based sequences. Equivalent patterns without a regular beat every 600 ms were created by using non-integer ratio durations, while keeping the number of intervals and grouping structure the same (random condition). For the pattern-based sequences, only pattern 1 was used, to allow for learning of the intervals. B) Four semi-randomly chosen patterns were concatenated to form rhythmic sequences. In both the beat-based and random sequences, the last pattern was always pattern 1 or 2, to equate the acoustic context preceding the silent period. C) To measure behavioral effects of expectations, a probe tone could appear at various temporal positions in the silent period (indicated by the dashed red lines), predictable based on a beat (B+, light orange), predictable based on the pattern (P+, light purple), or unpredictable based on the beat (B-) or pattern (P-). Subjects had to indicate how well the probe tone fitted the preceding rhythm. In Experiment 1, all 6 probe tone positions were used. In Experiment 2, only the last 3 probe positions were used.

Beat-based expectations. At both 600 ms and 1200 ms (expected in terms of a beat), probes in the beat-based condition were rated as better fitting than probes in the random condition (both *p*s < 0.001 as evident from the full model). At 780, 885, 990, and 1485 ms (unexpected in terms of a beat), probes in the beat-condition were rated as worse fitting than probes in the random condition (all *p*s < 0.001). Moreover, within the beat-based condition, at 600 ms, baseline corrected ratings were higher than at any other probe position (all *p*s < 0.001), and at 1200 ms baseline corrected ratings were higher than at 780, 885, 990, or 1485 ms (all *p*s < 0.001). Baseline corrected ratings for probes at 780, 885, and 990 ms (all unexpected in terms of the beat) did not differ from each (all *p*s > 0.38). Probes at 1485 ms (unexpected in terms of the beat) were rated as better fitting than probes at 780, 885, and 990 ms (all *p*s < 0.001).

As can be seen in Figure 2C and 2D, higher scores on the Musical Training questionnaire were associated with higher fitness ratings in the beat-based condition at 600 and 1200 ms (expected in terms of the beat), but lower fitness ratings at 780, 885, 990, and 1485 ms (unexpected in terms of the beat). Slopes reach significance at all positions except 1485 ms (all *p*s < 0.002). Also, the association between Musical Training and ratings differed between beat-based and pattern-based conditions, at 600, 780, 885, and 1200 ms (all *p*s < 0.001).

To sum up, for the beat-based sequences, we could observe a clear pattern in the results indicating that beat-based expectations were used to rate the probes well into the silence window, affecting ratings up to 1200 ms after the onset of the last sound. Beat-based expectations lead to higher ratings for expected probes (600 and 1200 ms), and lower ratings for unexpected probes (780, 885, 990, and 1485 ms), both when comparing ratings for each position to the random condition, and when comparing ratings for each position within the beat-based condition. At 1200 ms, these effects resulted in a classic inverted U-curve, as previously associated with beat-based processing (Bauer et al., 2015; Jones et al., 2002), with optimal performance on the beat, and diminished performance on either side (e.g., both earlier and

later). The effects of beat-based expectations did diminish over time, as is apparent from differences between ratings at 600 and 1200 ms, and at 1485 ms and other unexpected time points. Both the enhancing and attenuating effects of beat-based expectations were correlated with musical training. It is worth noting that the longer lasting effects of beat-based expectations (at 1200 ms) were very heterogenous in our participant pool. Out of 32 participants in Experiment 1, only 18 showed the inverted U, with higher ratings at 1200 than at 990 and 1485 ms.

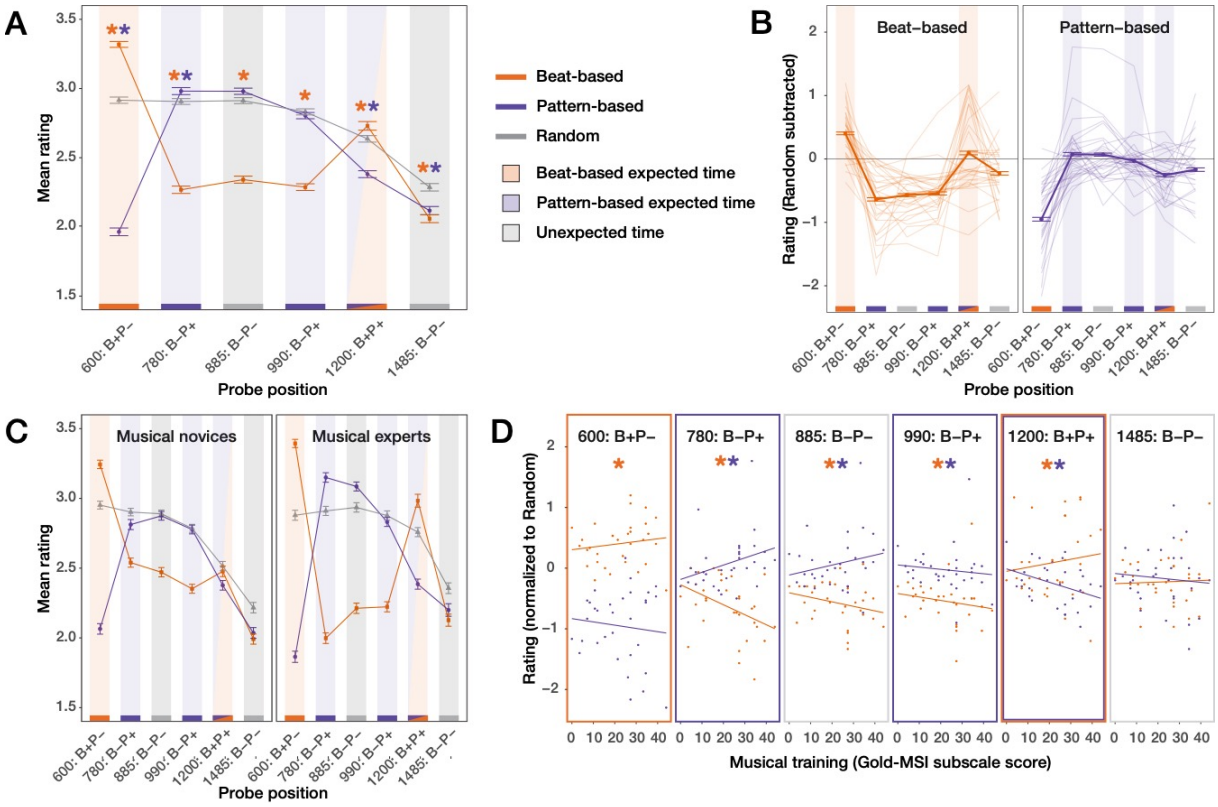

Figure 2. The effects of beat-based expectations on fitness ratings outlast those of pattern-based expectations, and are associated with musical training. A) Mean ratings for all conditions and positions. Colored stars indicate positions where ratings in the beat-based (orange) and pattern-based (purple) conditions differed from the random condition. B) Single participant data, with the random condition subtracted to account for serial position effects. For the beat-based but not the pattern-based condition, ratings follow the expectedness pattern, indicated by colored lines on the bottom of the plots (orange: expected based on the beat; purple: expected based on pattern; grey: neither). C) Data median split based on scores on the musical training questionnaire. The pattern of results, while present for both groups of participants, is enhanced for the group of participants with most musical training ("experts"). Note: the median split is for visualization purposes only, the models were run with musical training as a covariate. Error bars in panels A-C are 2 standard errors (note: these are computed on the complete dataset, not the participant averages, as the ordinal model is ran on trial-level data). D) Association between musical training and rating for each condition and position. A positive association was observed for the beat-based condition at 600 and 1200 ms (expected positions) and for the pattern-based condition at 780 (expected) and 885 ms (unexpected). Negative associations were observed in the beat-based condition at 780, 885, and 990 ms (all unexpected), and in the pattern-based condition at 990 (unexpected) and 1200 ms (expected).

Pattern-based expectations. For pattern-based sequences, ratings at 600 ms (unexpected based on the pattern) were lower than for the random sequences ($p < 0.001$) and lower than at any other position (all ps < 0.001, baseline corrected model), showing that participants also formed predictions based on the sequences. In line with this, at 780 ms (expected in terms of the pattern), ratings were higher in the pattern-based condition than in the random condition ($p = 0.009$). After this point, ratings did not differ between pattern-based and random conditions for probes at 885 (unexpected) and 990 (expected) ms, suggesting that the pattern-based expectations only exerted an influence on the rating in the beginning of the silent period. In line with this, the ratings continued to deviate from what would be predicted based on the pattern, with lower ratings for the pattern-based than random condition at 1200 ms (expected; $p < 0.001$). Ratings were also lower for the pattern-based than random condition at 1485 ms (unexpected, $p < 0.001$), and ratings at 780 ms (expected) did not differ from ratings at 885 ms (unexpected), while being higher than at 990 (expected), 1200 (expected), and 1485 (unexpected) ms (all ps < 0.032). In addition, ratings at 885 and 990 ms were higher than at 1200 and 1485 ms (all ps < 0.001).

As for beat-based expectations, for pattern-based expectations, there was a positive association between ratings and musical training at an expected time point (780 ms; $p < 0.001$), and a negative, albeit nonsignificant, association at an unexpected time point (600 ms; $p = 0.08$). However, at 885 ms (unexpected in terms of the pattern), the results behaved like at 780 ms, with higher ratings associated with more Musical Training ($p < 0.001$). At 990 and 1200 ms, Musical Training was associated with lower ratings (both ps < 0.004), but these results are counterintuitive, as these are expected positions based on the pattern.

The results for pattern-based expectations suggest that just like for beat-based expectations, participants were able to predict the timing of probes based on the preceding sequence. However, the results show that while this was still the case at 780 ms after the onset

of the last tone, at later probe positions, the effects of pattern-based expectations did not reflect the preceding sequence. At 885 ms, the results, both in terms of the ratings and how they were associated with musical training, behaved similar to at 780 ms. This may indicate that the pattern-based expectations at this point were not precise enough to differentiate between positions, even when the deviation in time is half the learned interval (105 of 210 ms). After this point, the results suggest that participants did not correctly represent the preceding sequence anymore.

Figure 3 shows the behavioral results obtained from the EEG experiment. Replicating Experiment 1, we found main effects of Condition ($\chi^2(2) = 92.30$, $p < 0.001$) and Position ($\chi^2(2) = 49.36$, $p < 0.001$), accompanied by interactions between Condition and Position ($\chi^2(4) = 96.62$, $p < 0.001$), and Condition, Position, and Musical Training ($\chi^2(4) = 11.04$, $p = 0.03$). Following the analysis strategy from Experiment 1, we subtracted the ratings from the Random condition from the ratings for the other two conditions, yielding a baseline corrected model with similar interactions (Condition and Position: $\chi^2(2) = 31.46$, $p < 0.001$; Condition, Position, and Musical Training: $\chi^2(2) = 8.29$, $p = 0.02$).

In line with the preceding beat-based sequences, probes at 990 and 1485 ms (both unexpected times based on the beat) were rated lower in the beat-based condition than in the random and pattern-based conditions (all *p*s < 0.004), and within the beat-based condition, probes at 1200 ms were rated as better fitting than at 990 ($p < 0.001$) and 1485 ms, though the latter difference did not reach significance. Thus, like in Experiment 1, we found an inverted U-curve at 1200 ms after the final tone, suggestive of beat-based expectations lasting at least two beat cycles. As in Experiment 1, at 1200 ms (expected based on the beat), higher ratings in the beat-based condition were associated with more musical training ($p = 0.008$), suggesting that the effects of beat-based expectations correlate with musical expertise. Additionally, probes at 990 ms were rated lower than at 1485 ms ($p < 0.001$), possibly because the effects of beat-based expectations diminished over the course of the silence.

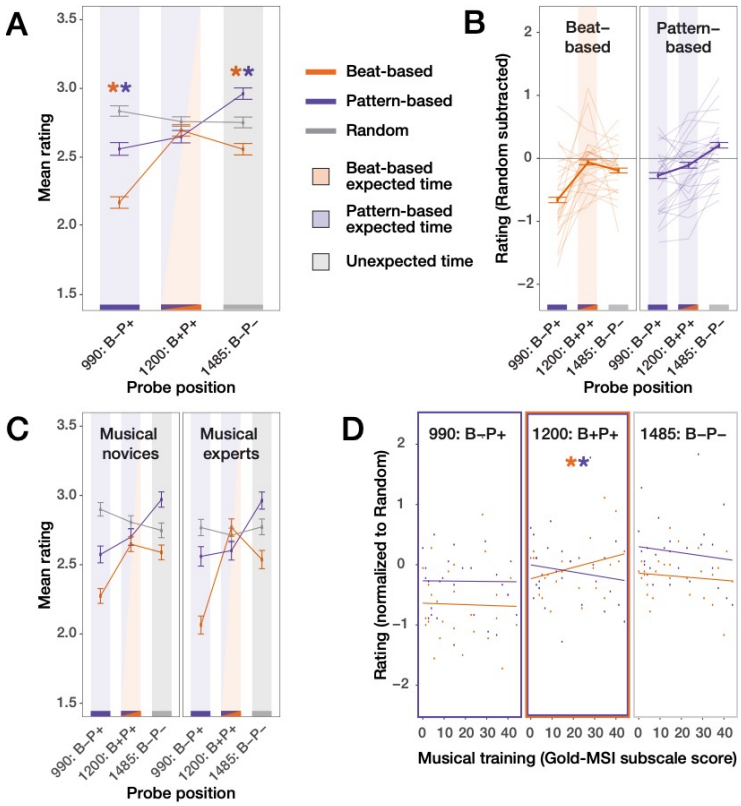

Figure 3. Behavioral results during the EEG experiment replicate findings from Experiment 1. Even with only 18 trials per participant per condition and position, we could replicate the inverted U-curve for beat-based sequences at the end of the silence epoch. Like in Experiment 1, the results for the pattern-based condition do not follow the pattern, but instead, are consistent with building expectations for the next trial. A) Mean ratings for all conditions and positions. Colored stars indicate positions where ratings in the beat-based (orange) and pattern-based (purple) conditions differed from the random condition. B) Single participant data, with the random condition subtracted to account for serial position effects. C) Data median split based on scores on the musical training questionnaire. Error bars in panels A-C are 2 standard errors D) Association between musical training and rating for each condition and position. See Figure 2 for more details.

In the pattern-based condition, ratings did not follow the pattern of the preceding sequences. At 990 ms (expected based on the pattern), probes were rated as worse fitting than in the random condition ($p < 0.001$), and as worse fitting than at 1200 (expected) and 1485 (unexpected) ms (both ps < 0.02), while at 1485 ms (unexpected), probes were rated as better fitting than in the random condition, and as better fitting than at 1200 (expected) ms (both ps $<$ 0.001). Also, at 1200 ms (expected), higher ratings were associated with less musical training for the pattern-based condition ($p = 0.01$), contrary to what would be expected if the effects of expectations are enlarged in musical experts.

Thus, the behavioral results from Experiment 2, though based on less trials than Experiment 1, suggest a similar pattern as found in Experiment 1: while beat-based expectations exert their effect well into the silent period, with participant faithfully following the beat in their goodness-of-fit ratings, pattern-based expectations fail to affect ratings in the second half of the silent period in a manner consistent with the learned pattern. Albeit speculatively, the results for the pattern-based expectations may be more in line with expectations for the start of the next trial leading to higher ratings for probe positions closer to the end of the silent period, as in Experiment 2, after non-probe trials, the next trial followed each silent period at a somewhat predictable time.

Contingent Negative Variation (CNV) is elicited by pattern-based, but not beat-based sequences

To assess possible differences in time-domain data caused by expectations, we first looked at ERPs in the silence window. In line with previous research, we may expect climbing neuronal activity, or a CNV, flexibly adapting its slope to peak at the moment that participants expect the next event (Breska & Deouell, 2017; Breska & Ivry, 2020; Damsma, Schlichting, & van Rijn, 2021; Mento, 2013). For beat-based sequences, the strongest expectation for the next event would be at 600 ms after the onset of the last sound (see Figure 1). For pattern-based expectations, this would be at 780 ms. To assess the presence of differences between conditions in the time-domain waveforms in the silent period, we used cluster-based permutation testing (Oostenveld, Fries, Maris, & Schoffelen, 2011), comparing all timepoints from 300 till 1200 ms after the onset of the final sound (see Figure 1). This excluded a large portion of the ERP response to the final sound, and included both the first expected moments for beat-based (600 ms) and pattern-based (780 ms) expectations, and additional time to allow for an evaluation of possible return to baseline of the CNV (Breska & Deouell, 2017).

Figure 4 shows the average ERPs for each condition, without baseline correction (see Materials and Methods), and scalp topographies for windows in which we found significant clusters. A CNV-like frontocentral negativity was observed early in the silence window in the pattern-based and random conditions, but not the beat-based condition. Without baseline correction, the random condition elicited a significantly more negative deflection (p = 0.011) than the beat-based condition in a frontocentral cluster between 300 and 614 ms after the onset of the last sound (though note that we did not include timepoints preceding 300 ms in the cluster-based tests). Likewise, in the same latency range (300 – 473 ms), there was a trend for the pattern-based condition to elicit a larger negative deflection than the beat-based condition (p = 0.077). Thus, in a window between approximately 300 and 450 ms, we found tentative evidence for more negative-going waveforms in both the random and pattern-based condition compared to the beat-based condition, with scalp topographies consistent with a CNV (Figure 4B). In a later window, the pattern-based condition elicited a second negative deflection, which showed a trend to be larger than in the beat-based condition (p = 0.077, 864 – 1005 ms).

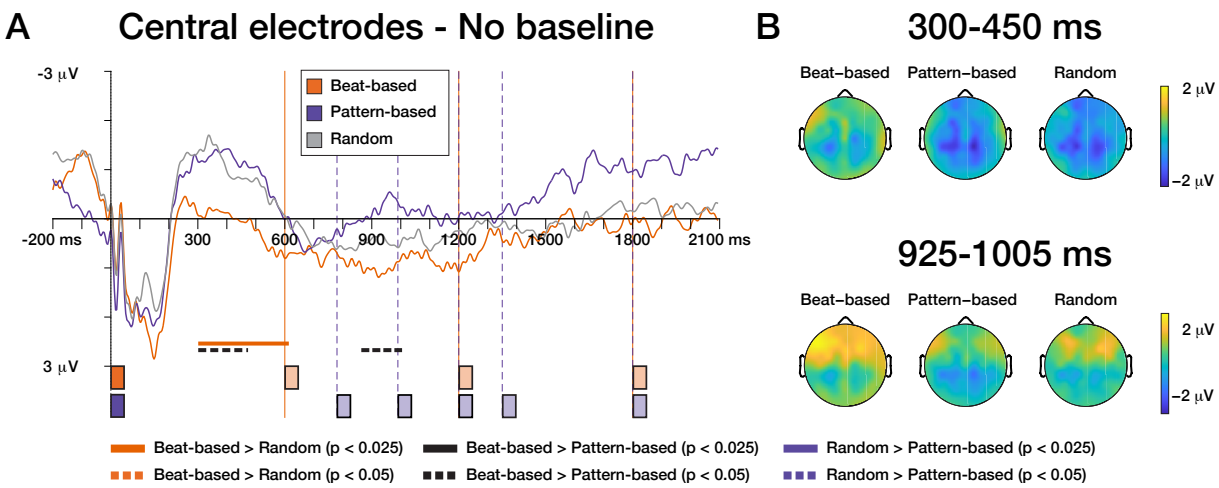

Figure 4. A CNV is present for pattern-based and random conditions, but not for the beat-based condition. A) Left panel show the grand average waveforms for the silence window for a central electrode cluster (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2). Time 0 is the onset of the last tone of the sequence. Colored bars on the bottom of the plots, and vertical orange and purple lines, indicate at which times a tone would be expected based on the beat (light orange) and memory (light purple). Note that these are expected times, but no sounds were played during the window shown after time 0. B) The scalp distributions for windows in which a significant cluster was observed.

With the more traditional pre-cue baseline (Figure S1), there was a trend for the random condition to elicit a more negative deflection than the beat-based condition (p = 0.058, 300 – 446 ms) and for the pattern-based condition to elicit a more negative deflection than the random condition (early window: p = 0.083, 380 – 551 ms; later window: p = 0.022, 774 – 1200 ms). Note that a significant cluster spanning the entire analysis window of 300 – 1200 ms was found when comparing between the pattern-based and beat-based condition with baseline correction (p = 0.011), indicative of a possible overall shift due to the baseline correction. However, while the results were somewhat different depending on the choice of baseline, the overall picture is the same, with significant clusters in an early and later window, and negative deflections with a frontocentral scalp distribution consistent with a CNV for the pattern-based and random but not the beat-based condition. As an exploratory analysis, since musical training was related to larger effects of entrainment in behavior, we also examined whether there was a relationship between the amplitude of the CNV (extracted from the time-electrode clusters depicted in Figure 4, see Materials and Methods for details) and musical expertise. However, we found no difference in the effects of expectations on CNV amplitude between musically trained and untrained subjects (all ps > 0.3).

Frequency tagging shows persistent power at the beat frequency following beat-based sequences

A strong prediction of entrainment theories is that the entrainment outlasts stimulation with the entraining stimulus. Therefore, we next looked at the frequency content of the EEG signal in the silent period. Specifically, we predicted that if entrainment occurs, we would find enhanced power at the frequencies associated with the rhythmic sounds to be found in the silence. First, we performed a spectral analysis on the sound signal, to find which frequencies were most prominent in the sound within the delta range (0.5-4 Hz). As can be seen in Figure

5, the beat-based condition contained two clear peaks at 1.67 Hz (the beat) and 3.33 Hz (subdivisions of the beat). For the pattern-based and random condition, peaks were more distributed, in line with the more irregular nature of these rhythms, and the highest peaks in the delta range were at 2.22 and 3.89 Hz. Next, we performed an FFT on the EEG data from both the control window and the silence (see Figure 1), averaged for each condition separately, to assess evoked (e.g., phase-locked) activity (Lenc et al., 2021; Nozaradan et al., 2011; Nozaradan, Peretz, & Mouraux, 2012). Using cluster-based permutation tests on the normalized spectral data from all electrodes, we assessed differences between conditions for the frequencies of interest (e.g., those most prominent in the sound signal). We included the frequencies that showed the highest peaks in the spectral analysis of the sound (i.e. 1.67 Hz, 2.22 Hz, and 3.33 Hz), except for 3.89 Hz, since a peak at this frequency did was absent on visual inspection in the spectral decomposition of the EEG data.

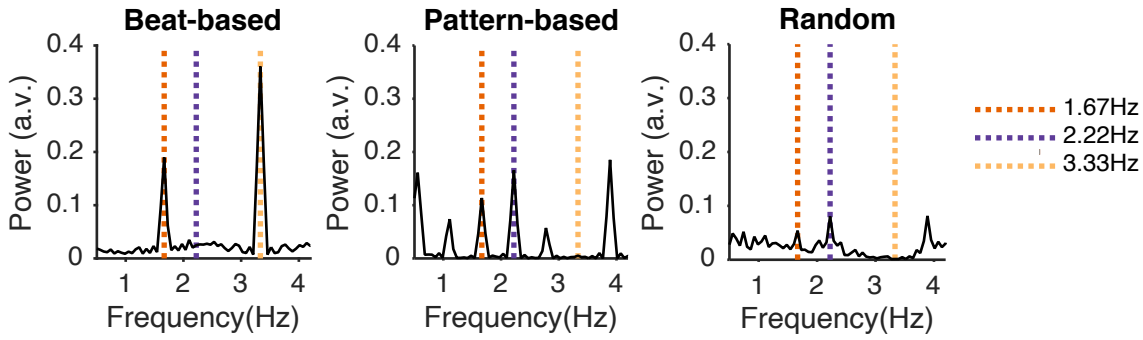

Figure 5. Spectral analysis of the sound signal of the different rhythmic sequences. Ten sequences per condition were generated to base the spectral analysis on. Note that these sequences were identical for the pattern-based condition, but semi-random for the beat-based and random conditions. The envelope of each sequence was obtained by performing a Hilbert transform, and subsequently, a fast fourier transform (fft) was used to obtain the spectral decomposition and power values were averaged over ten sequences.

In Figure 6, the average power for all electrodes, separated for each condition in the control (i.e. during the auditory sequence) and silence windows is depicted as a function of frequency. In the control window (Figure 6, left, and Figure 7, top), the frequency response

followed the sound input. That is, at the beat frequency (1.67 Hz), significant clusters indicated higher power in the control window for the beat-based sequences than the pattern-based ($p < 0.001$) and random ($p = 0.006$) sequences. At 2.22 Hz, prominent in the pattern-based and random sound sequences, higher power was observed in the EEG signal in the pattern-based than beat-based ($p = 0.02$) and random ($p = 0.023$) sequences, and higher power was observed in the random than beat-based sequences ($p = 0.035$). Finally, at 3.33 Hz (subdivisions of the beat), power in the control window was larger for the beat-based than pattern-based ($p = 0.013$) and random ($p = 0.005$) conditions. Thus, in the control window, the EEG signal reflected the spectral properties of the sound signal, as can be expected, since each sound will have elicited an ERP, which are represented in steady-state potentials, and thus picked up by the frequency analysis (Keitel, Obleser, Jessen, & Henry, 2021).

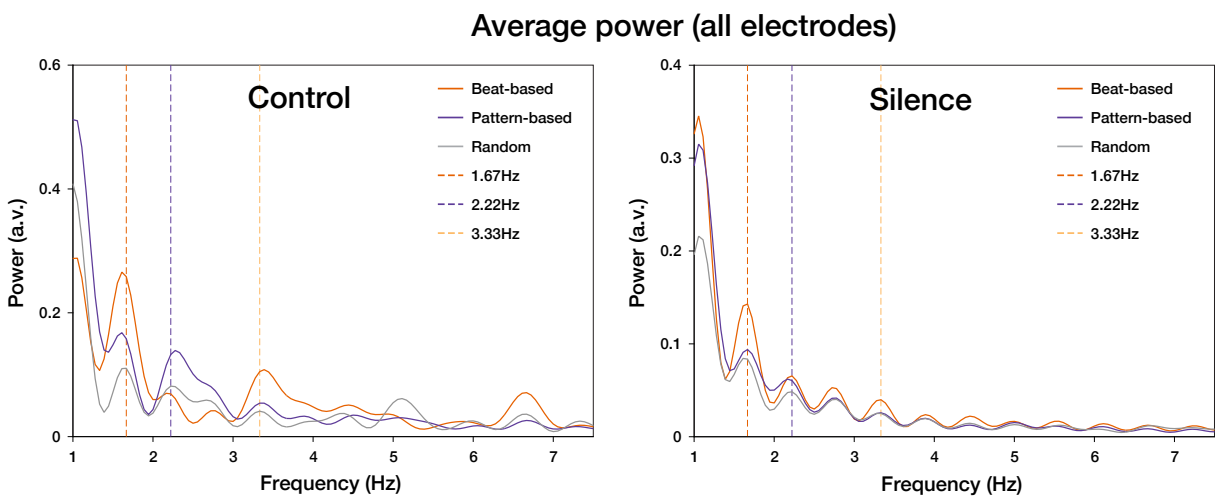

Figure 6. Oscillatory power at the beat frequency persists during the silence window. In the control window, during auditory stimulation, peaks can be observed at all frequencies of interest, and for the relevant conditions (1.67 and 3.33 Hz in the beat-based condition, and 2.22 Hz in the pattern-based condition). In the silence window, the peaks at 1.67 and 3.33 Hz are larger in the beat-based condition than in the pattern-based and random conditions, while peaks at 2.22 Hz do not differ between conditions. This suggests that only beat-based expectations persisted in the silence window.

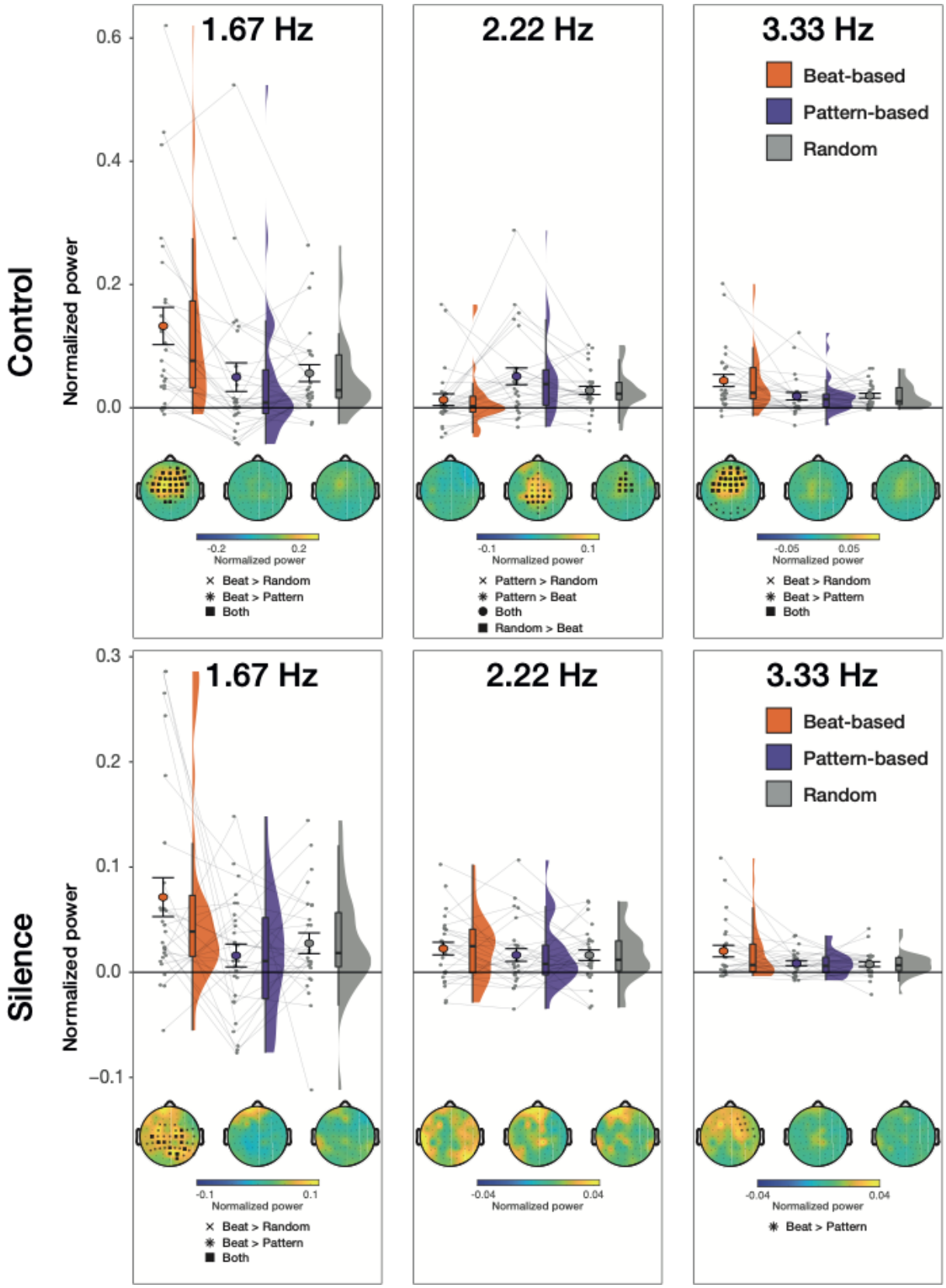

Figure 7. Individual participant data shows heterogeneity of the effects of temporal expectations on spectral power. All plots depict the spectral power at the frequency of interest, averaged over all electrodes, and normalized to account for the 1/f distribution (see Materials and Methods). Electrodes contributing to significant clusters are highlighted below each plot.

Importantly, during the silence (Figure 6, right, and Figure 7, bottom), no significant clusters were found at 2.22 Hz, but at the beat frequency (1.67 Hz), power was significantly larger for the beat-based than the pattern-based ($p = 0.01$) and random ($p = 0.012$) conditions. In addition, at 3.33 Hz, power for the beat-based condition was larger than for the pattern-based condition ($p = 0.038$), with a trend when comparing the beat-based with the random condition ($p = 0.077$). Thus, while the pattern-based and random conditions showed tracking of the sound during stimulation (which is sometimes considered entrainment "in the broad sense" (Obleser & Kayser, 2019)), in the silence, entrainment was only present for the beat-based condition. This finding fits our behavioral observations that beat-based expectations persisted longer in the silence than pattern-based expectations. Note however that like for the behavioral results indicative of entrainment, there was large heterogeneity between participants (see Figure 7). While the power differences were significant in the overall cluster-based analyses, out of 27 participants, only 16 showed on average (i.e. over all electrodes) numerically larger power in the beat-based condition at the beat frequency when compared to both the random and pattern-based condition. As for the CNV, musical training did not affect the difference in power between the beat-based and random condition at 1.67 Hz, nor the difference in power between the pattern-based and random condition at 2.22 Hz (both ps > 0.54).

Multiscale entropy as a non-stationary measure of temporal expectations

Several measurement challenges are associated with studying entrainment of low frequency oscillations. The most prominent one, the presence of evoked activity in the same frequency range as putative oscillations, we circumvent by measuring the EEG signal in the silence. Here, we used frequency tagging to probe the spectral properties of the EEG signal in the silence, and since this technique takes the entire silence window into account, we can achieve enough frequency specificity to disentangle the frequency of the beat (1.67 Hz) and that of the pattern (2.22 Hz). However, the assumption of stationarity in the signal is of course

a simplification that may not hold for the actual brain signal. Once the rhythmic sensory input ceases, the oscillatory activity in the brain may quickly return to an intrinsic eigenfrequency (Doelling & Florencia Assaneo, 2021). Moreover, the non-sinusoidal aspects of brain activity may themselves be important for cognition (Waschke, Kloosterman, Obleser, & Garrett, 2021). To account for these observations, here, we looked at the neural variability during the presentation of the rhythm and during the silence using multiscale entropy (MSE). Note that we consider this analysis exploratory in nature, and results should be interpreted as such. These results may provide a useful tool to study the neural dynamics underlying rhythm processing and temporal expectations, which are often hard to study due to methodological issues with analyzing the EEG signal (Zoefel et al., 2018).

MSE is a measure of signal irregularity. To compute MSE, the EEG signal is divided into patterns of a certain length, and throughout the signal, the number of repeating patterns is counted. More repetitions indicate a more regular signal, and yield a lower entropy value. By calculating entropy for patterns of different lengths ("multiscale"), the contributions of slower and faster timescales in the signal can be assessed. However, the mapping between entropy timescales and spectral frequencies is not absolute, especially since entropy is not per se related to a signal being oscillatory in nature (Kloosterman, Kosciessa, Lindenberger, Fahrenfort, & Garrett, 2020; Kosciessa et al., 2020). The advantage of using MSE is that it does not require filtering of the data, and it does not assume stationarity (e.g., it can pick up on regularities that are asymmetrical, or that do not have a fixed amplitude or period). Here, we computed MSE on the control and silence epochs separately. Given that MSE indexes signal irregularity, we would expect entropy to be higher for the random and pattern-based conditions than for the beat-based condition in the silence, given our above results that suggest persistent entrainment in the beat-based condition only.

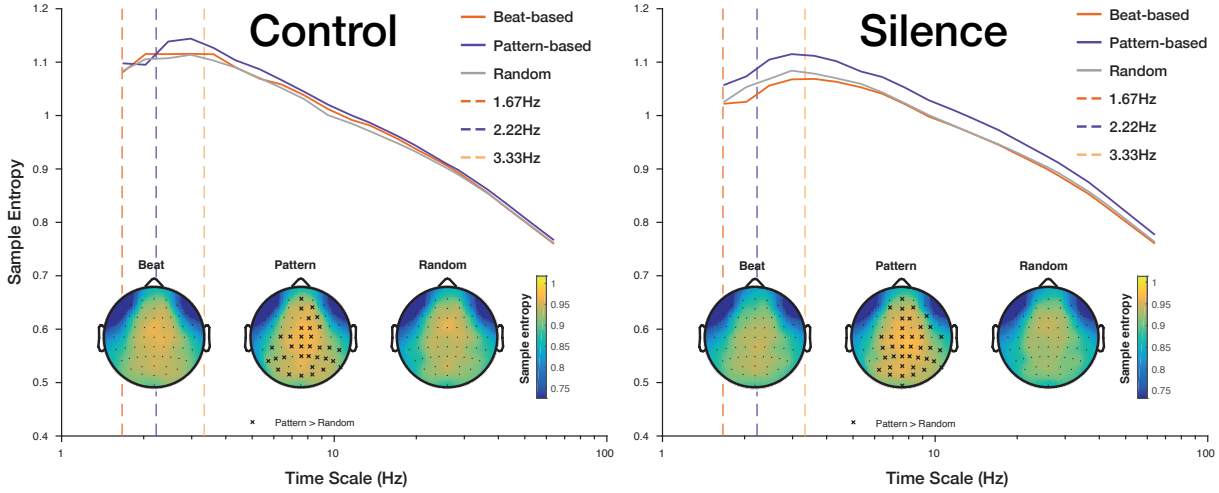

Figure 8. Entropy was higher for pattern-based than random sequences in both control and silence windows. Entropy in the Control window (left) and Silence window (right), averaged over frontocentral electrodes, and scalp distributions averaged over all timescales, depicting electrodes contributing to significant clusters.

A cluster-based test on all electrodes and timescales showed that both in the control window, and in the silence window, entropy was higher for the pattern-based than for the random condition (control: $p = 0.03$; silence: $p = 0.021$). Electrodes contributing to significant clusters are indicated in Figure 8. Note that in the silence window, all but the two highest timescales were included in the cluster. In the control window, the cluster spanned all timescales from 35 till 406 ms (see Materials and Methods for an explanation of the timescales). These results suggest that the signal was more irregular in the pattern-based than the random condition, over a broad range of timescales. Neither the beat-based condition compared to pattern-based nor the beat-based compared to the random condition reached significant differences in entropy, in either control or silence windows (all ps > 0.24).

Entropy has been related to various other EEG measures, such as spectral power and overall differences in signal variability (Kosciessa et al., 2020). To account for these, we ran several additional analyses. First, to check whether the differences in signal variability may have been caused by differences in low frequency activity, we repeated the MSE analysis on high-pass filtered data (Kloosterman et al., 2020), using a 5 Hz filter. This completely removed

the effects (all *p*s > 0.43), suggesting that the differences between conditions were caused by low frequency activity in the signal (see Figure S2).

Second, we checked whether differences in signal variability caused the differences in entropy, since entropy is calculated relative to the overall signal standard deviation (e.g., a pattern is considered a match at a lower threshold when overall signal variability is high). As can be seen in Figure S3, the similarity bounds used to compute entropy (derived from the time-domain signal standard deviation) differed between conditions, and this difference mirrored the differences in entropy, suggesting that at least some of the variance we observed was due to overall signal variability, and not necessarily signal irregularity. Finally, musical training did not affect the difference in sample entropy between the beat-based and random condition, nor the difference in entropy between the pattern-based and random condition (both *p*s > 0.66).

Multivariate decoding as a time-resolved method for studying entrainment

While MSE provides us with a method to look at possible non-sinusoidal contributions to the EEG signal related to temporal expectations, because of our interest in the low frequency components of the signal, we cannot look at how entrainment evolves over time. While both the frequency tagging approach and the MSE take the whole silence window into account, in reality, oscillations may quickly return to baseline after the sensory input stops (Doelling & Florencia Assaneo, 2021), and there is a limited time window to actually measure persistent oscillations in the signal. Therefore, in a final analysis, we explored the use of multivariate decoding as a time-resolved measure of neural activity. Our exploratory approach is based on the assumption that temporal expectations are always coupled with feature or spatial expectations (e.g., we cannot predict "when" without also predicting "what"), as suggested by studies showing that we only use temporal expectations to improve perception if we can also predict the content (Morillon et al., 2016) or location (O'Reilly et al., 2008) of an upcoming event. Thus, we expected to be able to decode the representation of the expected sound at

expected moments. As the expected moments are different for each condition, this then allows us to decode in the silence window whether participants were previously listening to a beat-based, pattern-based, or random sequence. As decoding is done on a time point by time point basis, we can also look at how temporal expectations affect the EEG signal over time.

Using a linear discriminant classifier and 10-fold cross-validation (Fahrenfort, van Driel, van Gaal, & Olivers, 2018), we decoded all combinations of the three conditions against each other (i.e., beat-based vs. pattern-based, beat-based vs. random, and pattern-based vs. random). We ran the decoding algorithm while training and testing on each time point in a window starting at 1800 ms before the onset of the final sound, and ending at 2100 ms after the onset of the final sound (see Figure 1), creating a temporal generalization matrix of classification accuracy at each possible combination of training and testing time points (King & Dehaene, 2014). We expected that training at expected times would yield above chance performance when testing at expected times, regardless of whether these time points were the same (e.g., when training at 600 ms, we expected to be able to accurately distinguish the beat-based from the random condition when testing at not just 600 ms, but also at 1200 and 1800 ms, as all these times were on the beat, or similarly expected). Classification accuracies for each comparison of two conditions at all time points in the silence window (300-2100 ms after the onset of the last sound) were submitted to cluster-based permutation tests to determine above-chance decoding.

Figure 9A shows the temporal generalization matrices for each comparison, with significant clusters indicated by a black contour. In the silence window we found above chance decoding when decoding the beat-based against the random condition ($p = 0.043$), the pattern-based against the random condition ($p < 0.001$), and the beat-based and pattern-based conditions against each other ($p = 0.009$), indicating that based on the EEG signal in the silence, we could classify which type of rhythm participants had heard just before. However, looking at the

temporal generalization matrices, it becomes apparent that this above-chance decoding was not due to recurrent activity for expected events. Only clusters on the diagonal were significant for each comparison. In addition, decoding for all three comparisons was best in the second halve of the silence window, which is where the probe tones were presented. This suggests that the decoding mainly picked up on task-related differences. While the probes were physically identical across conditions, participants may have had different strategies to perform the task, depending on the type of sequence, and this could have resulted in above-chance decoding related to the task and probes, even when only analyzing the silence window.

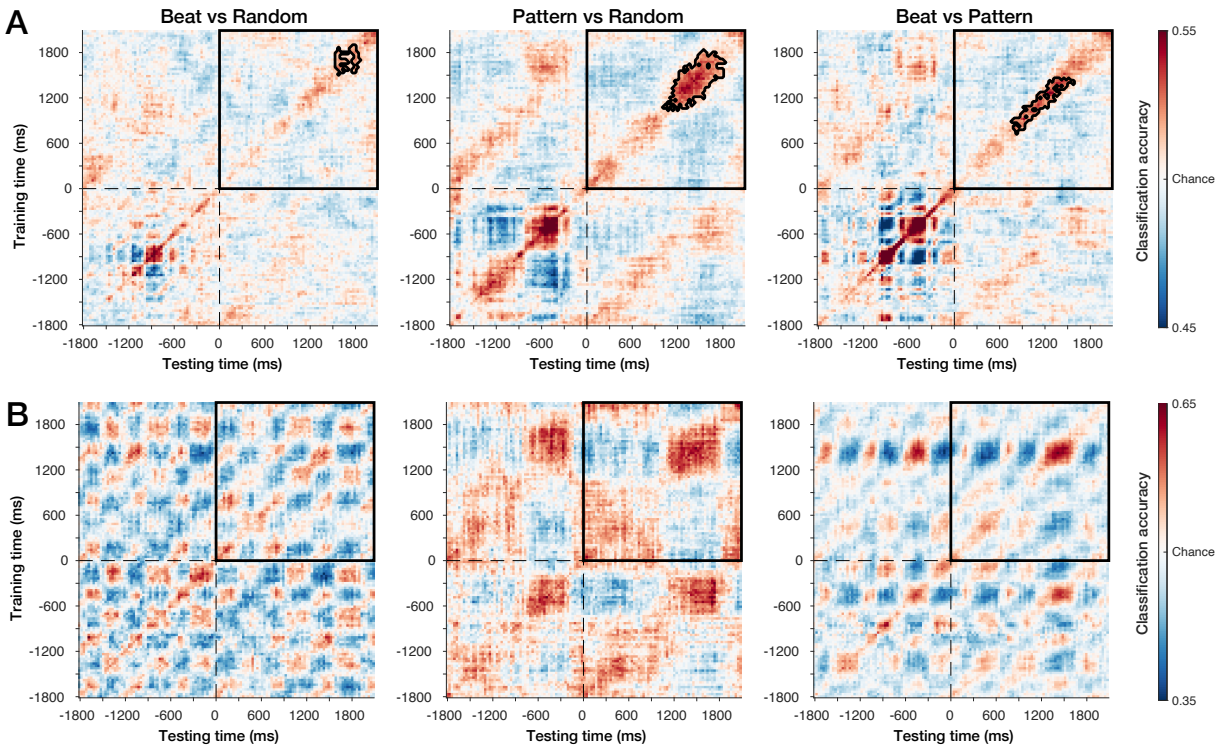

Figure 9. No recurrent activation in group average decoding, example of recurrency in single participant. A) Temporal generalization matrices for each comparison for the group analysis. The Y-axis shows the training time points, and the X-axis testing time points. The color scale indicates the classification accuracy, with 0.5 being chance level (all analyses are based on decoding two conditions against each other). Significance of classification accuracy was assessed by using cluster-based permutation tests. The black contour indicates significant clusters. Note that the upper right quadrant, highlighted by the black box, is the silence window. B) Temporal generalization matrix for a participant showing recurrent activity following the expected beats. When decoding the beat-based vs. the other two conditions (left- and rightmost plots), peaks in accuracy follow a clear oscillatory pattern, with a phase consistent with the beat-based sequence (600 ms between beats and decoding peaks).

Of note, the decoding results varied considerably between participants. In one participant in particular (Figure 9B), we observed a pattern in the temporal generalization matrix that was consistent with the hypothesized result for an oscillatory process (King & Dehaene, 2014). However, even though some other participants also showed recurrent activity, the exact times of most accurate decoding, and the exact period of the recurrent process, differed widely between participants, obscuring these effects in the grand averages. These individual differences may be caused by individual preferences for a level of regularity in the beat-based stimuli (Drake, Jones, & Baruch, 2000), with some people attending mostly to the beat level (1.67 Hz), but others possibly attending to subdivisions (3.33 Hz) or the level of the meter (0.83 Hz). Also, for different people, the optimal phase of delta oscillations (i.e., the phase that aligns with expected moments) may differ (Breska & Deouell, 2017; Henry & Obleser, 2012), possibly causing optimal decoding at different time points.

To circumvent the large task-related effects apparent in the decoding results, and the possible individual differences in phase and metrical level attended to, we ran an additional exploratory decoding analysis looking at decoding within instead of between conditions. Here, we defined time points in the silence as expected or unexpected based on the rhythmic sequences. For beat-based sequences, time points 600 and 1200 were expected (on the beat), while we used time points 780 and 885 as unexpected (offbeat). Likewise, the same time points take on a different meaning if preceded by the pattern-based sequences. Then, 780 and 1200 ms are considered expected (predictable based on the pattern), while both 600 and 885 ms are unexpected (unpredictable based on the pattern). For all conditions, we examined whether we could classify above chance whether a time window was centered on an expected moment or an unexpected moment in time. Logically, the beat-offbeat comparison should yield better decoding if preceded by the beat-based sequences, while the predicted-unpredicted comparison should be decoded better if preceded by the pattern-based sequences. To optimize for expectancy-related, but not task-related effects, for this analysis we used a region of interest consisting of fronto-central channels (i.e., auditory cortex).

The results of the within-condition decoding analysis are depicted in Figure 10. Only decoding of beat vs offbeat positions after the beat-based sequences yielded above-chance decoding ($t_{26} = 2.84$, $p = 0.009$). None of the other comparisons lead to decoding above chance (all ps > 0.18), but decoding of predicted vs unpredicted positions in the beat-based condition did yield below-chance decoding ($p = 0.026$). However, a subsequent ANOVA comparing decoding of beat against offbeat positions in all three conditions showed that decoding was not significantly better for the beat-based condition than the other conditions (effect of condition in the silence: $\chi^2 = 2.1$, $p = 0.35$). As can be seen in Figure 10, like for the other analyses, there were larger differences between participants, and decoding, while above chance, did not exceed 0.55 accuracy. As for the other EEG measures, the difference in decoding accuracy between the beat-based and random condition, and the patten-based and random condition did not depend on musical training (both ps > 0.21).

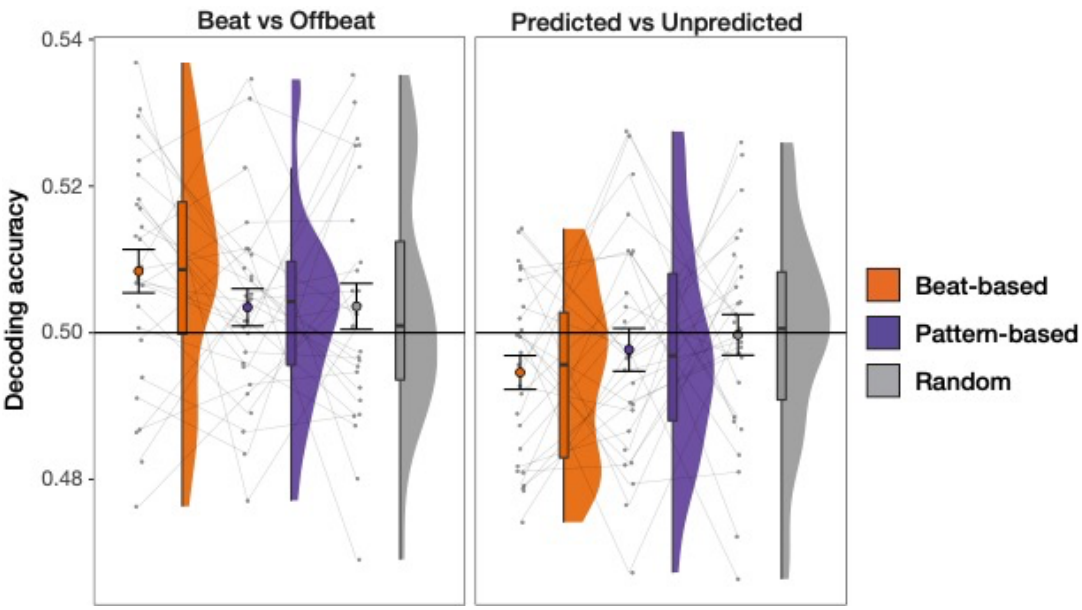

Figure 10. Above chance decoding for beat vs offbeat time points only after listening to beat-based sequences. For the within-condition decoding analysis, we used a classification algorithm to differentiate between times that were expected or unexpected based on the beat (on the beat: 600 and 1200 ms after the onset of the last tone; offbeat: 780 and 885 ms after the onset of the last tone) or pattern (predicted: 780 and 1200 ms after the onset of the last tone; unpredicted: 600 and 880 ms after the onset of the last tone). The graphs depict the average classification accuracy in a 100 ms time window centered on these expected and unexpected time points in the silence window. Only decoding of beat vs offbeat times in the beat-based sequences yielded above-chance decoding (accuracy > 0.5), but note that the analysis did not yield differences in decoding accuracy between conditions.

Discussion

In the current study, we aimed to identify and directly compare the neural mechanisms underlying temporal expectations based on a regular beat, as well as temporal expectations based on learning a predictable pattern, by examining the development of climbing activity and the persistence of neural entrainment after cessation of rhythmic input with either a regular beat, a predictable pattern of temporal intervals, or random timing. Instead of relying on isochrony to elicit beat-based expectations, which can also elicit different forms of memory-based expectations, we used varying rhythmic patterns to clearly separate beat-based and pattern-based expectations. Moreover, we assessed responses in a silent period after the auditory input ceased, side stepping the many possible confounds associated with acoustic differences between conditions (Capilla, Pazo-Alvarez, Darriba, Campo, & Gross, 2011; Haegens & Zion Golumbic, 2018; Novembre & Iannetti, 2018; Zoefel et al., 2018).

We found several indicators of separate mechanisms for beat-based and pattern-based expectations. First, behaviorally, we found that while the effects of beat-based expectations lasted for at least two beat cycles, the effects of pattern-based expectations subsided after the first expected moment in time. Second, while pattern-based expectations elicited a CNV component in the evoked potential, this negative deflection was absent for beat-based expectations. Third, we observed significantly more power at the beat frequency in the silence window when participants were previously listening to a beat-based sequence as compared to a pattern-based or random sequence. The observed longer lasting behavioral effects of beat-based compared to pattern-based expectations, and increased oscillatory power at the beat frequency following beat-based but not pattern-based sequences both point towards entrainment underlying beat-based, but not pattern-based expectations. The time-domain results showing a CNV for pattern-based, but not beat-based sequences suggests that the former may rely on climbing neuronal activity. Together, these findings provide support for the notion

that beat-based and pattern-based expectations rely on different neural mechanisms, as has previously been suggested for cue-based expectations (Breska & Deouell, 2017; Breska & Ivry, 2020), but our results do not support models that assume shared mechanisms, be it entrainment (Tichko & Large, 2019), or a general top-down mechanism for temporal expectations (Rimmele et al., 2018).

Behaviorally, for the beat-based condition, we observed a pattern clearly in line with entrainment models, with an inverted U-curve, an indication of entrainment (Bauer et al., 2015; Jones et al., 2002), present as late as 1200 ms after the final tone of each sequence in both experiments. Beat-based expectations thus affected the fitness ratings for at least two beat cycles after the end of the rhythmic pattern, as predicted by nonlinear oscillator models, that assume oscillations are self-sustaining (Large, 2008; Large & Palmer, 2002). Expectations in the beat-based condition not only lead to higher ratings for expected events, but also to lower ratings for unexpected events, when compared to the random and pattern-based conditions. Suppression of unexpected events may be metabolically beneficial (van Atteveldt et al., 2015), and as such, has been suggested to be a hallmark of entrainment and the associated "rhythmic" mode of processing (Schroeder & Lakatos, 2009a; Zoefel & Vanrullen, 2017). Indeed, suppression off the beat has even been proposed to be a better indication of beat-based expectations than facilitation on the beat (Bouwer et al., 2020; Breska & Deouell, 2017), in line with the current results, where the effects of beat-based expectations at unexpected time points exceeded those at expected time points. Pattern-based expectations similarly affected fitness ratings, with enhanced ratings at the first expected time point, and lower ratings at the first unexpected time point, showing that participants did form expectations based on the predictable rhythmic pattern. However, importantly, for the pattern-based condition, the results are not suggestive of sustained oscillations, as the effects of expectations subsided after the first expected moment.

Crucial to entrainment models of beat-based expectations (Haegens & Zion Golumbic, 2018; Henry & Herrmann, 2014; Large, 2008; Large & Jones, 1999; Obleser & Kayser, 2019), we found that power at the beat frequency (1.67 Hz) and its harmonic (3.33 Hz) in the EEG signal during the silence window was larger following beat-based than pattern-based or random sequences. Such enhanced power was not found for a frequency inherent to the pattern-based sequence (2.22 Hz). Methodologically, measuring phase locking during rhythmic stimulation can lead to confounding contributions from tone-evoked responses (Capilla et al., 2011; Novembre & Iannetti, 2018; Zoefel et al., 2018). Ongoing oscillations in silence, after sensory input has stopped, are therefore regarded as strong evidence for entrainment (Breska & Deouell, 2017; Haegens & Zion Golumbic, 2018; Obleser & Kayser, 2019; van Bree et al., 2021; Zoefel et al., 2018). As such, our observation of enhanced power at the beat frequency during silence provides important novel support for the notion that entrainment of low-frequency neural oscillations underlies beat-based perception, and our design, using non-isochronous rhythms, allows us to separate beat-based aspects from other structure present in natural rhythm (Bouwer, Nityananda, Rouse, & ten Cate, 2021).

Interestingly, a recent paper that assessed ongoing oscillations following auditory rhythmic input did not find any evidence for persistent entrainment at the frequency of the rhythm (Pesnot Lerousseau et al., 2021). Two differences between this study and our work may provide directions for future work. First, as mentioned before, isochronous rhythm allows for several different ways of forming temporal expectations, including not only beat-based, but also cue-based and pattern-based expectations. In the current study, this was the motivation to design stimuli that allowed for differentiating between these types of expectations. However, an additional concern could be that when presented with stimuli that do not require beat-based expectations to perform the task of tracking the temporal structure, the brain may not engage in forming such expectations. Thus, in addition to variation between individuals (Assaneo et al.,

2019), variation in the input signal may also determine which mechanism is used to form temporal expectations. Finally, the presence of persistent entrainment may also depend on task demands (Shalev, Nobre, & van Ede, 2019). In our study, the auditory sequences were task relevant, and the task itself was rhythm-related. In the study by Pesnot Lerousseau et al., participants listened to rhythms passively. The influence of contextual factors on entrainment, be it person, stimulus, or task, are an important topic for future research.

The absence of power at 2.22 Hz following the pattern-based sequences suggests that entrainment does not underlie expectations based on learning a pattern, contrary to a recently proposed oscillator model that can capture aspects of pattern-based expectations (Tichko & Large, 2019). In the time-domain, we found a CNV in the silence window following the pattern-based and random sequences, but not the beat-based sequences. The observed differences in the CNV suggest that rather than one entrainment mechanism for both regular and irregular rhythms, an alternative mechanism based on climbing neuronal activity specifically supports formation of pattern-based expectations. Contrary to this notion, previous studies did find a CNV peaking at expected times not only for cue-based, but also for beat-based expectations (Breska & Deouell, 2012, 2017; Breska & Ivry, 2020; Praamstra et al., 2006). Importantly, the latter studies used isochronous rhythmic sequences to probe beat-based expectations, and therefore, memory-based expectations, be it pattern-based or cue-based, could also have been formed in response to these sequences (Bouwer et al., 2020, 2021; Bouwer, Werner, Knetemann, & Honing, 2016; Breska & Ivry, 2016; Keele et al., 1989). Here, using a beat-based sequence that did not allow for expectations based on simply learning transitional probabilities, we did not observe a CNV, suggesting that a CNV is specific to pattern-based, and likely cue-based (Mento, 2013, 2017), temporal expectations.

While previous studies showed that the peak of the CNV coincided with an expected time (Breska & Deouell, 2017; Mento, 2017; Praamstra et al., 2006), here, the peak latency of

the CNV was earlier, peaking around 400 ms for the pattern-based condition, while the first expected time point in the silence was at 780 ms. This can be explained by assuming that the time course of the CNV indexes temporal expectations in a probabilistic way. While the pattern was fully predictable in the pattern-based sequences, it contained several temporal intervals with durations between 150 and 780 ms. The peak at 400 ms may have indexed the average interval presented in the sequence (~360 ms). This explanation is also supported by the presence of a CNV with a similar time course for the random condition. Like in the pattern-based condition, in the random condition, participants could not use a beat-based strategy to perform the task. Thus, they may have attempted to predict the timing of an upcoming sound based on the distribution of the absolute intervals, which while random in terms of transitional probabilities, was on average identical to the pattern-based condition. Interestingly, the average interval in the beat-based condition was also 360 ms, but here, the CNV was absent, suggesting that in the presence of a possible beat-based strategy, the brain may operate in a rhythmic mode of processing (Rimmele et al., 2018; Schroeder & Lakatos, 2009a). A second CNV-like peak, around 950 ms, was only present for the pattern-based condition, but not the random condition. Speculatively, this second peak may be indicative of expectations based on, not just the distribution of possible intervals, but also their transitional probabilities, which would predict a sound at 780 ms. Probabilistic models incorporating statistical regularities in inter-onset intervals at different levels have indeed been used to explain aspects of temporal processing (Cannon, 2021; Elliott, Wing, & Welchman, 2014; van der Weij, Pearce, & Honing, 2017). In future work, linking such models directly to neural markers of pattern-based expectations may provide more insight in the mechanisms underlying pattern-based expectations.

Some methodological caveats remain, especially when considering the result from the frequency tagging analysis. The Fourier transform assumes stationarity in the oscillating signal, while entrainment models propose a dampening factor to account for decreasing oscillatory

power over time (Large, Herrera, & Velasco, 2015). To assess power at specific frequencies in a time-resolved way, wavelet convolution is often used as an alternative. But, in the current study, differentiating between the specific frequencies of the beat and the pattern would require wavelet parameters that would result in a temporal resolution too low to disentangle activity during and before the silence (i.e., many wavelet cycles would be needed). Recently, a promising alternative to assess oscillatory activity in the time domain was proposed in the form of cycle-by-cycle analysis (Cole & Voytek, 2019). However, this approach requires filtering the signal at the specific frequency of interest, again posing problems for disentangling low frequency oscillations due to the beat, the pattern, and ongoing ERPs. Assessing the time course of low frequency oscillations thus remains a challenge for future research.

Multi scale entropy and multivariate pattern analysis may provide alternative ways to examine neural entrainment with high temporal precision. At the group level, entropy was higher following the pattern-based than random sequences. This could be explained by assuming that the brain uses a vigilance mode to track the pattern-based regularities, which at the neural level, translates in more irregular patterns of activity. Attention and arousal, which are both associated with temporal expectations (Schroeder & Lakatos, 2009b), have indeed been linked to neural variability as well (Waschke et al., 2021). However, the group level results can be explained at least to some extent by overall differences in signal variability (e.g., the signal variance), so this hypothesis remains to be confirmed in future research.

Using multivariate decoding, we were able to decode above chance from the silence window which type of sequence participants had just heard. However, this decoding seemed to primarily reflect general task-related activity. The best decoding was observed in the second half of the silence window, where probes could be presented. Also, in the group-average decoding results, above chance decoding was limited to training and testing on the same time points. In other words, we did not observe recurrent activity, as reflected in stronger decoding

accuracy at expected time points. Yet, we did show a proof of concept for our approach in at least one participant, who showed a clear oscillatory pattern in decoding accuracy when decoding the beat-based against the other two conditions. This pattern of activity is in line with the strength or sharpness of the neural representation of tones varying over time as a function of temporal expectations (Auksztulewicz et al., 2019, 2018).

At the group level, the behavioral results, along with the results obtained using frequency tagging and ERPs provide evidence for differential processing of the beat and the pattern in rhythm, while the results from the decoding and entropy analyses are less clear. One reason for this is heterogeneity between individuals. This heterogeneity may be present at several levels. First, even if we assume that entrainment is present in most people, there may be qualitative differences in the implementation. For example, people differ in their preferences for a level of regularity in beat-based perception (Drake et al., 2000). In the current experiment, some individuals may have been attending mostly to the beat level (1.67 Hz), and others to subdivisions (3.33 Hz) or the level of the meter (0.83 Hz). Also, the optimal phase of delta oscillations (e.g., the phase that aligns with expected moments) may differ across individuals (Breska & Deouell, 2017; Henry & Obleser, 2012; Sun et al., 2021), possibly causing optimal decoding at different time points. A second layer of individual differences may be introduced by differences in the propensity to rely on beat-based processing. While it is often assumed that most people automatically form beat-based expectations (Honing, 2012), recent evidence showed phase locking to speech in only about half of the population (Assaneo et al., 2019). Indeed, in our study, only about two-thirds of the participants behaviorally showed evidence for beat-based expectations in the second half of the silence window. Also, we only observed enhanced power at the beat frequency following beat-based sequences in about half of the participants, in line with individual differences in reliance on beat-based perception.

Finally, individual differences may be due to differences in musical expertise. The behavioral effects of beat-based and pattern-based expectations were associated with musical training. This is consistent with a large body of research looking at the association between musical training and beat-based perception (Bouwer et al., 2018, 2016; Cameron & Grahn, 2014; Matthews et al., 2016; Vuust et al., 2005). The finding that pattern-based expectations were also associated with training is in line with a previous study showing that percussionists outperformed non-musicians on a tapping task for both rhythms with and without a regular beat (Cameron & Grahn, 2014), and is not surprising given that musical training affects many aspects of auditory perception (Moreno & Bidelman, 2014). Some previous studies have failed to show differences between musicians and non-musicians, however (Bouwer, Van Zuijen, & Honing, 2014; Geiser et al., 2009; Grahn & Brett, 2007). As argued previously (Bouwer et al., 2018), these differences may be due to differences in task design. In the current study, the task required explicit ratings of how well probes fitted with the preceding rhythm. Performance on such an explicit task may be particularly improved by musical training, as it may provide musically trained participants with additional strategies to perform the task. The use of implicit timing tasks may be a better probe of innate differences in timing abilities, which need not necessarily be related to musical training (Law & Zentner, 2012).

We did not find an association between musical expertise and the neural results. This may partly be due to neural activity being more reflective of innate abilities, which do not per se correlate with musical training, as argued above. However, we must also note that the analyses concerning the effects of musical training were exploratory, and with the current sample size, we may have simply lacked power to detect associations between neural data and musical expertise. Also, our sample scored relatively low on the musical training scale, with a mean of 20.5 on the GMSI subscale, which is comparable to the 32nd percentile of the norm scores, which have a mean of 26.5 (Müllensiefen, Gingras, Musil, & Stewart, 2014). This may

have made it hard to find the effects of musical training. Indeed, even with larger N, brain-behavior relationships may be hard to detect given the noise added by anatomical and physiological differences between people (e.g., the EEG signal may vary wildly due to a different projection from sensory regions onto the scalp, which has nothing to do with behavior). Undoubtedly, given the heterogeneity we and others (Assaneo et al., 2019; Bauer et al., 2015; Sun et al., 2021) have observed in tasks probing temporal expectations, understanding individual differences is an important direction for future research, with significant implications for applications of musical rhythm, such as in motor rehabilitation (Dalla Bella, Dotov, Bardy, & Cochen De Cock, 2018).

Conclusion

In summary, we have shown that beat-based and pattern-based expectations can be differentiated in terms of their behavioral and neurophysiological effects once sensory input has ceased. These findings provide novel, more conclusive evidence for the notion that different mechanisms implement temporal expectations based on periodic and aperiodic input streams, with the former based on entrainment of low frequency neural oscillations, and the latter on climbing neural activity indexing a memorized interval.

Materials and methods

Participants

Thirty-two participants (18 women), aged between 18 and 44 years old (M = 24, SD = 5.6) took part in the behavioral Experiment 1, and 32 participants (26 women), aged between 19 and 28 years old (M = 23, SD = 2.5) took part in the EEG experiment (Experiment 2), in exchange for course credit or monetary compensation. Due to technical problems, the EEG data from five participants was not recorded correctly, hence we report the results for 27 participants (21 women, between 19 and 28 years old, M = 23, SD = 2.4). For the behavioral experiments, we used mixed-effects models, which need both the number of participants and the number of items to be taken into account to assess power (Brysbaert & Stevens, 2018). In two similar experiments in which ratings in response to rhythms of varying complexity were analyzed, small-sized effects were replicated with a total number of around 275 responses per condition (Bouwer et al., 2018). For our new experimental paradigm, we here included a multiple of this amount of trials (960 responses per condition in Experiment 1 and, after loss of data was accounted for, 486 responses in Experiment 2). Previous EEG experiments examining persistent entrainment in the auditory domain used sample sizes ranging from fifteen (Pesnot Lerousseau et al., 2021) to twenty-one (van Bree et al., 2021), similar to the sample size used in a study looking at different types of temporal expectations (twenty-one, Breska & Deouell, 2017). To obtain robust power, we here tested thirty-two participants, which, even when loss of data is accounted for, thus exceeded typical sample sizes as used previously. None of the participants reported a history of hearing or neurological problems, and all provided written informed consent prior to the onset of the study. The experiment was approved by the Ethics Review Board of the Faculty of Social and Behavioral Sciences of the University of Amsterdam.

Stimuli

We used rhythmic patterns marked by woodblock sounds of 60 ms length, generated in GarageBand (Apple Inc.), as previously used in (Bouwer et al., 2020) to elicit beat-based and pattern-based expectations (Figure 1). Each pattern was 1800 ms long and consisted of five temporal intervals. The number of tones was chosen to allow for learning of the rhythmic surface structure in the pattern-based condition, while the length of the pattern was such that the formation of beat-based expectations with a period of the entire pattern would be unlikely in the pattern-based sequences (Bouwer et al., 2020; Honing & Bouwer, 2019; London, 2012; Schultz, Stevens, Keller, & Tillmann, 2013). Sequences were constructed by concatenating four patterns and a final tone, for a total sequence length of 7260 ms (four patterns of 1800 ms, plus 60 ms for the final tone).

In the twelve beat-based patterns used (Figure 1), temporal intervals were related to each other with integer-ratio durations. The shortest interval had a length of 150 ms, with the relation between the five intervals used of 1:2:2:3:4 (i.e., 150, 300, 300, 450, and 600 ms). The sounds were grouped such that a perceptually accented sound (Povel & Okkerman, 1981) occurred every 600 ms (every unit length 4), giving rise to a beat at 100 beats per minute, or 1.67 Hz, within the range of preferred tempo for humans (London, 2012). All beat-based patterns were strictly metric, with the beat always marked by a sound (Grahn & Brett, 2007). Sequences of beat-based patterns were constructed from four semi-randomly chosen patterns, with the restriction that the last pattern of the sequences was always pattern 1 or 2 (see Figure 1). This way, the final 600 ms preceding the silence epoch was equated in terms of the acoustic context, to make the bleed of auditory ERPs into the silence as similar between conditions as possible. Note that in beat-based sequences, a sound could be expected every 600 ms based on the beat, but the surface structure of the rhythmic pattern was unpredictable, due to the random concatenation of patterns.

To create patterns that did not allow for beat-based expectations ("aperiodic" patterns, see Figure 1), the ratios by which the temporal intervals were related were changed to be non-integer (1:1.4:1.4:3:5.2, or 150, 210, 450, and 780 ms respectively). In these patterns, no marked beat was present at unit length four, nor at any other subdivision of the 1800 ms pattern (Bouwer et al., 2020), while the patterns were matched to their periodic counterparts in terms of overall length, event density, number of sounds, and grouping structure.

From the aperiodic patterns, two types of sequences were created: pattern-based and random sequences. To create sequences allowing for pattern-based expectations, we concatenated four identical patterns. To be able to use the data with an EEG-based decoding analysis (Experiment 2), we needed the timing of expectations in the silence to be identical for each sequence, hence we restricted the pattern-based sequences to only pattern 1. The use of a single pattern was not only necessary for decoding, but also optimized the experiment for pattern-based expectations, since participants only had to memorize one pattern, allowing them to easily form expectations, even if the single sequences were only four patterns long.

For the random sequences, four semi-randomly chosen aperiodic patterns were concatenated. Like for the beat-based sequences, the final pattern was always pattern 1 or 2, equating the final 600 ms of the sequences in terms of acoustics. In the random sequences, the timing of sounds could not be predicted based on the surface structure of the rhythm, nor on the basis of an underlying beat.

A spectral analysis of the stimuli in the range in which a beat can normally be perceived (the delta-range, 0.5-4 Hz) confirmed that in the beat-based sequences, a peak was present at the beat frequency of 1.67 Hz as well as at 3.33 Hz (see Figure 5). The 3.33 Hz peak is a harmonic of the beat frequency, but also the frequency at which participants may perceive subdivisions of the beat (e.g., an extra layer of perceived metrical regularity with a period of 300 ms). In the pattern-based and random sequences, the peak at 1.67 Hz was not just smaller

in an absolute sense, but also in a relative sense: in the pattern-based and, to a lesser extent, random condition, peaks at 2.22 Hz and 3.89 Hz were most prominent in the sound signal.

In Experiment 1, on each trial, a probe tone was presented at 600, 780, 885, 990, 1200, or 1485 ms after the onset of the last tone of the sequence (see Figure 1C). These positions were carefully chosen to represent times at which a tone could be expected based on the beat (600, 1200 ms), based on memory for the pattern (780, 990, 1200 ms), or neither (885, 1485 ms). Note that the latter two probe tones that were unexpected based on the beat (780, 885, 990, and 1485 ms) did not fall on subdivisions of the beat.

Experiment 2 contained both trials in which a probe tone was presented (using only the last three probe positions), and trials in which a 7260 ms sequence was followed by a silence period without a probe tone. The latter were used for EEG analyses, uncontaminated by a probe presentation. The silent period lasted for 2100 ms after the onset of the last sound, providing 300 ms immediately following the final sound to allow for ERPs to mostly return to baseline, and 1800 ms (three full beat cycles) of silence for the analysis. After the silence, the onset of the next sequence was jittered between 25 and 75 ms to prevent carryover of the beat from previous sequences (e.g., the next trial started between 325 and 375 ms after the last beat in the silence, which is not on the beat, nor at a subdivision of the beat). On trials that contained a probe tone, we chose to only use the last three probe positions, to 1) limit the time for the EEG experiment to prevent fatigue, 2) still provide participants with the incentive to form expectations well into the silence period, as a probe tone could appear as late as 1485 ms into the trial, and 3) obtain some measure of whether participants formed expectations, by including positions that were expected based on the beat and the pattern (1200 ms), based on the pattern only (990 ms) or neither (1485 ms).

Procedure

Participants were tested individually in a dedicated lab at the University of Amsterdam. Upon arrival, participants were provided with information about the experiment, provided informed consent, and were allowed to practice the task. On probe tone trials (all trials in Experiment 1, 25% of the trials in Experiment 2), participants were asked to judge on a four-point scale ("very poorly", "poorly", "well", "very well") how well the probe tone fitted the preceding sequence, similar to previous studies investigating the perception of musical meter (Manning, Harris, & Schutz, 2017; Manning & Schutz, 2013; Palmer & Krumhansl, 1990). Participants could respond with four buttons on the armrest of their chair, two on each side. The order of the answer options on the screen in front of the participants was randomized on each trial, to avoid any artefacts of motor preparation in participants that anticipated which answer they would provide. There was no time limit for responses and the next trial started after a response was made.

In Experiment 1, each participant was presented with 18 blocks of 30 trials, amounting to 540 trials in total, 180 trials per condition, and 30 trials per probe position for each condition. In Experiment 2, participants were presented with 18 blocks of 36 trials, for a total of 648 trials, 216 per condition. In Experiment 2, for each condition, 162 trials were silence trials, and did not contain a probe tone. Fifty-four trials contained a probe tone. In both experiments, in each block, only one type of sequence (beat-based, pattern-based, or random) could appear to optimize for the formation of expectations. Blocks were semi-randomized, with each type appearing once in a set of three blocks. In each block of Experiment 2, the number of probe trials was varied between 3 and 11, for an average of 25% probe trials, and 75% silent trials.

Sounds were presented through a single speaker positioned in front of the participant using Presentation® software (version 14.9, www.neurobs.com). After completion of the experiment, participants performed the Beat Alignment Task (Iversen & Patel, 2008;

Müllensiefen et al., 2014) to assess their beat perception abilities, and completed the musical training subscale from the Goldsmith Musical Sophistication Index (GMSI) questionnaire to assess their musical training (Müllensiefen et al., 2014). In total, a behavioral session lasted two hours, and the EEG session lasted between 3.5 and 4 hours.

Behavioral analysis

A total of 17280 responses was included in the analysis of Experiment 1 (32 participants, 3 conditions, 6 probe positions, 30 responses each), and 4374 in the analysis of Experiment 2 (27 participants, 3 conditions, 3 probe positions, 18 responses each). To account for the ordinal nature of the Likert-scale responses (Bouwer et al., 2018; Carifio & Perla, 2008; Jamieson, 2004), we used a mixed ordinal regression model. With this model, the ordinal responses are normalized, to correct for potential unequal distances between rating points. The results can subsequently be interpreted similar to the results from a normal mixed model regression. Two independent variables and their interaction were included in the model as fixed factors: Condition (beat-based, pattern-based, or random), and Probe Position (600, 780, 885, 990, 1200, or 1485 ms). Additionally, the score on the GMSI musical training questionnaire was included as a continuous variable (Musical Training, as well as its interactions with the two fixed factors. We used a random intercept for each subject to account for between-subject variation.

The initial model showed a significant effect of Probe Position in the random condition, most likely due to recency effects. To assess the effect of Probe Position in the beat-based and pattern-based conditions while accounting for recency effects, for each participant we subtracted the mean response in the random condition at each position from the responses in the beat-based and pattern-based condition. Subsequently, we submitted the random-baseline-corrected ratings to a second ordinal regression model, with only two levels for the factor Condition (beat-based and pattern-based) and without the random intercept for each participant

(as the baseline correction already corrected for between-subject variability). For both the original model, and the baseline corrected model, significant interactions were followed up by tests of simple effects, corrected for multiple comparisons using Tukey's HSD.

The statistical analysis was conducted in R (R Development Core Team, 2008). The ordinal mixed model was implemented using the clmm() function from the ordinal package (Christensen, 2019). Subsequently, we used the Anova() function from the car package (Fox & Weisberg, 2019) to look at omnibus effects for the main factors of interest, and the emmeans package (Lenth, 2019) to assess simple effects and compare slopes between conditions.

EEG recording

EEG was recorded at 1 kHz using a 64-channel Biosemi Active-Two acquisition system (Biosemi, Amsterdam, The Netherlands), with a standard 10/20 configuration and additional electrodes for EOG channels, on the nose, on both mastoids, and on both earlobes.

EEG analysis

Preprocessing was performed in MATLAB, version 2015a (Mathworks) and EEGLAB, version 14.1.1 (Delorme & Makeig, 2004). Data were offline down-sampled to 256 Hz, re-referenced to linked mastoids, bad channels were removed, and eye-blinks were removed using independent component analysis. Subsequently, bad channels were replaced by values interpolated from the surrounding channels.

ERPs. For the ERP analysis, the continuous data were filtered using 0.1 Hz high-pass and 40 Hz low-pass finite impulse response filters. Epochs were extracted from the data from -1800 till 2100 ms relative to the onset of the last sound. Epochs with a voltage change of more than 150 microvolts in a 200 ms sliding window were rejected from further analysis. For each participant and condition, epochs were averaged to obtain the ERPs, and ERPs were averaged over participants to obtain grand average waveforms for plotting. All waveforms were initially baseline corrected using the average voltage of a 50 ms window preceding the onset of the last

sound of the sequence. This baseline can be regarded as preceding the "cue" (the last event before the onset of the expectation). Such a baseline is customary in CNV analyses. However, visual inspection suggested that this baseline was biased, as baseline correction resulted in an overall shift of the waveform amplitude relative to each other, as also reflected in a significant cluster when comparing beat-based and pattern-based conditions that spanned the entire analysis epoch (see Figure S1). This was likely caused by the rapid succession of sounds preceding the onset of the silence, which made it impossible to find a clean, unbiased baseline. Therefore, we repeated the analysis without baseline correction, to confirm that the results were not caused by a noisy baseline.

Three cluster-based tests were used to compare all three conditions against each other (i.e., beat-based vs. random; pattern-based vs. random; beat-based vs. pattern-based). Clusters were formed based on adjacent time-electrode samples that were significantly different between conditions, based on dependent samples t-tests and a threshold of $p < 0.05$. Clusters were subsequently evaluated statistically using permutation tests with 2000 permutations of the data. We report corrected *p*-values to account for two-sided testing (multiplied by a factor of two).

Frequency tagging. To obtain the spectrum of the EEG signal in the silence, we used the raw, unfiltered data. Epochs were extracted from the continuous data both from -1800 till 0 ms relative to the onset of the last sound (control window, see Figure 1), and from 300 till 2100 ms relative to the onset of the last sound (silence window, see Figure 1), the latter starting at 300 ms to avoid contamination from the final ERPs. Both windows thus had equal length, both spanning three full cycles of the beat. Epochs with an amplitude change of 250 microvolts or more in a sliding 200 ms window were rejected from further analysis. All epochs were baseline corrected using the mean of the entire epoch. Subsequently, epochs were averaged for each condition separately to obtain the evoked signal, phase locked to the onset of the final sound,

and similar to previous studies using frequency tagging to look at beat-based perception (Lenc et al., 2021; Nozaradan et al., 2011, 2012).

For each participant and condition separately, the average waveforms were transformed into the frequency domain using an FFT, with the data zero-padded to 4608 samples (NFFT) to obtain a better frequency resolution (0.056 Hz), and importantly, be able to extract data at exactly the frequencies of interest. Note that the zero-padding can only improve the frequency resolution, but not the frequency precision, which by definition with the 1800 ms epochs is limited to 0.56 Hz. While the design of the experiment simply does not allow for a better resolution, the 0.56 Hz does allow us to differentiate between the frequencies of interest, which are 0.56 Hz or more apart. The obtained power values at each frequency were normalized to account for the 1/f distribution of noise, and to account for bleeding into neighboring frequency bins (Nozaradan et al., 2011, 2012). For each frequency, we averaged over 5 bins centered on the frequency of interest (1.56 – 1.78 Hz for the beat frequency, 2.11 – 2.33 Hz for the pattern-based frequency, and 3.22 – 3.44 Hz for the beat subdivisions). Subsequently, we subtracted the average of neighboring bins four to six on either side (1.33 – 1.44 Hz and 1.89 – 2.00 Hz for the beat frequency, 1.89 – 2.00 Hz and 2.44 – 2.56 Hz for the pattern-based frequency, and 3.00 – 3.11 Hz and 3.56 – 3.67 Hz for the beat subdivisions).

To statistically test differences between conditions in the evoked power at the frequencies of interest, we used cluster-based permutation tests. First, this avoided bias by selecting only a subset of electrodes, as we used all scalp electrodes, as was done in previous research (Lenc, Keller, Varlet, & Nozaradan, 2018; Nozaradan et al., 2011; Tal et al., 2017). Second, the permutation tests accounted for the non-normal distribution of the data. Like for the ERPs, we ran t-tests comparing the normalized data for all conditions. For each frequency of interest, clusters were formed based on adjacent electrodes with significant differences between conditions based on dependent samples t-tests and a threshold of $p < 0.05$. Clusters

were evaluated using permutation tests with 2000 permutations of the data. We report corrected *p*-values to account for two-sided testing (multiplied by a factor of two).

Multiscale entropy (MSE). We computed MSE on high-pass filtered data (0.5 Hz). Epochs were extracted from the continuous data from -1800 till 0 and 300 till 2100 ms relative to the onset of the last sound. Epochs with an amplitude change of 250 microvolts or more in a sliding 200 ms window were rejected from further analysis. All epochs were baseline corrected using the mean of the entire epoch.

To compute MSE, we used the mMSE toolbox, a plugin to the Fieldtrip toolbox (Kloosterman et al., 2020). To compute entropy, epochs were concatenated. Sample entropy was then calculated by taking the following steps:

1. In each time series, a template is selected, consisting of m samples. The calculation of entropy is an iterative process, using each sample in the time series as a starting point for the template once.
2. Throughout the time series, the algorithm searches for patterns that match the template pattern. A section of samples is considered a match if it resembles the template pattern enough to fall within a set boundary, which is defined as r x SD (the similarity bound).
3. The number of pattern matches is counted.
4. Subsequently, the same procedure is followed for patterns of $m + 1$ samples long.
5. For the total counts of pattern matches throughout the time series, sample entropy is then calculated as the logarithm of the ratio between pattern matches of length m and pattern matches of length $m + 1$.

Thus, sample entropy reflects the proportion of patterns in the time series that stays similar when an extra sample is added to the pattern. Here, we used $m = 2$ and $r = 0.5$, as was done previously for EEG data (Kloosterman et al., 2020).

Sample entropy is then repeated for multiple timescales, to account for contributions of both low and high frequency neural activity. The time series is coarsened step by step, by taking the average of a group of adjacent samples with step-wise increasing group size. This means that long, or coarse, timescales are equivalent to low frequency activity. For example, at a sampling rate of 256 Hz, a timescale of 4 (e.g., averaged over four adjacent samples, or 15.6 ms) is roughly the equivalent of looking at activity at 64 Hz, while a timescale of 153 at that sample rate corresponds to averaging over 598 ms, or the equivalent of activity at roughly 1.67 Hz. Here, we used twenty timescales ranging from 4 till 153, with 153 being the maximum timescale given the length of an epoch of 1800 ms (e.g., one epoch equals 460 samples, but since entropy is calculated on patterns of 2 and 3 samples, the maximum coarsening to retain the possibility of having a 3-sample pattern is by averaging over 153 samples). Similarity bounds were recomputed for each time scale (Kloosterman et al., 2020; Kosciessa et al., 2020). To control for the contribution of delta, we repeated the analysis on high-pass filtered data (5 Hz).

For the frequency tagging analysis, we used cluster-based permutation tests to assess statistical significance. For each comparison between conditions, we used paired t-tests comparing each electrode-timescale combination (note that the above computation of entropy yields one value per condition for each electrode and timescale). Clusters were formed based on adjacent electrodes and timescales with significant differences between conditions at a threshold of $p < 0.05$. Clusters were evaluated using permutation tests with 2000 permutations of the data and we report corrected *p*-values to account for two-sided testing (multiplied by a factor of two).

Multivariate decoding. The decoding was conducted on data that was preprocessed in a similar way as for the MSE analysis, but with epochs extending from -1800 to 2100 ms relative to the onset of the last sound. Since the decoding is done in a time resolved way (e.g.,

sample by sample), there is no need to leave out the response to the ERPs in the analysis. Additionally, the data were resampled to 32 Hz to increase signal to noise, using shape-preserving piecewise cubic interpolation, as implemented in the Fieldtrip toolbox (Oostenveld et al., 2011). Using the ADAM toolbox (Fahrenfort et al., 2018), we applied a classification algorithm to the preprocessed data for each participant. Using all electrodes, each dataset was split into 10 equally sized subsets, for 10-fold cross-validation of the decoding. For each subset, a linear discriminant classifier trained on the remaining 9 subsets was tested. Each condition was decoded against both other conditions (e.g., beat-based vs. pattern-based, beat-based vs. random, and pattern-based vs. random). Subsequently, we examined whether we could observe a pattern of recurrent activity (King & Dehaene, 2014). Classification accuracies for each comparison of two conditions for the silence window (300-2100 ms after the onset of the last sound) were submitted to cluster-based permutation tests to assess whether they exceeded the chance level of 0.5. T-tests with a threshold of 0.05 were performed for each training-testing time point combination, comparing the accuracy to 0.5. Subsequently, observed cluster-sizes were compared to a null distribution of cluster-sizes obtained by 2000 random permutations of the data to assess significance. We report corrected *p*-values to account for two-sided testing (multiplied by a factor of two).

We ran an additional decoding analysis in which we decoded expected and unexpected positions against each other within each condition. For this analysis, we only included frontocentral electrodes (C1, C2, C3, C4, Cz, FC1, FC2, FC3, FC4, FCz, F1, F2, F3, F4, Fz), which were the same electrodes that showed the largest P1 responses to the initial sounds of the sequences, indicative of representing the auditory cortex. To equate the choice of expected and unexpected time points as much as possible in terms of number of time points included and distance between time points, we used times that were also used in the probe tone experiment: 600, 780, 885, and 1200 ms. Of these, we constructed two sets of time points: 600 and 1200 ms

were combined to represent on-beat times (expected), and were decoded against 780 and 885 ms, representing off-beat times (unexpected). In a second decoding analysis, 780 and 1200 ms were combined to represent predicted times in terms of the pattern, and were decoded against 600 and 885 ms as unpredicted times. For each time point, an epoch of 100 ms centered on the time point of interest was extracted and used for the decoding analysis. As for the first decoding analysis, the data were resampled to improve signal to noise, here to 128 Hz to retain enough data points in the 100 ms window for analysis. After the decoding, we extracted the average classification accuracy for the 100 ms time window to test significance against chance level, using t-tests against 0.5, and between conditions in the silence, using a repeated measures ANOVA, with condition as an independent factor, a random intercept for participant, and decoding accuracy as the dependent variable.

Musical expertise. In an exploratory analysis, we assessed the relationship between the EEG data and musical expertise. We divided participants in two groups, based on a median split on the GMSI questionnaire scores. For each EEG marker, we extracted the relevant values indexing beat-based and pattern-based expectations, and we performed a t-test, comparing the two musical expertise groups. First, for the CNV, we extracted average amplitudes for each condition in the two windows in which the cluster-based analysis yielded a significant cluster (300-450 ms and 925-1005 ms, see also Figure 4), from central electrodes (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2). To index the effects of temporal expectations, we then followed the subtraction logic we also used in the analysis of the behavioral results: the effects of beat-based expectations were quantified as the amplitude in the beat-based condition minus that in the random condition, and the effects of pattern-based expectations were quantified as the amplitude in the pattern-based condition minus that in the random condition. Likewise, for the frequency tagging analysis, we extracted the difference in power at 1.67 Hz between the beat-based and random conditions, and the difference in power at 2.22 Hz between the pattern-based

and random conditions. Here, we used only electrodes that contributed to the significant cluster we had found in the silence for the frequency tagging analysis. For the MSE, we similarly extracted differences between entropy in the beat-based and pattern-based conditions when compared to the random condition, using only electrodes and timescales that contributed to the significant cluster found in the silence for the MSE analysis. For decoding, we extracted the classification accuracy when decoding expected against unexpected times for each participant from the beat-based and pattern-based conditions, and again contrasted those with the random condition.

Code and data availability. All datafiles, and code used for data acquisition, data analysis, and figure creation, are available through https://osf.io/uwny8/.

Acknowledgements

We would like to thank Babak Chawoush and Charlotte van Roijen for their assistance with the data collection, and fruitful discussions during the project.

References

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2014). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, *25*(9), 3077–3085. https://doi.org/10.1093/cercor/bhu103

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, *22*(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Auksztulewicz, R., Myers, N. E., Schnupp, J. W., & Nobre, A. C. (2019). Rhythmic temporal expectation boosts neural activity by increasing neural gain. *The Journal of Neuroscience*, *39*(49), 9806–9817. https://doi.org/10.1523/JNEUROSCI.0925-19.2019

Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., … Melloni, L. (2018). Not all predictions are equal: "What" and "when" predictions modulate activity in auditory cortex through different mechanisms. *The Journal of Neuroscience*, *38*(40), 8680–8693. https://doi.org/10.1523/JNEUROSCI.0369-18.2018

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

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

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

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

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

Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: An ERP study. *PLoS ONE*, *9*(5), e97467. https://doi.org/10.1371/journal.pone.0097467

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

Breska, A., & Deouell, L. Y. (2012). Automatic bias of temporal expectations following temporally regular input independently of high-level temporal expectation. *Journal of Cognitive Neuroscience*, *26*(7), 1555–1571. https://doi.org/10.1162/jocn

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

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

Breska, A., & Ivry, R. B. (2018). Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proceedings of the National Academy of Sciences*, *115*(48), 12283–12288. https://doi.org/10.1073/pnas.1810596115

Breska, A., & Ivry, R. B. (2020). Context-specific control over the neural dynamics of temporal attention by the human cerebellum. *Science Advances*, *6*(49), 1–16. https://doi.org/10.1126/sciadv.abb1141

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

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

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

Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., & Gross, J. (2011). Steady-State Visual Evoked Potentials Can Be Explained by Temporal Superposition of Transient Event-Related Responses. *PLoS ONE*, *6*, e14543. https://doi.org/10.1371/journal.pone.0014543

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*(12), 1150–1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x

Christensen, R. H. B. (2019). ordinal - Regression models for Ordinal Data. R package version 2019.4-25. Retrieved from http://www.cran.r-project.org/package=ordinal/

Cole, S., & Voytek, B. (2019). Cycle-by-cycle analysis of neural oscillations. *Journal of Neurophysiology*, *122*, 849–861. https://doi.org/10.1152/jn.00273.2019

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. *The Journal of Neuroscience*, *33*(9), 4002–4010. https://doi.org/10.1523/JNEUROSCI.4675-12.2013

Dalla Bella, S., Dotov, D., Bardy, B., & Cochen De Cock, V. (2018). Individualization of music-based rhythmic auditory cueing in Parkinson's disease. *Annals of the New York Academy of Sciences*, *1423*, 308–317. https://doi.org/10.1111/nyas.13859

Damsma, A., Schlichting, N., & van Rijn, H. (2021). Temporal context actively shapes EEG signatures of time perception. *The Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0628-20.2021

de Graaf, T. A., Gross, J., Paterson, G., Rusch, T., Sack, A. T., & Thut, G. (2013). Alpha-Band Rhythms in Visual Task Performance: Phase-Locking by Rhythmic Sensory Stimulation. *PLoS ONE*, *8*(3), e60035. https://doi.org/10.1371/journal.pone.0060035

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

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

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, *116*(20), 10113–10121. https://doi.org/10.1073/pnas.1816414116

Doelling, K. B., & Florencia Assaneo, M. (2021). Neural oscillations are a start toward understanding brain activity rather than the end. *PLoS Biology*, *19*(5), 1–12. https://doi.org/10.1371/journal.pbio.3001234

Donoghue, T., Schaworonkow, N., & Voytek, B. (2021). Methodological considerations for studying neural oscillations. *European Journal of Neuroscience*, (January), 1–26. https://doi.org/10.1111/ejn.15361

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

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2014). Moving in time: Bayesian causal inference explains movement coordination to auditory beats. *Proceedings of the Royal Society B*, *281*(May), 20140751.

Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA Using the Amsterdam Decoding and Modeling Toolbox (ADAM). *Frontiers in Neuroscience*, *12*, 368. https://doi.org/10.3389/fnins.2018.00368

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Thousand Oaks (CA): Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

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

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

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, *86*(March), 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Heideman, S. G., van Ede, F., & Nobre, A. C. (2018). Early Behavioural Facilitation by Temporal Expectations in Complex Visual-motor Sequences. *Neuroscience*, *389*, 74–84. https://doi.org/10.1016/J.NEUROSCIENCE.2018.05.014

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

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences*, *111*(41), 14935–14940. https://doi.org/10.1073/pnas.1408741111

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, *109*(49), 20095–20100. https://doi.org/10.1073/pnas.1213390109

Hickok, G., Farahbod, H., & Saberi, K. (2015). The Rhythm of Perception: Entrainment to Acoustic Rhythms Induces Subsequent Perceptual Oscillation. *Psychological Science*, *26*(7), 1006–1013. https://doi.org/10.1177/0956797615576533

Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*(1), 85–91. https://doi.org/10.1111/j.1749-6632.2011.06402.x

Honing, H., & Bouwer, F. L. (2019). Rhythm. In P. J. Rentfrow & D. Levitin (Eds.), *Foundations in Music Psychology: Theory and Research* (pp. 33–69). Cambridge, MA: MIT Press.

Iversen, J. R., & Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), *10th Intl. Conf. on Music Perception and Cognition* (pp. 465–468). Adelaide: Causal Productions.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

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

Keele, S. W., Nicoletti, R., Ivry, R. B., & Pokorny, R. A. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgements? *Psychological Research*, *50*, 251–256. https://doi.org/10.1007/BF00309261

Keitel, C., Obleser, J., Jessen, S., & Henry, M. J. (2021). Frequency-Specific Effects in Infant Electroencephalograms Do Not Require Entrained Neural Oscillations: A Commentary on Köster et al . (2019). *Psychological Science*, *32*(6), 966–971. https://doi.org/10.1177/09567976211001317

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/J.TICS.2014.01.002

Kloosterman, N. A., Kosciessa, J. Q., Lindenberger, U., Fahrenfort, J. J., & Garrett, D. D. (2020). Boosts in brain signal variability track liberal shifts in decision bias. *ELife*, *9*, 1–22. https://doi.org/10.7554/ELIFE.54201

Kosciessa, J. Q., Kloosterman, N. A., & Garrett, D. D. (2020). Standard multiscale entropy reflects neural dynamics at mismatched temporal scales: What's signal irregularity got to do with it? *PLoS Computational Biology*, *16*(5), e1007885. https://doi.org/10.1371/journal.pcbi.1007885

Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed.), *Psychology of time* (pp. 189–231). Bingley, UK: Emerald Group Publishing. https://doi.org/10.1016/B978-0-08046-977-5.00006-5

Large, E. W., Herrera, J. A., & Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, *9*(November), 159. https://doi.org/10.3389/fnsys.2015.00159

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review*, *106*(1), 119–159. https://doi.org/10.1037/0033-295X.106.1.119

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*(1), 1–37. https://doi.org/10.1207/s15516709cog2601_1

Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLoS ONE*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Lenc, T., Keller, P. E., Varlet, M., & Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proceedings of the National Academy of Sciences*, *115*(32), 8221–8226. https://doi.org/10.1073/pnas.1801421115

Lenc, T., Merchant, H., Keller, P. E., Honing, H., Varlet, M., & Nozaradan, S. (2021). Mapping between sound, brain and behaviour: Four-level framework for understanding rhythm processing in humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*, 20200325. https://doi.org/10.1098/rstb.2020.0325

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4. Retrieved from https://cran.r-project.org/package=emmeans

Lin, W. M., Oetringer, D. A., Bakker-Marshall, I., Emmerzaal, J., Wilsch, A., ElShafei, H. A., … Haegens, S. (2021). No behavioural evidence for rhythmic facilitation of perceptual discrimination. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.15208

London, J. (2012). *Hearing in time: Psychological aspects of musical meter.* (2nd ed.). Oxford: Oxford University Press.

Manning, F. C., Harris, J., & Schutz, M. (2017). Temporal prediction abilities are mediated by motor effector and rhythmic expertise. *Experimental Brain Research*, *235*(3), 861–871. https://doi.org/10.1007/s00221-016-4845-8

Manning, F. C., & Schutz, M. (2013). "Moving to the beat" improves timing perception. *Psychonomic Bulletin & Review*, *20*, 1133–1139. https://doi.org/10.3758/s13423-013-0439-7

Mathewson, K. E., Prudhomme, C., Fabiani, M., Beck, D. M., Lleras, A., & Gratton, G. (2012). Making Waves in the Stream of Consciousness: Entraining Oscillations in EEG Alpha and Fluctuations in Visual Awareness with Rhythmic Visual Stimulation. *Journal of Cognitive Neuroscience*, *24*(12), 2321–2333. https://doi.org/10.1162/jocn_a_00288

Matthews, T. E., Thibodeau, J. N. L., Gunther, B. P., & Penhune, V. B. (2016). The Impact of Instrument-Specific Musical Training on Rhythm Perception and Production. *Frontiers in Psychology*, *7*(February), 69.

https://doi.org/10.3389/fpsyg.2016.00069
Mento, G. (2013). The passive CNV: Carving out the contribution of task-related processes to expectancy. *Frontiers in Human Neuroscience*, *7*(December), 827. https://doi.org/10.3389/fnhum.2013.00827
Mento, G. (2017). The role of the P3 and CNV components in voluntary and automatic temporal orienting: A high spatial-resolution ERP study. *Neuropsychologia*, *107*, 31–40. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2017.10.037
Merchant, H., Grahn, J. A., Trainor, L. J., Rohrmeier, M. A., & Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society B*, *370*, 20140093. https://doi.org/10.1098/rstb.2014.0093
Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 1–11. https://doi.org/10.1080/23273798.2019.1693050
Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research*, *308*, 84–97. https://doi.org/10.1016/J.HEARES.2013.09.012
Morillon, B., Schroeder, C. E., Wyart, V., & Arnal, L. H. (2016). Temporal prediction in lieu of periodic stimulation. *The Journal of Neuroscience*, *36*(8), 2342–2347. https://doi.org/10.1523/JNEUROSCI.0836-15.2016
Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642
Nave-Blodgett, J. E., Hannon, E., & Snyder, J. S. (2021). Hierarchical beat perception develops throughout childhood and adolescence and is enhanced in those with musical training. *Journal of Experimental Psychology: General*, *150*(2), 314–339. https://doi.org/10.1037/xge0000903
Nobre, A. C., & van Ede, F. (2018). Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience*, *19*, 34–48. https://doi.org/10.1038/nrn.2017.141
Novembre, G., & Iannetti, G. D. (2018). Tagging the musical beat: Neural entrainment or event-related potentials? *Proceedings of the National Academy of Sciences*, *115*(47), E11002–E11003. https://doi.org/10.1073/pnas.1815311115
Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, *31*(28), 10234–10240. https://doi.org/10.1523/JNEUROSCI.0411-11.2011
Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience*, *32*(49), 17572–17581. https://doi.org/10.1523/JNEUROSCI.3203-12.2012
O'Reilly, J. X., McCarthy, K. J., Capizzi, M., & Nobre, A. C. (2008). Acquisition of the temporal and ordinal structure of movement sequences in incidental learning. *Journal of Neurophysiology*, *99*, 2731–2735. https://doi.org/10.1152/jn.01141.2007
Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/J.TICS.2019.08.004
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869
Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(4), 728–741. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2148588
Pesnot Lerousseau, J., Trébuchon, A., Morillon, B., & Schön, D. (2021). Frequency selectivity of persistent cortical oscillatory responses to auditory rhythmic stimulation. *The Journal of Neuroscience*, *41*(38), 7991–8006. https://doi.org/10.1523/JNEUROSCI.0213-21.2021
Povel, D.-J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, *30*(6), 565–572. https://doi.org/10.3758/BF03202011
Praamstra, P., Kourtis, D., Kwok, H. F., & Oostenveld, R. (2006). Neurophysiology of implicit timing in serial choice reaction-time performance. *The Journal of Neuroscience*, *26*(20), 5448–5455. https://doi.org/10.1523/JNEUROSCI.0440-06.2006
R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org
Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882. https://doi.org/10.1016/J.TICS.2018.08.003
Schroeder, C. E., & Lakatos, P. (2009a). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012
Schroeder, C. E., & Lakatos, P. (2009b). The Gamma Oscillation: Master or Slave? *Brain Topography*, *22*, 24–

26. https://doi.org/10.1007/s10548-009-0080-y
Schultz, B. G., Stevens, C. J., Keller, P. E., & Tillmann, B. (2013). The implicit learning of metrical and nonmetrical temporal patterns. *The Quarterly Journal of Experimental Psychology*, *66*(2), 360–380. https://doi.org/10.1080/17470218.2012.712146
Shalev, N., Nobre, A. C., & van Ede, F. (2019). Time for What? Breaking Down Temporal Anticipation. *Trends in Neurosciences*, *42*(6), 373–374. https://doi.org/10.1016/J.TINS.2019.03.002
Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase Entrainment of Human Delta Oscillations Can Mediate the Effects of Expectation on Reaction Speed. *The Journal of Neuroscience*, *30*(41), 13578–13585. https://doi.org/10.1523/JNEUROSCI.0703-10.2010
Sun, Y., Michalareas, G., & Poeppel, D. (2021). The impact of phase entrainment on auditory detection is highly variable: Revisiting a key finding. *European Journal of Neuroscience*, (June), 1–18. https://doi.org/10.1111/ejn.15367
Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Golumbic, E. Z. (2017). Neural entrainment to the beat: The "missing-pulse" phenomenon. *The Journal of Neuroscience*, *37*(26), 6331–6341. https://doi.org/10.1523/JNEUROSCI.2500-16.2017
Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *The Journal of Neuroscience*, *31*(10), 3805–3812. https://doi.org/10.1523/JNEUROSCI.5561-10.2011
Tichko, P., & Large, E. W. (2019). Modeling infants' perceptual narrowing to musical rhythms: neural oscillation and Hebbian plasticity. *Annals of the New York Academy of Sciences*, *1453*(April), 125–139. https://doi.org/10.1111/nyas.14050
van Atteveldt, N., Musacchia, G., Zion-Golumbic, E., Sehatpour, P., Javitt, D. C., & Schroeder, C. E. (2015). Complementary fMRI and EEG evidence for more efficient neural processing of rhythmic vs. unpredictably timed sounds. *Frontiers in Psychology*, *6*(OCT), 1–11. https://doi.org/10.3389/fpsyg.2015.01663
van Bree, S., Sohoglu, E., Davis, M. H., & Zoefel, B. (2021). Sustained neural rhythms reveal endogenous oscillations supporting speech perception. *PLOS Biology*, *19*(2), e3001142. https://doi.org/10.1371/journal.pbio.3001142
van der Weij, B., Pearce, M. T., & Honing, H. (2017). A probabilistic model of meter perception: simulating enculturation. *Frontiers in Psychology*, *8*(May), 824. https://doi.org/10.3389/fpsyg.2017.00824
Vuust, P., Pallesen, K. J., Bailey, C., Van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, *24*(2), 560–564. https://doi.org/10.1016/j.neuroimage.2004.08.039
Waschke, L., Kloosterman, N. A., Obleser, J., & Garrett, D. D. (2021). Behavior needs neural variability. *Neuron*, *109*(5), 751–766. https://doi.org/10.1016/j.neuron.2021.01.023
Wilsch, A., Mercier, M. R., Obleser, J., Schroeder, C. E., & Haegens, S. (2020). Spatial Attention and Temporal Expectation Exert Differential Effects on Visual and Auditory Discrimination. *Journal of Cognitive Neuroscience*, *32*(8), 1562–1576. https://doi.org/10.1162/jocn_a_01567
Zoefel, B., ten Oever, S., & Sack, A. T. (2018). The Involvement of Endogenous Neural Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of Evoked Neural Responses. *Frontiers in Neuroscience*, *12*, 95. https://doi.org/10.3389/fnins.2018.00095
Zoefel, B., & Vanrullen, R. (2017). Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: State-of-the-art, speculations and suggestions. *Frontiers in Neuroscience*, *11*(May), 296. https://doi.org/10.3389/fnins.2017.00296

Supporting Information

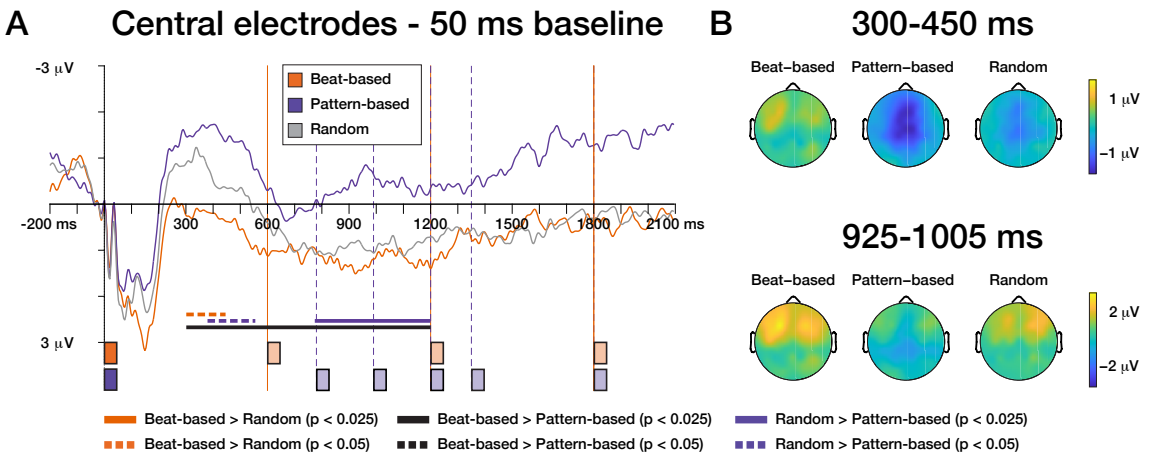

Figure S1. CNV with traditional 50 ms baseline. The traditional baseline causes overall shifts in the waveforms, suggestive of baseline differences between conditions. However, the overall picture stays the same as when not using a baseline, a CNV can be seen around 300-450 ms after the last sound onset for the pattern-based and random, but not the beat-based condition.

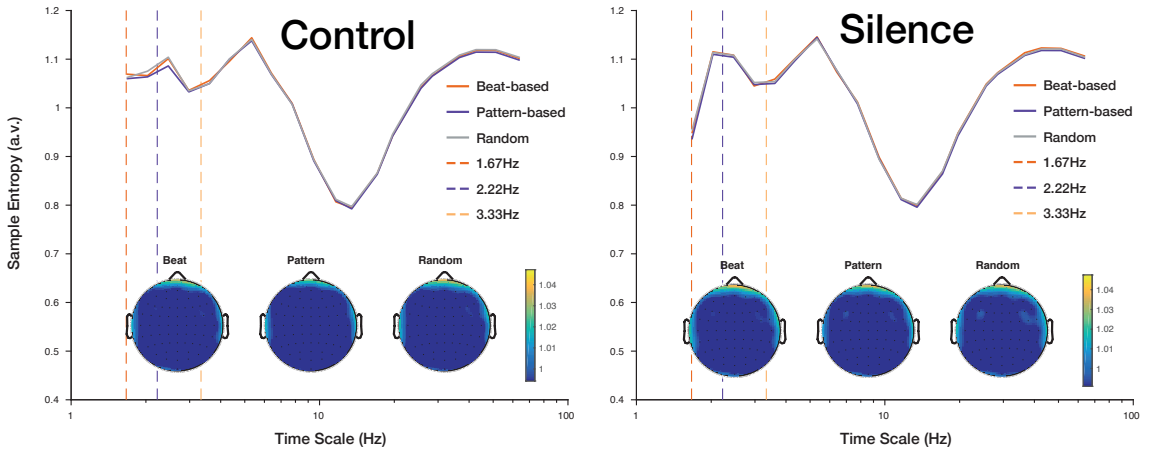

Figure S2. MMSE after highpass filtering at 5 Hz. Here, all differences between conditions are eliminated by the highpass filtering.

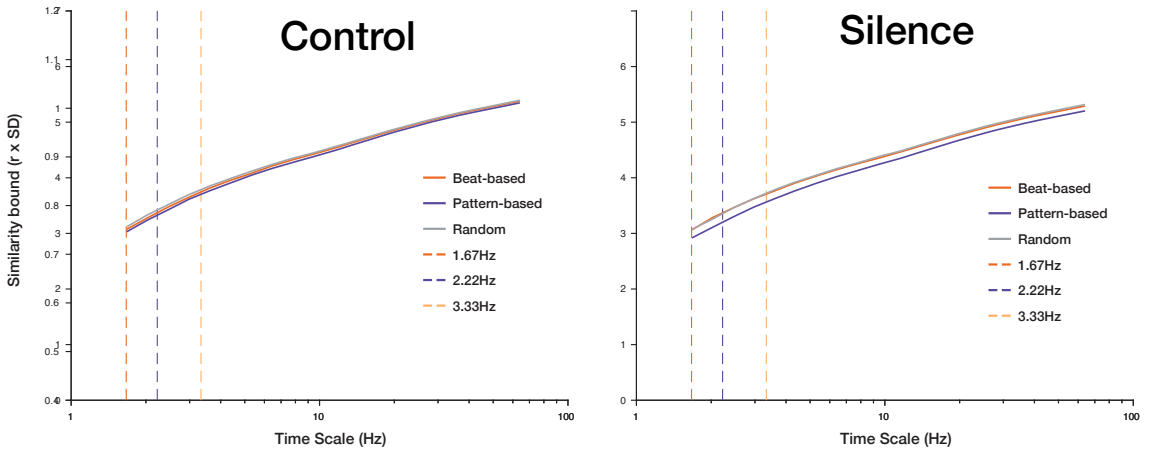

Figure S3. Similarity bounds used to calculate MMSE values mirror the condition differences. This finding suggests that condition differences found in the entropy measure, with higher entropy for the pattern-based than random condition in the silence window, may be due to overall variability differences, with lower variability for the pattern-based than random condition, rather than differences in entropy of the signal.