

1 PIPPET: A Bayesian framework for generalized
2 entrainment to stochastic rhythms

3 Jonathan Cannon

4 November 5, 2020

5 Department of Brain and Cognitive Science, Massachusetts Institute of
6 Technology, Cambridge, MA, USA

7 Tel.: +314-749-6902

8 jcan@mit.edu

9 **Abstract**

10 When presented with complex rhythmic auditory stimuli, humans are
11 able to track underlying temporal structure (e.g., a “beat”), both covertly
12 and with their movements. This capacity goes far beyond that of a simple
13 entrained oscillator, drawing on contextual and enculturated timing ex-
14 pectations and adjusting rapidly to perturbations in event timing, phase,
15 and tempo. Here we propose that the problem of rhythm tracking is
16 most naturally characterized as a problem of continuously estimating an
17 underlying phase and tempo based on precise event times and their cor-
18 respondence to timing expectations. We formalize this problem as a case
19 of inferring a distribution on a hidden state from point process data in
20 continuous time: either Phase Inference from Point Process Event Tim-
21 ing (PIPPET) or Phase And Tempo Inference (PATIPPET). This ap-
22 proach to rhythm tracking generalizes to non-isochronous and multi-voice

rhythms. We demonstrate that these inference problems can be approximately solved using a variational Bayesian method that generalizes the Kalman-Bucy filter to point-process data. These solutions reproduce multiple characteristics of overt and covert human rhythm tracking, including period-dependent phase corrections, illusory contraction of unexpectedly empty intervals, and failure to track excessively syncopated rhythms, and could be plausibly approximated in the brain. PIPPET can serve as the basis for models of performance on a wide range of timing and entrainment tasks and opens the door to even richer predictive processing and active inference models of rhythmic timing.

Keywords: Bayesian Inference, Active Inference, Timing, Rhythm, Entrainment

1 Introduction

The human brain is remarkably proficient at identifying and exploiting temporal structure in its environment, especially in the auditory domain. This phenomenon is most easily observed in the case of auditory stimuli with underlying periodicity: humans adeptly and often spontaneously synchronize their movements with such auditory rhythms [1], and human brain activity in auditory and motor regions aligns to auditory stimulus periodicity even in the absence of movement [2]. Both of these phenomena are cases of “entrainment” (sensorimotor and neural, respectively), where we define “entrainment” as in [3]: the temporal alignment of a biological or behavioral process with the regularities in an exogenously occurring stimulus.

A simple sinusoidal phase oscillator can entrain to a periodic stimulus; however, it is difficult to discuss the flexible entrainment of human behavior and cognitive processes to variable and sometimes aperiodic patterns such as speech without invoking the cognitive concept of “temporal expectation.” Expecta-

50 tions for event timing can be used to achieve a range of behavioral goals. They
51 can help us hone our sensory detection, our sensory discrimination, and our
52 response time for behaviorally important stimuli at the anticipated time [4, 5].
53 In some situations, temporal expectations attenuate neural responses [6], which
54 may help to conserve neural resources. And timing expectations bias our per-
55 ception of time, allowing us to use prior experience to supplement noisy sensory
56 data as we make temporal judgments [7].

57 Entrainment in humans involves an interplay of stimulus and temporal ex-
58 pectation [8]. Nowhere is this clearer than in interaction with music, hu-
59 mankind’s playground for auditory temporal expectation and entrainment [9].
60 But the precise nature of this interplay is an open question. The framework
61 of Dynamic Attending Theory characterizes temporal expectancy as pulses of
62 “attentional energy” issued by entrained neural oscillators, and mathematical
63 models based on these ideas describe bidirectional interactions between tempo-
64 ral expectation and entrainment that reproduce aspects of human behavior and
65 perception [10, 11]. But although the behavior of these models may be satis-
66 fying, the groundwork underlying them is less so: key high-level concepts like
67 the “attentional pulse” are difficult to define mechanistically, so the implemen-
68 tations of these concepts in models remain impressionistic. Moreover, recent
69 results have emphasized the relevance and neural correlates of aperiodic modes
70 of temporal expectation [12, 5, 13], but dynamic attending models are designed
71 to describe entrainment to periodicity and cannot account for aperiodic forms
72 of structured temporal expectation such as entrainment to memorized temporal
73 patterns, irregular musical meters, and the loose temporal regularities of speech
74 [14].

75 Here, we propose a normative framework for understanding the interaction
76 of entrainment and expectation. The goal is to first suggest a formal problem

77 that is being solved by general entrainment – namely, the problem of inferring
 78 the state of the exogenous process giving rise to a series of events in time – and
 79 then use mathematics to describe an optimal solution to that problem. This
 80 teleological approach to entrainment complements previous approaches based on
 81 cognitive constructs like dynamic attending. It brings to the table a concrete and
 82 mathematically precise link between the phenomenon of expectation-informed
 83 entrainment and the statistical structure of the stimuli that entrainment is used
 84 to exploit. If such a solution bears sufficient similarities to observations in
 85 humans, then we can begin to discuss human entrainment as a precise reflection
 86 of the temporal structure of the sensory world. Moreover, this approach is
 87 sufficiently general to describe entrainment to “stochastic” rhythms (rhythms in
 88 which some expected events may omitted) based on either periodic or aperiodic
 89 temporal expectations.

90 In the next section, we discuss previous models of expectation in cognition
 91 and where they fall short for our purposes. We then formulate three versions
 92 of the problem of entrainment that are amenable to precise solutions. In the
 93 first, “Phase Inference from Point Process Event Timing” (PIPPET), a hidden
 94 phase variable advances steadily with added noise, and the observer is tasked
 95 with continuously inferring the phase based on the observation of events emit-
 96 ted probabilistically at certain phases with certain degrees of precision. The
 97 variational Bayesian solution to this inference problem provides a continuous
 98 estimate of phase that entrains to the actual phase, as well as an estimated level
 99 of certainty about that phase. In the second, “Phase And Tempo Inference from
 100 Point Process Event Timing” (PATIPPET), the rate of phase advance (tempo)
 101 is also a dynamic variable with drift, and the solution simultaneously estimates
 102 phase, tempo, and certainty about both. The third (multi-PIPPET) general-
 103 izes the first two to incorporate the observation of multiple types of events, each

104 with distinct characteristic phases and precisions, into the inference process.

105 In the following section, we simulate these solutions, drawing on music as
106 a rich source of intuitive examples of entrainment informed by expectation. In
107 doing so, we provide intuition into the range of behaviors of these solutions,
108 and show how they reproduce key aspects of human sensorimotor entrainment
109 behavior that are not explained by other entrainment models. These include:

- 110 1. Failure to track phase through excessive syncopation (events occurring at
111 weakly expected times but omitted at strongly expected times).
- 112 2. Illusory contraction of intervals when expected events are omitted.
- 113 3. Near-linear corrections to phase after event timing perturbations, with
114 larger (and even over-) corrections for stimulus trains with longer inter-
115 onset intervals.

116 In the final section, we discuss the potential contributions of PIPPET and
117 PATIPPET to our understanding of human entrainment.

118 2 Mathematical framework

119 The framework of “predictive processing” has emerged as the preferred lens for
120 modeling the role of expectations in the brain [15, 16]. According to this con-
121 stellation of ideas, expectations (or, interchangeably, “predictions”) from higher
122 levels of the sensory processing hierarchy are sent to lower levels, where they
123 are compared to incoming sensory information and used to compute “predic-
124 tion errors.” These prediction errors are used to inform dynamic adjustments
125 to the expectations at all levels of processing, as well as slower adjustments to
126 the learned models upon which predictions are based. This is formalized as
127 a process of variational Bayesian inference based on a hierarchical generative
128 model.

Predictive processing would be a natural modeling framework for understanding rhythmic expectation and entrainment as inference [17, 18, 19] except for one key limitation: existing predictive coding models that operate in continuous time are structured to perform inference based on continuous observation, characterizing prediction errors in terms of deviation between a true level of input and a mean expected level of input [20, 21]. They describe predictions about “what” rather than “when,” and are therefore ill-suited to characterizing moment-by-moment errors in *timing* prediction, which arrive sporadically, separated by intervals largely devoid of informative prediction error. This may be a fundamental shortcoming in modeling inference in the brain: behavior and neurophysiology suggests that information about “when” is carried by its own distinctive pathways and represented separately from “what,” both in perceptual and motor tasks [22, 5, 9]. Bayesian methods have been applied to describe inferences about timing in the brain [23, 24, 25], but in these cases the problem the brain solves has been formulated as discrete inferences about consecutive intervals rather than a continuous inference process.

Here, we use event timing to inform a continuous variational inference process using the mathematical tool of point processes. The result approximates an ideal observer with respect to a generative process in continuous time that describes the probabilistic generation of a time series of events.

2.1 Phase Inference from Point Process Event Timing (PIPPET)

PIPPET is a simple generative model of a homogeneous, temporally structured series of instantaneous sensory events. This model consists of a phase $\phi \in \mathbb{R}$

153 that advances as a drift-diffusion process:

$$d\phi = dt + \sigma dW_t \quad (1)$$

154 and an inhomogeneous point process that generates events with probability
 155 $\lambda(\phi)$, a function of phase. We will refer to $\lambda(\phi)$ as a “temporal expectation
 156 template,” though it can also be understood as a hazard function for events. To
 157 achieve both analytical tractability and flexible descriptive power, we assume
 158 that $\lambda(\phi)$ is a sum of a constant λ_0 and a countable set of scaled Gaussian
 159 functions indexed by $i = 1, 2, \dots$ etc. Each Gaussian i is centered at a mean
 160 phase ϕ_i with variance v_i and scale λ_i :

$$\lambda(\phi) = \lambda_0 + \sum_i \lambda_i N(\phi | \phi_i, v_i) \quad (2)$$

161 where $N(x|m, v)$ denotes a normalized Gaussian distribution with mean m and
 162 variance v . Each Gaussian mean ϕ_i represents a phase at which an event is
 163 expected; λ_i represents the strength of that expectation; and v_i^{-1} is the tem-
 164 poral precision of that expectation. $\lambda_0 > 0$ represents the rate of events being
 165 generated as part of a uniform noise background unrelated to phase. Together,
 166 $\lambda(\phi)$ constitutes a likelihood function for an event occurring at phase ϕ . See
 167 Figure 1 for illustration.

168 Note that ϕ is assumed to be on the real line, not the circle. This design
 169 decision allows PIPPET to entrain to temporally patterned expectations with
 170 or without periodic structure by choosing a periodic or aperiodic temporal ex-
 171 pectation template λ . We discuss this decision further in the Discussion section.

172 Given a series of event times $[t_n]$, a temporal expectation template $\lambda(\phi)$, and
 173 a prior distribution $p_0(\phi)$ describing the distribution of phase at time $t = 0$, the
 174 observer’s goal is to infer a posterior distribution $p_t(\phi)$ describing an estimate

$$\lambda(\phi) = \lambda_0 + \sum_i \lambda_i N(\phi | \phi_i, v_i)$$

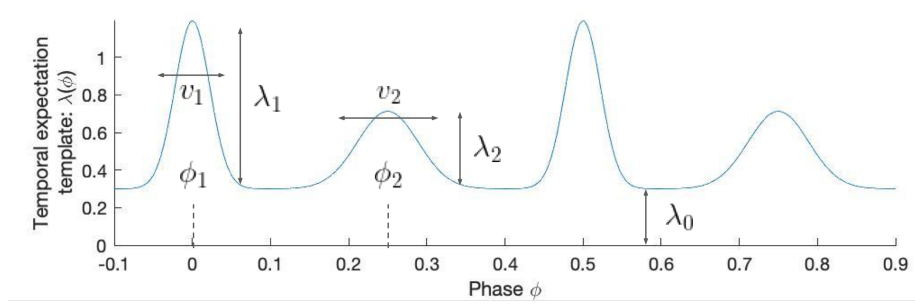


Figure 1: **The temporal expectation template.** In the PIP-PET/PATIPPET generative model, $\lambda(\phi)$ represents the instantaneous rate of events occurring when the underlying temporal process is at phase ϕ . This is assumed to be a sum of Gaussian-shaped functions with means ϕ_i representing the phases at which specific events are expected, variances v_i representing the (inverse of) the temporal precision expected of those events, and scales λ_i representing the strength of the expectations. A constant λ_0 is also added, representing the instantaneous rate of events unrelated to the underlying phase.

175 of phase ϕ at every time $t > 0$.

176 In [26], Snyder derives exact equations for the evolution of this posterior
177 distribution over time. Following the predictive processing ansatz of maintaining
178 Gaussian posterior distributions (the Laplace assumption), which provides both
179 computational tractability and neurophysiological plausibility by reducing the
180 representation of the posterior to a mean and a variance, we project the posterior
181 onto a Gaussian at each dt time-step. We do this by moment-matching: we use
182 Snyder’s solution to determine the evolution of the mean and variance of the
183 posterior, and then replace the true posterior with a Gaussian of the same mean
184 and variance. This choice of Gaussian is the choice with minimum KL divergence
185 from the true posterior [27], and therefore also minimizes the free energy of the
186 solution within the family of possible Gaussian posteriors, in accordance with
187 the Free Energy Principle of predictive processing [28].

188 The result of this derivation is a generalization of a Kalman-Bucy filter with
 189 Poisson observation noise. Eden and Brown [29] have derived an explicit form
 190 for this filter for any λ ; however, for λ a mixture of Gaussians, we find it easier
 191 to arrive at a clear and intuitive expression for the filter by deriving it directly
 192 from Synder's solution in [26]. Derivation is presented in Appendix 6.1.

193 **Solution: the PIPPET filter** At any time t , let μ_t denote the mean and Σ_t
 194 denote the variance of the Gaussian posterior. At each event time t , we let μ_{t-}
 195 and Σ_{t-} denote the left-hand limits of μ and Σ before the event, and we write
 196 μ_{t+} and Σ_{t+} to denote their right-hand limit values after the event. μ_t and Σ_t
 197 evolve according to the ODE

$$\begin{cases} \dot{\mu} = & 1 - \bar{\Lambda}(\bar{\mu} - \mu) \\ \dot{\Sigma} = & \sigma^2 - \bar{\Lambda}(\bar{\Sigma} - \Sigma) \end{cases} \quad (3)$$

198 and at each event $\mu_{t+} = \bar{\mu}$ and $\Sigma_{t+} = \bar{\Sigma}$, where we define

$$\begin{aligned} \bar{\mu} &:= \frac{\lambda_0}{\bar{\Lambda}} \mu_{t-} + \sum_i \frac{\Lambda_i}{\bar{\Lambda}} \bar{\mu}_i \\ \bar{\Sigma} &:= \frac{\lambda_0}{\bar{\Lambda}} \Sigma_{t-} + \sum_i \frac{\Lambda_i}{\bar{\Lambda}} (K_i + (\bar{\mu}_i - \mu_{t-})^2) \\ \bar{\mu}_i &:= K_i (\Sigma_{t-}^{-1} \mu_{t-} + v_i^{-1} \phi_i) \\ \Lambda_i &:= \lambda_i N(\phi_i | \mu_{t-}, v_i + \Sigma_{t-}) \\ K_i &:= \frac{1}{\Sigma_{t-}^{-1} + v_i^{-1}} \\ \bar{\Lambda} &:= \sum_i \Lambda_i \end{aligned}$$

199 Intuitively,

- 200 • μ_t is the estimated phase at time t , and Σ_t is the level of uncertainty about
201 the phase estimate.
- 202 • At each event time t , $\lambda(\phi)$ serves as a likelihood function for phase, and
203 the role of prior is played by a Gaussian with mean μ_{t-} and variance Σ_{t-} .
- 204 • At any time t , $\bar{\mu}_i$ would be the mean of the posterior if an event occurred
205 and was known to come from Gaussian i . It is a weighted sum of the
206 current mean estimated phase μ_t and the mean ϕ_i of Gaussian i , weighted
207 by the precision $\frac{1}{\Sigma_t}$ on estimated phase and the temporal precision $\frac{1}{v_i}$ of
208 the Gaussian generating the event, respectively.
- 209 • At any time t , $\bar{\mu}$ and $\bar{\Sigma}$ would be the mean and variance of the posterior if
210 an event occurred and its source was not known. These are weighted sums
211 of the influences of each Gaussian, weighted by Λ_i , the relative likelihood
212 that the event is drawn from Gaussian i .
- 213 • Between events, each dt time step is taken as a Bayesian inference with
214 likelihood $1 - \lambda(\phi)dt$ and with a Gaussian prior consisting of the posterior
215 of the previous time step carried forward by dt according to the Fokker-
216 Planck evolution associated with the ODE (3).
- 217 • In the absence of an event, this continuous inference process pushes μ and
218 Σ away from $\bar{\mu}$ and $\bar{\Sigma}$ with a strength proportionate to $\bar{\Lambda}$, the current
219 strength of the expectation of an event – thus, the absence of an event
220 continuously pushes the posterior in the opposite directing as would the
221 occurrence of an event.

2.2 Phase And Tempo Inference from Point Process Event Timing (PATIPPET)

PATIPPET is generative model of homogeneous point process events in time that extends PIPPET by making the rate of phase advancement itself a noisy dynamic variable subject to ongoing inference. The dynamic state of the system is now a two-dimensional vector $\phi = \begin{pmatrix} \phi \\ \theta \end{pmatrix}$, where ϕ is the phase as above, T is the rate of phase advancement (or tempo), and σ and σ_θ are the levels of phase and tempo noise, respectively:

$$d\phi = \begin{pmatrix} \theta \\ 0 \end{pmatrix} dt + \begin{pmatrix} \sigma dW_t \\ \sigma_\theta dW_t^\theta \end{pmatrix} \quad (4)$$

As above, an inhomogeneous point process generates events with probability $\lambda(\phi_1)$, where λ is a sum of Gaussians and a constant:

$$\lambda(\phi) = \lambda_0 + \sum_i \lambda_i N(\phi | \phi_i, v_i) \quad (5)$$

Given a series of event times $\{t_n\}$, a temporal expectation template $\lambda(\phi)$, and a prior distribution $p_0(\phi)$ describing the distribution of phase and tempo at time $t = 0$, the observer's goal is to infer a posterior distribution $p_t(\phi)$ describing an estimate of phase and tempo at every time $t > 0$. A similar derivation provides a point-process Kalman-Bucy filter that optimally serves this function within the constraint of Gaussian posteriors, providing a running estimate of a mean phase and tempo μ_t and a phase/tempo covariance matrix Σ_t . The solution and its derivation are presented in 6.1.

The resulting PATIPPET filter generalizes the PIPPET filter, and is identical if the initial tempo distribution is set to a delta distribution at $\theta = 1$ and

242 σ_θ is set to zero. At each event, the distribution of phase and tempo is dis-
 243 continuously updated to a 2D Gaussian posterior, which evolves continuously
 244 between events. This scheme is similar to [30], which estimates phase and tempo
 245 by updating a 2D Gaussian posterior, but is updated in continuous time and
 246 is significantly more flexible in its capacity to track phase based on arbitrary
 247 temporal expectation templates.

248 **2.3 PIPPET with multiple event streams (multi-PIPPET)**

249 Finally, we generalize PIPPET to include multiple types of events (indexed by
 250 j), each generated as point processes with rates determined by functions $\lambda^j(\phi)$
 251 of a single underlying phase:

$$d\phi = dt + \sigma dW_t \quad (6)$$

252

$$\lambda^j(\phi) = \lambda_0^j + \sum_i \lambda_i^j N(\phi | \phi_i^j, v_i^j) \quad (7)$$

253 The Kalman-Bucy estimate of phase for this model is described by mean μ
 254 and variance Σ evolving according to the ODE

$$\begin{cases} \dot{\mu} = 1 - \sum_j \bar{\Lambda}^j (\bar{\mu}^j - \mu) \\ \dot{\Sigma} = \sigma^2 - \sum_j \bar{\Lambda}^j (\bar{\Sigma}^j - \Sigma) \end{cases} \quad (8)$$

255 and resetting to $\mu_{t+} = \bar{\mu}^j$ and $\Sigma_{t+} = \bar{\Sigma}^j$ when an event occurs in stream j ,
 256 where we define $\bar{\Lambda}^j$, $\bar{\mu}^j$, and $\bar{\Sigma}^j$ as we defined $\bar{\Lambda}$, $\bar{\mu}$, and $\bar{\Sigma}$ above but in reference
 257 only to event stream j .

258 The same adjustment can be made to the PATIPPET generative model, and
 259 the PATIPPET filter can be similarly generalized to account for multiple event
 260 streams.

261 3 Results

262 In this section we conduct a series of simulations to explore parallels between the
263 behavior of the the PIPPET and PATIPPET filters and human entrainment.
264 Parameters for these simulations are listed in Appendix 6.2.

265 3.1 Response to events: phase and variance correction

266 We simulated PIPPET filter with simple metronomic expectations to illustrate
267 its basic behavior. Events occurring near an expected event phase ϕ_i cause the
268 mean phase estimate μ to shift linearly toward ϕ_i , as indicated by the plateaus
269 in the phase transition function (Figure 2A). Events occurring far from any
270 expected event phase ϕ_i caused negligible adjustment in the phase estimate
271 because they were attributed to the background rate λ_0 of events occurring
272 unrelated to any specific expectation. This leads to a phase response curve
273 that crosses zero with negative slope near each expected event phase and sits
274 uniformly near zero away from expected event phases (Figure 2A).

275 If the estimated phase μ_{t-} just before an event time t was very close to an
276 expected event phase ϕ_i , the phase uncertainty Σ decreased at the event, which
277 effectively “corroborated” the phase estimate (Figure 2B). Events occurring
278 when μ_{t-} was far from any expected event phase had no impact on Σ , as they
279 were effectively attributed to the background noise rate λ_0 and thus contained
280 no new information about phase. Events occurring in the liminal zone near but
281 not very near an expected event phase ϕ_i caused uncertainty Σ to increase.

282 3.2 Stochastic rhythms with uneven subdivision

283 The PIPPET framework describes entrainment to “stochastic” rhythms in which
284 each expected event phase may or may not be populated by an event. Fur-
285 ther, PIPPET is formulated in sufficient generality to describe entrainment to

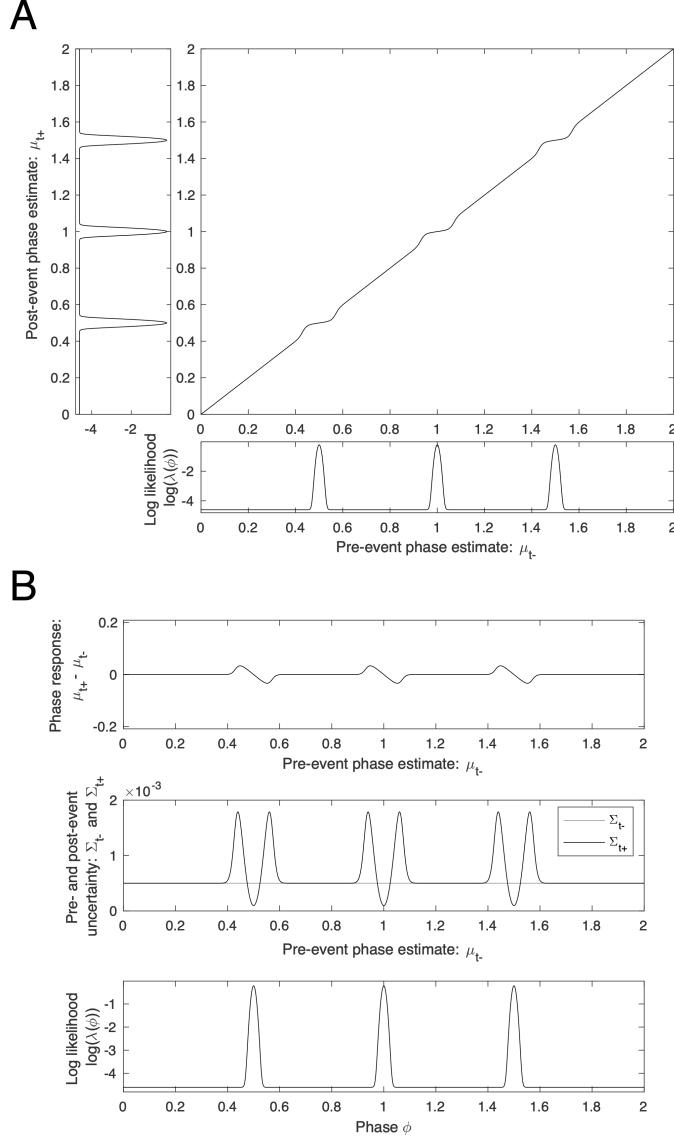


Figure 2: **Characterizing PIPPET's behavior at events** A) Phase transition curve for PIPPET with expectation of three isochronous events. Note that events occurring when the phase estimate μ_{t-} is between expected event phases ϕ_i have little corrective effect on the posterior mean phase μ_{t+} , as indicated by a diagonal phase transition curve, whereas events occurring when the estimated phase is near an expected event phase tend draw the phase estimate toward the expected phase, as indicated by plateaus in the phase transition curve. B) Phase and variance response curves. Note that events occurring when estimated phase is very close to an expected event phase cause the variance of the posterior on phase to decrease, whereas events occurring slightly offset from an expected event phase cause the variance to increase. Events occurring far from any expected event phase have little effect on posterior variance.

286 rhythms based on timing expectations with complex, non-isochronous stress
 287 patterns [31] and with non-integer duration ratios using suitably designed (or,
 288 presumably, learned) temporal expectation templates $\lambda(\phi)$. Such rhythmic pat-
 289 terns have been shown to support highly precise synchronization in musicians
 290 with appropriate training and enculturated expectations [32], and should there-
 291 fore be accounted for by any plausible model of human entrainment. Thus,
 292 PIPPET is equipped to model entrainment to a very wide range of rhythmic
 293 structures with any degree of predictability.

294 As an example of entrainment to a stochastic rhythm based on a temporal
 295 structure with non-integer duration ratios, we simulated entrainment to a swing
 296 rhythm. The rhythm is based on an underlying grid of “swung” eighth notes,
 297 where the first eighth note of every pair is given a slightly longer duration than
 298 the second. Though the “swing” feel is often caricatured using eighth note
 299 pairs with a 2:1 duration ratio, this value has been shown to vary by player
 300 and tempo and is certainly not limited to small integer ratios [33]. We used a
 301 temporal expectation template with a swing ratio close to 3:2 and associated the
 302 first eighth note in each pair with a stronger expectation than the second. The
 303 simulation entrained to a complex, syncopated rhythm based on this template,
 304 and corrected the phase estimate when a phase shift was introduced into the
 305 rhythm (Figure 3).

306 **3.3 Failure mode: too much syncopation**

307 Another attractive aspect of the PIPPET framework is that it can account for
 308 realistic failures in tracking perfectly timed rhythms. In addition to failures
 309 due to time warping described above, failures may occur due to interference
 310 between expectations packed closely together in time. Every expected event
 311 phase ϕ_i exerts an influence on the evolution of the posterior at all times. This

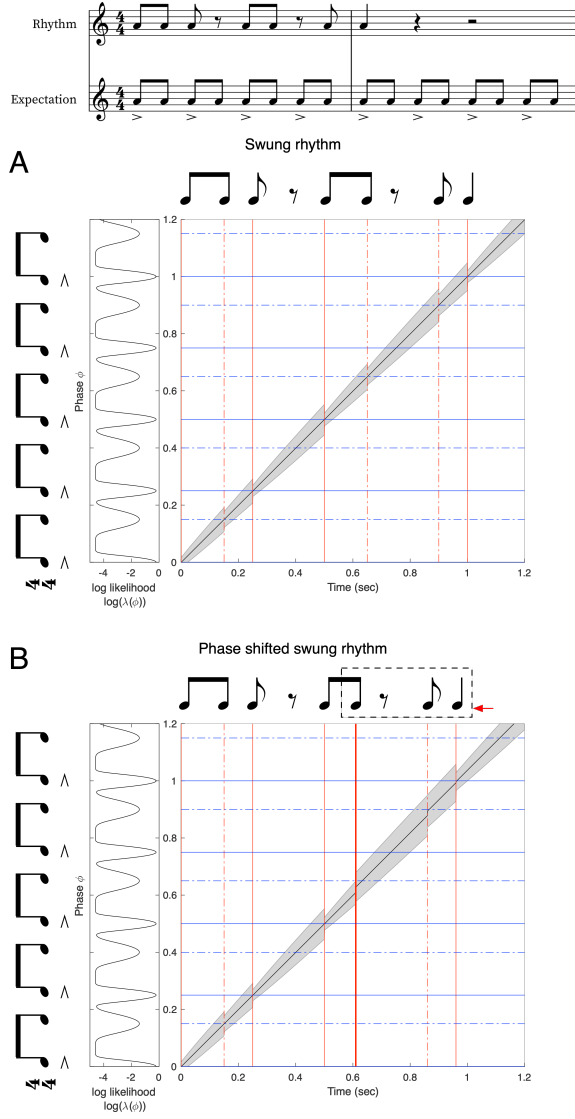


Figure 3: **Tracking phase through swung rhythms.** (Same color key as 5.) A: Phase is estimated over the course of a rhythm. Temporal expectations are not isochronous, but instead represent a swing pattern in which the first eighth note of every pair is slightly longer and more strongly expected than the second. Dotted lines correspond to weak expectations and solid lines correspond to strong expectations. B: A phase shift is introduced into the rhythm, moving all subsequent events earlier in time. When the first early event arrives, uncertainty Σ increases. Mean estimated phase μ is corrected over the first few events after the shift, and Σ decreases most substantially when the estimate μ is corroborated by a strongly expected event happening at the appropriate estimated phase.

influence is very weak if the current phase estimate is far from ϕ_i . However, if the uncertainty Σ of the phase estimate is large enough to encompass several expected event phases, or if several events are expected at neighboring phases with insufficient precision, the event may not be fully “attributed” to a single expected event phase. As a result, the adjustment to the phase estimate at an event may reflect an amalgam of these multiple influences, with stronger expectations exerting more influence than weaker ones.

A prime example of this failure mode in human rhythm tracking is tracking overly syncopated rhythms (rhythms with a predominance of events at time points with weaker expectations). Listeners tend to “re-hear” such rhythms by attributing events to metrical positions where events are more strongly expected [34]. We created an expectation template with a swing grid as in the previous section but with weakened expectations for the second eighth note in each pair. Against this background, we simulated a strongly syncopated rhythm (Figure 4). The rhythm’s phase was not tracked successfully due to a convergence of factors. Phase uncertainty Σ was only slightly reduced when events occurred at weakly expected phases, so it accumulated over the course of the rhythm, and especially during the long silence. Once Σ was large, strongly expected event phases ϕ_i began to exert more influence at each event, until eventually events that should have been attributed to weak phase points were instead attributed primarily to adjacent strong phase points. This type of attribution error in syncopated rhythm perception is described in [35].

3.4 In the absence of events: time warping

When an event is strongly expected but no event occurs, an optimal Bayesian observer should initially be biased to believe that in spite of their current estimate, the stimulus may not have reached the expected event phase yet. When

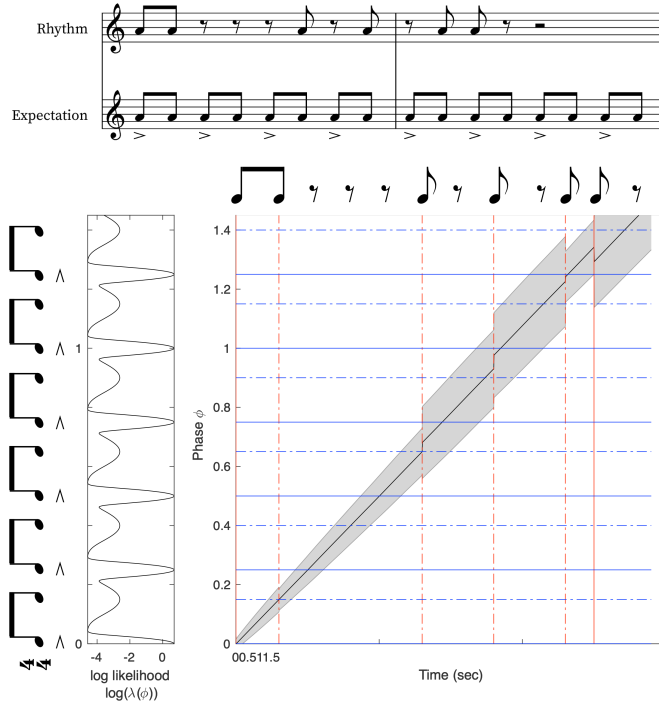


Figure 4: **Too much syncopation causes rhythm tracking failure.** Syncopation combined with imprecise and weak timing expectations on at weak time points can lead to a failure to track phase accurately. In this example, phase uncertainty Σ increases over a long silence. At the next event, this high uncertainty leads the model to partially attribute a weakly expected event to the nearby phase at which an event is strongly expected. As a result, the model ends up aligning the fifth event with a strong phase rather than a weak one.

we stimulated PIPPET with sufficiently strong metronomic expectations by scaling up λ , PIPPET’s behavior at each event was unchanged; however, when strongly expected events were omitted, the mean phase estimate slowed down at each expected event phase, leading to an overall slowing in estimated phase advance (Figure 5).

There is evidence of such an effect in human perception. The “filled duration” illusion is the impression that an isochronous sequence has changed tempo when it is initially subdivided by additional predictable events and then subdivisions are eliminated. According to multiple reports, the magnitude of this effect is reduced or eliminated if the empty intervals precede the filled intervals [36, 37, 38, 39] (though there is some disagreement about this [40]), suggesting that the established expectation of continuing subdivision interferes with perceived timing when subdivisions cease. In PIPPET, this effect is created when the slowing of phase advance causes a properly timed event at the end of the empty interval to arrive at an earlier apparent phase than expected, causing the interval to “seem” shorter.

A second result that could similarly be accounted for by this aspect of PIPPET is the surprising finding in [41] that a participant tapping along with a subdivided beat delays their tap following the omission of an expected subdivision. If taps are planned to coincide with the arrival of a specific mean estimated phase, then the slowing of phase induced by an omission of a strongly expected event in PIPPET would delay the subsequent tap.

3.5 Tempo inference

We simulated the PATIPPET filter with basic metronomic expectations to observe its capacity to infer phase and tempo at once. We gave the model a wide initial range of possible tempi and a simple metronomic stimulus with actual

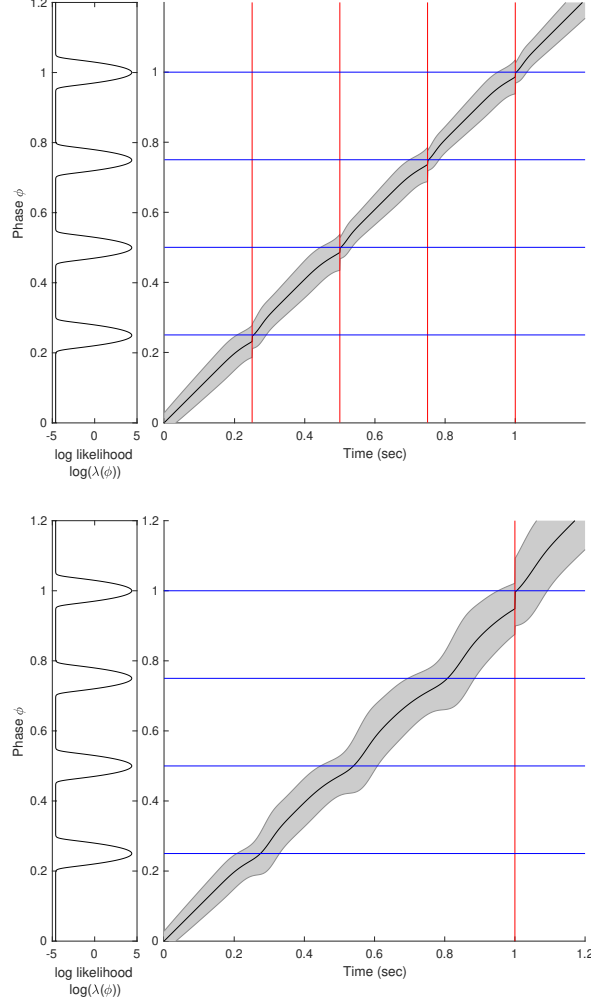


Figure 5: Time warping by the omission of strongly expected events. Black curve tracks the estimated mean phase μ over time. Red lines mark event times; blue lines mark expected event phases. Grey shading represents uncertainty about phase, quantified in the model as variance Σ and displayed by shading two standard deviations up and down. PIPPET is given strong expectations for four isochronous events. Above: when the strongly expected events occur as expected, mean phase stays on track, advancing (on average) at a rate of 1. Below: the first three expected events are omitted. When the strongly expected events do not occur, the advance of μ slows around the expected event phase and then speeds back up. On average over the interval, μ advances at a rate slower than 1. As a results, when the fourth event does occur at time $t = 1$, it occurs when μ_t is still substantially short of $\mu = 1$. The event is thus perceived as occurring at an earlier phase than expected.

tempo near the upper end of that range. In these conditions and with the parameter set we chose, the model established the appropriate tempo and phase to within a tight range over the course of the first two events (Figure 6).

In addition to its value as a model of human rhythmic cognition, the PATIP-PET filter shows promise as a general-purpose tempo tracking algorithm for musical applications. This would require a principled method of choosing values for the various free parameters of the generative model, which might be done a priori based on a labeled corpus, adaptively over the course of listening, or through some combination of the two. We leave a more thorough exploration of the relative performance of this model to future work.

3.6 Period-dependent corrections

In entrainment literature, finger taps entrained to a metronome generally shift to correct a certain fraction of an event timing perturbation on the next tap. This fraction is called α . In human subjects, α has repeatedly been observed to increase linearly with metronome period (“inter-onset interval,” or IOI), exceeding 1 (i.e., over-correction) for sufficiently long IOIs [42, 43].

The PIPPET framework offers a principled explanation for α increasing with IOI. During an event-free interval, phase uncertainty increases over time. When an event does occur, the precision of the prior distribution on phase and tempo is weighed against the precision of the likelihood function associated with the expectation of that event. If the prior is less precise due to accumulated uncertainty, the precision of the likelihood weighs more heavily against it and the adjustment in phase is more thorough. Thus, all else being equal, events spaced more widely apart in time induce more extensive phase corrections.

Since the strongest phase correction PIPPET can make at an event is to fully update the phase estimate to the expected event time, it cannot account

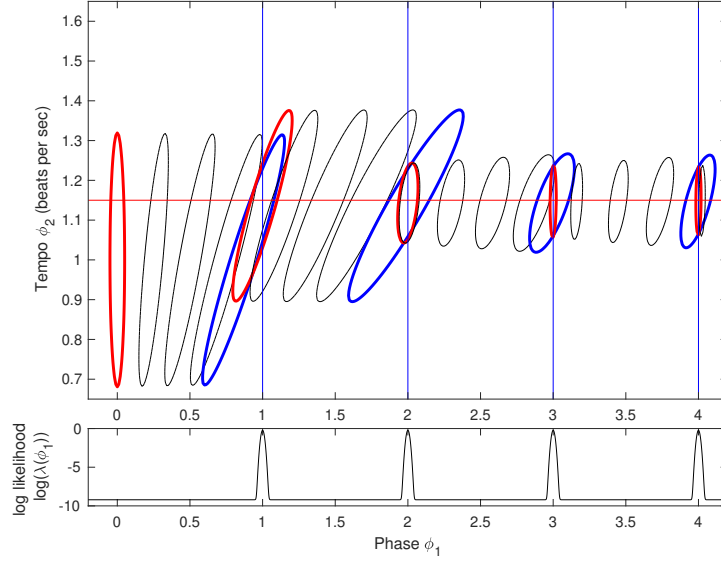


Figure 6: **A point process Kalman-Bucy Filter estimates phase and tempo.** Ellipses trace the contours of the Gaussian posterior distributions on phase and tempo. Black ellipses show a strobed visualization of the evolution of the posterior between events. Blue ellipses are the posterior distributions just before each event, and red ellipses are the posterior distributions just after each event. Here, PATIPPET is initialized with a high variance in its estimate of tempo. The first event occurs relatively early, causing the posterior mean tempo μ_θ to increase. Each subsequent event occurs close to the time expected based on the mean estimated phase μ and tempo μ_θ , causing, the posterior to contract in both the phase and variance direction as its prediction of event time is fulfilled and its phase and tempo estimates are corroborated. Ultimately, PATIPPET settles on a narrow distribution around the appropriate tempo as it continues to accurately estimate phase.

for α values above 1. However, it has been previously suggested that α may exceed 1 for long metronome periods due to some period correction occurring in addition to phase correction [42]. We were therefore curious to see whether PATIPPET could reproduce the linear increase of α with increasing IOI up to and beyond $\alpha = 1$.

In Figure 7, we show that with appropriate parameters, PATIPPET can indeed reproduce the experimental observation of a linear increase in α from below to above 1 as IOI increases. In PATIPPET, this phenomenon is a natural consequence of optimal inference in the context of phase and tempo uncertainty that accumulates between observations.

3.7 Multiple event streams

Multi-PIPPET generalizes the PIPPET/PATIPPET framework to cases of multiple distinguishable event types, each with its own set of expectations as a function of phase. One example could be listening, tapping, or dancing to a kit drum track with bass drum, snare, and hi-hat cymbal. Timing perturbations of different instruments in drum rhythms have been shown to differently affect human entrainment [44]. By letting j take values from $\{bass, snare, hihat\}$ and choosing appropriate values for ϕ_i^j , v_i^j , and λ_i^j for each event i on the metrical grid, we can create a set of timing expectations with strength and precision dependent on the specific drum and metrical position that could then be used to optimally track underlying phase and tempo through a complex kit drum rhythm. We illustrate such a template in Figure 8. A similar setup could be used to implement the assumption that pitches in a melody match the harmonic context more often in strong metrical positions, allowing event attribution and timing correction during melody listening to be influenced by scale degree.

Multi-PIPPET with $j \rightarrow \infty$ can be used to account for a continuum of event

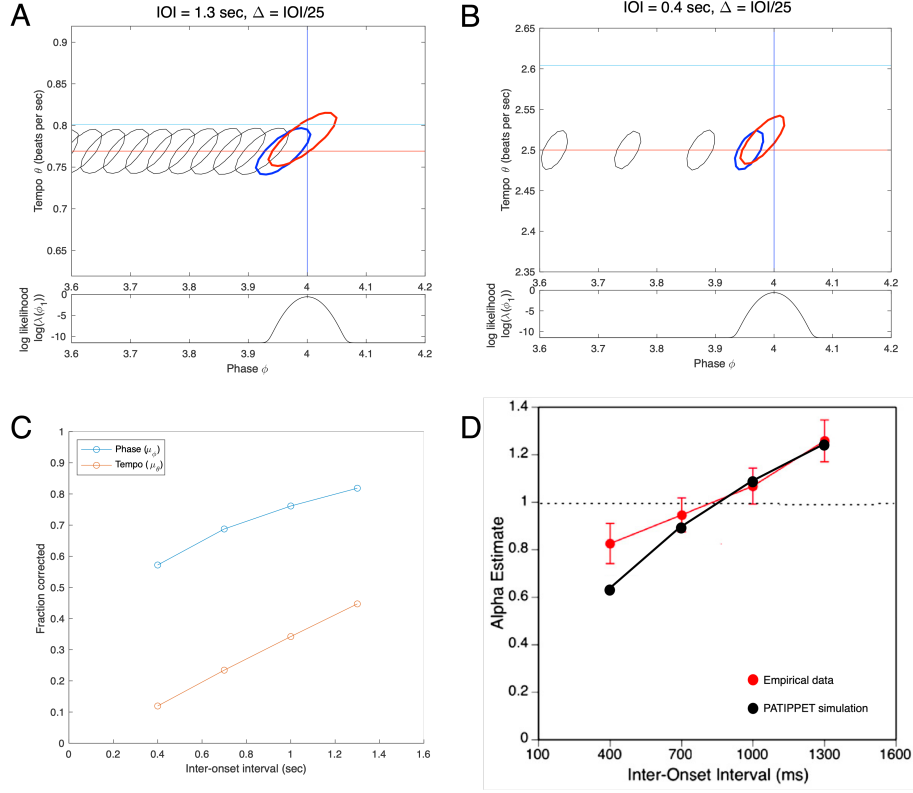


Figure 7: **PATIPPET reproduces human tapping data showing over-correction after timing perturbations to slow metronomes.** A and B) The distribution on phase and tempo leading up to and following a phase shift at the fourth event in an isochronous sequence for two different metronome tempi (i.e., two different inter-onset intervals). See Figure 6 for color key. Note that when the IOI is short, PATIPPET arrives at the phase-shifted event with a high degree of phase and tempo certainty. C) PATIPPET makes a proportionally larger correction to phase and tempo for long IOIs than for short IOIs due to the greater degree of uncertainty preceding each event. D) Alpha (α) is the proportion of a phase shift that is corrected at the next tap time. With this set of parameters, PATIPPET reproduces the empirical observation from [43] that the phase shift is undercorrected when IOIs are short and overcorrected $\alpha > 1$ when IOIs are long.

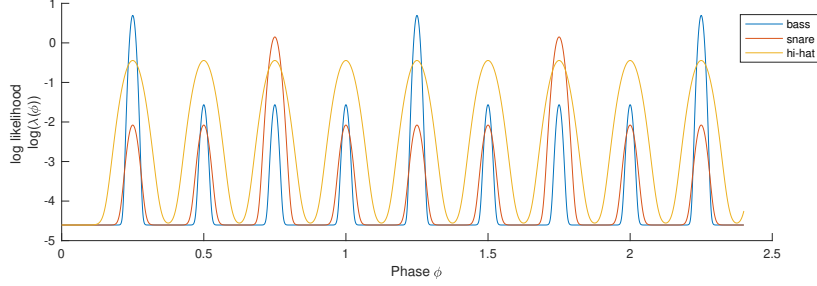


Figure 8: **Example expectation template for a basic rock beat.** In this illustration, bass drum hits are expected more strongly on the first of each cycle of four eighth notes, and are expected with high timing precision such that misplaced bass drum hits will exert a strong influence on phase. Snare drum hits are expected more strongly on the third eighth note of each cycle, and are expected with higher variance such that a misplaced snare hit exerts less influence on estimated phase. Hi-hat hits are evenly expected across all eighth note positions, but they are expected with low precision, so misplaced hi-hat hits will not exert a strong influence on estimated phase.

416 types. Thus, we could create a forward model in which it is more likely for notes
 417 played with stronger accents to fall on strong beats, or in which lower pitches
 418 are expected with higher timing precision and therefore exert greater influence
 419 on synchronization (as observed in [45]).

420 Multi-PIPPET could also be useful in flexibly modeling tapping data. Ex-
 421 periments have shown that the presence of entrained tapping prior to temporal
 422 perturbations in a metronomic stimulus reduces the phase correction response
 423 [46], indicating that the estimate of moment-by-moment phase is influenced by
 424 the proprioceptive and auditory feedback from tapping. Given working assump-
 425 tions about how taps are planned and executed based on an underlying phase
 426 estimate, the taps themselves could provide a second stream of input to the
 427 ongoing phase estimation that would bias it toward making smaller corrections
 428 to timing perturbations.

429 Importantly, using tap times to inform an estimate of underlying phase chal-
 430 lenges our interpretation of this phase representing a purely external source of

431 temporally patterned events. Instead, the inferred phase would be a hybrid of
 432 an external phase and the phase of one’s own motor cycle. Functionally, this
 433 is similar to the perceptual oscillator forced by both an external stimulus and
 434 one’s own periodic action proposed by [47]. This may be an especially useful
 435 way to think about synchronization with another agent, where one can adopt
 436 strategies ranging from following (assigning high precision to input from the
 437 other) to leading (assigning low precision to input from the other, and possibly
 438 higher precision to self-generated events). See [48] for a discussion of such a
 439 coding strategy as a means of minimizing representational neural resources.

440 The PIPPET framework could be further generalized to take into consider-
 441 ation additional stream of continuous input. This could be visual input from
 442 watching a pendulum, auditory input from a continuously modulated sound,
 443 or proprioceptive feedback from continuous entrained motion (as opposed to
 444 discrete, timed proprioceptive feedback like tapping). This goes beyond the
 445 scope of the mathematics presented here, but is a straightforward application
 446 of results proven in [26].

447 4 Discussion

448 Here we have presented PIPPET, a framework representing entrainment to
 449 a time series of discrete events based on a template of temporal expectations.
 450 PIPPET treats the event stream as the output of a point process modulated
 451 by the state of a hidden phase variable. The PIPPET filter uses variational
 452 Bayes to continuously estimate phase and track phase uncertainty based on
 453 this generative model. PATIPPET extends PIPPET to include a generative
 454 model of tempo change, and the PATIPPET filter simultaneously estimates
 455 phase, tempo, and the covariance matrix representing their uncertainty and
 456 their codependence. This framework is intended to serve as a hypothesis for

457 how the human brain integrates auditory event timing to inform and update an
458 estimate of the state and rate of an underlying temporal process.

459 Our chosen examples have been auditory rhythms based on cyclical (met-
460 ric) patterns of temporal expectations. But PIPPET is sufficiently general to
461 describe entrainment based on non-isochronous and even aperiodic temporal
462 expectations, an area that has been largely neglected in entrainment model-
463 ing. Further, it can describe the integration of multiple event streams into an
464 entrainment process, each with its own associated timing expectations.

465 PIPPET and PATIPPET reproduce several qualitative features of human
466 entrainment, including realistic failures to track overly perfectly-timed but over-
467 syncopated rhythms, perceived acceleration of a metronomic pulse when strongly
468 expected events are omitted, and error correction after metronome timing per-
469 turbations that increases with increasing inter-onset interval. We show that
470 these phenomena all follow naturally from our framing of entrainment as a pro-
471 cess of Bayesian inference based on specific phase-based temporal expectations.

472 4.1 Relationship to other models of timing

473 The dynamics of PIPPET and PATIPPET in response to sensory events are
474 similar to dynamics of other entrainment models that correct phase and period
475 based on event timing, e.g., [49, 50]. Models based on dynamic attending the-
476 ory, e.g., [10, 11], are also similar in explicitly modeling timing expectations
477 and their effect on phase and period adjustment. Our frameworks differ from
478 these in three key ways. First, they are derived as optimal solutions to specific
479 inference problems, and therefore all modeling decisions can be justified within
480 a normative framework. Second, they explicitly track uncertainty in phase and
481 tempo – without this feature, they would not account for observed dependence
482 of phase shift response on inter-onset interval or mimic human failures to track

483 overly-syncopated rhythms. Finally, they allow expectations to influence the
484 inferred phase even in the absence of sensory events, creating the time-warping
485 effect of disappointed expectations evidenced in humans by the “filled duration”
486 illusion.

487 Bayesian methods have been used elsewhere to analyze rhythmic structure
488 as time series of point events. Some of these are application-focused methods
489 that require offline analyses [51, 52] and therefore do not serve as satisfying
490 models of real-time behavior. Cemgil et al (2000) [30] use a Kalman filter that
491 tracks a distribution on phase and tempo similarly to PATIPPET. However,
492 this model is structured to infer phase and tempo event-by-event rather than in
493 continuous time, and is not equipped to handle stochastic rhythms or temporal
494 structures more complex than approximate isochrony.

495 Bayesian inference has also been used to model timing estimation in the
496 brain (e.g., [23, 24]), but it is generally used to describe inferences about discrete
497 variables like interval durations and event times, whereas PIPPET describes a
498 continuous inference process underlying predictions about event times. One
499 such model leading to particularly PIPPET-like results was presented in Elliot
500 et al 2014 [25]. The authors created a Bayesian model to explain the results of
501 an experiment that had participants tap along to a stimulus consisting of two
502 jittered metronomes. The model behaves similarly to PIPPET in that it esti-
503 mates the next event time using a weighted average of previous event times and
504 prior beliefs, with weights informed by expected timing precision. However, like
505 [30], their model infers the anticipated timing of discrete, metronomic events,
506 whereas PIPPET predicts and updates an underlying phase in continuous time
507 and can therefore generalize to non-isochronous and stochastic rhythms and ac-
508 count for the effects of event omissions. Additionally, in order to account for
509 participants ignoring events far from predicted time points, they introduce the

510 assumption that participants repeatedly test the hypotheses that events come
511 from one or two separate streams, whereas PIPPET naturally accounts for this
512 phenomenon by attributing stray events to a background event rate λ_0 .

513 4.2 Motor, perceptual, and neural entrainment

514 Throughout this work, we have made mention of perceptual and motor expres-
515 sions of entrainment, but have remained agnostic as to how we would expect
516 to observe an expression of phase and tempo inference in humans. These two
517 readouts sometimes give conflicting results: for example, exposure to musical
518 performance with expressively irregular timing affects perceptual reports of tim-
519 ing in subsequent stimuli [53], but does not affect phase correction in tapping
520 to subsequent stimuli [54].

521 We expect that both physical entrainment and perceptual report are in-
522 formed by a neural process of estimating underlying phase. Further, principles
523 of economy suggest that they should share in such an estimate rather than draw-
524 ing on separately instantiated processes of neural inference. However, neither
525 motor nor perceptual experiments will necessarily give a straightforward readout
526 of this inference process. Both readouts may be affected by independent sources
527 of additional noise, and also potential biases: certain perceptual responses may
528 be implicitly considered less likely than others, and certain motor errors may be
529 implicitly considered more costly than others. Thus, an attempt at a normative
530 Bayesian model at a specific task should be prepared to take into account this
531 additional layer of complexity.

532 4.3 PIPPET in the brain

533 If the brain is indeed performing an optimal estimation of phase and tempo,
534 then this estimate should be legible in neural activity somewhere in the brain.

535 At the scalp level and in intracortical electrodes, slow electrical oscillations do
 536 seem to anticipatorily track the structure of periodic auditory stimuli [55, 56],
 537 and this tracking is associated with the subjective passage of time [57]; these os-
 538 cillations could be explored as possible estimates of mean underlying phase. In
 539 monkeys, the supplementary motor area appears to track the phase underlying
 540 periodic visual events [58]; recordings from this region could be another candi-
 541 date for reading out mean phase. Nigrostriatal dopaminergic signaling has been
 542 identified as a possible marker of timing certainty [59, 60], so those dopaminer-
 543 gic populations might be a good place to look for a readout of phase variance.
 544 The temporal expectation template is a hazard function, and may therefore be
 545 observable by using techniques recently applied to decode the temporal hazard
 546 function from EEG data [61], or through its correlation with beta oscillations
 547 [62].

548 Though PIPPET and PATIPPET are not committed to a particular brain-
 549 based implementation, advances in the brain basis of timing and beat-keeping
 550 combined with the hypothesized neural bases of predictive processing suggest
 551 the beginnings of a plausible implementation of PIPPET in the brain. A de-
 552 tailed discussion of a possible neural basis of beat maintenance is presented in
 553 [63]. Briefly, supplementary motor area may maintain an ongoing estimate of
 554 mean phase through some combination of intrinsic dynamics and interaction
 555 with the basal ganglia, while dopaminergic signaling in striatum may maintain
 556 an estimate of phase uncertainty. The phase estimate may be used to inform
 557 auditory timing expectancy via learned models in premotor cortex [64]. These
 558 expectations may be delivered to the early stages of audition via the top-down
 559 connections along the dorsal auditory pathway, where they can be used to eval-
 560 uate timing prediction error [65]. These errors, weighted by their precisions,
 561 may be transmitted back to the supplementary motor area via the bottom-up

connectivity of the dorsal auditory pathway and used to update the estimate of phase.

4.4 Learning and inference outside of PIPPET

If the brain does treat entrainment as a process of inference based on a generative model, this raises the question of how the properties of the generative model are established in the first place. The PIPPET framework does not address this question directly, but by examining the parameters necessary to formulate PIPPET, we can clearly see what components need to be in place before a process of continuous phase and tempo updating can begin.

First, the brain must learn the temporal structures of the expectation template for rhythmic expectation. Learning these underlying structures from an experiential corpus of noisy, stochastic rhythms is not trivial. It seems likely to involve some type of bootstrapping in which a recognition of some degree of temporal structure allows for attribution of events to positions in that structure, allowing for deeper structure learning. Earlier exposure to simpler, less stochastic rhythms would likely help with such a bootstrapping process. For a discussion of the challenges of this type of simultaneous learning and filtering and a proposed solution for non-point-process data, see [66].

The brain must also learn noise and precision parameters for the model. Note that neither the temporal expectation variance parameters v_i nor the noise parameters σ and σ_θ necessarily correspond to the actual precision of the neural or external timing mechanisms in play. The brain may underestimate the noisiness (σ) of the timing process it uses to track underlying phase, leading to under-adjustment to auditory event timing and minimal time-warping between events, or do the opposite. Presumably, these parameters must be learned through experience and prediction error.

588 The precision parameters v_i may be informed by several factors. First, an
 589 upper bound on the precision of expected event timing is the precision of sensory
 590 timing perception, which is, for example, high for human audition and signifi-
 591 cantly lower for human vision¹. Second, expected event timing precision may
 592 also be informed by the observed relative timing distributions of event streams.
 593 These observations may inform expectations on time scales ranging from a single
 594 sitting to a lifetime of listening. Expected timing may be learned separately for
 595 different sensory modalities, different musical genres (e.g., techno vs. funk), or
 596 even different instruments (e.g., kick drum, snare, hi-hat, as discussed above).
 597 The precision of a beat-based temporal expectation is closely related to the
 598 width of a “beat bin,” the window of time (rather than a single time point) that
 599 is proposed to constitute the “beat” in [67], and to the width of the temporal
 600 “expectancy region” described in dynamic attending theory [10]; in both cases,
 601 this width is increased by imprecision in the immediately preceding stimulus.

602 When the brain is exposed to a rhythmic stimulus, it must first recognize
 603 that a predictable pattern exists and select an appropriate temporal expectation
 604 template from its learned repertoire. This is its own process of inference, and
 605 may be amenable to a Bayesian description. Since the PIPPET filter maintains
 606 a unimodal posterior, it is not well-suited to model this initial inference process,
 607 which may require maintaining a distribution over multiple distinct possible
 608 starting phases and temporal expectation templates. This problem might be
 609 partially addressed at a modeling level by incorporating a model of meter in-
 610 ference based on prior probabilities of hearing specific meters at specific tempi,
 611 e.g. [68], as an additional level of inference in parallel with phase and tempo

¹An event can only be experienced after it occurs, so (as pointed out in [24]) the likelihood function on underlying phase associated with this type of uncertainty should be asymmetrical. The analytically tractable incarnation of our framework presented here uses Gaussian likelihood peaks, so cannot account for the effect of asymmetrical likelihoods; however, we could posit a λ function with asymmetrical peaks and use numerical methods rather than the explicit solution derived here to estimate underlying phase at each time step.

inference.

Finally, aspects of the temporal expectation template are likely changing even as a rhythm plays out in time. This is evidenced by the grammar-like structure of music rhythm [69]: certain patterns of events are more expected than others regardless of their metrical positions. PIPPET and PATIPPET take a template of expected event time points as an input, and thus do not take into account immediate stimulus history in creating expectations. However, such effects could be incorporated into a model based on this framework by adding a history dependence to the expectation template λ . The precise details of this history dependence could be based on any suitable formal model for rhythmic grammar (e.g., [70] or [69]).

4.5 Future directions

In evaluating future directions, it is important to be clear that PIPPET and PATIPPET are not “models” but “frameworks.” Directly testing their validity as models of human behavior would require setting values for many free parameters, and it is not yet clear to what extent the parameters of individual expected events should be based on empirical data collected over a lifetime or empirical data collected trial by trial.

However, there is a certain extent to which these frameworks can be validated as descriptions of human cognition. First, these models predict certain qualitative effects such as the slowing of perceived phase advance as strong expectations are disappointed. Second, although the parameters in the forward models are not directly empirically measurable values, changes in stimulus history should influence them in predictable ways. For example, if a certain type of event occurs consistently at a particular metrical position within an extended stimulus presentation or within the music the listener has experienced in a life-

time of listening, then it should induce stronger phase corrections than an event that occurs inconsistently as if it has been given a higher value of λ_i . Parameters may also be influenced by long term listening experience, but they should at least respond to recent empirical experience by changing in the direction predicted by PIPPET.

If we find situations in which human behavior differs from solutions to the inference problems posed by PIPPET and PATIPPET, this suggests that the tasks being performed in those situations are being performed with a different objective than optimal inference of phase and tempo based on these generative models. In this case, we would be challenged to articulate the true nature of the problem being solved. This might require modifications of the generative model, e.g., introducing the belief that tempo changes occur in jumps or ramps rather than as random drift, or modification of the objective of the task, e.g., by including additional cost functions or priors associated with perceptual report or motor output as discussed above.

Once we are satisfied with the PIPPET framework’s utility in describing to human behavior, we can use it to model and analyze experimental data. Given a perceptual or behavioral task, we can suppose that motor or perceptual human entrainment behavior is optimally solving an inference problem, and determine the parameters of that problem by fitting them with appropriate methods. We can study the changes in these parameters over the course of an experiment, over different variations on the same experiment, over the human lifespan, across cultures, etc. This approach could add an additional level of insight to the analysis of a wide range of timing tasks.

One specific question that the PIPPET framework might help resolve is how periodic and nonperiodic entrainment differ. PIPPET has no specific machinery to account for ways in which the two situations differ (for neural and behavioral

evidence of differences between memory-based and periodicity based entrainment, see, e.g., [13, 5]. However, since it is sufficiently general to model both, it could guide an exploration of parameter differences between the performance of similar tasks in periodic and aperiodic contexts.

We can also let the PIPPET framework guide a search for the brain bases of entrainment. Even if perceptual and motor outputs are subject to different biases and costs, they would both be well-served by an optimal estimate of a ground truth, so there is reason to expect to find such an estimate represented in the brain. Such a search could proceed by looking for covariates for PIPPET’s phase and uncertainty estimates in neural data during the performance of tasks that require non-trivial updating of these estimates.

Finally, the PIPPET framework can serve as a cog in larger predictive processing models. The generative models we describe here allow for the evaluation of joint and marginal distributions on specific timing patterns and hidden states underlying them. By introducing additional levels of hidden states and additional sources of sensory input, we can create Bayesian inference models that use event timing to infer higher-order contextual states, e.g. meter, and predict other aspects of sensory input, e.g. pitch, creating a unified picture of human musical expectation.

5 Acknowledgments

Thanks to Tom Kaplan for extensive discussions and insights motivating this manuscript, and to Darren Rhodes and Nori Jacoby for helpful feedback.

687 6 Appendix

688 6.1 Derivation of differential equations and update equa- 689 tions.

690 Snyder [26] provides this general solution for the probability distribution on a
691 continuously stochastically evolving state

$$d\phi = F(\phi)dt + \sigma dW_t \quad (9)$$

692 which generates observable point process events at rate $\lambda(\phi)$:

$$dp_t(\phi) = \mathcal{L}[p_t(\phi)]dt + p_t(\phi) (\lambda(\phi) - \mathbb{E}_p[\lambda(\phi)]) \cdot (\mathbb{E}_p[\lambda(\phi)]dN_t - dt) \quad (10)$$

693 where dN_t is the increment in the event count over each dt time step (assumed
694 to be either 1 or 0 with probability 1), \mathbb{E}_p denotes expectation under distribution
695 $p_t(\phi)$, and \mathcal{L} is the Kolmogorov forward operator associated with (9):

$$\mathcal{L}[p(x)] = - \sum_i \partial_i [(Fx_t)_i p(x)] + \frac{1}{2} \sum_{i,j} \partial^2 [\sigma \sigma' p(x)]_{ij} / \partial x_i \partial x_j$$

696 Here we project p onto a Gaussian distribution at each time step by matching
697 moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which is also the projection with minimal KL divergence.
698 We can do this by finding the moments of dp , which are $d\boldsymbol{\mu}$ and $d\boldsymbol{\Sigma}$, and using
699 these to drive the evolution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

$$d\boldsymbol{\mu} = \int_{\phi} \phi \mathcal{L}[p_t(\phi)] d\phi dt + (\mathbb{E}_p[\phi \lambda(\phi)] - \boldsymbol{\mu} \mathbb{E}_p[\lambda(\phi)]) \cdot (\mathbb{E}_p[\lambda(\phi)]^{-1} dN_t - dt) \quad (11)$$

$$\begin{aligned}
d\mathbf{\Sigma} &= \int_{\phi} (\phi - \boldsymbol{\mu})(\phi - \boldsymbol{\mu})^T \mathcal{L}[p_t(\phi|N_t)] d\phi dt \\
&+ (\mathbb{E}_p [(\phi - \boldsymbol{\mu})(\phi - \boldsymbol{\mu})^T \lambda(\phi)] - \mathbf{\Sigma} \mathbb{E}_p [\lambda(\phi)]) \cdot (\mathbb{E}_p [\lambda(\phi)]^{-1} dN_t - dt)
\end{aligned} \tag{12}$$

Let $\|x\|_A^2$ denote $x^T A x$. For both PIPPET and PATIPPET, we can write

$$p(\phi) = \frac{1}{\sqrt{2\pi|\mathbf{\Sigma}|}} e^{-\frac{1}{2}\|\phi - \boldsymbol{\mu}\|_{\mathbf{\Sigma}^{-1}}^2}$$

701

$$\lambda(\phi) = \lambda_0 + \sum_i \frac{\lambda_i}{\sqrt{2\pi v_i}} e^{-\frac{1}{2}\|\phi - \phi_i\|_{P_i}^2}$$

702 where in PIPPET we set

$$\mathbf{P}_i = v_i^{-1}, \boldsymbol{\mu} = \mu, \phi = \phi, \text{ and } \phi_i = \phi_i$$

703 with scalar-valued $\mathbf{\Sigma} = \Sigma$, and in PATIPPET we set

$$\mathbf{P}_i = \begin{pmatrix} v_i^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu \\ \mu_\theta \end{pmatrix}, \phi = \begin{pmatrix} \phi \\ \theta \end{pmatrix}, \text{ and } \phi_i = \begin{pmatrix} \phi_i \\ 0 \end{pmatrix}$$

704 with matrix-valued $\mathbf{\Sigma} = \begin{pmatrix} \Sigma & s_{21} \\ s_{12} & s_{22} \end{pmatrix}$.

705 We will make use of the following result, a generalized form of a well-known
706 result about quadratic forms (see [71] for proof and similar application):

$$\|x - a\|_A^2 + \|x - b\|_B^2 = \|a - b\|_{A(A+B)^{-1}B}^2 + \|x - (A+B)^{-1}(Aa+Bb)\|_{A+B}^2 \tag{13}$$

In order to calculate the expectations in (11) and (12), we derive a simple

expression for $p(\phi)\lambda(\phi)$:

$$\begin{aligned} p(\phi)\lambda(\phi) &= \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}\|\phi-\mu\|_{\Sigma^{-1}}^2} \left(\lambda_0 + \sum_i \frac{\lambda_i}{\sqrt{2\pi v_i}} e^{-\frac{1}{2}\|\phi-\phi_i\|_{P_i}^2} \right) \\ &= \frac{\lambda_0}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}\|\phi-\mu\|_{\Sigma^{-1}}^2} + \sum_i \frac{\lambda_i}{2\pi\sqrt{v_i|\Sigma|}} e^{-\frac{1}{2}\|\phi-\phi_i\|_{P_i}^2 - \frac{1}{2}\|\phi-\mu\|_{\Sigma^{-1}}^2} \end{aligned}$$

Applying (13),

$$\begin{aligned} p(\phi)\lambda(\phi) &= \frac{\lambda_0}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}\|\phi-\mu\|_{\Sigma^{-1}}^2} \\ &\quad + \sum_i \lambda_i \left(\frac{1}{\sqrt{2\pi(v_i^{-1} + \Sigma)}} e^{-\frac{1}{2}\|\phi_i-\mu\|_{P_i K_i \Sigma^{-1}}^2} \right) \left(\frac{1}{\sqrt{2\pi \frac{v_i|\Sigma|}{v_i^{-1} + \Sigma}}} e^{-\frac{1}{2}\|\phi - K_i(P_i \phi_i + \Sigma^{-1}\mu)\|_{K_i^{-1}}^2} \right) \end{aligned} \quad (14)$$

707 where we define $K_i := (P_i + \Sigma^{-1})^{-1}$. For both PIPPET and PATIPPET, we
708 have

$$\|\phi_i - \mu\|_{P_i K_i \Sigma^{-1}}^2 = \|\phi_i - \mu\|_{(v_i^{-1} + \Sigma)^{-1}}^2$$

and $|K_i| = \frac{v_i|\Sigma|}{v_i^{-1} + \Sigma}$, so (14) can be written in terms of normal distributions:

$$p(\phi)\lambda(\phi) = \lambda_0 N(\phi|\mu, \Sigma) + \sum_i \lambda_i N(\phi_i|\mu, v_i^{-1} + \Sigma) N(\phi|K_i(P_i \phi_i + \Sigma^{-1}\mu), K_i) \quad (15)$$

709 Setting $\Lambda_0 := \lambda_0$, $\Lambda_i := \lambda_i N(\phi_i|\mu, v_i + \Sigma)$, and $\bar{\mu}_i := K_i(P_i \phi_i + \Sigma^{-1}\mu)$, we can
710 write

$$p(\phi)\lambda(\phi) = \Lambda_0 N(\phi|\mu, \Sigma) + \sum_i \Lambda_i N(\phi|\bar{\mu}_i, K_i)$$

We use this expression and the moments of normal distributions to calculate

the following expectations and define $\bar{\Lambda}$, $\bar{\mu}$, and $\bar{\Sigma}$:

$$\begin{aligned}\bar{\Lambda} &:= \mathbb{E}_p [\lambda(\phi)] = \sum_i \Lambda_i \\ \bar{\mu} &:= \frac{1}{\bar{\Lambda}} \mathbb{E}_p [\phi \lambda(\phi)] = \frac{\Lambda_0}{\bar{\Lambda}} \mu + \sum_i \frac{\Lambda_i}{\bar{\Lambda}} \bar{\mu}_i \\ \bar{\Sigma} &:= \frac{1}{\bar{\Lambda}} \mathbb{E}_p [(\phi - \mu)(\phi - \mu)^T \lambda(\phi)] = \frac{\Lambda_0}{\bar{\Lambda}} \Sigma + \sum_i \frac{\Lambda_i}{\bar{\Lambda}} (\mathbf{K}_i + (\bar{\mu}_i - \mu)(\bar{\mu}_i - \mu)^T)\end{aligned}\tag{16}$$

711 Substituting into (11) and (12), we have

$$d\mu = \int_{\phi} \phi \mathcal{L}[p_t(\phi)] d\phi dt + (\bar{\mu} - \mu) \cdot (dN_t - \bar{\Lambda} dt) \tag{17}$$

$$d\Sigma = \int_{\phi} (\phi - \mu)(\phi - \mu)^T \mathcal{L}[p_t(\phi|N_t)] d\phi dt \tag{18}$$

$$+ (\bar{\Sigma} - \Sigma) \cdot (dN_t - \bar{\Lambda} dt) \tag{19}$$

712 Calculating the moments of $\mathcal{L}[p_t(\phi)]$ for the PIPPET SDE (1), we derive
713 the PIPPET filter:

$$\begin{cases} d\mu = dt - (\bar{\mu} - \mu)(dN_t - \bar{\Lambda} dt) \\ d\Sigma = \sigma^2 dt - (\bar{\Sigma} - \Sigma)(dN_t - \bar{\Lambda} dt) \end{cases} \tag{20}$$

714 which is equivalent to equation (3) with its accompanying reset rule at events.

715 Similarly, calculating the moments for the PATIPPET SDE (4), we derive the

716 PATIPPET filter:

$$\begin{cases} d\boldsymbol{\mu} = \begin{pmatrix} \mu_\theta \\ 0 \end{pmatrix} dt - (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})(dN_t - \bar{\Lambda}dt) \\ d\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 + 2s_{12} & s_{22} \\ s_{22} & \sigma_\theta^2 \end{pmatrix} dt - (\bar{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})(dN_t - \bar{\Lambda}dt) \end{cases} \quad (21)$$

717 For multiple event streams j ,:

$$dp_t(\phi) = \mathcal{L}[p_t(\phi)]dt + p_t(\phi) \sum_j (\lambda_j(\phi) - \mathbb{E}_p[\lambda_j(\phi)]) \cdot (\mathbb{E}_p[\lambda_j(\phi)]^{-1} dN_j - dt) \quad (22)$$

718 This follows directly from application of the derivation above to equation
 719 (5) in [72] with a discrete spatial dimension. By the methods above, it yields
 720 the multi-PIPPET filter:

$$\begin{cases} d\mu = dt - \sum_j (\bar{\mu}^j - \mu)(dN_t^j - \bar{\Lambda}^j dt) \\ d\Sigma = \sigma^2 dt - \sum_j (\bar{\Sigma}^j - \Sigma)(dN_t^j - \bar{\Lambda}^j dt) \end{cases} \quad (23)$$

721 and the multi-PATIPPET filter:

$$\begin{cases} d\boldsymbol{\mu} = \begin{pmatrix} \mu_\theta \\ 0 \end{pmatrix} dt - \sum_j (\bar{\boldsymbol{\mu}}^j - \boldsymbol{\mu})(dN_t^j - \bar{\Lambda}^j dt) \\ d\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 + 2s_{12} & s_{22} \\ s_{22} & \sigma_\theta^2 \end{pmatrix} dt - \sum_j (\bar{\boldsymbol{\Sigma}}^j - \boldsymbol{\Sigma})(dN_t^j - \bar{\Lambda}^j dt) \end{cases} \quad (24)$$

722 6.2 Simulation parameters.

723 All code used to create figures in this manuscript is available at [https://](https://github.com/joncannon/PIPPET)
 724 github.com/joncannon/PIPPET.

725 PIPPET simulations were conducted by numerical simulation of (1) with
726 $dt = 0.001$ and initialized with $\mu_0 = 0$ and $\Sigma_0 = 0.0002$. Parameters for
727 the simulations shown in each figure are listed below, with t_i used to denote
728 simulated event times. (ϕ_i and t_i are given in units of seconds, and v_i is given
729 in units of s^2 .)

730 *Figure 1:* $\phi_i = t_i = \{0.5, 1, 1.5\}$, $v_i = 0.0001$, $\lambda_i = 0.02$, $\lambda_0 = 0.01$, $\sigma = 0.05$

731 *Figure 2A:* $\phi_i = t_i = \{0.25, 0.5, 0.75, 1\}$, $v_i = 0.0001$, $\lambda_i = 2$, $\lambda_0 = 0.01$,
732 $\sigma = 0.05$.

733 *Figure 2B:* Same as Figure 2A, but with $t_i = \{1\}$.

Figure 3A:

$$t_i = \{0, 0.150, 0.25, 0.5, 0.65, 0.9, 1\}$$

$$\phi_i = \{0, 0.15, 0.25, 0.4, 0.5, 0.65, 0.75, 0.9, 1, 1.15\}$$

$$v_i = \{.0001, .0003, .0001, .0003, .0001, .0003, .0001, .0003\}$$

$$\lambda_i = \{.02, .01, .02, .01, .02, .01, .02, .01\}$$

$$\lambda_0 = 0.01$$

$$\sigma = 0.05$$

734 *Figure 3B:* Same as Figure 3A, but with $t_i = \{0, 0.150, 0.25, 0.5, 0.61, 0.86, 0.96\}$.

Figure 4:

$$\begin{aligned}
t_i &= \{0, 0.15, .65, .9, 1.15, 1.25\} \\
\phi_i &= \{0, 0.15, 0.25, 0.4, 0.5, 0.65, 0.75, 0.9, 1, 1.15\} \\
v_i &= \{.0001, .001, .0001, .001, .0001, .001, .0001, .001\} \\
\lambda_i &= \{.05, .005, .05, .005, .05, .005, .05, .005\} \\
\lambda_0 &= 0.01 \\
\sigma &= 0.05
\end{aligned}$$

Figure 5: (No numerical simulation was performed for this figure.)

$$\begin{aligned}
\phi_i^j &= 0.25i \text{ for } j = \text{bass, snare, hihat} \\
v_i^{\text{bass}} &= .0001, v_i^{\text{snare}} = .0003, v_i^{\text{hihat}} = .001 \\
\lambda_i^{\text{bass}} &= \{.05, .005, .005, .005, \dots\} \\
\lambda_i^{\text{snare}} &= \{.005, .005, .05, .005, \dots\} \\
\lambda_i^{\text{hihat}} &= \{.05, .05, .05, .05, \dots\} \\
\lambda_0 &= 0.01
\end{aligned}$$

735 PATIPPET simulations were conducted by numerical simulation of (4) with
736 $dt = 0.001$. Parameters for the simulations shown in each figure are listed below.

Figure 6:

$$t_i = \frac{i}{1.15}$$

$$\phi_i = i$$

$$v_i = \{.0001, .0003, .0001, .0003, .0001, .0003, .0001, .0003\}$$

$$\lambda_i = \{.02, .01, .02, .01, .02, .01, .02, .01\}$$

$$\lambda_0 = 10^{-4}$$

$$\sigma = 0.05$$

$$\sigma_\theta = 0.05$$

$$\mu_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\Sigma_0 = \begin{pmatrix} .001 & 0 \\ 0 & .04 \end{pmatrix}$$

Figure 7: In four simulations, we set the inter-onset interval Δ to $0.4s$, $0.7s$,

1.0s, and 1.3s. In each simulation, we set the perturbation δ to $\frac{\Delta}{25}$.

$$t_i = \{\Delta, 2\Delta, 3\Delta, 4\Delta + \delta\}$$

$$\phi_i = i$$

$$v_i = 0.0002$$

$$\lambda_i = \{.02, .01, .02, .01, .02, .01, .02, .01\}$$

$$\lambda_0 = 10^{-5}$$

$$\sigma = 0.01$$

$$\sigma_\theta = 0.01$$

$$\mu_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\Sigma_0 = \begin{pmatrix} 10^{-4} & 0 \\ 0 & 10^{-4} \end{pmatrix}$$

References

1. Repp BH and Su YH. Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin and Review* 2013; 20:403–52. DOI: 10.3758/s13423-012-0371-2. arXiv: NIHMS150003
2. Merchant H, Grahn J, Trainor L, Rohrmeier M, and Fitch WT. Finding the beat: a neural perspective across humans and non-human primates. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2015; 370. DOI: 10.1098/rstb.2014.0093. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25646516>
3. Obleser J and Kayser C. Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences* 2019; 23:1–14. DOI:

- 748 10.1016/j.tics.2019.08.004. Available from: <https://doi.org/10.1016/j.tics.2019.08.004>
- 749
- 750 4. Nobre AC and Van Ede F. Anticipated moments: Temporal structure in
751 attention. *Nature Reviews Neuroscience* 2018; 19:34–48. DOI: 10.1038/
752 nrn.2017.141. Available from: <http://dx.doi.org/10.1038/nrn.2017.141>
- 753
- 754 5. Morillon B, Schroeder CE, Wyart V, and Arnal LH. Temporal prediction
755 in lieu of periodic stimulation. *Journal of Neuroscience* 2016; 36:2342–7.
756 DOI: 10.1523/JNEUROSCI.0836-15.2016
- 757 6. Lange K. Brain correlates of early auditory processing are attenuated by
758 expectations for time and pitch. *Brain and Cognition* 2009; 69:127–37.
759 DOI: 10.1016/j.bandc.2008.06.004. Available from: <http://dx.doi.org/10.1016/j.bandc.2008.06.004>
- 760
- 761 7. Jazayeri M and Shadlen MN. Temporal context calibrates interval timing.
762 *Nature Neuroscience* 2010; 13:1020–6. DOI: 10.1038/nn.2590
- 763 8. Herrmann B, Henry MJ, Haegens S, and Obleser J. Temporal expectations
764 and neural amplitude fluctuations in auditory cortex interactively influence
765 perception. *NeuroImage* 2016; 124:487–97. DOI: 10.1016/j.neuroimage.2015.09.019
- 766
- 767 9. Rajendran VG, Teki S, and Schnupp JW. Temporal Processing in Audi-
768 tion: Insights from Music. *Neuroscience* 2018; 389:4–18. DOI: 10.1016/
769 j.neuroscience.2017.10.041. Available from: <https://doi.org/10.1016/j.neuroscience.2017.10.041>
- 770
- 771 10. Large EW and Jones MR. The dynamics of attending: How people track
772 time-varying events. *Psychological Review* 1999; 106:119–59. DOI: 10.
773 1037//0033-295x.106.1.119

- 774 11. Large EW and Palmer C. Perceiving temporal regularity in music. *Cognitive Science* 2002; 26:1–37. DOI: 10.1016/S0364-0213(01)00057-X
- 775
- 776 12. Breska A and Deouell LY. Neural mechanisms of rhythm-based tempo-
777 ral prediction: Delta phase-locking reflects temporal predictability but not
778 rhythmic entrainment. *PLoS Biology* 2017; 15:1–30. DOI: 10.1371/journal.
779 pbio.2001665
- 780 13. Bouwer FL, Honing H, and Slagter HA. Beat-based and memory-based
781 temporal expectations in rhythm: similar perceptual effects, different un-
782 derlying mechanisms. 2019; 8:55
- 783 14. Rimmele JM, Morillon B, Poeppel D, and Arnal LH. Proactive Sensing of
784 Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*
785 2018; 22:870–82. DOI: 10.1016/j.tics.2018.08.003. Available from:
786 <https://doi.org/10.1016/j.tics.2018.08.003>
- 787 15. Friston K. A theory of cortical responses. *Philosophical Transactions of*
788 *the Royal Society B: Biological Sciences* 2005; 360:815–36. DOI: 10.1098/
789 rstb.2005.1622
- 790 16. Friston K. Does predictive coding have a future? *Nature Neuroscience*
791 2018; 21:1019–21. DOI: 10.1038/s41593-018-0200-7
- 792 17. Vuust P and Witek MA. Rhythmic complexity and predictive coding: A
793 novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology* 2014; 5:1–14. DOI: 10.3389/fpsyg.2014.01111
- 794
- 795 18. Vuust P, Dietz MJ, Witek M, and Kringelbach ML. Now you hear it: A
796 predictive coding model for understanding rhythmic incongruity. *Annals*
797 *of the New York Academy of Sciences* 2018; 1423:19–29. DOI: 10.1111/
798 nyas.13622

- 799 19. Proksch S, Comstock DC, Médé B, Pabst A, and Balasubramaniam R.
800 Motor and Predictive Processes in Auditory Beat and Rhythm Perception.
801 2020; 14. DOI: 10.3389/fnhum.2020.578546
- 802 20. Friston K, Stephan K, Li B, and Daunizeau J. Generalised filtering. Math-
803 ematical Problems in Engineering 2010; 2010. DOI: 10.1155/2010/621670
- 804 21. Buckley CL, Kim CS, McGregor S, and Seth AK. The free energy principle
805 for action and perception: A mathematical review. Journal of Mathemati-
806 cal Psychology 2017; 81:55–79. DOI: 10.1016/j.jmp.2017.09.004. arXiv:
807 1705.09156. Available from: [http://dx.doi.org/10.1016/j.jmp.2017.](http://dx.doi.org/10.1016/j.jmp.2017.09.004)
808 09.004
- 809 22. Schwartz M and Kotz SA. A dual-pathway neural architecture for spe-
810 cific temporal prediction. Neuroscience and Biobehavioral Reviews 2013;
811 37:2587–96. DOI: 10.1016/j.neubiorev.2013.08.005. Available from:
812 <http://dx.doi.org/10.1016/j.neubiorev.2013.08.005>
- 813 23. Egger SW and Jazayeri M. A nonlinear updating algorithm captures subop-
814 timal inference in the presence of signal-dependent noise. Scientific Reports
815 2018 ;18–20. DOI: 10.1038/s41598-018-30722-0
- 816 24. DI Luca M and Rhodes D. Optimal Perceived Timing: Integrating Sensory
817 Information with Dynamically Updated Expectations. Scientific Reports
818 2016; 6:1–15. DOI: 10.1038/srep28563
- 819 25. Elliott MT, Wing AM, and Welchman AE. Moving in time: Bayesian causal
820 inference explains movement coordination to auditory beats. Proceedings
821 of the Royal Society B: Biological Sciences 2014; 281. DOI: 10.1098/rspb.
822 2014.0751

- 823 26. Snyder DL. Filtering and Detection for Doubly Stochastic Poisson Pro-
824 cesses. *IEEE Transactions on Information Theory* 1972; 18:91–102. DOI:
825 10.1109/TIT.1972.1054756
- 826 27. Oppen M. A Bayesian Approach to On-line Learning. *On-Line Learning in*
827 *Neural Networks* 2010 :363–78. DOI: 10.1017/cbo9780511569920.017
- 828 28. Friston K. The free-energy principle: A unified brain theory? *Nature Re-*
829 *views Neuroscience* 2010; 11:127–38. DOI: 10.1038/nrn2787
- 830 29. Eden UT and Brown EN. CONTINUOUS-TIME FILTERS FOR STATE
831 ESTIMATION FROM POINT PROCESS MODELS OF NEURAL DATA.
832 *Statistica Sinica* 2008; 18:1293–310
- 833 30. Cemgil AT, Kappen B, Desain P, and Honing H. On tempo tracking:
834 Tempogram representation and Kalman filtering. *Journal of New Music*
835 *Research* 2000; 29:259–73. DOI: 10.1080/09298210008565462
- 836 31. London J, Polak R, and Jacoby N. Rhythm histograms and musical meter:
837 A corpus study of Malian percussion music. *Psychonomic Bulletin and*
838 *Review* 2017; 24:474–80. DOI: 10.3758/s13423-016-1093-7
- 839 32. Polak R, London J, and Jacoby N. Both isochronous and non-isochronous
840 metrical subdivision afford precise and stable ensemble entrainment: A
841 corpus study of malian jembe drumming. *Frontiers in Neuroscience* 2016;
842 10:1–11. DOI: 10.3389/fnins.2016.00285
- 843 33. Friberg A and Sundström A. Swing Ratios and Ensemble Timing in Jazz
844 Performance: Evidence for a Common Rhythmic Pattern. *Music Percep-*
845 *tion* 2002; 19:333–49. DOI: 10.1525/mp.2002.19.3.333
- 846 34. Fitch WT and Rosenfeld AJ. Perception and Production of Syncopated
847 Rhythms. *Music Perception* 2007; 25:43–58

- 848 35. Warren RM and Gregory RL. An Auditory Analogue of the Visual Re-
849 versible Figure. *The American Journal of Psychology* 1958; 71:612-3
- 850 36. HALL GS and JASTROW J. STUDIES OF RHYTHM. *Mind* 1886 Jan;
851 os-XI:55-62. DOI: 10.1093/mind/os-XI.41.55. eprint: [https://](https://academic.oup.com/mind/article-pdf/os-XI/41/55/9358438/os-XI\41\55.pdf)
852 [academic.oup.com/mind/article-pdf/os-XI/41/55/9358438/os-](https://academic.oup.com/mind/article-pdf/os-XI/41/55/9358438/os-XI\41\55.pdf)
853 [XI\41\55.pdf](https://academic.oup.com/mind/article-pdf/os-XI/41/55/9358438/os-XI\41\55.pdf). Available from: [https://doi.org/10.1093/mind/os-](https://doi.org/10.1093/mind/os-XI.41.55)
854 [XI.41.55](https://doi.org/10.1093/mind/os-XI.41.55)
- 855 37. Nakajima Y. A psychophysical investigation of divided time intervals shown
856 by sound bursts. *Journal of the Acoustical Society of Japan* 1979; 35:145-
857 51
- 858 38. Meumann E. Beiträge zur Psychologie des Zeitbewußtseins [contributions
859 to the psychology of time consciousness]. *Philosophische Studien* 1896;
860 12:128-254
- 861 39. Grimm K. der einfluß der Zeitform auf die Wahrnehmung der Zeitdauer
862 [the influence of time-form on the perception of duration]. *Zeitschrift für*
863 *Psychologie* 1934; 132:104-32
- 864 40. Repp BH and Bruttomesso M. A filled duration illusion in music: Effects
865 of metrical subdivision on the perception and production of beat tempo.
866 *Advances in Cognitive Psychology* 2009; 5:114-34. DOI: 10.2478/V10053-
867 008-0071-7
- 868 41. Repp B and Jendoubi H. Flexibility of temporal expectations for triple
869 subdivision of a beat. *Advances in Cognitive Psychology* 2009; 5:27-41.
870 DOI: 10.2478/v10053-008-0063-7
- 871 42. Repp BH. Tapping in synchrony with a perturbed metronome: The phase
872 correction response to small and large phase shifts as a function of tempo.

- Journal of Motor Behavior 2011; 43:213–27. DOI: 10.1080/00222895.2011.561377
43. Repp BH, Keller PE, and Jacoby N. Quantifying phase correction in sensorimotor synchronization: Empirical comparison of three paradigms. *Acta Psychologica* 2012; 139:281–90. DOI: 10.1016/j.actpsy.2011.11.002. Available from: <http://dx.doi.org/10.1016/j.actpsy.2011.11.002>
44. Witek MA, Clarke EF, Kringelbach ML, and Vuust P. Effects of Polyphonic Context, Instrumentation, and Metrical Location on Syncopation in Music. *Music Perception* 2014; 32:201–17
45. Hove MJ, Marie C, Bruce IC, and Trainor LJ. Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proceedings of the National Academy of Sciences of the United States of America* 2014; 111:10383–8. DOI: 10.1073/pnas.1402039111
46. Repp BH. Phase Correction , Phase Resetting , and Phase Shifts After Subliminal Timing Perturbations in Sensorimotor Synchronization. *Journal of Experimental Psychology: Human Perception and Performance* 2001; 27:600–21. DOI: 10.1037//0096-1523.27.3.600
47. Heggli OA, Cabral J, Konvalinka I, Vuust P, and Kringelbach ML. A Kuramoto model of self-other integration across interpersonal synchronization strategies. *PLoS Computational Biology* 2019; 15:1–17. DOI: 10.1371/journal.pcbi.1007422
48. Koban L, Ramamoorthy A, and Konvalinka I. Why do we fall into sync with others? Interpersonal synchronization and the brain’s optimization principle. *Social Neuroscience* 2019; 14:1–9

- 897 49. Wing AM and Kristofferson AB. Response delays and the timing of discrete
898 motor responses. *Perception & Psychophysics* 1973; 14:5–12. DOI: 10 .
899 3758/BF03198607
- 900 50. Mates J. A model of synchronization of motor acts to a stimulus sequence
901 - II. Stability analysis, error estimation and simulations. *Biological Cyber-*
902 *netics* 1994; 70:475–84. DOI: 10.1007/BF00203240
- 903 51. Fox C, Rezek I, and Roberts S. Drum ' N ' Bayes : on-Line Variational
904 Inference for Beat Tracking and Rhythm Recognition. *International Com-*
905 *puter Music Conference* 2007. DOI: 10.1016/j.chieco.2016.10.003
- 906 52. Pesek M, Leonardis A, and Marolt M. An Analysis of Rhythmic Pat-
907 terns with Unsupervised Learning. *Applied Sciences* 2019. DOI: 10.3390/
908 app10010178
- 909 53. Repp BH. Obligatory "expectations" of expressive timing induced by per-
910 ception of musical structure. *Psychological Research* 1998; 61:33–43. DOI:
911 10.1007/s004260050011
- 912 54. Repp BH. Compensation for subliminal timing perturbations in perceptual-
913 motor synchronization. *Psychological Research* 2000; 63:106–28. DOI: 10 .
914 1007/PL00008170
- 915 55. Schroeder CE and Lakatos P. Low-frequency neuronal oscillations as in-
916 struments of sensory selection. *Trends in neurosciences* 2009; 32. DOI:
917 10.1016/j.tins.2008.09.012.Low-frequency
- 918 56. Arnal LH and Giraud AL. Cortical oscillations and sensory predictions.
919 *Trends in Cognitive Sciences* 2012; 16:390–8. DOI: 10.1016/j.tics .
920 2012.05.003. Available from: [http://dx.doi.org/10.1016/j.tics.](http://dx.doi.org/10.1016/j.tics.2012.05.003)
921 2012.05.003

- 922 57. Arnal LH and Kleinschmidt AK. Entrained delta oscillations reflect the
923 subjective tracking of time. *Cerebral Cortex* 2017 :e1349583. DOI: 10 .
924 1093/cercor/bhu103
- 925 58. Gámez J, Mendoza G, Prado L, Betancourt A, and Merchant H. The am-
926 plitude in periodic neural state trajectories underlies the tempo of rhythmic
927 tapping. *PLoS biology* 2019; 17:e3000054
- 928 59. Tomassini A, Ruge D, Galea JM, Penny W, and Bestmann S. The Role
929 of Dopamine in Temporal Uncertainty. *Journal of Cognitive Neuroscience*
930 2016. DOI: 10 . 1162/jocn. arXiv: 1511 . 04103. Available from: [http :
931 //dx.doi.org/10.1162/jocn%7B%5C_%7Da%7B%5C_%7D00409%7B%5C_%7D5Cnhttp://www.mitpressjournals.org/doi/abs/10.1162/jocn%
932 %7D5Cnhttp://www.mitpressjournals.org/doi/abs/10.1162/jocn%
933 7B%5C_%7Da%7B%5C_%7D00409](http://dx.doi.org/10.1162/jocn%7B%5C_%7Da%7B%5C_%7D00409%7B%5C_%7D5Cnhttp://www.mitpressjournals.org/doi/abs/10.1162/jocn%7B%5C_%7Da%7B%5C_%7D00409)
- 934 60. Sarno S, De Lafuente V, Romo R, and Parga N. Dopamine reward predic-
935 tion error signal codes the temporal evaluation of a perceptual decision re-
936 port. *Proceedings of the National Academy of Sciences of the United States*
937 *of America* 2017; 114:E10494–E10503. DOI: 10.1073/pnas.1712479114
- 938 61. Herbst SK, Fiedler L, and Obleser J. Tracking temporal hazard in the hu-
939 man electroencephalogram using a forward encoding model. *eNeuro* 2018;
940 5:1–17. DOI: 10.1523/ENEURO.0017–18.2018
- 941 62. Tavano A, Schröger E, and Kotz SA. Beta power encodes contextual esti-
942 mates of temporal event probability in the human brain. *PLoS ONE* 2019;
943 14. DOI: 10.1371/journal.pone.0222420
- 944 63. Cannon J and Patel AD. How beat perception coopts motor neurophysiol-
945 ogy: a proposal. *bioRxiv* 2020. DOI: <https://doi.org/10.1101/805838>

- 946 64. Schubotz RI. Prediction of external events with our motor system: towards
947 a new framework. *Trends in Cognitive Sciences* 2007; 11:211–8. DOI: 10.
948 1016/j.tics.2007.02.006
- 949 65. Rauschecker JP. An expanded role for the dorsal auditory pathway in
950 sensorimotor control and integration. *Hearing Research* 2011; 271:16–25.
951 DOI: 10.1016/j.heares.2010.09.001. Available from: [http://dx.doi.](http://dx.doi.org/10.1016/j.heares.2010.09.001)
952 [org/10.1016/j.heares.2010.09.001](http://dx.doi.org/10.1016/j.heares.2010.09.001)
- 953 66. Kneissler J, Drugowitsch J, Friston K, and Butz MV. Simultaneous learn-
954 ing and filtering without delusions: A bayes-optimal combination of pre-
955 dictive inference and adaptive filtering. *Frontiers in Computational Neu-*
956 *roscience* 2015; 9:1–12. DOI: 10.3389/fncom.2015.00047
- 957 67. Danielsen A. Here, There, and Everywhere: three accounts of pulse in
958 D’Angelo’s ‘Left and Right’. 2010 Jan :19–36. DOI: 10.4324/9781315596983-
959 2
- 960 68. Weij B van der, Pearce MT, and Honing H. A probabilistic model of meter
961 perception: Simulating enculturation. *Frontiers in Psychology* 2017; 8:1–
962 18. DOI: 10.3389/fpsyg.2017.00824
- 963 69. Rohrmeier M. Towards a formalization of musical rhythm. *Proc. of the*
964 *21st Int. Society for Music Information Retrieval Conf.* 2020
- 965 70. Pearce MT. The construction and evaluation of statistical models of melodic
966 structure in music perception and composition. PhD thesis. City Univer-
967 sity, London, 2005
- 968 71. Harel Y, Meir R, and Oppel M. A tractable approximation to optimal
969 point process filtering: Application to neural encoding. *Advances in Neural*
970 *Information Processing Systems* 2015; 2015-Janua:1603–11

- 971 72. Snyder DL and Fishman P. How to track a swarm of fireflies by observing
972 their flashes. IEEE Transactions on Information Theory 1975; 21