Expectancy-based rhythmic entrainment as continuous Bayesian inference

Jonathan Cannon

March 5, 2021

Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA, USA
Tel.: +314-749-6902
jcan@mit.edu

Abstract

When presented with complex rhythmic auditory stimuli, humans are able to track underlying temporal structure (e.g., a "beat"), both covertly and with their movements. This capacity goes far beyond that of a simple entrained oscillator, drawing on contextual and enculturated timing expectations and adjusting rapidly to perturbations in event timing, phase, and tempo. Previous modeling work has described how entrainment to rhythms may be shaped by event timing expectations, but sheds little light on any underlying computational principles that could unify the phenomenon of expectation-based entrainment with other brain processes. Inspired by the predictive processing framework, we propose that the problem of rhythm tracking is naturally characterized as a problem of continuously estimating an underlying phase and tempo based on precise event times and their correspondence to timing expectations. We present two inference problems formalizing this insight: PIPPET (Phase Inference from Point Process Event Timing) and PATIPPET (Phase and Tempo Inference). Variational solutions to these inference problems resemble previous "Dynamic Attending" models of perceptual entrainment, but introduce new terms representing the dynamics of uncertainty and the influence of expectations in the absence of sensory events. These terms allow us to model multiple characteristics of covert and motor human rhythm tracking not addressed by other models, including sensitivity of error corrections to inter-event interval and perceived tempo changes induced by event omissions. We show that positing these novel influences in human entrainment yields a range of testable behavioral predictions. Guided by recent neurophysiological observations, we attempt to align the phase inference framework with a specific brain implementation. We also explore the potential of this normative framework to guide the interpretation of experimental data and serve as building blocks for even richer predictive processing and active inference models of 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." Expectations for event timing can be used to achieve a range of behavioral goals. They can help us hone our sensory detection, our sensory discrimination, and our response time for behaviorally important stimuli at the anticipated time [4, 5, 6]. In some situations, temporal expectations attenuate neural responses [7], which may help to conserve neural resources. And timing expectations bias our perception of time, allowing us to use prior experience to supplement noisy sensory data as we make temporal judgments [8].

Entrainment in humans involves an interplay of stimulus and temporal expectation [9]. Nowhere is this clearer than in interaction with music, humankind's playground for auditory temporal expectation and entrainment [10]. But the precise nature of this interplay is an open question. The framework of Dynamic Attending Theory characterizes temporal expectancy as pulses of "attentional energy" issued by entrained neural oscillators, and mathematical models based on these ideas describe bidirectional interactions between temporal expectation and entrainment that reproduce aspects of human behavior and perception [11, 12]. But although the behavior of these models may be satisfying in certain applications, the groundwork underlying them is less so: key high-level concepts like the "attentional pulse" are difficult to define mechanistically or computationally, so the implementations of these concepts in models remain impressionistic.

An alternative approach to modeling the role of expectations in the brain is the "predictive processing" framework [13]. This framework posits that the brain engages in a continuous process of inferring the hidden causes of sensory events based on a learned understanding of how those causes produce sensation. Unlike the terms in Dynamic Attending Theory models, the terms in predictive processing models are directly linked to the formal inference problem being solved: the solution to the problem demands that certain quantities be computed, giving us reason to expect to find those quantities represented in the brain. In particular, "precision" or certainty plays a key role, determining how new sensory information is weighted relative to existing beliefs about the hidden causes.

Here, we apply the predictive processing approach to the process of expectancy-based entrainment by formalizing it as an inference problem: namely, the problem of inferring the state of the exogenous process giving rise to a series of events in time. We use the mathematical tool of point processes to formulate a model of precise event timing. We derive an optimal solution to the inference problem, which we hypothesize corresponds with the brain's mechanisms for entrainment. The resulting models resemble Dynamic Attending Theory models, but introduce two key novel elements:

1. Dynamically estimated phase uncertainty moderates the balance between top-down and bottom-up influences on estimated phase.
2. Event expectations influence estimated phase *even in the absence of actual events.*

These elements allow them to reproduce aspects of human entrainment unaddressed by existing models, including:

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

They are also significantly more flexible than Dynamic Attending Theory models in their descriptive power, allowing us to describe entrainment based on either periodic or aperiodic expectation patterns, and, as predictive processing models, they recast entrainment in a formal language that links it a the wide range of other cognitive phenomena.

In the next section, we formulate three versions of the problem of expectancy-based entrainment that are amenable to precise solutions, which we refer to collectively as the "phase inference framework." In the first, "Phase Inference from Point Process Event Timing" (PIPPET), a hidden phase variable advances steadily with added noise, and the observer is tasked with continuously inferring the phase based on the observation of events emitted probabilistically at certain phases with certain degrees of precision. In the second version, "Phase And Tempo Inference from Point Process Event Timing" (PATIPPET), the rate of phase advance (tempo) is also a dynamic variable with drift, and the solution simultaneously estimates phase, tempo, and certainty about both. The third version (mPIPPET) generalizes the first two to incorporate the observation of multiple types of events, each with distinct characteristic phases and precisions, into the inference process. We present variational filtering equations that approximate perfect Bayesian solutions to these problems.

In the Results section, we simulate these filters, drawing on music as a rich source of intuitive examples of entrainment informed by expectation. In doing so, we provide intuition into the range of behaviors of these solutions, and show how novel features introduced by the normative framework reproduce key aspects of human entrainment behavior that are not explained by other models. In the Discussion, we discuss the potential contributions of PIPPET and PATIPPET to the analysis of experimental data, to richer and more detailed models, and to our understanding of entrainment in the brain.

2 Mathematical framework

Predictive processing should be a natural modeling framework for understanding rhythmic expectation and entrainment [14, 15, 16]. However, 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 [17, 18]. In other words, they describe predictions about "what" rather than "when." They are therefore ill-suited to characterizing moment-by-moment errors in *timing* prediction, which are made sporadically and separated by intervals mostly 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 [19, 6, 10]. Bayesian methods have been applied to describe inferences about timing in the brain [20, 21, 22], 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 by first creating a generative model describing the probabilistic generation of precisely timed events and then variationally inverting that model. To model event generation, we use the mathematical tool of point processes.

2.1 Phase Inference from Point Process Event Timing (PIPPET)

PIPPET is the problem of dynamically estimating a hidden noisy phase variable based on the timing of events generated as a point process whose rate is modulated as a specific function of phase. The generative model consists of a phase $\phi \in \mathbb{R}$ that advances as a drift-diffusion process:

$$d\phi = dt + \sigma dW_t \tag{1}$$

and an inhomogeneous point process that generates events with probability $\lambda(\phi)$. This function is known to the observer. We will refer to $\lambda(\phi)$ as an "expectation template" because it describes the temporal structure of the observer's event expectations, though it can also be understood as a hazard rate for events. To achieve both analytical tractability and flexible descriptive power, we assume that $\lambda(\phi)$ is a sum of a constant λ_0 and a set of scaled Gaussian peaks indexed by $i = 1, 2, \ldots$ etc. Each Gaussian peak i is centered at a mean phase ϕ_i with variance v_i and scale λ_i:

$$\lambda(\phi) = \lambda_0 + \sum_i \lambda_i \varphi(\phi|\phi_i, v_i) \tag{2}$$

where $\varphi(\cdot|m, v)$ denotes the pdf of a Gaussian distribution with mean m and variance v.

- Each Gaussian mean ϕ_i represents a phase at which an event is expected;
- λ_i represents the strength of that expectation;
- and v_i^{-1} is the temporal precision of that expectation.
- $\lambda_0 > 0$ represents the rate of events being generated as part of a uniform noise background unrelated to phase.

The point process with rate described by (2) can be understood as a sum of independent point processes i, one for each expectation peak and one for the uniform background process with rate λ_0, whose events are indistinguishable. The mathematics of updating a phase estimate at an event can be understood to involve a causal inference on which of these processes caused each event.

$\lambda(\phi)dt$ is the likelihood function over ϕ associated with the occurrence of an event, so $\lambda(\phi)$ is a rescaled likelihood function. See Figure 1A for illustration.

Note that ϕ is assumed to be on the real line, not the circle. This design decision allows PIPPET to entrain to temporally patterned expectations with or without periodic structure by choosing a periodic or aperiodic expectation template λ.

Given a series of event times $\{t_n\}$ tallied by an event-counting function $N_t : \mathbb{R} \to \mathbb{Z}^{0+}$, an expectation template $\lambda(\phi)$, and a prior distribution $p_0(\phi)$ describing the distribution of phase at time $t = 0$, the observer's goal is to infer a posterior distribution $p_t(\phi) = p(\phi|N_{\tau<t})$ describing an estimate of phase ϕ at any time t based on the event history up to t.

In [23], Snyder derives an exact PDE for the evolution of this posterior distribution over time. Following the predictive processing ansatz of maintaining Gaussian posterior distributions (the Laplace assumption), which provides both computational tractability and neurophysiological plausibility by reducing the representation of the posterior to a mean and a variance, we project the posterior onto a Gaussian at each dt time-step. We do this by moment-matching: we use Snyder's solution to determine the evolution of the mean and variance of the posterior, and then replace the true posterior with a Gaussian with the same mean and variance. This choice of Gaussian is the choice with minimum KL divergence from the true posterior [24], and therefore also minimizes the free energy of the solution within the family of possible Gaussian posteriors in accordance with the Free Energy Principle [25].

The result of this derivation is a generalization of a Kalman-Bucy filter with Poisson observation noise. Eden and Brown [26] have derived an explicit form for this filter, but it relies on a local approximation of the rate function λ that hides some of the interesting effects of events expected at nearby time points. For λ a mixture of Gaussians, we derive a filter directly from Snyder's solution in [23] that more accurately approximates the optimal (Bayesian) solution. The derivation is presented in Appendix 6.2.

Solution: the PIPPET filter At any time t, let μ_t denote the mean and V_t denote the variance of the Gaussian posterior. At each event time t, we let μ_t and V_t equal the left-hand limits of μ and V before the event, and we write μ_{t+} and V_{t+} to denote their right-hand limit values after the event (μ and V are left-continuous). Let dN_t denote the increment in the event-counting process at time t, which is either 0 or 1 with probability one. μ_t and V_t evolve according to the stochastic differential equation:

$$\begin{cases} d\mu = & dt + (\hat{\mu} - \mu)(dN_t - \Lambda dt) \\ dV = & \sigma^2 dt + (\hat{V} - V)(dN_t - \Lambda dt) \end{cases} \tag{3}$$

or, equivalently, they evolve between events according to the ODE: $\begin{cases} \dot{\mu} = & 1 - \Lambda(\hat{\mu} - \mu) \\ \dot{V} = & \sigma^2 - \Lambda(\hat{V} - V) \end{cases}$ and reset at each event to $\mu_{t+} = \hat{\mu}$ and $V_{t+} = \hat{V}$, where we define

$$\hat{\mu} := \frac{\lambda_0}{\Lambda}\mu_t + \sum_{i=1,\cdots} \frac{\Lambda_i}{\Lambda}\hat{\mu}_i$$

$$\hat{V} := \frac{\lambda_0}{\Lambda}\left(V_t + (\mu_t - \mu_{t+})^2\right) + \sum_{i=1,\cdots} \frac{\Lambda_i}{\Lambda}\left(\hat{V}_i + (\hat{\mu}_i - \mu_{t+})^2\right)$$

(Note that in this formulation, μ_{t+} must be calculated before V_{t+}.)

$$\hat{\mu}_i := \frac{V_t^{-1}\mu_t + v_i^{-1}\phi_i}{V_t^{-1} + v_i^{-1}} \qquad \text{and} \qquad \hat{V}_i := \frac{1}{V_t^{-1} + v_i^{-1}}$$

$$\Lambda_i := \lambda_i \varphi(\mu_t | \phi_i, v_i + V_t) \qquad \text{and} \qquad \Lambda := \sum_i \Lambda_i$$

These terms are illustrated in Figure 1. Intuitively,

- Λ (implicitly a function of μ_t and V_t) is the degree to which an event is anticipated at t while taking into account uncertainty about underlying phase, also known as the "subjective hazard rate". Λ_i is the degree to which an event is anticipated from peak i (the "conditional subjective hazard rate").

- At each event time t, $\lambda(\phi)$ serves as a (rescaled) likelihood function for phase, and the role of prior is played by the phase distribution p_t, a Gaussian with mean μ_t and variance V_t. Each peak i of λ is a possible "cause" of the event, as is the background event rate λ_0. Each peak is associated with a "candidate posterior" with mean $\hat{\mu}_i$ and variance $\hat{V}_i$ – this would be the posterior on phase if the event were known to be caused by peak i. $\hat{\mu}_i$ is a weighted sum of the current mean estimated phase μ_t and the center ϕ_i of expectation peak i, weighted by their respective precisions. Note that, following the predictive processing ansatz, this is the phase that minimizes precision-weighted prediction error with respect to predicted event timing and predicted phase.

- At an event, the phase distribution resets to a Gaussian with mean $\hat{\mu}$ and variance $\hat{V}$. These are weighted sums of the influences of each candidate posterior, each weighted by conditional subjective hazard rate Λ_i. The expression for $\hat{V}$ contains additional terms $(\hat{\mu}_i - \mu_{t+})^2$ and $(\mu_t - \mu_{t+})^2$, which cause the variance of the posterior to increase if the cause of the event is ambiguous.

- The background rate λ_0 acts as an alternative possible cause for any event. It serves to weight the posterior phase distribution toward the prior distribution before the event, and gives rise to causal ambiguity for any event and a resulting increase in posterior variance.

- Between events, each dt time step is taken as a Bayesian inference with likelihood $1 - \lambda(\phi)dt$ and with a Gaussian prior consisting of the posterior of the previous time step carried forward by dt according to the Fokker-Planck evolution associated with equation (1). This prior causes μ_t to increase steadily and V_t to grow at rate σ^2. The likelihood pushes μ and V away from $\hat{\mu}$ and $\hat{V}$ with a strength proportionate to subjective hazard rate Λ. Thus, the absence of an event continuously pushes the posterior in the opposite direction as would the occurrence of an event.

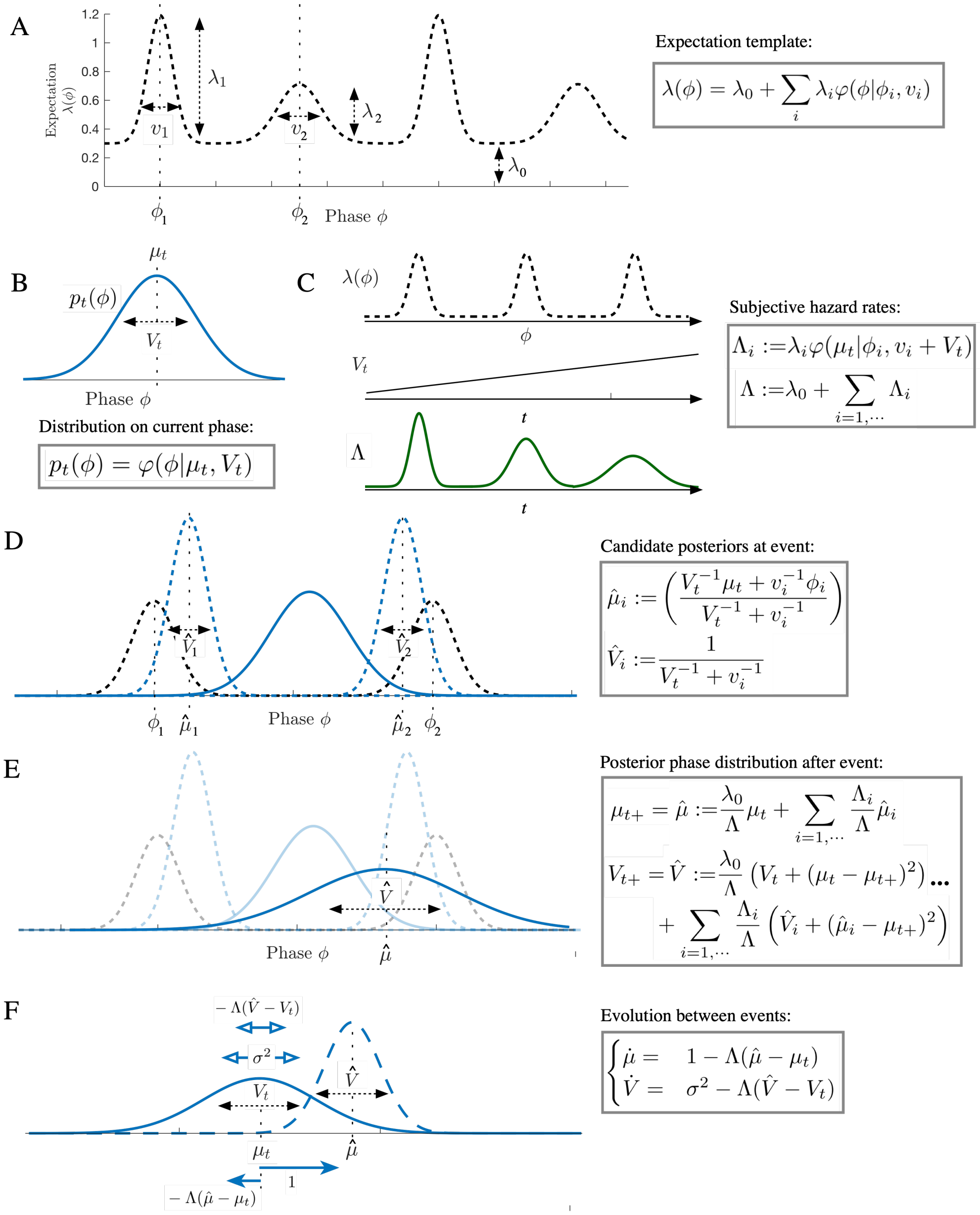
A
Expectation $\lambda(\phi)$
1.2
1
0.8
0.6
0.4
0.2
0
λ_1
v_1
v_2
λ_2
λ_0
ϕ_1
ϕ_2
Phase ϕ
Expectation template:
$\lambda(\phi) = \lambda_0 + \sum_i \lambda_i \varphi(\phi|\phi_i, v_i)$
B
μ_t
$p_t(\phi)$
V_t
Phase ϕ
Distribution on current phase:
$p_t(\phi) = \varphi(\phi|\mu_t, V_t)$
C
$\lambda(\phi)$
ϕ
V_t
t
Λ
t
Subjective hazard rates:
$\Lambda_i := \lambda_i \varphi(\mu_t|\phi_i, v_i + V_t)$
$\Lambda := \lambda_0 + \sum_{i=1,\cdots} \Lambda_i$
D
$\hat{V}_1$
$\hat{V}_2$
ϕ_1
$\hat{\mu}_1$
Phase ϕ
$\hat{\mu}_2$
ϕ_2
Candidate posteriors at event:
$\hat{\mu}_i := \left(\frac{V_t^{-1}\mu_t + v_i^{-1}\phi_i}{V_t^{-1} + v_i^{-1}}\right)$
$\hat{V}_i := \frac{1}{V_t^{-1} + v_i^{-1}}$
E
$\hat{V}$
Phase ϕ
$\hat{\mu}$
Posterior phase distribution after event:
$\mu_{t+} = \hat{\mu} := \frac{\lambda_0}{\Lambda}\mu_t + \sum_{i=1,\cdots} \frac{\Lambda_i}{\Lambda}\hat{\mu}_i$
$V_{t+} = \hat{V} := \frac{\lambda_0}{\Lambda}\left(V_t + (\mu_t - \mu_{t+})^2\right)$...
$+ \sum_{i=1,\cdots} \frac{\Lambda_i}{\Lambda}\left(\hat{V}_i + (\hat{\mu}_i - \mu_{t+})^2\right)$
F
$-\Lambda(\hat{V} - V_t)$
σ^2
V_t
$\hat{V}$
μ_t
$\hat{\mu}$
1
$-\Lambda(\hat{\mu} - \mu_t)$
Evolution between events:
$\begin{cases} \dot{\mu} = & 1 - \Lambda(\hat{\mu} - \mu_t) \\ \dot{V} = & \sigma^2 - \Lambda(\hat{V} - V_t) \end{cases}$

Figure 1: **Illustration of the PIPPET filter.** A) In the PIPPET 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 of the expectations, 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 phase. B) At any time t, the filter's estimate of current phase $p_t(\phi)$ is forced to be a Gaussian with mean μ_t (the estimated phase at time t) and variance V_t (the level of uncertainty about the phase estimate). D) These allow us to define a subjective hazard rate Λ (implicitly a function of time) representing the degree to which an event is anticipated at t, and conditional subjective hazard rates Λ_i representing the degree to which an event is anticipated from peak i. These hazard rates become less precise as phase uncertainty V_t increases. D) Each peak i of λ is associated with a "candidate posterior" with mean $\hat{\mu}_i$ and variance $\hat{V}_i$ – this would be the posterior on phase if the event were known to come from peak i. E) At an event, the phase distribution resets to a Gaussian with mean $\hat{\mu}$ and variance $\hat{V}$. These incorporate the influences of each candidate posterior, and $\hat{V}$ can increase if the cause of the event is ambiguous (as dramatically illustrated above). F) Between events, μ_t increases at rate 1 and V_t grows at rate σ^2. Additionally, μ and V are pushed away from $\hat{\mu}$ and $\hat{V}$ with a strength proportionate to subjective hazard rate Λ.

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

PATIPPET 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 $\mathbf{x} = \begin{pmatrix} \phi \\ \theta \end{pmatrix}$, where ϕ is the phase as above, θ is the rate of phase advancement (or tempo), and σ and σ_θ are the levels of phase and tempo noise, respectively:

$$d\mathbf{x} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} dt + \begin{pmatrix} \sigma dW_t \\ \sigma_\theta dW_t^\theta \end{pmatrix} \tag{4}$$

As above, an inhomogeneous point process generates events with probability based on an expectation template λ, which in this case is a function of both phase ϕ and tempo θ. In this formulation, we want events to occur with a certain probability in each $d\phi$ phase bin regardless of tempo, which we can accomplish by scaling the event rate by θ:

$$\lambda(\phi, \theta) = \theta \left(\lambda_0 + \sum_i \lambda_i \varphi(\phi | \phi_i, v_i) \right) \tag{5}$$

Note that this is the same as the PIPPET expression for event rate if we set $\theta = 1$.

As before, the observer's goal is to infer a posterior distribution at any time t using preceding event times; now the distribution $p_t(\mathbf{x})$ describes an estimate of both phase and tempo. 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 $\boldsymbol{\mu}_t$ and a phase/tempo covariance matrix $\mathbf{V}_t$. The solution is presented in 6.1 and its derivation is presented in 6.2.

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 σ_θ is set to zero. At each event, the distribution of phase and tempo is discontinuously updated to a 2D Gaussian posterior, which evolves continuously between events. This scheme is similar to [27], which estimates phase and tempo by updating a 2D Gaussian posterior, but is updated in continuous time and is significantly more flexible in its capacity to track phase based on arbitrary expectation templates.

2.3 PIPPET with multiple event streams (mPIPPET)

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

$$d\phi = dt + \sigma dW_t \tag{6}$$

$$\lambda^j(\phi) = \lambda_0^j + \sum_i \lambda_i^j \varphi(\phi|\phi_i^j, v_i^j) \tag{7}$$

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

$$\begin{cases} \dot{\mu} = & 1 - \sum_j \Lambda^j(\hat{\mu}^j - \mu) \\ \dot{V} = & \sigma^2 - \sum_j \Lambda^j(\hat{V}^j - V) \end{cases} \tag{8}$$

and resetting to $\mu_{t+} = \hat{\mu}^j$ and $V_{t+} = \hat{V}^j$ when an event occurs in stream j, where we define Λ^j, $\hat{\mu}^j$, and $\hat{V}^j$ as we defined Λ, $\hat{\mu}$, and $\hat{V}$ above but in reference only to event stream j.

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

3 Results

In this section we conduct a series of simulations to illustrate how the novel terms representing dynamic tracking of uncertainty and the influence of expectations in the absence of events allow the PIPPET and PATIPPET filters to reproduce perceptual and behavioral observations during human entrainment to auditory rhythms. Parameters for these simulations are listed in Appendix 6.3.

3.1 Updating posterior in response to events

We simulated the PIPPET filter with a single expectation peak and varied parameters to illustrate its basic behavior (Figure 2). Figure Figure 2, column i illustrates the effect of an event on the phase estimate as a function of initial estimated phase μ_t. Events occurring when μ_t is near an expected event phase ϕ_1 caused μ to shift linearly toward ϕ_1. When we set the uniform rate of background events $\lambda_0 > 0$, events occurring far from the expected event phase ϕ_1 were attributed to the background and therefore caused negligible adjustment to the phase estimate. Phase uncertainty V_t decreased at events except when λ_0 was positive and μ was not sufficiently close to ϕ_1; in this case, V_t increased due to causal ambiguity, or stayed the same if the cause was unambiguously the uniform background source.

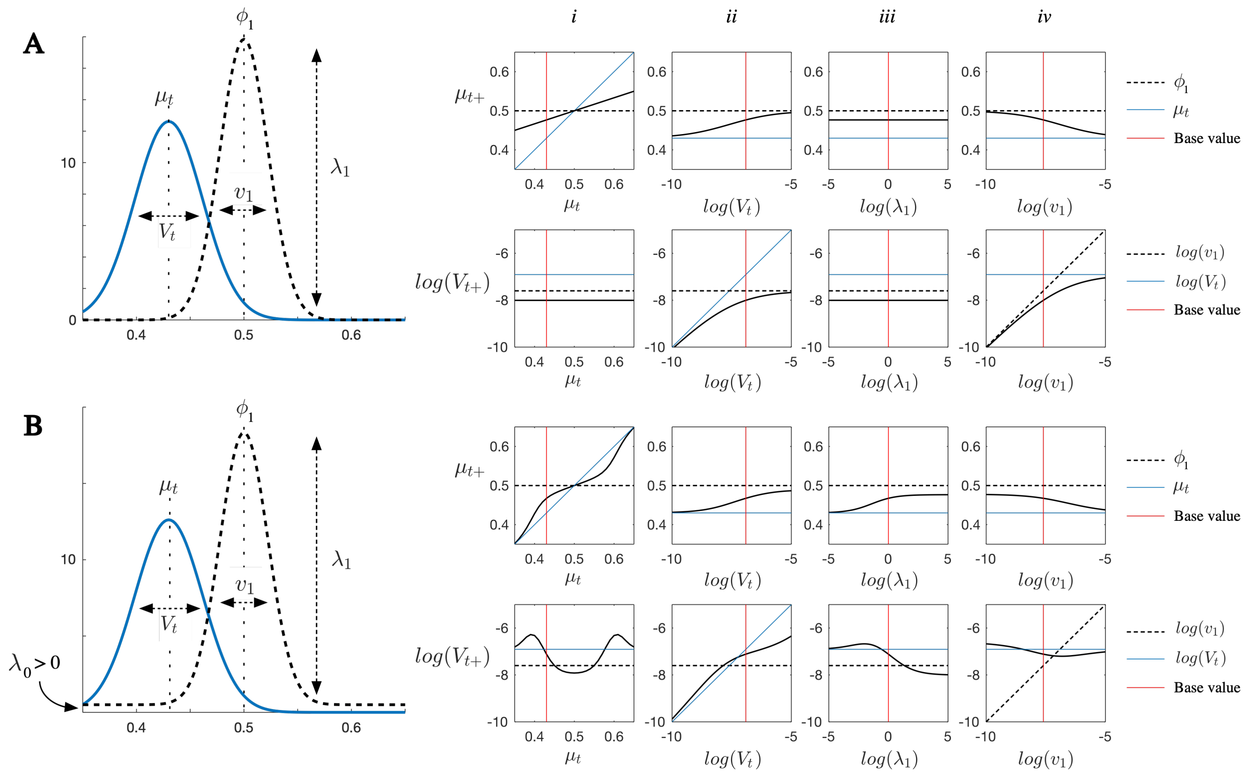

Figure 2: **Characterizing PIPPET's behavior at events.** A) An event is expected at phase $\phi_1 = 0.5$ with variance v_1 and expectation strength λ_1. The expected background event rate is set to $\lambda_0 = 0$. An event occurs when the phase estimate is at μ_t with uncertainty V_t. Panels in columns *i*-*iv* show the resulting mean μ_{t+} and variance V_{t+} of the posterior on phase as the parameters μ_t, V_t, λ_1, and v_1 are varied. *i*) μ is corrected linearly toward ϕ_1, while V decreases uniformly regardless of initial phase. *ii*) Corrections to μ are more thorough when V_t is large. *iii*) These corrections do not depend on λ_1. *iv*) These corrections are more thorough for smaller v_1. B) The same simulations are carried out with background event rate $\lambda_0 = 0.5$. *i*) If μ_t is close to ϕ_1, it is linearly corrected toward ϕ_1 and V_t decreases; if it is far, no correction is made. In the liminal zone, V_t increases due to the ambiguity of whether the event was related to the expectation peak or due to the background source. *ii*) V_{t+} is larger due to the effect of ambiguity as to whether the event is associated with ϕ_1 or with the background rate. *iii*) Now the correction depends on λ_1: stronger expectations make this peak the favored cause relative to the background source. *iv*) Note that if the expectation peak is extremely narrow, V_{t+} may still be large after the event and μ_t may not fully reset to ϕ_1 due to the aforementioned causal ambiguity.

3.2 Tracking complex rhythms with uneven subdivision

The PIPPET framework describes entrainment to rhythms in which each expected event phase may or may not be populated by an event. It is formulated in sufficient generality to describe entrainment to rhythms based on timing expectations with complex, non-isochronous stress patterns [28] and with non-integer duration ratios using suitably constructed (presumably learned) expectation templates $\lambda(\phi)$. Such rhythmic patterns have been shown to support highly precise synchronization in musicians with appropriate training and enculturated expectations [29], and should therefore be accounted for by models of human entrainment.

As an example of entrainment to a complex rhythm based on a temporal structure with non-integer duration ratios, we simulated entrainment to a swing rhythm. The rhythm is based on an underlying grid of "swung" eighth notes, where the first event of every pair is followed by a slightly longer inter-event duration than the second. Though the "swing" feel is often caricatured using eighth note pairs with a 2:1 duration ratio, this value has been shown to vary by with style and tempo and is certainly not limited to small integer ratios [30]. We used an expectation template with a swing ratio of 3:2 (though the exact ratio is not important) and associated the first eighth note in each pair with a stronger expectation than the second. The PIPET filter entrained to a complex, syncopated rhythm based on this template, drawing on the timing of both strongly and weakly expected events (Figure 3A). It corrected its phase estimate when an event timing shift or a phase shift was introduced into the rhythm (Figure 3B and 3C).

3.3 Failure mode: too much syncopation

The phase inference framework can account for human failures to track perfectly timed rhythms, i.e., rhythms in which every event falls at a peak of the expectation template. 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 [31, 32].

In PIPPET, these failures consist of inferring the presence of phase noise where none actually occurred. Such behavior is a necessary consequence of Bayesian optimality: a given stimulus may be generated by different combinations of phase noise and point process event generation noise, and the inference process is concerned only with the most likely explanation for the stimulus, which may include phase noise even if the stimulus was actually generated without it.

Using the expectation template with a swing grid as in the previous section, we simulated a strongly syncopated rhythm (Figure 4A). The rhythm's phase was not tracked successfully due to a convergence of two

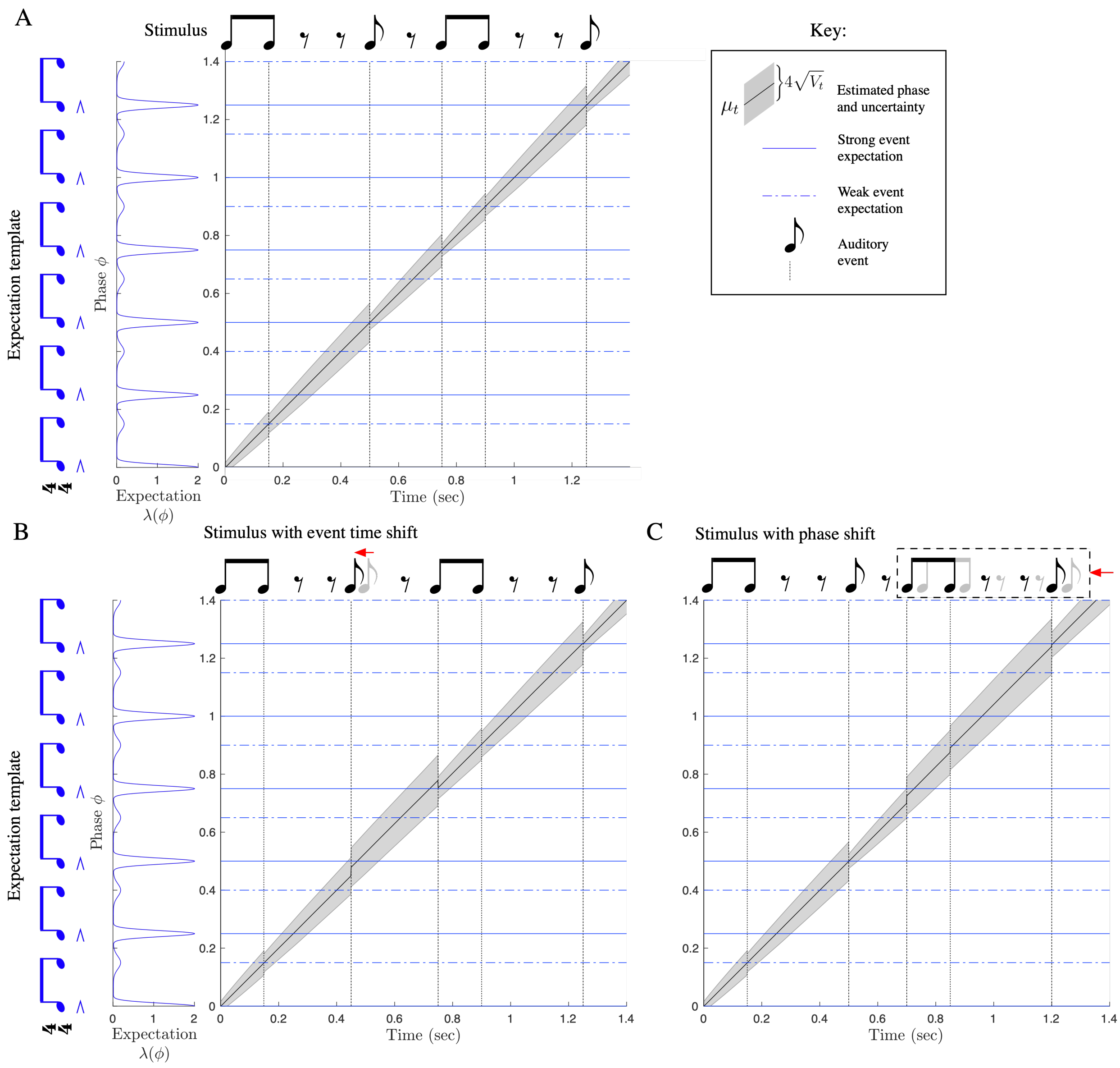

Figure 3: **Tracking phase through swung rhythms.** PIPPET is given a pattern of expectations representing "swung" eighth notes, with alternating longer and shorter inter-event durations and stronger, more precise expectations on the first of every pair. Dotted lines correspond to weaker expectations and solid lines correspond to stronger expectations. A) Phase is successfully tracked over the course of a rhythmic stimulus, with phase uncertainty growing between events and contracting at events. B) One event in the rhythm is shifted earlier in time. Estimated phase μ_t adjusts partially to compensate for the timing shift, and then adjusts back at the subsequent event. Uncertainty V_t is not as effectively reined in by these unpredictably-timed events, but decreases as later events corroborate the corrected phase estimate. C) A phase shift is introduced into the rhythm, moving all subsequent events earlier in time. When the first early event arrives, uncertainty increases. Estimated phase is corrected over the first few events after the shift, and V_t decreases most substantially when the estimate μ_t is corroborated by a strongly expected event happening at the appropriate estimated phase.

factors: the disproportionate influence of the higher peaks of the expectation template, and the accumulation of phase uncertainty V_t. Phase uncertainty was only slightly reduced by events occurring at weakly expected phases, so it accumulated over the course of the rhythm, and especially during the long silence. Once V_t was large, indicating the possibility of substantial phase noise having accumulated, the higher expectation peaks ϕ_i became the most likely explanations for events that were actually perfectly timed to coincide with nearby lower peaks – since precise event timing was no longer a reliable indicator of the source of an event, local peak height became the best indicator, and higher peaks won out. Thus, at each event, the estimated phase was adjusted to better align the higher peaks with the events.

The same rhythm could be successfully tracked in two alternate conditions. First, it was successfully tracked when we decreasing the rate of accumulation of phase uncertainty σ^2 (Figure 4B), demonstrating the key role of uncertainty in making the system susceptible to the disruptive effect of syncopation. Second, it was successfully tracked when an additional stream of sensory input was added by simulating an isochronous finger tap (Figure 4C). We used mPIPPET to create a second expectation template for tapping. As phase tracking was simulated, we planned new tap events just before μ reached expected tap phases by extrapolating μ forward. When taps occurred, phase uncertainty decreased, reducing the disruptive effects of syncopation. Note that planning actions specifically to fulfill sensory expectations and using this sensory feedback to inform inference about the outside world is an example of "active inference", the principle framework for understanding action in the literature on predictive processing [25].

3.4 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 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 5).

In addition to its value as a model of human rhythmic cognition, the PATIPPET 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.

3.5 Period-dependent corrections

In sensorimotor 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

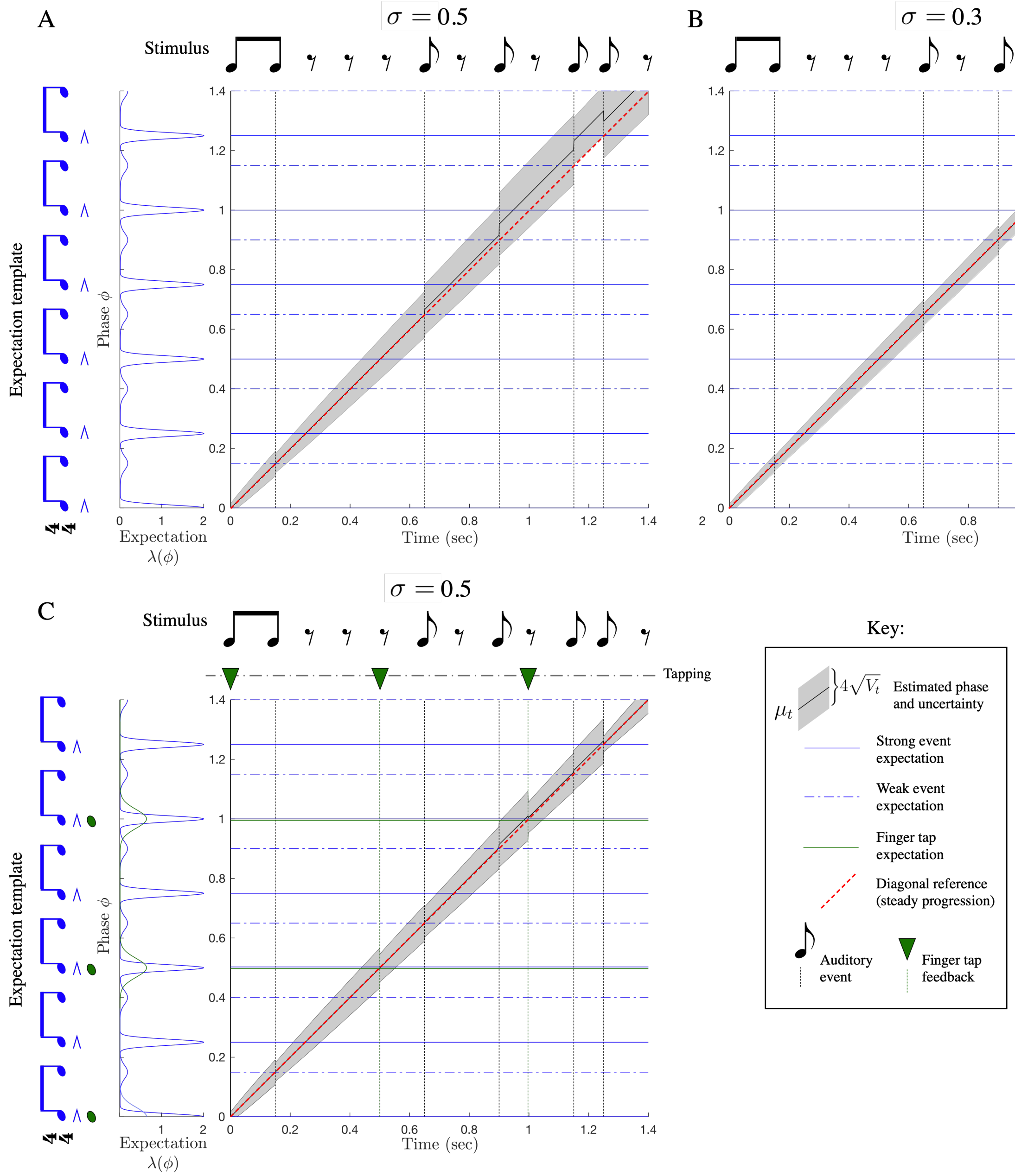

Figure 4: **Too much syncopation causes rhythm tracking failure.** A predominance of events associated with weak expectations combined with accumulated phase uncertainty can lead to a failure to track phase accurately. A) In this example, phase uncertainty V increases over a long silence. At the next 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, and overestimating phase at the final event (correct phase marked with yellow dot). B) When the rate of accumulation of phase uncertainty (i.e., the expected phase noise σ^2) is decreased, phase is tracked correctly. C) Alternatively, phase can be tracked successfully by inserting an isochronous stream of finger taps and a suitable template for the alignment between expected auditory feedback from the taps and phase. We use mPIPPET to simulate an expectation for isochronous taps (green notes and trace on the left). For simplicity, taps are placed every 0.5 sec; however, even noisy taps generated based on estimated phase could serve to reduce phase uncertainty and avoid a total phase tracking failure.

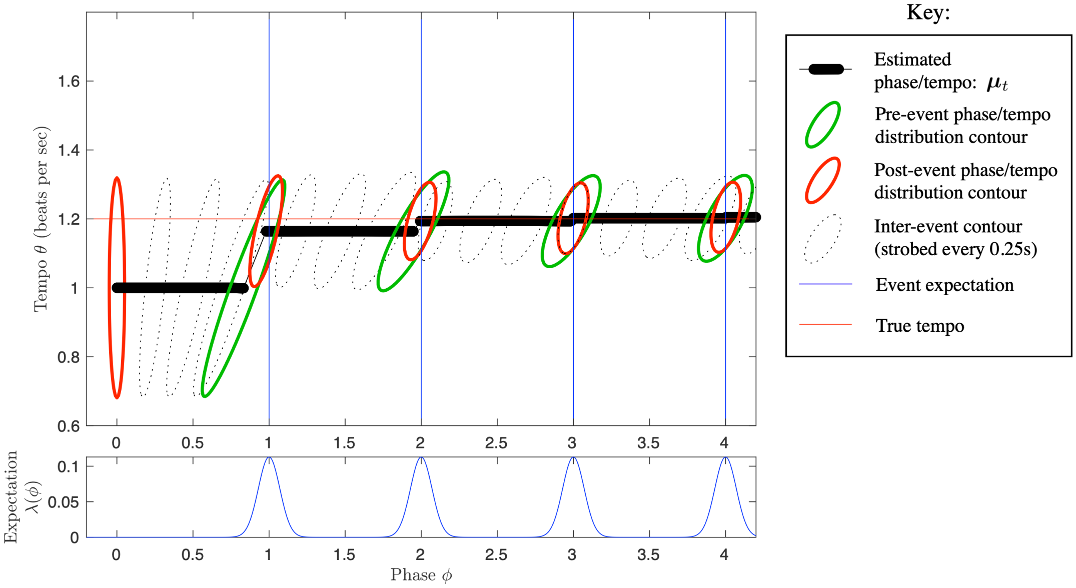

Figure 5: **The PATIPPET filter estimates phase and tempo.** PATIPPET is initialized with high tempo uncertainty. The first event occurs relatively early, causing the estimated tempo to increase. Each subsequent event occurs close to the time expected based on the estimated phase and tempo, causing the posterior to contract in both the phase and tempo 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.

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 [33, 34].

The phase inference 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 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 [33]. 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 6, we show that with appropriate parameters, PATIPPET can indeed reproduce the experimental observation of a near-linear increase in α from below to above 1 as IOI increases. In PATIPPET, this

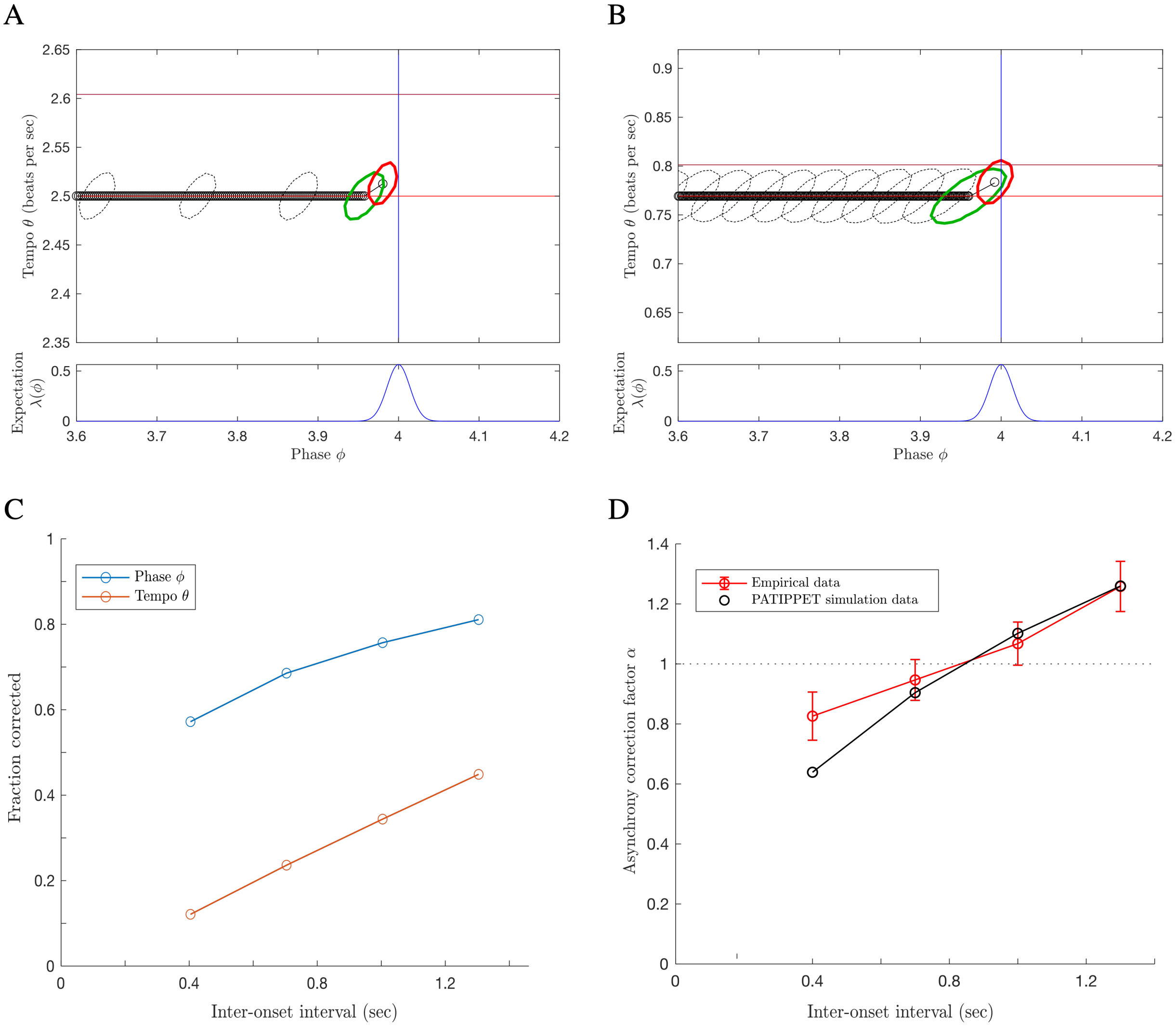

Figure 6: **PATIPPET reproduces human tapping data showing stronger error correction for longer inter-onset intervals.** A and B) The distribution on phase and tempo leading up to and following a phase shift at the fourth event in an isochronous sequence for two different metronome tempi, i.e., two different inter-onset intervals. (Same color key as Figure 5, but with phase/tempo distribution contours strobed every .05 sec.) 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 [34] that the phase shift is undercorrected when IOIs are short and overcorrected $\alpha > 1$ when IOIs are long.

phenomenon is a natural consequence of optimal inference in the context of phase and tempo uncertainty that accumulates between observed events.

3.6 Time warping in the absence of expected events

When an event in a rhythmic stimulus is strongly expected but no event occurs, an optimal Bayesian observer should initially be biased to believe that in spite of their current phase estimate, the stimulus may not have reached the expected event phase yet. The result should be that a perfectly timed event later in the stimulus will seem to be arriving earlier than expected: in other words, the tempo of the stimulus will seem to accelerate. The degree of this effect will depend on the observer's degree of phase and tempo uncertainty.

There is evidence of such an effect in human rhythm 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 [35, 36, 37, 38] (though there is some disagreement about this [39]), suggesting that it is indeed the established expectation of continuing subdivision that interferes with the perceived passage of time when the subdivisions cease. A second result that could be similarly accounted for is the surprising finding in [40] 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 estimated phase induced by an omission of a strongly expected event should indeed delay the subsequent tap.

We stimulated PATIPPET with a strong isochronous expectation template by scaling up λ and presented it with a "filled duration" in which all expected events occurred and an "empty duration" in which events occurred only at the beginning and end of the interval (Figure 7). PATIPPET loyally tracked phase through the filled duration; 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 and an unexpectedly early onset of the event marking the end of the empty duration (Figure 7A).

Specifically timed event expectations are not necessary to produce a filled duration illusion: random raindrop sounds were sufficient to lengthen produced intervals during audiomotor synchronization task [41]. In PATIPPET, a filled duration effect was also produced when the expectation template consisted only of a high expected background rate of events λ_0. In this case, estimated phase advance slowed during the empty interval because estimated tempo dropped. The PATIPPET filter effectively noted that not as many events were occurring as expected, and in response it lowered estimated tempo because a lower event rate is expected at a lower tempo. This type of explanation could be invoked to offer a normative account for other non-rhythmic filled interval illusions, though doing so is beyond the scope of this work.

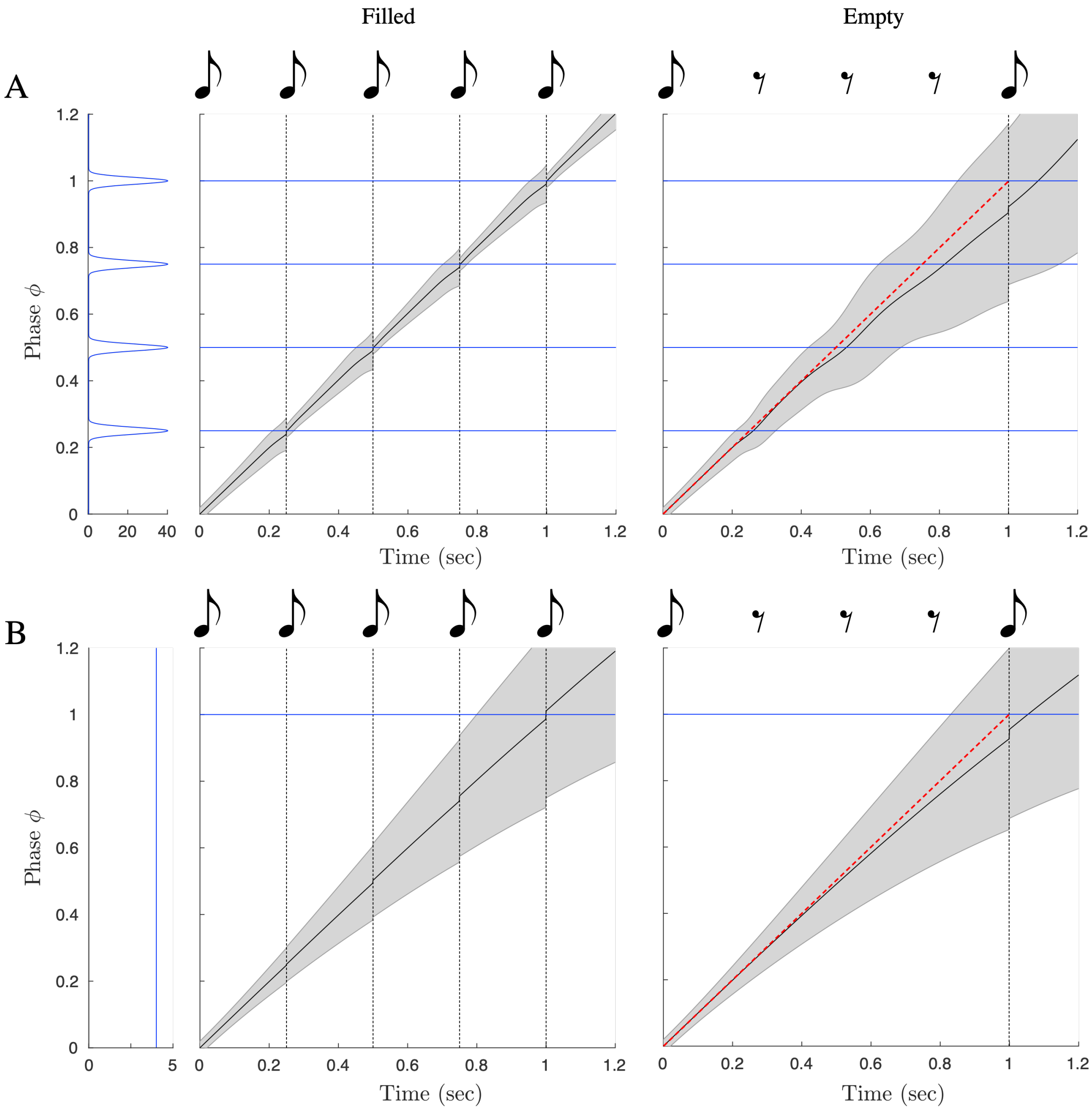

Figure 7: **The filled duration illusion: time warping by the omission of strongly expected events.** (Same image key as 4, with shading displaying PATIPPET phase variance.) A) PATIPPET is simulated with strong expectations for isochronous events. Left: When a set of strongly expected events occur as expected (a filled duration), estimated phase stays on track, advancing (on average) at a rate of 1. Right: When the duration is empty, estimated phase deviates from steady progression (red diagonal) by dragging as each expected event point approaches and passes, leading to the illusion that the event marking the end of the interval has arrived earlier than expected. B) PATIPPET is simulated with a high expected background rate of events λ_0, but no phase-specific event expectations ϕ_i. In this case, too, an empty duration leads to dragging estimated phase and an unexpectedly early final event.

4 Discussion

Here were have presented PIPPET, a framework representing entrainment to a time series of discrete events based on a template of temporal expectations. PIPPET treats the event stream as the output of a point process modulated by the state of a hidden phase variable. The PIPPET filter uses variational Bayes to continuously estimate phase and track phase uncertainty based on this generative model. PATIPPET extends PIPPET to include a generative model of tempo change, and the PATIPPET filter simultaneously estimates phase, tempo, and the covariance matrix representing their uncertainty and their codependence. This framework is intended to serve as a hypothesis for how the human brain integrates auditory event timing to inform and update an estimate of the state and rate of an underlying temporal process.

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

4.1 Relationship to other models of timing

The dynamics of PIPPET and PATIPPET in response to sensory events are similar to dynamics of other entrainment models that correct phase and period based on event timing, e.g., [42, 43]. Models based on Dynamic Attending Theory, e.g., [11, 12], are also similar in explicitly modeling timing expectations and their effect on phase and period adjustment. The phase inference framework differ from these existing models in four key ways. First, they are derived as optimal solutions to specific inference problems, and therefore all modeling decisions can be justified within a normative framework. Second, they are formulated in sufficient generality to describe entrainment based on non-isochronous and even aperiodic temporal expectations, an area that has lately received increasing experimental attention [6, 44, 45] but has been largely neglected in entrainment modeling. Third, they allow expectations to influence the inferred phase even in the absence of sensory events, creating the time-warping effect of disappointed expectations evidenced in humans by the "filled duration" illusion. Finally and most critically, they explicitly track uncertainty in phase and tempo, providing a system for moderating between assimilation of new timing data and loyalty to an internal sense of time.

Bayesian methods have been used elsewhere to analyze rhythmic structure as time series of point events. Some of these are application-focused methods that require offline analyses [46, 47] and therefore do not serve as satisfying models of real-time behavior. Cemgil et al (2000) [27] use a Kalman filter that tracks a distribution on phase and tempo similarly to PATIPPET. However, this model is structured to infer phase and tempo event-by-event rather than in continuous time, and is not equipped to handle complex rhythms or temporal structures more complex than approximate isochrony.

Bayesian inference has also been used to model timing estimation in the brain (e.g., [20, 21]), but it is generally used to describe inferences about discrete variables like interval durations and event times, whereas PIPPET describes a continuous inference process underlying predictions about event times. One such model leading to particularly PIPPET-like results was presented in Elliot et al 2014 [22]. The authors created a Bayesian model to explain the results of an experiment that had participants tap along to a stimulus consisting of two jittered metronomes. The model behaves similarly to PIPPET in that it estimates the next event time using a weighted average of previous event times and prior beliefs, with weights informed by expected timing precision. However, like [27], their model infers the anticipated timing of discrete, metronomic events, whereas PIPPET predicts and updates an underlying phase in continuous time and can therefore generalize to non-isochronous and complex rhythms and account for the effects of event omissions. Additionally, in order to account for participants ignoring events far from predicted time points, they introduce the assumption that participants repeatedly test the hypotheses that events come from one or two separate streams, whereas PIPPET naturally accounts for this phenomenon by attributing stray events to a uniform background event rate λ_0.

4.2 Interpreting the generative model

The PIPPET generative model is formulated as though it implements perfect variational Bayesian inference on inherently stochastic stimuli. However, Bayesian computations in the brain are often invoked to compensate for internal as well as external sources of stochasticity [48], and in the case of PIPPET the most reasonable interpretation may be a combination of the two possibilities. In reality, we do not often listen to musical rhythms with random timing and phase jitter; however, neural noise and interaction with other ongoing processes may introduce timing variability into the processing of sensory events and give rise to variability in the process of tracking estimated phase. This interpretation also allows for changes in generative model parameters based on internal states that might affect internal noise levels, e.g., attentiveness (which has been shown to affect tempo correction but not phase correction [49], and which therefore might be modeled through its effect on σ_θ). Ideally, the phase inference framework could be reconstructed based on assumptions of a combination of internal and external noise; however, that is beyond the scope of the current work.

Given this ambiguity, the generative model parameters may ultimately reflect some combination of the empirical statistics of rhythmic stimuli and internal factors. We briefly discuss the precision parameters v_i as an example. First, an upper bound on the precision of expected event timing is the precision of sensory timing perception, which is, for example, high for human audition and significantly lower for human vision[1]. Second, expected event timing precision may further reflect the observed relative timing distributions of event streams. These observations may inform expectations on time scales ranging from a single sitting to a lifetime of listening. Expected timing may be learned separately for different sensory modalities, different musical genres (e.g., techno vs. funk), or even different instruments (e.g., kick drum, snare, hi-hat, as discussed below). The precision of a beat-based temporal expectation is closely related to the width of a "beat bin," the window of time (rather than a single time point) that is proposed to constitute the "beat" in [50], and to the width of the temporal "expectancy region" described in dynamic attending theory [11]; in both cases, this width is increased by imprecision in the immediately preceding stimulus.

4.3 Testable behavioral predictions

Given the ambiguous interpretation of the generative model discussed above, the question of whether human expectation-based entrainment is truly described by a normative framework may be ill-posed. However, two key qualitative elements of this framework can be tested directly: the tracking of phase uncertainty and the influence of expectations in the absence of events. Seeking further experimental evidence of these two phenomena would help determine the value of phase-inference-based models in describing human entrainment behavior.

The phase inference framework predicts that the accumulation of uncertainty over the course of empty time has a critical effect on the perceptual interpretation of subsequent events. In Figure 4, we show a rhythm that is perceptually misinterpreted due in part to empty time preceding syncopation. An experiment could be designed along the lines of [32] to test this aspect of the phase inference framework by measuring the effect of empty time on the interpretation of rhythmic stimuli that follow.

A second prediction along these lines is that various measurable perceptual phenomena, including period-dependent error correction in motor entrainment, perceptual parsing of ambiguous rhythms, and susceptibility to temporal illusions such as the filled duration illusion, should depend critically on levels of phase and tempo uncertainty. Assuming that the parameters of uncertainty tracking vary across individuals, the PIPPET/PATIPPET framework would predict correlations in measurements across these domains: certain

[1] An event can only be experienced after it occurs, so (as pointed out in [21]) 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.

individuals should show increased sensitivity to temporal illusion, misleading rhythms, and the effect of period on error correction. Further, stimulus manipulations that affect phase and tempo uncertainty, including the temporal precision of the auditory events and the length of the click train establishing an initial tempo estimate, should have direct and predictable effects on these perceptual and behavioral measures.

Third, the phase inference framework predicts that omissions of strongly expected events should systematically distort estimates of phase and tempo, or, perhaps indistinguishably, of elapsed time. These effects could be explored by parametrically manipulating event expectations through priming stimuli and then measuring distortions induced by event omissions through perceptual report or timed motor response.

If we find situations in which human behavior qualitatively differs from solutions to the inference problems posed by PIPPET and PATIPPET, these can be interpreted in two perfectly valid ways: either human behavior has not been optimally tuned for the task at hand, or we have not correctly identified and encapsulated the task and its survival-relevant objective. If we follow the latter interpretation, we might attempt to refine the generative model, e.g., by introducing the belief that tempo changes occur in jumps or ramps rather than as random drift, or to modify the objective of the task, e.g., by including additional cost functions or priors associated with perceptual report or motor output as discussed above.

4.4 Application to analysis of behavioral data

The phase inference framework offers a predictive processing lens for understanding the results of rhythm perception and production experiments. 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. These parameters come with natural interpretations in the language of prediction and precision. We can then 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.

For some experimental data, the many parameters available in PIPPET may prove redundant. For example, the observation of weak error correction in entrained tapping could be explained by imprecise auditory timing expectations (high v_i), an overly precise internal model of phase (low V_t, caused perhaps by low σ), or overly precise tap feedback timing expectations (as discussed below). However, we believe these to be meaningful distinctions that call for disambiguation through carefully designed experiments – for example, skipping taps to separate out the precision effects of tapping feedback or varying silent durations within the stimulus to separate the accumulating effects of phase uncertainty V_t from the history-independent effects of timing expectation uncertainty v_i. For experiments that do not take such measures, redundant parameter sets that fit the data may be interpreted as meaningfully different possible interpretations of the results.

4.4.1 Multiple event characteristics

mPIPPET 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 [51]. 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, one could 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. 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 rhythm parsing during melody listening to be influenced by scale degree.

Alternatively, the j index may be used to treat events over multiple sensory modalities. Visual event timing is judged with less precision than auditory event timing in perceptual report [21] and in timing-sensitive sensory pathways [52], and might therefore be modeled with a less precise expectation template. (Note, however, that visual information may not have the same access to motor-related brain regions used for auditory entrainment [53], so the same modeling framework may not be appropriate.)

mPIPPET with $j \to \infty$ can be used to account for a continuum of event types. Thus, we could create a forward model in which it is more likely for notes played with stronger accents to fall on strong beats, or in which lower pitches are expected with higher timing precision [54] and therefore exert greater influence on neural entrainment [55].

The phase inference framework could be further generalized to take into consideration additional stream of continuous input. This could be visual input from watching a pendulum, auditory input from a continuously modulated sound, or proprioceptive feedback from continuous entrained motion (as opposed to discrete, timed proprioceptive feedback like tapping). This goes beyond the scope of the mathematics presented here, but is a straightforward application of results proven in [23].

4.4.2 Tapping

As illustrated in Figure 4, mPIPPET can be used to describe entrained tapping data. Experiments have shown that the presence of entrained tapping prior to temporal perturbations in a metronomic stimulus reduces the phase correction response [56], indicating that the estimate of moment-by-moment phase is influenced by the proprioceptive, tactile, and auditory feedback from tapping. The phase inference framework is well-suited to modeling this influence as its own separate stream of informative input, though a thorough tapping model would require introducing noise into tap execution and into the phase tracking process itself.

Importantly, using tap times to inform an estimate of underlying phase challenges our interpretation of this phase representing a purely external source of temporally patterned events. Instead, the inferred phase would be a hybrid of an external phase and the phase of one's own motor cycle. Functionally, this is similar to the perceptual oscillator forced by both an external stimulus and one's own periodic action proposed by [57]. This may be an especially useful way to think about synchronization with another agent, where one can adopt strategies ranging from following (assigning high precision to input from the other) to leading (assigning low precision to input from the other, and possibly higher precision to self-generated events). See [58] for a discussion of such a coding strategy as a means of minimizing representational neural resources.

4.4.3 Aperiodic rhythm, speech, and musical grammar

One specific question that the phase inference framework might help resolve is how periodic and nonperiodic entrainment differ. PIPPET does not intrinsically differentiate between these two processes; 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. (For neural and behavioral evidence of differences between memory-based and periodicity based entraiment, see, e.g., [45, 6].)

By accommodating aperiodic expectations with any degree of precision or imprecision, the phase inference framework may be especially well-suited to modeling the loose temporal regularities of speech [59]. However, as currently formulated, it is limited in that expectations are not history-dependent: the occurrence or absence of an event does nothing to the expectancy of an event at a later timepoint. This is appropriate for modeling the metrical aspect of rhythmic expectancy, but does not address the grammar-like structure of music rhythm [60], i.e., the expectation of certain temporal patterns of events over others regardless of their metrical positions. Speech, of course, is even more thoroughly grammatical, with certain sound events strongly shaping the temporal and spectral patterns expected in the immediate future.

Such effects could be readily incorporated into the phase inference framework by adding history dependence to the expectation template λ, though that is beyond the scope of this work. The precise details of this history dependence in rhythm parsing could be based on any suitable formal model for rhythmic grammar (e.g., [61, 62, 60]), and for speech applications could include whatever aspects of the co-dependence of timing and content expectations were appropriate for the task at hand.

4.5 Limitations and possible extensions of the phase inference framework

4.5.1 Perceptual vs. motor entrainment

PIPPET is formulated as a perceptual process, without specific reference to how entrained movement is produced by this process. In presenting the PIPPET framework and using it to explain tapping results, we have posited that perceptual and motor entrainment are rooted in the same internal tracking of the phase of an external process. However, perceptual and motor measures of entrainment sometimes give conflicting results: for example, exposure to musical performance with expressively irregular timing affects perceptual reports of timing in subsequent stimuli [63], but does not affect phase correction in tapping to subsequent stimuli [64].

We expect that both physical entrainment and perceptual report are informed by a neural process of estimating underlying phase. Principles of economy suggest that they should share in such an estimate rather than drawing on separately instantiated processes of neural inference, and experimental correlations between motor and perceptual results tentatively support this conclusion (e.g., [65]). However, it is possible that rapid, automatic audiomotor adjustment mechanisms have been selected to prioritize speed over precision (e.g., the spinocerebellar vermis [66]), especially in the case of entrainment to simple isochronous stimuli, and thus may not take uncertainty into account. If this is the case, then motor entrainment experiments not be clean indicators of perceptual management of uncertainty until the effects of these mechanisms are separated out.

4.5.2 Learning expectation templates

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, complex 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 complex 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 [67].)

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 parameter σ necessarily correspond to the actual precision of the neural or external timing mechanisms in play. The brain may underestimate the noisiness 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.

4.5.3 Selecting and updating expectation templates

When the brain is exposed to a rhythmic stimulus, it must first recognize that a predictable pattern exists and select an appropriate expectation template from its learned repertoire. This is its own process of inference, and may be amenable to a Bayesian description. Since the PIPPET filter maintains a unimodal posterior, it is not well-suited to model this initial inference process, which may require maintaining a distribution over multiple distinct possible starting phases and expectation templates. This problem might be partially addressed by incorporating a model that evaluates multiple distinct hypotheses for beat or meter (e.g. [68, 69], or [70] with appropriate probabilistic interpretation) as an additional level of inference in parallel with ongoing phase and tempo inference.

4.6 PIPPET in the brain

Though PIPPET and PATIPPET are abstract models not committed to a particular brain-based implementation, advances in the brain basis of timing and beat-keeping combined with the hypothesized neural bases of predictive processing suggest the beginnings of a plausible approximation of PIPPET in the brain, described below.

The essential aspect of the PIPPET framework that qualitatively differentiates its behavior from previous models is the explicit tracking of uncertainty over time for the purpose of informing the relative weights of sensory event timing and internal state estimates. There have been various proposals of how uncertainty is represented and utilized in the brain, and the system likely differs by task and type of uncertainty [71, 48]. One proposal is of particular interest in relation to timing: uncertainty about a hidden state may be computed in medial frontal cortex and signalled via dopaminergic neurons in the ventral tegmental area [72]. In this case, the hidden state would be the phase and tempo of the stimulus. This proposal is consistent with the observations that dopaminergic neurons encode of certainty in the temporal expectation of sensory cues [73] and that dopamine receptor antagonism in humans causes increased timing uncertainty [74].

In the predictive processing literature, dopamine is often given the role of signaling certainty ("expected precision") across levels of hierarchical processing [75]. In this framework, it participates in probabilistic computations by weighting the input to error-calculating neural populations, causing these errors to be weighted more heavily in the ongoing process of error-minimization that implements variational Bayesian estimation of hidden states. Different dopaminergic populations may signal precision at different levels of processing; in particular, dopamine may signal precision of both higher-level state estimates and lower-level sensory expectations. Thus, phase certainty V_t^{-1} and expected timing precision v_i^{-1} may both influence computation through dopaminergic signalling.

Experiments with non-human primates have shown neural trajectories in medial premotor cortex (MPC, encompassing the supplementary and pre-supplementary motor areas) that represent progress through self-generated behavioral processes. The author hypothesizes in [76] that similar trajectories represent rhythmic phase in human MPC. A representation of a linear phase ϕ, used in the phase inference framework for flexibility and mathematical tractability, would seem to be a limiting factor for implementation in the brain. For shorter, aperiodic learned patterns of temporal expectation, phase could be represented by short, aperiodic trajectories [77], as observed in primates in timed response tasks; for simple periodic patterns, phase could be represented circularly [78], as observed in isochronous tapping tasks; and for longer, hierarchical patterns, phase could be represented by hierarchically structured trajectories that loop but also evolve in other dimensions, as observed in cyclic behaviors whose sensory components change from one cycle to the next [79].

Guided by the "Action Simulation for Auditory Prediction" (ASAP) hypothesis presented in [80] and further developed in [76], the theory of hierarchical predictive processing [81], and the predictive functions proposed for the dorsal auditory pathway [82, 83], we propose a neural implementation of PIPPET's phase estimation in Figure 8. An essential aspect of this account is that it does not insist on the mathematical convenience of instantaneous phase updates, which are obviously implausible in the brain. Instead, precise timing predictions are issued with appropriate timing to intercept rising sensory signals, and the resulting timing errors are then be used to update phase through an error minimization process over the next few hundred milliseconds.

Briefly, phase is represented by stereotyped trajectories of population firing rates in MPC, and phase uncertainty is also represented locally in medial frontal cortex [72]. Basal ganglia selects and activates an expectation template appropriate to the context. This template is combined with phase and phase uncertainty estimates in MPC to compute a momentary subjective hazard rate Λ. The hazard rate is sent to parietal cortex as a prediction of event-based input, where it meets ascending pulses from the auditory system associated with auditory events (which may be relayed rapidly from the dorsal cochlear nucleus via cerebellum [19]). "Event prediction error" from parietal cortex returns to MPC, where it pushes μ in the direction that reduces error: toward expected event phases at events and away from them between events.

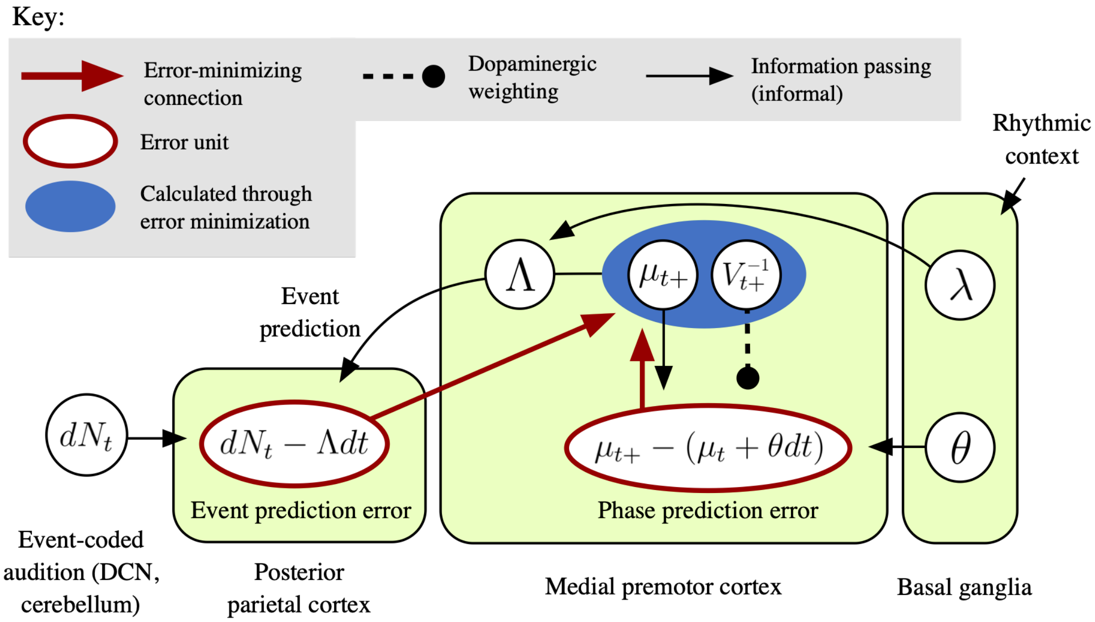

Figure 8: **A possible implementation of PIPPET in the brain.** This diagram embeds a formal predictive coding error minimization scheme is embedded within an informal information-passing schematic to outline how estimated phase μ_t might be calculated and updated on each dt time step by a network of interacting brain regions. Estimated phase μ_t and phase uncertainty V_t are represented in medial premotor cortex (MPC). These estimates are used to calculate instantaneous subjective hazard rate Λ with the help of basal ganglia, which has selected an expectation template λ based on recent rhythmic context. The hazard rate is sent to parietal cortex, where it acts as a prediction of pulses rising from the event-based auditory pathway. An "event prediction error" signal comparing pulses to their prediction is sent back up to MPC, where it pushes μ_{t+} in the direction that reduces prediction error – strongly toward local expectancy peaks when events occur, and weakly away from them when there are no events. (Note that phase updating at events is assumed to be rapid but not instantaneous as represented in the PIPPET filter.) The event prediction error is counterbalanced by a local "phase prediction error" signal generated through local interactions within MPC that pushes μ_{t+} to continue its steady forward progress. Phase prediction error is weighted by dopaminergic signaling of state precision V_t^{-1} through VTA.

This influence is opposed by a "phase prediction error" signal within MPC that pulls μ to progress steadily at tempo θ. This error signal is weighted by phase precision V^{-1}.

Note that this is not a full, formal error-minimization scheme for implementing PIPPET, which is beyond the scope of this manuscript. In particular, it leaves out an updating scheme for V; see [84] for a discussion of the neurophysiology of precision updating. Further, it does not yet include an appropriate scheme for weighting event prediction error.

Although it would be difficult to directly test this neurophysiological setting of PIPPET in humans, it may be possible to indirectly observe a PIPPET-like process in neural data. At the scalp level and in intracortical electrodes, slow electrical oscillations do seem to anticipatorily track the structure of periodic auditory stimuli [85, 86], and this tracking is associated with the subjective passage of time [87]; these oscillations could be explored as possible estimates of mean underlying phase, with particular focus on those in motor areas. Ideally, timing prediction errors could be observed in the evoked EEG response to events (the ERP), allowing a direct measurement of event expectancy at each event time, and there are indeed indicators that the ERP is sensitive to temporal predictability (e.g., [88, 89]); however, the sensitivity of the ERP to recent stimulus history makes this approach unpromising. However, timing prediction errors may be observable in EEG/MEG through their effect on gamma oscillations [90, 91]. Further, the subjective hazard rate Λ itself may be observable by using techniques recently applied to decode the temporal hazard function from EEG data [92], or through its correlation with beta oscillations [93].

Although human-like beat-based perceptual and audio-motor entrainment seems to be unique to humans, other primates do show rudimentary rhythmic timing abilities, especially in the visual modality [94], and represent phase of self-generated cyclic behavioral processes in MPC [79, 78]. Experimental paradigms appropriately modified to engage mechanisms of self-action tracking might activate in non-human primates the same mechanisms of uncertainty-informed event-timing-based phase tracking that we hypothesize for auditory rhythm tracking in humans. Thus, primate neurophysiology in MPC and the dopaminergic system may be a promising avenue for indirectly testing the phase inference framework as a description of the human faculty of rhythmic entrainment.

5 Acknowledgments

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

6 Appendix

6.1 The PATIPPET filter

We let $\boldsymbol{\mu} = \begin{pmatrix} \bar{\phi} \\ \bar{\theta} \end{pmatrix}$ denote the posterior mean and $\mathbf{V} = \begin{pmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{pmatrix}$ denote the posterior covariance. The expressions for the evolution of the PATIPPET filter, which we derive in the following section, are:

$$\begin{cases} d\boldsymbol{\mu} = & \begin{pmatrix} \bar{\theta} \\ 0 \end{pmatrix} dt + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_t) \cdot (dN_t - \Lambda dt) \\ d\mathbf{V} = & \begin{pmatrix} 2V^{12} + \sigma^2 & V^{22} \\ V^{22} & \sigma_\theta^2 \end{pmatrix} dt + \left(\hat{\mathbf{V}} - \mathbf{V}_t\right) \cdot (dN_t - \Lambda dt) \end{cases} \tag{9}$$

where we define

$$\begin{cases} \Lambda := & \sum_{i=0,1,\dots} \Lambda_i \hat{\theta}_i \\ \hat{\boldsymbol{\mu}} = & \frac{1}{\Lambda} \sum_{i=0,1,\dots} \Lambda_i \begin{pmatrix} K_i^{12} + \hat{\phi}_i \hat{\theta}_i \\ K_i^{22} + \hat{\theta}_i^2 \end{pmatrix} \\ \hat{\mathbf{V}} := & \frac{1}{\Lambda} \sum_{i=0,1,\dots} \Lambda_i \Big(\hat{\theta}_i \mathbf{K}_i + \hat{\theta}_i (\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+})(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+})^T \\ & + (\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+}) \begin{pmatrix} K_i^{21} & K_i^{22} \end{pmatrix} + \begin{pmatrix} K_i^{12} \\ K_i^{22} \end{pmatrix} (\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+})^T \Big) \end{cases} \tag{10}$$

and where

$\mathbf{K}_0 := \mathbf{V}$, $\mathbf{K}_i := \left(\mathbf{P}_i + \mathbf{V}^{-1}\right)^{-1}$ for $i > 0$.

K_i^{kl} denotes the entries in $\mathbf{K}_i$.

$\Lambda_0 := \lambda_0$, $\Lambda_i := \lambda_i \varphi(\phi_i | \bar{\phi}, v_i^{-1} + (V^{11})^{-1})$ for $i > 0$.

$\hat{\boldsymbol{\mu}}_i = \begin{pmatrix} \hat{\phi}_i \\ \hat{\theta}_i \end{pmatrix} := \mathbf{K}_i \left(\begin{pmatrix} v_i^{-1} \phi_i \\ 0 \end{pmatrix} + \mathbf{V}^{-1} \boldsymbol{\mu} \right)$ for $i > 0$, and $\hat{\boldsymbol{\mu}}_0 := \boldsymbol{\mu}$.

$\mathbf{P}_i := \begin{pmatrix} v_i^{-1} & 0 \\ 0 & 0 \end{pmatrix}$

6.2 Derivation of differential equations and update equations.

Here we derive the PATIPPET filter; the PIPPET filter can be derived similarly or as a special case of PATIPPET.

Snyder [23] provides a partial differential equation describing the evolution of a probability distribution on a continuously stochastically evolving state that drives the emission of point process events. If the evolution of the underlying state is described by a Gauss-Markov diffusion process:

$$d\mathbf{x} = \mathbf{A}\mathbf{x}dt + \mathbf{B}d\mathbf{W}_t \tag{11}$$

and events are generated at rate $\lambda(\mathbf{x})$, then the evolution of the probability distribution $p_t(\mathbf{x})$ is described by

$$dp_t(\mathbf{x}) = \mathcal{L}[p_t(\mathbf{x})]dt + p_t(\mathbf{x}) \left(\frac{\lambda(\mathbf{x})}{\Lambda} - 1 \right) \cdot (dN_t - \Lambda dt) \tag{12}$$

where $\Lambda := \mathbb{E}[\lambda(\mathbf{x})]$ (with $\mathbb{E}$ denoting expectation under distribution $p_t(\mathbf{x})$), dN_t is the increment in the event count over each dt time step (assumed to be either 1 or 0 with probability 1), and $\mathcal{L}$ is the Kolmogorov forward operator associated with (11):

$$\mathcal{L}[p(\mathbf{x})] = -\sum_i \frac{\partial}{\partial x_i}[\mathbf{A}\mathbf{x}]_i p(\mathbf{x}) + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}[\mathbf{B}\mathbf{B}^T]_{ij} p(\mathbf{x}) \tag{13}$$

Here we project p onto a Gaussian distribution at each time step by matching mean $\boldsymbol{\mu}$ and covariance $\mathbf{V}$, which is also the projection with minimal KL divergence. We do this by finding the differentials of these moments of p_t and using them to drive the evolution of these two variables:

$$\begin{aligned} d\boldsymbol{\mu}_t =& \boldsymbol{\mu}_{t+} - \boldsymbol{\mu}_t = \int_{\mathbf{x}} \mathbf{x} p_{t+}(\mathbf{x}) d\mathbf{x} - \int_{\mathbf{x}} \mathbf{x} p_t(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{x}} \mathbf{x}\left(p_{t+}(\mathbf{x}) - p_t(\mathbf{x})\right)(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{x}} \mathbf{x}\, dp_t(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{x}} \mathbf{x}\mathcal{L}[p_t(\mathbf{x})] dt d\mathbf{x} + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_t) \cdot (dN_t - \Lambda dt) \end{aligned} \tag{14}$$

where we define $\hat{\boldsymbol{\mu}} := \mathbb{E}\left[\mathbf{x}\lambda(\mathbf{x})\right]$, and

$$d\mathbf{V}_t = \mathbf{V}_{t+} - \mathbf{V}_t = \int_{\mathbf{x}} [[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 p_{t+}(\mathbf{x}) d\mathbf{x} - \int_{\mathbf{x}} [[\mathbf{x} - \boldsymbol{\mu}_t]]^2 p_t(\mathbf{x}) d\mathbf{x}$$

where $[[\mathbf{x}]]^2$ denotes $\mathbf{x}\mathbf{x}^T$.

$$\begin{aligned} d\mathbf{V}_t =& \int_{\mathbf{x}} [[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 \left(p_{t+}(\mathbf{x}) - p_t(\mathbf{x})\right) d\mathbf{x} \\ &+ \int_{\mathbf{x}} \left([[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 - [[\mathbf{x} - \boldsymbol{\mu}_t]]^2\right) p_t(\mathbf{x}) d\mathbf{x} \\ =& \int_{\mathbf{x}} [[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 dp_t(\mathbf{x}) - [[\boldsymbol{\mu}_{t+} - \boldsymbol{\mu}_t]]^2 \\ =& \int_{\mathbf{x}} [[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 \mathcal{L}[p_t(\mathbf{x}|N_t)] dt d\mathbf{x} + \left(\hat{\mathbf{V}} - \mathbf{V}_t\right) \cdot (dN_t - \Lambda dt) \end{aligned} \tag{15}$$

where we define $\hat{\mathbf{V}} := \mathbb{E}\left[[[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 \lambda(\mathbf{x})\right]$.

Integrating by parts (or following [26]), we can calculate the appropriate integrals of $\mathcal{L}[p_t(\mathbf{x}|N_t)]$, arriving at a general expression for the variational Bayesian filter for point process data:

$$\begin{cases} d\boldsymbol{\mu}_t = & \mathbf{A}\boldsymbol{\mu}_t dt + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_t) \cdot (dN_t - \Lambda dt) \\ d\mathbf{V}_t = & (\mathbf{A}\mathbf{V}_t + \mathbf{V}_t\mathbf{A}^T + \mathbf{B}\mathbf{B}^T) dt + \left(\hat{\mathbf{V}} - \mathbf{V}_t\right) \cdot (dN_t - \Lambda dt) \end{cases} \tag{16}$$

From (4), the PATIPPET generative model is described by the Gauss-Markov diffusion process (11) with

$$\mathbf{x} = \begin{pmatrix} \phi \\ \theta \end{pmatrix} \text{ and } \boldsymbol{\mu} = \begin{pmatrix} \bar{\phi} \\ \bar{\theta} \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{pmatrix}$$

$$\mathbf{A} := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } \mathbf{B} := \begin{pmatrix} \sigma & 0 \\ 0 & \sigma_\theta \end{pmatrix}.$$

Plugging into (16), we have

$$\begin{cases} d\boldsymbol{\mu}_t = & \begin{pmatrix} \bar{\theta} \\ 0 \end{pmatrix} dt + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_t) \cdot (dN_t - \Lambda dt) \\ d\mathbf{V} = & \begin{pmatrix} 2V^{12} + \sigma^2 & V^{22} \\ V^{22} & \sigma_\theta^2 \end{pmatrix} dt + \left(\hat{\mathbf{V}} - \mathbf{V}_t\right) \cdot (dN_t - \Lambda dt) \end{cases} \tag{17}$$

We complete the derivation by calculating Λ, $\hat{\boldsymbol{\mu}}$, and $\hat{\mathbf{V}}$. This proceeds by first deriving a simple expression for $p(\mathbf{x})\lambda(\mathbf{x})$ as a sum of scaled normal distributions.

Let $\|x\|_A^2$ denote $x^T A x$. We will make use of the following result, a generalized form of a well-known result about quadratic forms that allows us to write products of multivariate normal distributions as normal distributions (see [95] for proof and similar application):

$$\|x - a\|_A^2 + \|x - b\|_B^2 = \|a - b\|_{A(A+B)^{-1}B}^2 + \|x - (A + B)^{-1}(Aa + Bb)\|_{A+B}^2 \tag{18}$$

In the PATIPPET generative model, events are generated at rate

$$\lambda(\mathbf{x}) = \theta \left(\lambda_0 + \sum_{i=1,2,\cdots} \frac{\lambda_i}{\sqrt{2\pi v_i}} e^{-\frac{1}{2}\|\mathbf{x}-\mathbf{x}_i\|_{\mathbf{P}_i}^2} \right)$$

$$\mathbf{P}_i = \begin{pmatrix} v_i^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \; \mathbf{x}_i = \begin{pmatrix} \phi_i \\ 0 \end{pmatrix}.$$

686 $p(\mathbf{x})$ is assumed (forced) to be Gaussian, so we can write:

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

We calculate:

$$\begin{aligned}
p(\mathbf{x})\lambda(\mathbf{x}) =& \frac{\theta}{\sqrt{2\pi|\mathbf{V}|}} e^{-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2_{\mathbf{V}^{-1}}} \left(\lambda_0 + \sum_{i=1,2,\cdots} \frac{\lambda_i}{\sqrt{2\pi v_i}} e^{-\frac{1}{2}\|\mathbf{x}-\mathbf{x}_i\|^2_{\mathbf{P}_i}} \right) \\
=& \frac{\lambda_0\theta}{\sqrt{2\pi|\mathbf{V}|}} e^{-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2_{\mathbf{V}^{-1}}} + \theta \sum_{i=1,2,\cdots} \frac{\lambda_i}{2\pi\sqrt{v_i|\mathbf{V}|}} e^{-\frac{1}{2}\|\mathbf{x}-\mathbf{x}_i\|^2_{\mathbf{P}_i}-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2_{\mathbf{V}^{-1}}}
\end{aligned}$$

Applying (18),

$$\begin{aligned}
p(\mathbf{x})\lambda(\mathbf{x}) =& \frac{\lambda_0\theta}{\sqrt{2\pi|\mathbf{V}|}} e^{-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2_{\mathbf{V}^{-1}}} \\
&+ \theta \sum_{i=1,2,\cdots} \lambda_i \left(\frac{1}{\sqrt{2\pi(v_i^{-1}+V^{-1})}} e^{-\frac{1}{2}\|\mathbf{x}_i-\boldsymbol{\mu}\|^2_{\mathbf{P}_i\mathbf{K}_i(\mathbf{V}^1 1)^{-1}}} \right) \left(\frac{1}{\sqrt{2\pi\frac{v_i|\mathbf{V}|}{v_i^{-1}+(\mathbf{V}^1 1)^{-1}}}} e^{-\frac{1}{2}\|\mathbf{x}-\mathbf{K}_i(\mathbf{P}_i\mathbf{x}_i+\mathbf{V}^{-1}\boldsymbol{\mu})\|^2_{\mathbf{K}_i^{-1}}} \right)
\end{aligned} \tag{19}$$

where we define $\mathbf{K}_i := (\mathbf{P}_i + \mathbf{V}^{-1})^{-1}$. These two final terms are both expressions for normal distributions, so we can rewrite (19) as

$$p(\mathbf{x})\lambda(\mathbf{x}) =\lambda_0\theta\varphi(\mathbf{x}|\boldsymbol{\mu},\mathbf{V}) + \theta \sum_{i=1,2,\cdots} \lambda_i\varphi(\phi_i|\bar{\phi}, v_i^{-1} + (V^{11})^{-1})\varphi(\mathbf{x}|\mathbf{K}_i(\mathbf{P}_i\mathbf{x}_i + \mathbf{V}^{-1}\boldsymbol{\mu}), \mathbf{K}_i) \tag{20}$$

We simplify this expression by defining $\Lambda_i := \lambda_i\varphi(\phi_i|\bar{\phi}, v_i^{-1} + (V^{11})^{-1})$ for $i > 0$, and setting $\Lambda_0 := \lambda_0$ and $\mathbf{K}_0 = \mathbf{V}$. We define $\hat{\boldsymbol{\mu}}_i := \begin{pmatrix} \hat{\phi}_i \\ \hat{\theta}_i \end{pmatrix} := \mathbf{K}_i(\mathbf{P}_i\mathbf{x}_i + \mathbf{V}^{-1}\boldsymbol{\mu})$ for $i > 0$ and set $\hat{\boldsymbol{\mu}}_0 := \boldsymbol{\mu}$. This lets us write

$$p(\mathbf{x})\lambda(\mathbf{x}) = \sum_{i=0,1,\cdots} \Lambda_i\theta\varphi(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \mathbf{K}_i) \tag{21}$$

We use this expression and the moments of normal distributions to calculate Λ, $\hat{\boldsymbol{\mu}}$, and $\hat{\mathbf{V}}$:

$$\Lambda :=\mathbb{E}_p\left[\lambda(\mathbf{x})\right] = \sum_{i=0,1,\cdots} \Lambda_i \int \theta\varphi(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \mathbf{K}_i)d\mathbf{x} = \sum_{i=0,1,\cdots} \Lambda_i\hat{\theta}_i \tag{22}$$

$$\hat{\boldsymbol{\mu}} :=\frac{1}{\Lambda}\mathbb{E}\left[\mathbf{x}\lambda(\mathbf{x})\right] = \frac{1}{\Lambda} \sum_{i=0,1,\cdots} \Lambda_i \int \begin{pmatrix} \phi\theta \\ \theta^2 \end{pmatrix} \varphi(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \mathbf{K}_i)d\mathbf{x}$$

This expression picks out non-central second moment terms of each normal distributions in (21), each of which can be written in terms of the covariance matrix and mean of the distribution. Using K_i^{kl} to denote the entries in $\mathbf{K}_i$, we can write

$$\hat{\boldsymbol{\mu}} = \frac{1}{\Lambda} \sum_{i=0,1,\cdots} \Lambda_i \begin{pmatrix} K_i^{12} + \hat{\phi}_i \hat{\theta}_i \\ K_i^{22} + \hat{\theta}_i^2 \end{pmatrix} \tag{23}$$

The third-order expression for $\hat{\mathbf{V}}$ can also be written in terms of covariance matrices and means since the central third moments of normal distributions are zero.

$$\begin{aligned}
\hat{\mathbf{V}} :=& \frac{1}{\Lambda} \mathbb{E}_p \left[[[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 \lambda(\mathbf{x}) \right] \\
=& \frac{1}{\Lambda} \sum_{i=0,1,\cdots} \Lambda_i \int [[\mathbf{x} - \boldsymbol{\mu}_{t+}]]^2 \theta \varphi(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \mathbf{K}_i) d\mathbf{x} \\
=& \sum_{i=0,1,\cdots} \Lambda_i \Big[\hat{\theta}_i \int [[\mathbf{x} - \hat{\boldsymbol{\mu}}_i]]^2 \varphi(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \mathbf{K}_i) d\mathbf{x} \cdots \\
& + \hat{\theta}_i [[\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+}]]^2 \cdots \\
& + (\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+}) \int (\mathbf{x} - \hat{\boldsymbol{\mu}}_i)^T (\theta - \hat{\theta}_i) \varphi(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \mathbf{K}_i) d\mathbf{x} \cdots \\
& + \left(\int (\mathbf{x} - \hat{\boldsymbol{\mu}}_i)(\theta - \hat{\theta}_i) \varphi(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \mathbf{K}_i) d\mathbf{x} \right) (\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+})^T \Big] \\
=& \frac{1}{\Lambda} \sum_{i=0,1,\cdots} \Lambda_i [\hat{\theta}_i \mathbf{K}_i + \hat{\theta}_i [[\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+}]]^2 \cdots \\
& + (\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+}) \begin{pmatrix} K_i^{21} & K_i^{22} \end{pmatrix} + \begin{pmatrix} K_i^{12} \\ K_i^{22} \end{pmatrix} (\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{t+})^T]
\end{aligned} \tag{24}$$

Expressions (22), (23), and (24) coupled with (17) constitute the PATIPPET filter.

The PIPPET filter can be derived as a special case of the PATIPPET filter by setting $\sigma_\theta = 0$, $\theta_0 = 1$, and all terms in $\mathbf{V}$ to zero except V. However, this requires finessing various degeneracies, e.g. wherever $\mathbf{V}$ is inverted. More straightforward is to follow the same process as above, starting from the PIPPET generative model (1) and (2). Either way ultimately yields the PIPPET filter (3).

For multiple event streams j,:

$$dp_t(\mathbf{x}) = \mathcal{L}[p_t(\mathbf{x})]dt + p_t(\mathbf{x}) \sum_j (\lambda_j(\mathbf{x}) - \mathbb{E}_p[\lambda_j(\mathbf{x})]) \cdot (\mathbb{E}_p[\lambda_j(\mathbf{x})]^{-1} dN_j - dt) \tag{25}$$

This follows directly from application of the derivation above to equation (5) in [96] with a discrete spatial dimension. By the methods above, it yields the mPIPPET filter (8) and the mPATIPPET filter:

$$\begin{cases} d\boldsymbol{\mu}_t = \begin{pmatrix} \bar{\theta} \\ 0 \end{pmatrix} dt + \sum_j \left(\hat{\boldsymbol{\mu}}^j - \boldsymbol{\mu}_t\right) \cdot (dN_t^j - \Lambda^j dt) \\ d\mathbf{V} = \begin{pmatrix} 2V^{12} + \sigma^2 & V^{22} \\ V^{22} & \sigma_\theta^2 \end{pmatrix} dt + \sum_j \left(\hat{\mathbf{V}}^j - \mathbf{V}_t\right) \cdot (dN_t^j - \Lambda^j dt) \end{cases} \tag{26}$$

695
6.3 Simulation parameters.

696 All code used to create figures in this manuscript is available at `https://github.com/joncannon/PIPPET`.

697 PIPPET simulations were conducted by numerical simulation of (3) with $dt = .001$ and initialized with 698 $\phi_0 = 0$ and $V_0 = .0002$. Parameters for the simulations shown in each figure are listed below, with t_n used 699 to denote simulated event times (in units of seconds).

700 *Figure 2:* $\phi_1 = .5$, $v_1 = .0005$, $\lambda_1 = 1$, $\mu_t = .43$, $V_t = .001$, $\lambda_0 = 0$ or $.5$, except as otherwise specified.

Figure 3A:

$$\begin{aligned} \{t_n\} =& \{0, .150, .5, .75, .9, 1.25\} \\ \{\phi_i\} =& \{0, .15, .25, .4, .5, .65, .75, .9, 1, 1.15, 1.25, 1.4\} \\ \{v_i\} =& \{.0001, .0005, .0001, .0005, .0001, .0005, .0001, .0005, .0001, .0005\} \\ \{\lambda_i\} =& \{.05, .01, .05, .01, .05, .01, .05, .01, .05, .01\} \\ \lambda_0 =& .01 \\ \sigma =& .05 \end{aligned}$$

701 *Figure 3B:* Same as Figure 3A, but with $t_3 = .45$ (50 ms negative event time shift).

702 *Figure 3C:* Same as Figure 3A, but with $\{t_n\} = \{0, .15, .5, .7, .85, 1.2\}$ (50 ms negative phase shift).

703 *Figure 4A:* Same as Figure 3, but with $\{t_n\} = \{0, .150, .65, .9, 1.15, 1.25\}$.

704 *Figure 4B:* Same as Figure 4A, but with $\sigma = .3$.

Figure 4A: Same as Figure 4A, but with additional tap times and tap feedback expectations:

$$\begin{aligned}\{t_n^{tap}\} =&\{\phi_i^{tap}\} = \{0, .5, 1\}\\ v_i^{tap} =&.0005\\ \lambda_i^{tap} =&.05\\ \lambda_0^{tap} =&.01\end{aligned}$$

705 PATIPPET simulations were conducted by numerical simulation of (4) with $dt = .001$. Parameters for 706 the simulations shown in each figure are listed below.

Figure 5:

$$\begin{aligned}t_n =&\frac{n}{1.2Hz}\\ \phi_i =&i\\ v_i =&.005\\ \lambda_i =&.02\\ \lambda_0 =&.0001\\ \sigma =&.05\\ \sigma_\theta =&.05\\ \boldsymbol{\mu}_0 =&\begin{pmatrix}0\\1\end{pmatrix}\\ \mathbf{V}_0 =&\begin{pmatrix}.001 & 0\\ 0 & .04\end{pmatrix}\end{aligned}$$

Figure 6: In four simulations, we set the inter-onset interval Δ to $.4s$, $0,7s$, $1.0s$, and $1.3s$. In each

simulation, we set the perturbation δ to $\frac{\Delta}{25}$.

$$\begin{aligned}
\{t_n\} =& \{\Delta, 2\Delta, 3\Delta, 4\Delta + \delta\}\\
\phi_i =& i\\
v_i =& .0002\\
\lambda_i =& .02\\
\lambda_0 =& 10^{-5}\\
\sigma =& .01\\
\sigma_\theta =& .01\\
\boldsymbol{\mu}_0 =& \begin{pmatrix} 0 \\ 1 \end{pmatrix}\\
\mathbf{V}_0 =& \begin{pmatrix} 10^{-4} & 0 \\ 0 & 10^{-4} \end{pmatrix}
\end{aligned}$$

Figure 7A:

$$\begin{aligned}
\phi_i =& .25i\\
v_i =& .0001\\
\lambda_i =& 1\\
\lambda_0 =& .0001\\
\sigma =& .015\\
\sigma_\theta =& .2\\
\boldsymbol{\mu}_0 =& \begin{pmatrix} 0 \\ 1 \end{pmatrix}\\
\mathbf{V}_0 =& \begin{pmatrix} .0001 & 0 \\ 0 & .005 \end{pmatrix}
\end{aligned}$$

Left: $\{t_n\} = \{.25, .5, .75, 1\}$. Right: $\{t_n\} = \{1\}$.

Figure 7B: Same as Figure 7A, but with $\lambda_i = 0$ and $\lambda_0 = 4$.

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: 10.1016/j.tics.2019.08.004. Available from: https://doi.org/10.1016/j.tics.2019.08.004
4. Lawrance ELA, Harper NS, Cooke JE, and Schnupp JWH. Temporal predictability enhances auditory detection. The Journal of the Acoustical Society of America 2014; 135:EL357–EL363. DOI: 10.1121/1.4879667. Available from: http://dx.doi.org/10.1121/1.4879667
5. Nobre AC and Van Ede F. Anticipated moments: Temporal structure in attention. Nature Reviews Neuroscience 2018; 19:34–48. DOI: 10.1038/nrn.2017.141. Available from: http://dx.doi.org/10.1038/nrn.2017.141
6. Morillon B, Schroeder CE, Wyart V, and Arnal LH. Temporal prediction in lieu of periodic stimulation. Journal of Neuroscience 2016; 36:2342–7. DOI: 10.1523/JNEUROSCI.0836-15.2016
7. Lange K. Brain correlates of early auditory processing are attenuated by expectations for time and pitch. Brain and Cognition 2009; 69:127–37. DOI: 10.1016/j.bandc.2008.06.004. Available from: http://dx.doi.org/10.1016/j.bandc.2008.06.004
8. Jazayeri M and Shadlen MN. Temporal context calibrates interval timing. Nature Neuroscience 2010; 13:1020–6. DOI: 10.1038/nn.2590
9. Herrmann B, Henry MJ, Haegens S, and Obleser J. Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. NeuroImage 2016; 124:487–97. DOI: 10.1016/j.neuroimage.2015.09.019
10. Rajendran VG, Teki S, and Schnupp JW. Temporal Processing in Audition: Insights from Music. Neuroscience 2018; 389:4–18. DOI: 10.1016/j.neuroscience.2017.10.041. Available from: https://doi.org/10.1016/j.neuroscience.2017.10.041
11. Large EW and Jones MR. The dynamics of attending: How people track time-varying events. Psychological Review 1999; 106:119–59. DOI: 10.1037//0033-295x.106.1.119

12. Large EW and Palmer C. Perceiving temporal regularity in music. Cognitive Science 2002; 26:1–37. DOI: 10.1016/S0364-0213(01)00057-X

13. Friston K. A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences 2005; 360:815–36. DOI: 10.1098/rstb.2005.1622

14. Vuust P and Witek MA. Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. Frontiers in Psychology 2014; 5:1–14. DOI: 10.3389/fpsyg.2014.01111

15. Vuust P, Dietz MJ, Witek M, and Kringelbach ML. Now you hear it: A predictive coding model for understanding rhythmic incongruity. Annals of the New York Academy of Sciences 2018; 1423:19–29. DOI: 10.1111/nyas.13622

16. Proksch S, Comstock DC, Médé B, Pabst A, and Balasubramaniam R. Motor and Predictive Processes in Auditory Beat and Rhythm Perception. 2020; 14. DOI: 10.3389/fnhum.2020.578546

17. Friston K, Stephan K, Li B, and Daunizeau J. Generalised filtering. Mathematical Problems in Engineering 2010; 2010. DOI: 10.1155/2010/621670

18. Buckley CL, Kim CS, McGregor S, and Seth AK. The free energy principle for action and perception: A mathematical review. Journal of Mathematical Psychology 2017; 81:55–79. DOI: 10.1016/j.jmp.2017.09.004. arXiv: 1705.09156. Available from: http://dx.doi.org/10.1016/j.jmp.2017.09.004

19. Schwartze M and Kotz SA. A dual-pathway neural architecture for specific temporal prediction. Neuroscience and Biobehavioral Reviews 2013; 37:2587–96. DOI: 10.1016/j.neubiorev.2013.08.005. Available from: http://dx.doi.org/10.1016/j.neubiorev.2013.08.005

20. Egger SW and Jazayeri M. A nonlinear updating algorithm captures suboptimal inference in the presence of signal-dependent noise. Scientific Reports 2018 :18–20. DOI: 10.1038/s41598-018-30722-0

21. DI Luca M and Rhodes D. Optimal Perceived Timing: Integrating Sensory Information with Dynamically Updated Expectations. Scientific Reports 2016; 6:1–15. DOI: 10.1038/srep28563

22. Elliott MT, Wing AM, and Welchman AE. Moving in time: Bayesian causal inference explains movement coordination to auditory beats. Proceedings of the Royal Society B: Biological Sciences 2014; 281. DOI: 10.1098/rspb.2014.0751

23. Snyder DL. Filtering and Detection for Doubly Stochastic Poisson Processes. IEEE Transactions on Information Theory 1972; 18:91–102. DOI: 10.1109/TIT.1972.1054756

24. Opper M. A Bayesian Approach to On-line Learning. On-Line Learning in Neural Networks 2010 :363–78. DOI: 10.1017/cbo9780511569920.017

25. Friston K. The free-energy principle: A unified brain theory? Nature Reviews Neuroscience 2010; 11:127–38. DOI: 10.1038/nrn2787

26. Eden UT and Brown EN. Continuous-time filters for state estimation from point-process models of neural data. Statistica Sinica 2008; 18:1293–310

27. Cemgil AT, Kappen B, Desain P, and Honing H. On tempo tracking: Tempogram representation and Kalman filtering. Journal of New Music Research 2000; 29:259–73. DOI: 10.1080/09298210008565462

28. London J, Polak R, and Jacoby N. Rhythm histograms and musical meter: A corpus study of Malian percussion music. Psychonomic Bulletin and Review 2017; 24:474–80. DOI: 10.3758/s13423-016-1093-7

29. Polak R, London J, and Jacoby N. Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of malian jembe drumming. Frontiers in Neuroscience 2016; 10:1–11. DOI: 10.3389/fnins.2016.00285

30. Friberg A and Sundström A. Swing Ratios and Ensemble Timing in Jazz Performance: Evidence for a Common Rhythmic Pattern. Music Perception 2002; 19:333–49. DOI: 10.1525/mp.2002.19.3.333

31. Warren RM and Gregory RL. An Auditory Analogue of the Visual Reversible Figure. The American Journal of Psychology 1958; 71:612–3

32. Fitch WT and Rosenfeld AJ. Perception and Production of Syncopated Rhythms. Music Perception 2007; 25:43–58

33. Repp BH. Tapping in synchrony with a perturbed metronome: The phase 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

34. 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

35. Hall GS and Jastrow J. Studies of Rhythm. Mind 1886 Jan; os-XI:55–62. DOI: 10.1093/mind/os-XI.41.55. eprint: 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-XI.41.55

36. Nakajima Y. A psychophysical investigation of divided time intervals shown by sound bursts. Journal of the Acoustical Society of Japan 1979; 35:145–51

37. Meumann E. Beiträge zur Psychologie des Zeitbewußtseins [contributions to the psychology of time consciousness]. Philosophische Studien 1896; 12:128–254

38. Grimm K. der einfluß der Zeitform auf die Wahrnehmung der Zeitdauer [the influence of time-form on the perception of duration]. Zeitschrift für Psychologie 1934; 132:104–32

39. Repp BH and Bruttomesso M. A filled duration illusion in music: Effects of metrical subdivision on the perception and production of beat tempo. Advances in Cognitive Psychology 2009; 5:114–34. DOI: 10.2478/V10053-008-0071-7

40. Repp B and Jendoubi H. Flexibility of temporal expectations for triple subdivision of a beat. Advances in Cognitive Psychology 2009; 5:27–41. DOI: 10.2478/v10053-008-0063-7

41. Wohlschläger A and Koch R. Synchronization error: An error in time perception. *Rhythm perception and production*. Ed. by Desain P and Winsdor L. Swets:115–27

42. Wing AM and Kristofferson AB. Response delays and the timing of discrete motor responses. Perception & Psychophysics 1973; 14:5–12. DOI: 10.3758/BF03198607

43. Mates J. A model of synchronization of motor acts to a stimulus sequence - II. Stability analysis, error estimation and simulations. Biological Cybernetics 1994; 70:475–84. DOI: 10.1007/BF00203240

44. Breska A and Deouell LY. Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. PLoS Biology 2017; 15:1–30. DOI: 10.1371/journal.pbio.2001665

45. Bouwer FL, Honing H, and Slagter HA. Beat-based and memory-based temporal expectations in rhythm: similar perceptual effects, different underlying mechanisms. 2019; 8:55

46. Fox C, Rezek I, and Roberts S. Drum ' N ' Bayes : on-Line Variational Inference for Beat Tracking and Rhythm Recognition. International Computer Music Conference 2007. DOI: 10.1016/j.chieco.2016.10.003

47. Pesek M, Leonardis A, and Marolt M. An Analysis of Rhythmic Patterns with Unsupervised Learning. Applied Sciences 2019. DOI: 10.3390/app10010178

48. Ma WJ and Jazayeri M. Neural coding of uncertainty and probability. Annual Review of Neuroscience 2014; 37:205–20. DOI: 10.1146/annurev-neuro-071013-014017

49. Repp BH and Keller PE. Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology 2004; 57:499–521. DOI: 10.1080/02724980343000369

50. Danielsen A. Here, There, and Everywhere: three accounts of pulse in D'Angelo's 'Left and Right'. 2010 Jan :19–36. DOI: 10.4324/9781315596983-2

51. 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

52. Rauschecker JP. Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. Cortex 2018; 98:262–8. DOI: 10.1016/j.cortex.2017.10.020. Available from: https://doi.org/10.1016/j.cortex.2017.10.020

53. Comstock DC, Hove MJ, and Balasubramaniam R. Sensorimotor synchronization with auditory and visual modalities: Behavioral and neural differences. Frontiers in Computational Neuroscience 2018; 12:1–8. DOI: 10.3389/fncom.2018.00053

54. 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

55. Lenc T, Keller PE, Varlet M, and Nozaradan S. Neural tracking of the musical beat is enhanced by low-frequency sounds. Proceedings of the National Academy of Sciences of the United States of America 2018; 115:8221–6. DOI: 10.1073/pnas.1801421115

56. 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

57. 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

58. 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

59. Rimmele JM, Morillon B, Poeppel D, and Arnal LH. Proactive Sensing of Periodic and Aperiodic Auditory Patterns. Trends in Cognitive Sciences 2018; 22:870–82. DOI: 10.1016/j.tics.2018.08.003. Available from: https://doi.org/10.1016/j.tics.2018.08.003

60. Rohrmeier M. Towards a formalization of musical rhythm. *Proc. of the 21st Int. Society for Music Information Retrieval Conf.* 2020

61. Pearce MT. The construction and evaluation of statistical models of melodic structure in music perception and composition. PhD thesis. City University, London, 2005

62. Sioros G, Davies ME, and Guedes C. A generative model for the characterization of musical rhythms. Journal of New Music Research 2018; 47:114–28. DOI: 10.1080/09298215.2017.1409769. Available from: http://doi.org/10.1080/09298215.2017.1409769

63. Repp BH. Obligatory "expectations" of expressive timing induced by perception of musical structure. Psychological Research 1998; 61:33–43. DOI: 10.1007/s004260050011

64. Repp BH. Compensation for subliminal timing perturbations in perceptual-motor synchronization. Psychological Research 2000; 63:106–28. DOI: 10.1007/PL00008170

65. Schwartze M and Kotz SA. The Timing of Regular Sequences: Production, Perception, and Covariation. Journal of Cognitive Neuroscience 2015; 27:139. DOI: 10.1162/jocn. Available from: https://www.apa.org/ptsd-guideline/ptsd.pdf%7B%5C%%7D0Ahttps://www.apa.org/about/offices/directorates/guidelines/ptsd.pdf

66. Chauvigné LaS, Gitau KM, and Brown S. The neural basis of audiomotor entrainment: an ALE meta-analysis. Frontiers in human neuroscience 2014 Jan; 8:776. DOI: 10.3389/fnhum.2014.00776. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4179708%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract

67. Kneissler J, Drugowitsch J, Friston K, and Butz MV. Simultaneous learning and filtering without delusions: A bayes-optimal combination of predictive inference and adaptive filtering. Frontiers in Computational Neuroscience 2015; 9:1–12. DOI: 10.3389/fncom.2015.00047

68. Weij B van der, Pearce MT, and Honing H. A probabilistic model of meter perception: Simulating enculturation. Frontiers in Psychology 2017; 8:1–18. DOI: 10.3389/fpsyg.2017.00824

69. Alejandro M, Id M, Sigman M, and Slezak DF. From beat tracking to beat expectation : Cognitive-based beat tracking for capturing pulse clarity through time. PLoS ONE 2020; 15:e0242207. DOI: 10.17605/OSF.IO/P3QTV

70. Large EW, Almonte FV, and Velasco MJ. A canonical model for gradient frequency neural networks. Physica D: Nonlinear Phenomena 2010; 239:905–11. DOI: 10.1016/j.physd.2009.11.015. Available from: http://dx.doi.org/10.1016/j.physd.2009.11.015

71. Pouget A, Beck JM, Ma WJ, and Latham PE. Probabilistic brains: Knowns and unknowns. Nature Neuroscience 2013; 16:1170–8. DOI: 10.1038/nn.3495

72. Gershman SJ and Uchida N. Believing in dopamine. Nature Reviews Neuroscience 2019; 20:703–14. DOI: 10.1038/s41583-019-0220-7. Available from: http://dx.doi.org/10.1038/s41583-019-0220-7

73. Sarno S, De Lafuente V, Romo R, and Parga N. Dopamine reward prediction error signal codes the temporal evaluation of a perceptual decision report. Proceedings of the National Academy of Sciences of the United States of America 2017; 114:E10494–E10503. DOI: 10.1073/pnas.1712479114

74. Tomassini A, Ruge D, Galea JM, Penny W, and Bestmann S. The Role of Dopamine in Temporal Uncertainty. Journal of Cognitive Neuroscience 2016. DOI: 10.1162/jocn. arXiv: 1511.04103. Available from: 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

75. Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, and Bestmann S. Dopamine, affordance and active inference. PLoS Computational Biology 2012; 8. DOI: 10.1371/journal.pcbi.1002327

76. Cannon J and Patel AD. How beat perception coopts motor neurophysiology: a proposal. bioRxiv 2020. DOI: https://doi.org/10.1101/805838

77. Wang J, Narain D, Hosseini EA, and Jazayeri M. Flexible timing by temporal scaling of cortical responses. Nature Neuroscience 2018; 21:102–12. DOI: 10.1038/s41593-017-0028-6. Available from: http://dx.doi.org/10.1038/s41593-017-0028-6

78. Gámez J, Mendoza G, Prado L, Betancourt A, and Merchant H. The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. PLoS biology 2019; 17:e3000054

79. Russo AA, Khajeh R, Bittner SR, Perkins SM, Cunningham JP, Abbott LF, and Churchland MM. Neural trajectories in the supplementary motor area and primary motor cortex exhibit distinct geometries, compatible with different classes of computation. Neuron 2020; 107. DOI: 10.1101/650002. Available from: https://www.biorxiv.org/content/10.1101/650002v1.abstract

80. Patel AD and Iversen JR. The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. Frontiers in Systems Neuroscience 2014; 8:1–14. DOI: 10.3389/fnsys.2014.00057. Available from: http://journal.frontiersin.org/article/10.3389/fnsys.2014.00057/abstract

81. Friston K. Hierarchical models in the brain. PLoS Computational Biology 2008; 4. DOI: 10.1371/journal.pcbi.1000211

82. Schubotz RI. Prediction of external events with our motor system: towards a new framework. Trends in Cognitive Sciences 2007; 11:211–8. DOI: 10.1016/j.tics.2007.02.006

83. Rauschecker JP. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hearing Research 2011; 271:16–25. DOI: 10.1016/j.heares.2010.09.001. Available from: http://dx.doi.org/10.1016/j.heares.2010.09.001

84. Kanai R, Komura Y, Shipp S, and Friston K. Cerebral hierarchies: Predictive processing, precision and the pulvinar. Philosophical Transactions of the Royal Society B: Biological Sciences 2015; 370. DOI: 10.1098/rstb.2014.0169

85. Schroeder CE and Lakatos P. Low-frequency neuronal oscillations as instruments of sensory selection. Trends in neurosciences 2009; 32. DOI: 10.1016/j.tins.2008.09.012.Low-frequency

86. Arnal LH and Giraud AL. Cortical oscillations and sensory predictions. Trends in Cognitive Sciences 2012; 16:390–8. DOI: 10.1016/j.tics.2012.05.003. Available from: http://dx.doi.org/10.1016/j.tics.2012.05.003

87. Arnal LH and Kleinschmidt AK. Entrained delta oscillations reflect the subjective tracking of time. Cerebral Cortex 2017 :e1349583. DOI: 10.1093/cercor/bhu103

88. Schwartze M, Farrugia N, and Kotz SA. Dissociation of formal and temporal predictability in early auditory evoked potentials. Neuropsychologia 2013; 51:320–5. DOI: 10.1016/j.neuropsychologia.2012.09.037. Available from: http://dx.doi.org/10.1016/j.neuropsychologia.2012.09.037

89. Ungan P, Karsilar H, and Yagcioglu S. Pre-attentive Mismatch Response and Involuntary Attention Switching to a Deviance in an Earlier-Than-Usual Auditory Stimulus: An ERP Study. Frontiers in Human Neuroscience 2019; 13:1–16. DOI: 10.3389/fnhum.2019.00058

90. Todorovic A, Ede F van, Maris E, and Lange FP de. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. Journal of Neuroscience 2011; 31:9118–23. DOI: 10.1523/JNEUROSCI.1425-11.2011

91. Bastos AM, Usrey WM, Adams Ra, Mangun GR, Fries P, and Friston KJ. Canonical microcircuits for predictive coding. Neuron 2012 Nov; 76:695–711. DOI: 10.1016/j.neuron.2012.10.038. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3777738%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract

92. Herbst SK, Fiedler L, and Obleser J. Tracking temporal hazard in the human electroencephalogram using a forward encoding model. eNeuro 2018; 5:1–17. DOI: 10.1523/ENEURO.0017-18.2018

93. Tavano A, Schröger E, and Kotz SA. Beta power encodes contextual estimates of temporal event probability in the human brain. PLoS ONE 2019; 14. DOI: 10.1371/journal.pone.0222420

94. Merchant H and Honing H. Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. Frontiers in neuroscience 2014 Jan; 7:274. DOI: 10.3389/fnins.2013.00274. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3894452%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract

95. Harel Y, Meir R, and Opper M. A tractable approximation to optimal point process filtering: Application to neural encoding. Advances in Neural Information Processing Systems 2015; 2015-Janua:1603–11

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