PIPPET: A Bayesian framework for generalized entrainment to stochastic rhythms

Jonathan Cannon

November 30, 2020

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. Here we propose that the problem of rhythm tracking is most 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 formalize this problem as a case of inferring a distribution on a hidden state from point process data in continuous time: either Phase Inference from Point Process Event Timing (PIPPET) or Phase And Tempo Inference (PATIPPET). This approach to rhythm tracking generalizes to non-isochronous and multi-voice

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

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

1 Introduction

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

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

tions 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, the groundwork underlying them is less so: key high-level concepts like the "attentional pulse" are difficult to define mechanistically, so the implementations of these concepts in models remain impressionistic. Moreover, recent results have emphasized the relevance and neural correlates of aperiodic modes of temporal expectation [13, 6, 14], but dynamic attending models are designed to describe entrainment to periodicity and cannot account for aperiodic forms of structured temporal expectation such as entrainment to memorized temporal patterns, irregular musical meters, and the loose temporal regularities of speech [15].

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

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

In the next section, we discuss previous models of expectation in cognition and where they fall short for our purposes. We then formulate three versions of the problem of entrainment that are amenable to precise solutions. 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. The variational Bayesian solution to this inference problem provides a continuous estimate of phase that entrains to the actual phase, as well as an estimated level of certainty about that phase. In the second, "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 (multi-PIPPET) generalizes the first two to incorporate the observation of multiple types of events, each

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

In the following section, we simulate these solutions, 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 they reproduce key aspects of human sensorimotor entrainment behavior that are not explained by other entrainment models. These include:

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.

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

2 Mathematical framework

The framework of "predictive processing" has emerged as the preferred lens for modeling the role of expectations in the brain [16, 17]. According to this constellation of ideas, expectations (or, interchangeably, "predictions") from higher levels of the sensory processing hierarchy are sent to lower levels, where they are compared to incoming sensory information and used to compute "prediction errors." These prediction errors are used to inform dynamic adjustments to the expectations at all levels of processing, as well as slower adjustments to the learned models upon which predictions are based. This is formalized as a process of variational Bayesian inference based on a hierarchical generative model.

Predictive processing would be a natural modeling framework for understanding rhythmic expectation and entrainment as inference [18, 19, 20] except for one key limitation: existing predictive coding models that operate in continuous time are structured to perform inference based on continuous observation, characterizing prediction errors in terms of deviation between a true level of input and a mean expected level of input [21, 22]. They describe predictions about "what" rather than "when," and are therefore ill-suited to characterizing moment-by-moment errors in *timing* prediction, which arrive sporadically, separated by intervals largely devoid of informative prediction error. This may be a fundamental shortcoming in modeling inference in the brain: behavior and neurophysiology suggests that information about "when" is carried by its own distinctive pathways and represented separately from "what," both in perceptual and motor tasks [23, 6, 10]. Bayesian methods have been applied to describe inferences about timing in the brain [24, 25, 26], but in these cases the problem the brain solves has been formulated as discrete inferences about consecutive intervals rather than a continuous inference process.

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

2.1 Phase Inference from Point Process Event Timing (PIPPET)

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

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 $\tau(\phi)$, a function of phase. We will refer to $\tau(\phi)$ as a "temporal expectation template," though it can also be understood as a hazard function for events. To achieve both analytical tractability and flexible descriptive power, we assume that $\tau(\phi)$ is a sum of a constant τ_0 and a countable set of scaled Gaussian functions indexed by $i = 1, 2, \ldots$ etc. Each Gaussian i is centered at a mean phase ϕ_i with variance v_i and scale τ_i:

$$\tau(\phi) = \tau_0 + \sum_i \tau_i N(\phi|\phi_i, v_i) \tag{2}$$

where $N(x|m, v)$ denotes a normalized Gaussian distribution with mean m and variance v. Each Gaussian mean ϕ_i represents a phase at which an event is expected; τ_i represents the strength of that expectation; and v_i^{-1} is the temporal precision of that expectation. $\tau_0 > 0$ represents the rate of events being generated as part of a uniform noise background unrelated to phase. Together, $\tau(\phi)$ constitutes a likelihood function for an event occurring at phase ϕ. See Figure 1 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 temporal expectation template τ. We discuss this decision further in the Discussion section.

Given a series of event times $\{t_n\}$, a temporal expectation template $\tau(\phi)$, 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 $p_t(\phi)$ describing an estimate

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

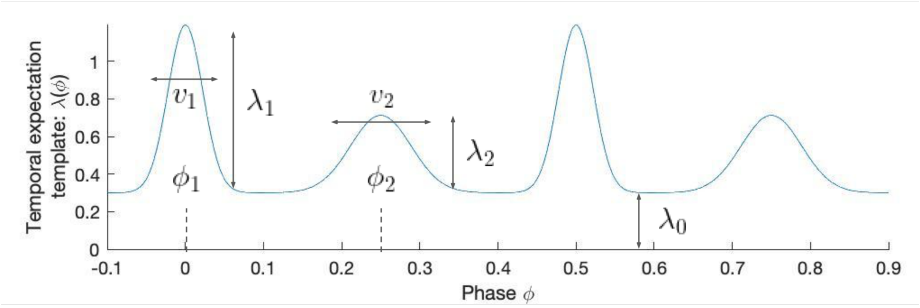

Figure 1: **The temporal expectation template.** In the PIPPET/PATIPPET generative model, $\tau(\phi)$ represents the instantaneous rate of events occurring when the underlying temporal process is at phase ϕ. This is assumed to be a sum of Gaussian-shaped functions with means ϕ_i representing the phases at which specific events are expected, variances v_i representing the (inverse of) the temporal precision expected of those events, and scales τ_i representing the strength of the expectations. A constant τ_0 is also added, representing the instantaneous rate of events unrelated to the underlying phase.

of phase ϕ at every time $t > 0$.

In [27], Snyder derives exact equations 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 of the same mean and variance. This choice of Gaussian is the choice with minimum KL divergence from the true posterior [28], and therefore also minimizes the free energy of the solution within the family of possible Gaussian posteriors, in accordance with the Free Energy Principle of predictive processing [29].

The result of this derivation is a generalization of a Kalman-Bucy filter with Poisson observation noise. Eden and Brown [30] have derived an explicit form for this filter, but it relies on a local approximation of the firing rate function τ that hides some of the interesting effects of events expected at nearby time points. For τ a mixture of Gaussians, we derive a filter that more accurately approximates the Bayesian directly from Synder's solution in [27]. Derivation is presented in Appendix 6.2.

Solution: the PIPPET filter At any time t, let $\bar{\phi}_t$ denote the mean and V_t denote the variance of the Gaussian posterior. At each event time t, we let $\bar{\phi}_{t-}$ and V_{t-} denote the left-hand limits of $\bar{\phi}$ and V before the event, and we write $\bar{\phi}_{t+}$ and V_{t+} to denote their right-hand limit values after the event. $\bar{\phi}_t$ and V_t evolve according to the ODE

$$\begin{cases} \dot{\bar{\phi}} = & 1 - \hat{\Lambda}(\hat{\phi} - \bar{\phi}) \\ \dot{V} = & \sigma^2 - \hat{\Lambda}(\hat{V} - V) \end{cases} \tag{3}$$

and at each event $\bar{\phi}_{t+} = \hat{\phi}$ and $V_{t+} = \hat{V}$, where we define

$$\hat{\phi} := \sum_{i=0,1,\cdots} \frac{T_i}{\hat{\Lambda}} \hat{\phi}_i$$
$$\hat{V} := \sum_{i=0,1,\cdots} \frac{T_i}{\hat{\Lambda}} \left(K_i + (\hat{\phi}_i - \bar{\phi}_{t+})^2 \right)$$

(Note that in this formulation, $\bar{\phi}_{t+}$ must be calculated before V_{t+}.)

$$\hat{\phi}_0 := \bar{\phi}_{t-} \text{ and } \hat{\phi}_i := K_i(V_{t-}^{-1}\bar{\phi}_{t-} + v_i^{-1}\phi_i) \text{ for } i > 0.$$

$$T_0 := \tau_0 \text{ and } T_i := \tau_i N(\phi_i|\bar{\phi}_{t-}, v_i + V_{t-}) \text{ for } i > 0.$$

$$K_0 := V_{t-} \text{ and } K_i := \frac{1}{V_{t-}^{-1} + v_i^{-1}} \text{ for } i > 0.$$

$$\hat{\Lambda} := \sum_i T_i$$

Intuitively,

- $\bar{\phi}_t$ is the estimated phase at time t, and V_t is the level of uncertainty about the phase estimate.
- At each event time t, $\tau(\phi)$ serves as a likelihood function for phase, and the role of prior is played by a Gaussian with mean $\bar{\phi}_{t-}$ and variance V_{t-}.
- At any time t, $\hat{\phi}_i$ would be the mean of the posterior if an event occurred and was known to come from Gaussian i. It is a weighted sum of the current mean estimated phase $\bar{\phi}_t$ and the mean ϕ_i of Gaussian i, weighted by the precision $\frac{1}{V_t}$ on estimated phase and the temporal precision $\frac{1}{v_i}$ of the Gaussian generating the event, respectively.
- At any time t, $\hat{\phi}$ and $\hat{V}$ would be the mean and variance of the posterior if an event occurred and its source was not known. These are weighted sums of the influences of each Gaussian, weighted by T_i, the relative likelihood that the event is drawn from Gaussian i.
- Between events, each dt time step is taken as a Bayesian inference with likelihood $1 - \tau(\phi)dt$ 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 the ODE (3).

- In the absence of an event, this continuous inference process pushes $\bar{\phi}$ and V away from $\hat{\phi}$ and $\hat{V}$ with a strength proportionate to $\hat{\Lambda}$, the current strength of the expectation of an event – thus, the absence of an event continuously pushes the posterior in the opposite directing as would the occurrence of an event.

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

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

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

As above, an inhomogeneous point process generates events with probability based on a temporal expectation template $\tau(\phi)$, where τ is a sum of Gaussians and a constant:

$$\tau(\phi) = \tau_0 + \sum_i \tau_i N(\phi|\phi_i, v_i) \tag{5}$$

However, in this formulation, we want events to occur with a certain probability in each $d\phi$ phase bin regardless of tempo, which is not the case if events are generated with probability $\tau(\phi)dt$; instead, we let events occur with probability

$$\tau(\phi)\mathbb{E}[d\phi] = \tau(\phi)\theta dt.$$

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

Given a series of event times $\{t_n\}$, a temporal expectation template $\tau(\phi)$, and a prior distribution $p_0(\boldsymbol{x})$ describing the distribution of phase and tempo at time $t = 0$, the observer's goal is to infer a posterior distribution $p_t(\boldsymbol{x})$ describing an estimate of phase and tempo at every time $t > 0$. A similar derivation provides a point-process Kalman-Bucy filter that optimally serves this function within the constraint of Gaussian posteriors, providing a running estimate of a mean phase and tempo $\hat{\boldsymbol{x}}_t$ and a phase/tempo covariance matrix $\boldsymbol{V}_t$. 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 [31], 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 temporal expectation templates.

2.3 PIPPET with multiple event streams (multi-PIPPET)

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

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

$$\tau^j(\phi) = \tau_0^j + \sum_i \tau_i^j N(\phi|\phi_i^j, v_i^j) \tag{7}$$

The Kalman-Bucy estimate of phase for this model is described by mean $\bar{\phi}$ and variance V evolving according to the ODE

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

and resetting to $\bar{\phi}_{t+} = \hat{\phi}^j$ and $V_{t+} = \hat{V}^j$ when an event occurs in stream j, where we define $\hat{\Lambda}^j$, $\hat{\phi}^j$, and $\hat{V}^j$ as we defined $\hat{\Lambda}$, $\hat{\phi}$, 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 explore parallels between the behavior of the the PIPPET and PATIPPET filters and human entrainment. Parameters for these simulations are listed in Appendix 6.3.

3.1 Response to events: phase and variance correction

We simulated PIPPET filter with simple metronomic expectations to illustrate its basic behavior. Events occurring near an expected event phase ϕ_i cause the mean phase estimate $\bar{\phi}$ to shift linearly toward ϕ_i, as indicated by the plateaus in the phase transition function (Figure 2A). Events occurring far from any expected event phase ϕ_i caused negligible adjustment in the phase estimate because they were attributed to the background rate τ_0 of events occurring unrelated to any specific expectation. This leads to a phase response curve that crosses zero with negative slope near each expected event phase and sits

uniformly near zero away from expected event phases (Figure 2A).

If the estimated phase $\bar{\phi}_{t-}$ just before an event time t was very close to an expected event phase ϕ_i, the phase uncertainty V decreased at the event, which effectively "corroborated" the phase estimate (Figure 2B). Events occurring when $\bar{\phi}_{t-}$ was far from any expected event phase had no impact on V, as they were effectively attributed to the background noise rate τ_0 and thus contained no new information about phase. Events occurring in the liminal zone near but not very near an expected event phase ϕ_i caused uncertainty V to increase.

3.2 Stochastic rhythms with uneven subdivision

The PIPPET framework describes entrainment to "stochastic" rhythms in which each expected event phase may or may not be populated by an event. Further, PIPPET is formulated in sufficient generality to describe entrainment to rhythms based on timing expectations with complex, non-isochronous stress patterns [32] and with non-integer duration ratios using suitably designed (or, presumably, learned) temporal expectation templates $\tau(\phi)$. Such rhythmic patterns have been shown to support highly precise synchronization in musicians with appropriate training and enculturated expectations [33], and should therefore be accounted for by any plausible model of human entrainment. Thus, PIPPET is equipped to model entrainment to a very wide range of rhythmic structures with any degree of predictability.

As an example of entrainment to a stochastic 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 eight note of every pair is given a slightly longer 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 player

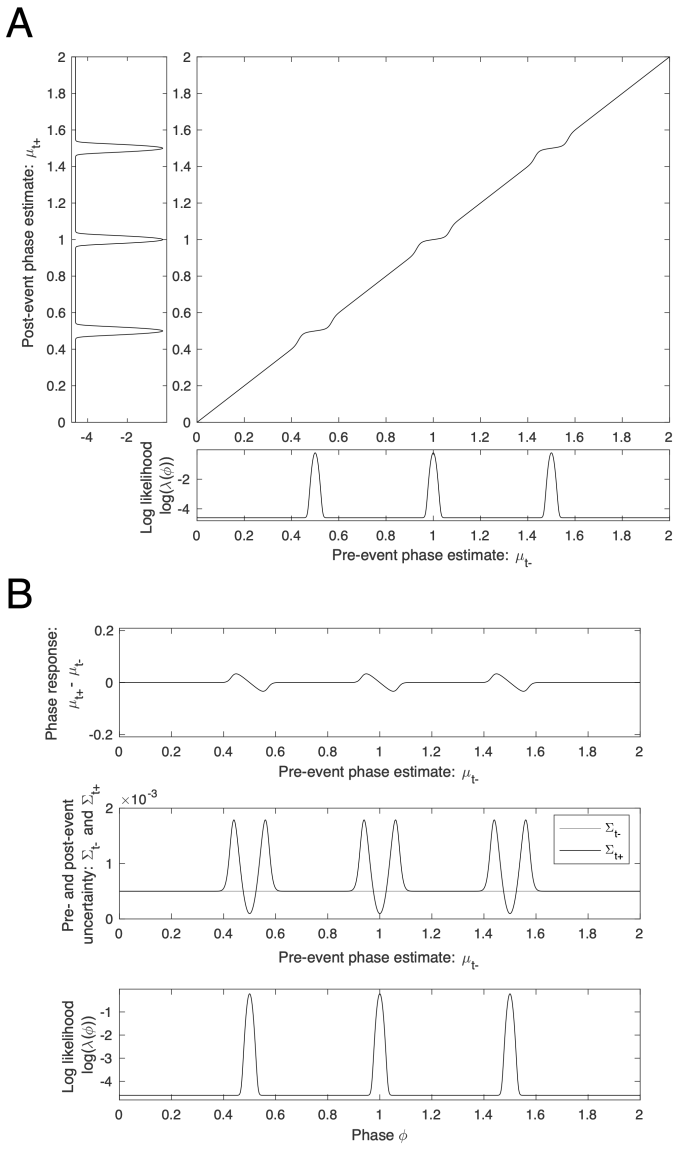

Figure 2: **Characterizing PIPPET's behavior at events** A) Phase transition curve for PIPPET with expectation of three isochronous events. Note that events occurring when the phase estimate $\bar{\phi}_{t-}$ is between expected event phases ϕ_i have little corrective effect on the posterior mean phase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

and tempo and is certainly not limited to small integer ratios [34]. We used a temporal expectation template with a swing ratio close to 3:2 and associated the first eighth note in each pair with a stronger expectation than the second. The simulation entrained to a complex, syncopated rhythm based on this template, and corrected the phase estimate when a phase shift was introduced into the rhythm (Figure 3).

3.3 Failure mode: too much syncopation

Another attractive aspect of the PIPPET framework is that it can account for realistic failures in tracking perfectly timed rhythms. In addition to failures due to time warping described above, failures may occur due to interference between expectations packed closely together in time. Every expected event phase ϕ_i exerts an influence on the evolution of the posterior at all times. This influence is very weak if the current phase estimate is far from ϕ_i. However, if the uncertainty V of the phase estimate is large enough to encompass several expected event phases, or if several events are expected at neighboring phases with insufficient precision, the event may not be fully "attributed" to a single expected event phase. As a result, the adjustment to the phase estimate at an event may reflect an amalgam of these multiple influences, with stronger expectations exerting more influence than weaker ones.

A prime example of this failure mode in human rhythm tracking is tracking overly syncopated rhythms (rhythms with a predominance of events at time points with weaker expectations). Listeners tend to "re-hear" such rhythms by attributing events to metrical positions where events are more strongly expected [35]. Using the expectation template with a swing grid as in the previous section, we simulated a strongly syncopated rhythm (Figure 4). The rhythm's phase was not tracked successfully due to a convergence of factors. Phase un-

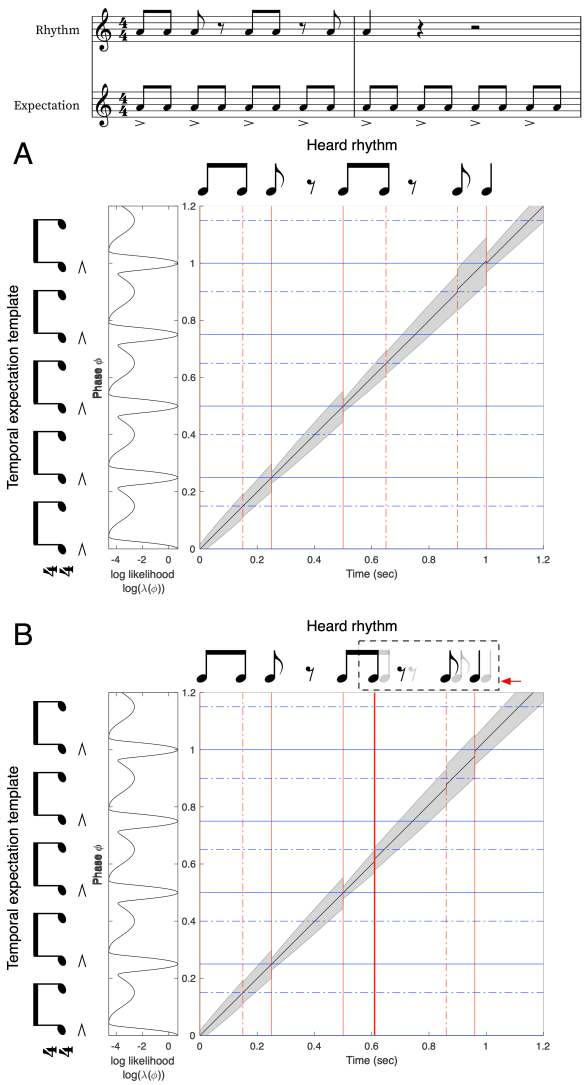

Figure 3: **Tracking phase through swung rhythms.** (Same color key as 5.) A: Phase is estimated over the course of a rhythm. Temporal expectations are not isochronous, but instead represent a swing pattern in which the first eighth note of every pair is slightly longer and more strongly expected than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

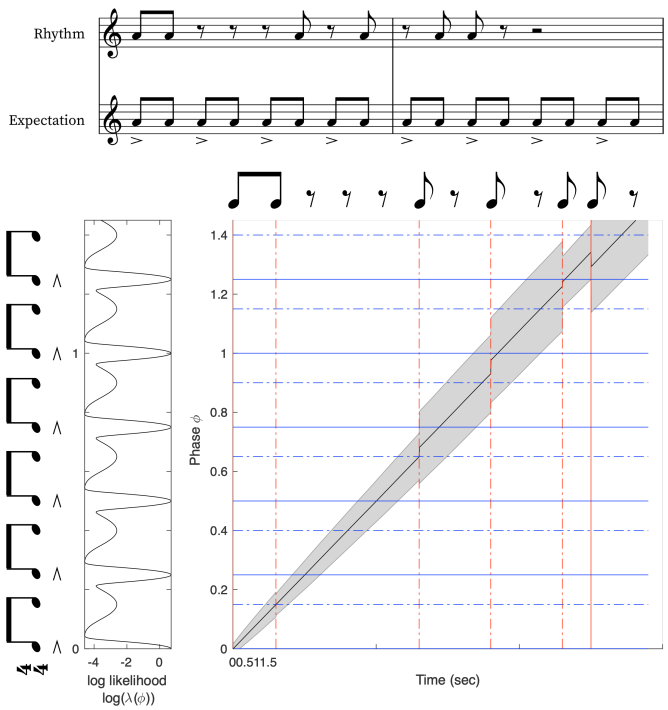

Figure 4: **Too much syncopation causes rhythm tracking failure.** Syncopation combined with imprecise and weak timing expectations on at weak time points can lead to a failure to track phase accurately. In this example, phase uncertainty V increases over a long silence. At the next event, this high 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.

certainty V was only slightly reduced when events occurred at weakly expected phases, so it accumulated over the course of the rhythm, and especially during the long silence. Once V was large, strongly expected event phases ϕ_i began to exert more influence at each event, until eventually events that should have been attributed to weak phase points were instead attributed primarily to adjacent strong phase points. This type of attribution error in syncopated rhythm perception is described in [36].

3.4 In the absence of events: time warping

When an event is strongly expected but no event occurs, an optimal Bayesian observer should initially be biased to believe that in spite of their current estimate, the stimulus may not have reached the expected event phase yet. When we stimulated PIPPET with sufficiently strong metronomic expectations by scaling up τ, PIPPET's behavior at each event was unchanged; however, when strongly expected events were omitted, the mean phase estimate slowed down at each expected event phase, leading to an overall slowing in estimated phase advance (Figure 5).

There is evidence of such an effect in human perception. The "filled duration" illusion is the impression that an isochronous sequence has changed tempo when it is initially subdivided by additional predictable events and then subdivisions are eliminated. According to multiple reports, the magnitude of this effect is reduced or eliminated if the empty intervals precede the filled intervals [37, 38, 39, 40] (though there is some disagreement about this [41]), suggesting that the established expectation of continuing subdivision interferes with perceived timing when subdivisions cease. In PIPPET, this effect is created when the slowing of phase advance causes a properly timed event at the end of the empty interval to arrive at an earlier apparent phase than expected, causing the interval to "seem" shorter.

A second result that could similarly be accounted for by this aspect of PIPPET is the surprising finding in [42] that a participant tapping along with a subdivided beat delays their tap following the omission of an expected subdivision. If taps are planned to coincide with the arrival of a specific mean estimated phase, then the slowing of phase induced by an omission of a strongly expected event in PIPPET would delay the subsequent tap.

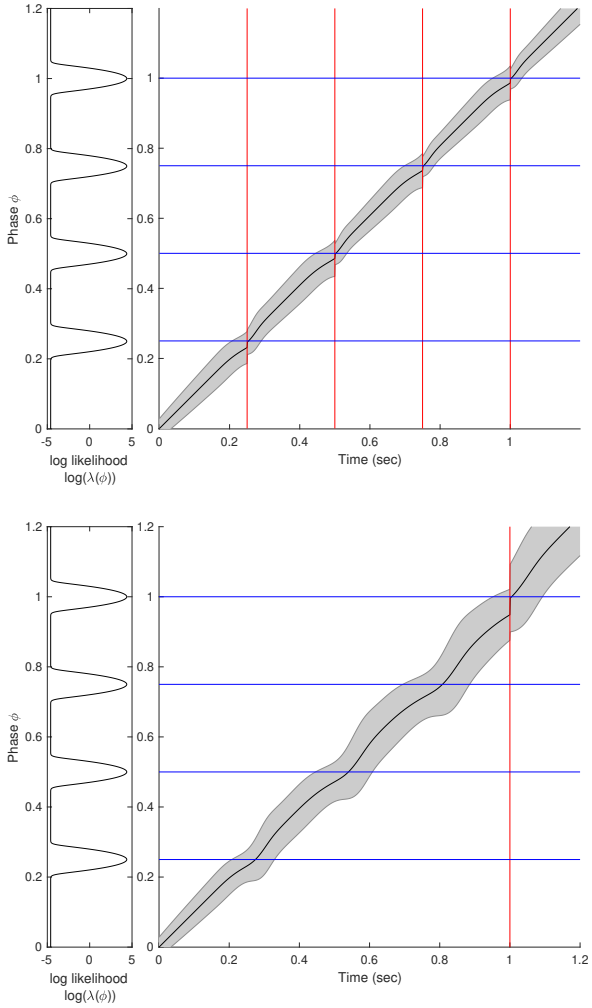

Figure 5: **Time warping by the omission of strongly expected events.** Black curve tracks the estimated mean phase $\bar{\phi}$ over time. Red lines mark event times; blue lines mark expected event phases. Grey shading represents uncertainty about phase, quantified in the model as variance *Sigma* and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

3.5 Tempo inference

We simulated the PATIPPET filter with basic metronomic expectations to observe its capacity to infer phase and tempo at once. We gave the model a wide initial range of possible tempi and a simple metronomic stimulus with actual tempo near the upper end of that range. In these conditions and with the parameter set we chose, the model established the appropriate tempo and phase to within a tight range over the course of the first two events (Figure 6).

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

3.6 Period-dependent corrections

In entrainment literature, finger taps entrained to a metronome generally shift to correct a certain fraction of an event timing perturbation on the next tap. This fraction is called α. In human subjects, α has repeatedly been observed to increase linearly with metronome period ("inter-onset interval," or IOI), exceeding 1 (i.e., over-correction) for sufficiently long IOIs [43, 44].

The PIPPET framework offers a principled explanation for α increasing with IOI. During an event-free interval, phase uncertainty increases over time. When an event does occur, the precision of the prior distribution on phase and tempo is weighed against the precision of the likelihood function associated with the expectation of that event. If the prior is less precise due to accumulated uncertainty, the precision of the likelihood weighs more heavily against it and

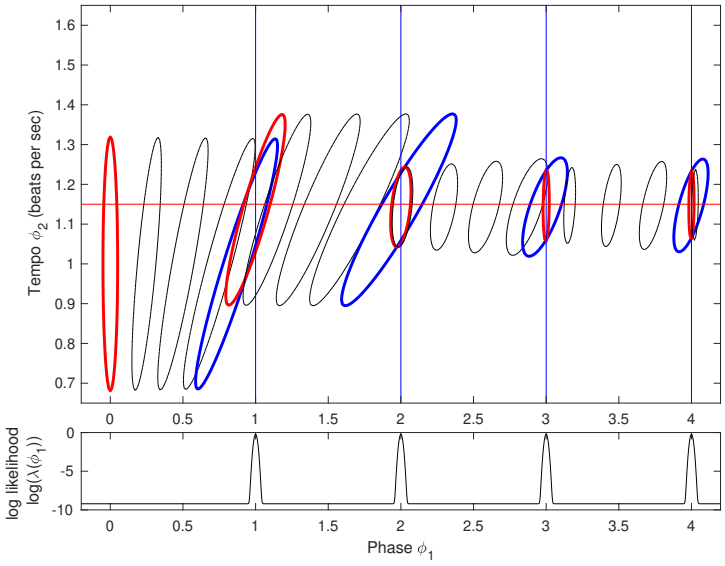

Figure 6: **A point process Kalman-Bucy Filter estimates phase and tempo.** Ellipses trace the contours of the Gaussian posterior distributions on phase and tempo. Black ellipses show a strobed visualization of the evolution of the posterior between events. Blue ellipses are the posterior distributions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

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 [43]. We were therefore curious to see whether PATIPPET could reproduce the linear increase of α with increasing IOI up to and beyond $\alpha = 1$.

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

3.7 Multiple event streams

Multi-PIPPET generalizes the PIPPET/PATIPPET framework to cases of multiple distinguishable event types, each with its own set of expectations as a function of phase. One example could be listening, tapping, or dancing to a kit drum track with bass drum, snare, and hi-hat cymbal. Timing perturbations of different instruments in drum rhythms have been shown to differently affect human entrainment [45]. By letting j take values from $\{bass, snare, hihat\}$ and choosing appropriate values for ϕ_i^j, v_i^j, and τ_i^j for each event i on the metrical grid, we can create a set of timing expectations with strength and precision dependent on the specific drum and metrical position that could then be used to optimally track underlying phase and tempo through a complex kit drum rhythm. We illustrate such a template in Figure 8. A similar setup could be

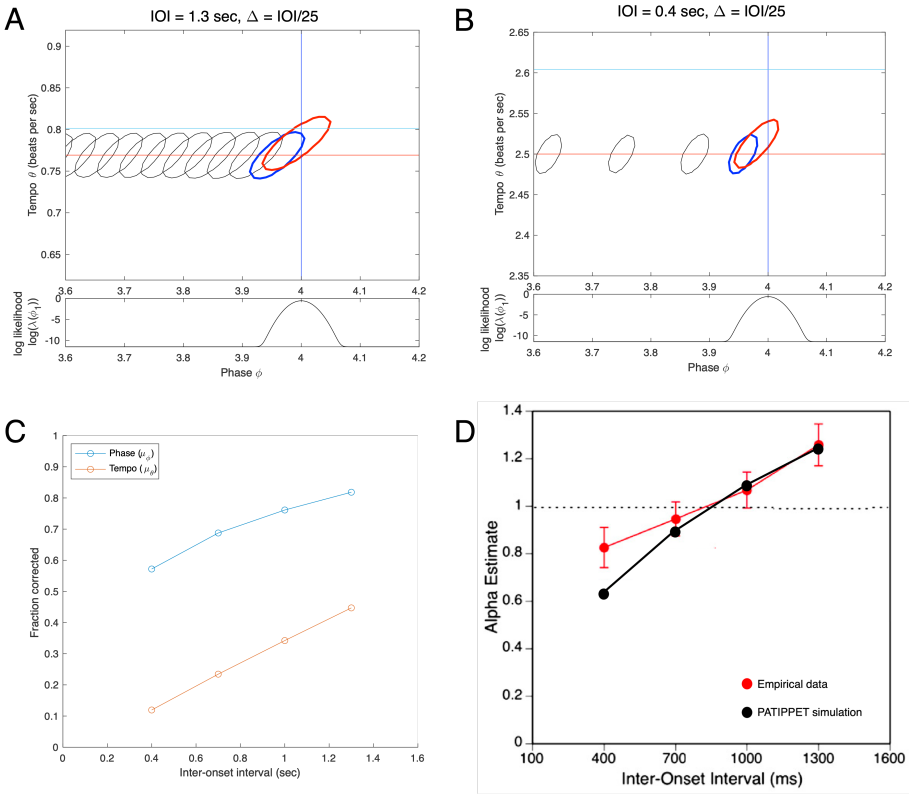

Figure 7: **PATIPPET reproduces human tapping data showing overcorrection after timing perturbations to slow metronomes.** A and B) The distribution on phase and tempo leading up to and following a phase shift at the fourth event in an isochronous sequence for two different metronome tempi (i.e., two different inter-onset intervals). See Figure 6 for color key. Note that when the IOI is short, PATIPPET arrives at the phase-shifted event with a high degree of phase and tempo certainty. C) PATIPPET makes a proportionally larger correction to phase and tempo for long IOIs than for short IOIs due to the greater degree of uncertainty preceding each event. D) Alpha (α) is the proportion of a phase shift that is corrected at the next tap time. With this set of parameters, PATIPPET reproduces the empirical observation from [44] that the phase shift is undercorrected when IOIs are short and overcorrected $\alpha > 1$ when IOIs are long.

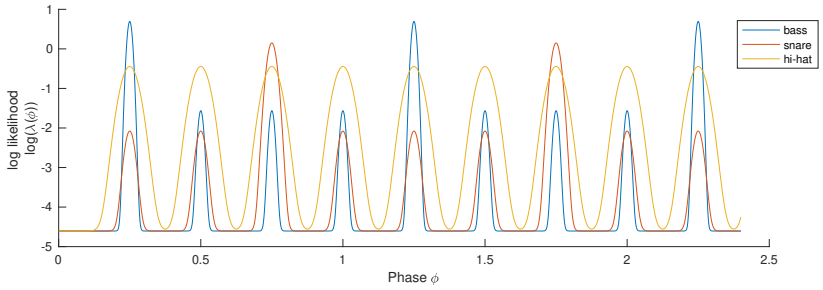

Figure 8: **Example expectation template for a basic rock beat.** In this illustration, bass drum hits are expected more strongly on the first of each cycle of four eighth notes, and are expected with high timing precision such that misplaced bass drum hits will exert a strong influence on phase. Snare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

used to implement the assumption that pitches in a melody match the harmonic context more often in strong metrical positions, allowing event attribution and timing correction during melody listening to be influenced by scale degree.

Multi-PIPPET with $j \to \infty$ 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 and therefore exert greater influence on synchronization (as observed in [46]).

Multi-PIPPET could also be useful in flexibly modeling tapping data. Experiments have shown that the presence of entrained tapping prior to temporal perturbations in a metronomic stimulus reduces the phase correction response [47], indicating that the estimate of moment-by-moment phase is influenced by the proprioceptive and auditory feedback from tapping. Given working assumptions about how taps are planned and executed based on an underlying phase estimate, the taps themselves could provide a second stream of input to the

ongoing phase estimation that would bias it toward making smaller corrections to timing perturbations.

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 [48]. 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 [49] for a discussion of such a coding strategy as a means of minimizing representational neural resources.

The PIPPET 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 [27].

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.

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

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 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., [50, 51]. 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. Our frameworks differ from these in three 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 explicitly track uncertainty in phase and tempo – without this feature, they would not account for observed dependence of phase shift response on inter-onset interval or mimic human failures to track overly-syncopated rhythms. Finally, 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.

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 [52, 53] and therefore do not serve as satisfying models of real-time behavior. Cemgil et al (2000) [31] 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 stochastic rhythms or temporal structures more complex than approximate isochrony.

Bayesian inference has also been used to model timing estimation in the brain (e.g., [24, 25]), 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 [26]. 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 [31], 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 stochastic 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 background event rate τ_0.

4.2 Motor, perceptual, and neural entrainment

Throughout this work, we have made mention of perceptual and motor expressions of entrainment, but have remained agnostic as to how we would expect to observe an expression of phase and tempo inference in humans. These two readouts sometimes give conflicting results: for example, exposure to musical performance with expressively irregular timing affects perceptual reports of timing in subsequent stimuli [54], but does not affect phase correction in tapping to subsequent stimuli [55].

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

4.3 PIPPET in the brain

If the brain is indeed performing an optimal estimation of phase and tempo, then this estimate should be legible in neural activity somewhere in the brain. At the scalp level and in intracortical electrodes, slow electrical oscillations do seem to anticipatorily track the structure of periodic auditory stimuli [56, 57], and this tracking is associated with the subjective passage of time [58]; these oscillations could be explored as possible estimates of mean underlying phase. In monkeys, the supplementary motor area appears to track the phase underlying periodic visual events [59]; recordings from this region could be another candidate for reading out mean phase. Nigrostriatal dopaminergic signaling has been identified as a possible marker of timing certainty [60, 61], so those dopaminergic populations might be a good place to look for a readout of phase variance. The temporal expectation template is a hazard function, and may therefore be observable by using techniques recently applied to decode the temporal hazard function from EEG data [62], or through its correlation with beta oscillations [63].

Though PIPPET and PATIPPET are 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 implementation of PIPPET in the brain. A detailed discussion of a possible neural basis of beat maintenance is presented in [64]. Briefly, supplementary motor area may maintain an ongoing estimate of mean phase through some combination of intrinsic dynamics and interaction with the basal ganglia, while dopaminergic signaling in striatum may maintain an estimate of phase uncertainty. The phase estimate may be used to inform auditory timing expectancy via learned models in premotor cortex [65]. These expectations may be delivered to the early stages of audition via the top-down

connections along the dorsal auditory pathway, where they can be used to evaluate timing prediction error [66]. These errors, weighted by their precisions, may be transmitted back to the supplementary motor area via the bottom-up connectivity of the dorsal auditory pathway and used to update the estimate of phase.

4.4 Learning and inference outside of PIPPET

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

First, the brain must learn the temporal structures of the expectation template for rhythmic expectation. Learning these underlying structures from an experiential corpus of noisy, stochastic rhythms is not trivial. It seems likely to involve some type of bootstrapping in which a recognition of some degree of temporal structure allows for attribution of events to positions in that structure, allowing for deeper structure learning. Earlier exposure to simpler, less stochastic rhythms would likely help with such a bootstrapping process. For a discussion of the challenges of this type of simultaneous learning and filtering and a proposed solution for non-point-process data, see [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 parameters σ and σ_θ necessarily correspond to the actual precision of the neural or external timing mechanisms in play. The brain may underestimate the noisiness (σ) of the timing process it uses to track underlying phase, leading to under-

adjustment to auditory event timing and minimal time-warping between events, or do the opposite. Presumably, these parameters must be learned through experience and prediction error.

The precision parameters v_i may be informed by several factors. 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 also be informed by 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 above). 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 [68], 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.

When the brain is exposed to a rhythmic stimulus, it must first recognize that a predictable pattern exists and select an appropriate temporal 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 temporal expectation templates. This problem might be

[1] An event can only be experienced after it occurs, so (as pointed out in [25]) 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 τ function with asymmetrical peaks and use numerical methods rather than the explicit solution derived here to estimate underlying phase at each time step.

partially addressed at a modeling level by incorporating a model of meter inference based on prior probabilities of hearing specific meters at specific tempi, e.g. [69], as an additional level of inference in parallel with phase and tempo inference.

Finally, aspects of the temporal expectation template are likely changing even as a rhythm plays out in time. This is evidenced by the grammar-like structure of music rhythm [70]: certain patterns of events are more expected than others regardless of their metrical positions. PIPPET and PATIPPET take a template of expected event time points as an input, and thus do not take into account immediate stimulus history in creating expectations. However, such effects could be incorporated into a model based on this framework by adding a history dependence to the expectation template τ. The precise details of this history dependence could be based on any suitable formal model for rhythmic grammar (e.g., [71, 72, 70]).

4.5 Future directions

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

However, there is a certain extent to which these frameworks can be validated as descriptions of human cognition. First, these models predict certain qualitative effects such as the slowing of perceived phase advance as strong expectations are disappointed. Second, although the parameters in the forward models are not directly empirically measurable values, changes in stimulus his-

tory should influence them in predictable ways. For example, if a certain type of event occurs consistently at a particular metrical position within an extended stimulus presentation or within the music the listener has experienced in a lifetime of listening, then it should induce stronger phase corrections than an event that occurs inconsistently as if it has been given a higher value of τ_i. Parameters may also be influenced by long term listening experience, but they should at least respond to recent empirical experience by changing in the direction predicted by PIPPET.

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

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

One specific question that the PIPPET framework might help resolve is how periodic and nonperiodic entrainment differ. PIPPET has no specific machinery to account for ways in which the two situations differ (for neural and behavioral evidence of differences between memory-based and periodicity based entraiment, see, e.g., [14, 6]. However, since it is sufficiently general to model both, it could guide an exploration of parameter differences between the performance of similar tasks in periodic and aperiodic contexts.

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

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

5 Acknowledgments

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

6 Appendix

6.1 The PATIPPET filter

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

$$\begin{cases} d\bar{\boldsymbol{x}}_t = & \begin{pmatrix} \bar{\theta} \\ 0 \end{pmatrix} dt + (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}_{t-}) \cdot (dN_t - \hat{\Lambda} dt) \\ d\boldsymbol{\Sigma} = & \begin{pmatrix} 2\Sigma^{12} + \sigma_\phi^2 & \Sigma^{22} \\ \Sigma^{22} & \sigma_\theta^2 \end{pmatrix} dt + \left(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{t-} \right) \cdot (dN_t - \hat{\Lambda} dt) \end{cases} \tag{9}$$

where we define

$$\begin{cases} \hat{\Lambda} := & \sum_{i=0,1,\cdots} T_i \hat{\theta}_i \\ \hat{\boldsymbol{x}} = & \frac{1}{\hat{\Lambda}} \sum_{i=0,1,\cdots} T_i \begin{pmatrix} K_i^{12} + \hat{\phi}_i \hat{\theta}_i \\ K_i^{22} + \hat{\theta}_i^2 \end{pmatrix} \\ \hat{\boldsymbol{\Sigma}} := & \frac{1}{\hat{\Lambda}} \sum_{i=0,1,\cdots} T_i \Big(\hat{\theta}_i \boldsymbol{K}_i + \hat{\theta}_i (\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+})(\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+})^T \\ & + (\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+}) \begin{pmatrix} K_i^{21} & K_i^{22} \end{pmatrix} + \begin{pmatrix} K_i^{12} \\ K_i^{22} \end{pmatrix} (\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+})^T \Big) \end{cases}$$

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

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

$T_0 := \tau_0$, $T_i := \tau_i N(\phi_i | \bar{\phi}, v_i^{-1} + V^{-1})$ for $i > 0$.

$\hat{\boldsymbol{x}}_i = \begin{pmatrix} \hat{\phi}_i \\ \hat{\theta}_i \end{pmatrix} := \boldsymbol{K}_i (\boldsymbol{P}_i \boldsymbol{x}_i + \boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{x}})$ for $i > 0$, and $\hat{\boldsymbol{x}}_0 := \bar{\boldsymbol{x}}$.

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

6.2 Derivation of differential equations and update equations.

We derive the PATIPPET filter first, and then derive the PIPPET filter as a special case.

Snyder [27] 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:

$$d\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}dt + \boldsymbol{B}d\boldsymbol{W}_t \tag{10}$$

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

$$dp_t(\boldsymbol{x}) = \mathcal{L}[p_{t-}(\boldsymbol{x})]dt + p_{t-}(\boldsymbol{x})\left(\frac{\lambda(\boldsymbol{x})}{\hat{\Lambda}} - 1\right) \cdot (dN_t - \hat{\Lambda}dt) \tag{11}$$

where $\hat{\Lambda} := \mathbb{E}[\lambda(\boldsymbol{x})]$ (with $\mathbb{E}$ denoting expectation under distribution $p_{t-}(\boldsymbol{x})$), 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 (10):

$$\mathcal{L}[p(\boldsymbol{x})] = -\sum_i \frac{\partial}{\partial x_i}[\boldsymbol{A}\boldsymbol{x}]_i p(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}[\boldsymbol{B}\boldsymbol{B}^T]_{ij} p(\boldsymbol{x}) \tag{12}$$

Here we project p onto a Gaussian distribution at each time step by matching mean $\bar{\boldsymbol{x}}$ and covariance $\boldsymbol{\Sigma}$, 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:

$$\begin{aligned} d\bar{\boldsymbol{x}}_t =& \bar{\boldsymbol{x}}_{t+} - \bar{\boldsymbol{x}}_{t-} = \int_{\boldsymbol{x}} \boldsymbol{x} p_{t+}(\boldsymbol{x}) d\boldsymbol{x} - \int_{\boldsymbol{x}} \boldsymbol{x} p_{t-}(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}} \boldsymbol{x} \left(p_{t+}(\boldsymbol{x}) - p_{t-}(\boldsymbol{x}) \right) (\boldsymbol{x}) d\boldsymbol{x} = \int_{\boldsymbol{x}} \boldsymbol{x} dp_t(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}} \boldsymbol{x} \mathcal{L}[p_t(\boldsymbol{x})] dt d\boldsymbol{x} + (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}_{t-}) \cdot (dN_t - \hat{\Lambda} dt) \end{aligned} \tag{13}$$

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

$$d\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t+} - \boldsymbol{\Sigma}_{t-} = \int_{\boldsymbol{x}} [[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 p_{t+}(\boldsymbol{x}) d\boldsymbol{x} - \int_{\boldsymbol{x}} [[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t-}]]^2 p_{t-}(\boldsymbol{x}) d\boldsymbol{x}$$

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

$$\begin{aligned} d\boldsymbol{\Sigma}_t =& \int_{\boldsymbol{x}} [[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 \left(p_{t+}(\boldsymbol{x}) - p_{t-}(\boldsymbol{x}) \right) d\boldsymbol{x} \\ &+ \int_{\boldsymbol{x}} \left([[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 - [[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t-}]]^2 \right) p_{t-}(\boldsymbol{x}) d\boldsymbol{x} \\ =& \int_{\boldsymbol{x}} [[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 dp_t(\boldsymbol{x}) - [[\bar{\boldsymbol{x}}_{t+} - \bar{\boldsymbol{x}}_{t-}]]^2 \\ =& \int_{\boldsymbol{x}} [[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 \mathcal{L}[p_t(\boldsymbol{x}|N_t)] dt d\boldsymbol{x} + \left(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{t-} \right) \cdot (dN_t - \hat{\Lambda} dt) \end{aligned} \tag{14}$$

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

Integrating by parts (or following [30]), we can calculate the appropriate integrals of $\mathcal{L}[p_t(\boldsymbol{x}|N_t)]$, arriving at a general expression for the variational Bayesian filter for point process data:

$$\begin{cases} d\bar{\boldsymbol{x}}_t = & \boldsymbol{A}\bar{\boldsymbol{x}}_{t-}dt + (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}_{t-}) \cdot (dN_t - \hat{\Lambda}dt) \\ d\boldsymbol{\Sigma}_t = & (\boldsymbol{A}\boldsymbol{\Sigma}_{t-} + \boldsymbol{\Sigma}_{t-}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{B}^T)dt + \left(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{t-}\right) \cdot (dN_t - \hat{\Lambda}dt) \end{cases} \tag{15}$$

From (4), the PATIPPET generative model is described by the Gauss-Markov diffusion process (10) with

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

$$\boldsymbol{\Sigma} = \begin{pmatrix} V & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}$$

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

Plugging into (15), we have

$$\begin{cases} d\bar{\boldsymbol{x}}_t = & \begin{pmatrix} \bar{\theta} \\ 0 \end{pmatrix} dt + (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}_{t-}) \cdot (dN_t - \hat{\Lambda}dt) \\ d\boldsymbol{\Sigma} = & \begin{pmatrix} 2\Sigma^{12} + \sigma_\phi^2 & \Sigma^{22} \\ \Sigma^{22} & \sigma_\theta^2 \end{pmatrix} dt + \left(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{t-}\right) \cdot (dN_t - \hat{\Lambda}dt) \end{cases} \tag{16}$$

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

Let $\|x\|_A^2$ denote $x^T Ax$. 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 [73] for proof and similar application):

$$\|x-a\|_A^2+\|x-b\|_B^2=\|a-b\|_{A(A+B)^{-1}B}^2+\|x-(A+B)^{-1}(Aa+Bb)\|_{A+B}^2 \quad (17)$$

In the PATIPPET generative model, events are generated at rate $\lambda(\boldsymbol{x}) = \theta\tau(\phi) = \theta\tau(\phi)$, where

$$\tau(\phi)=\tau_0+\sum_{i=1,2,\cdots}\frac{\tau_i}{\sqrt{2\pi v_i}}e^{-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_i\|_{\boldsymbol{P}_i}^2}$$

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

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

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

We calculate:

$$\begin{aligned}p(\boldsymbol{x})\tau(\boldsymbol{x}) =& \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}|}}e^{-\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{\boldsymbol{\Sigma}^{-1}}^2}\left(\tau_0+\sum_{i=1,2,\cdots}\frac{\tau_i}{\sqrt{2\pi v_i}}e^{-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_i\|_{\boldsymbol{P}_i}^2}\right)\\ =& \frac{\tau_0}{\sqrt{2\pi|\boldsymbol{\Sigma}|}}e^{-\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{\boldsymbol{\Sigma}^{-1}}^2}+\sum_{i=1,2,\cdots}\frac{\tau_i}{2\pi\sqrt{v_i|\boldsymbol{\Sigma}|}}e^{-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_i\|_{\boldsymbol{P}_i}^2-\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{\boldsymbol{\Sigma}^{-1}}^2}\end{aligned}$$

Applying (17),

$$\begin{aligned}p(\boldsymbol{x})\tau(\boldsymbol{x}) =& \frac{\tau_0}{\sqrt{2\pi|\boldsymbol{\Sigma}|}}e^{-\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{\boldsymbol{\Sigma}^{-1}}^2}\\ &+\sum_{i=1,2,\cdots}\tau_i\left(\frac{1}{\sqrt{2\pi(v_i^{-1}+V^{-1})}}e^{-\frac{1}{2}\|\boldsymbol{x}_i-\bar{\boldsymbol{x}}\|_{\boldsymbol{P}_i\boldsymbol{K}_i\boldsymbol{\Sigma}^{-1}}^2}\right)\left(\frac{1}{\sqrt{2\pi\frac{v_i|\boldsymbol{\Sigma}|}{v_i^{-1}+V^{-1}}}}e^{-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{K}_i(\boldsymbol{P}_i\boldsymbol{x}_i+\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}})\|_{\boldsymbol{K}_i^{-1}}^2}\right)\end{aligned} \quad (18)$$

where we define $\boldsymbol{K}_i := (\boldsymbol{P}_i + \boldsymbol{\Sigma}^{-1})^{-1}$.

We next calculate:

$$\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|^2_{\boldsymbol{P}_i\boldsymbol{K}_i\boldsymbol{\Sigma}^{-1}} = \frac{(\phi_i - \phi)^2}{v_i^{-1} + V^{-1}}$$

and

$$|\boldsymbol{K}_i| = \frac{v_i|\boldsymbol{\Sigma}|}{v_i^{-1} + V^{-1}}$$

We use these expressions to write (18) in terms of normal distributions:

$$p(\boldsymbol{x})\tau(\boldsymbol{x}) = \tau_0 N(\boldsymbol{x}|\bar{\boldsymbol{x}}, \boldsymbol{\Sigma}) + \sum_{i=1,2,\cdots} \tau_i N(\phi_i|\bar{\phi}, v_i^{-1} + V^{-1}) N(\boldsymbol{x}|\boldsymbol{K}_i(\boldsymbol{P}_i\boldsymbol{x}_i + \boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}), \boldsymbol{K}_i) \tag{19}$$

We simplify this expression by defining $T_i := \tau_i N(\phi_i|\bar{\phi}, v_i^{-1} + V^{-1})$ for $i > 0$, and setting $T_0 := \tau_0$ and $\boldsymbol{K}_0 = \boldsymbol{\Sigma}$. We define $\hat{\boldsymbol{x}}_i := \begin{pmatrix} \hat{\phi}_i \\ \hat{\theta}_i \end{pmatrix} := \boldsymbol{K}_i(\boldsymbol{P}_i\boldsymbol{x}_i + \boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}})$ for $i > 0$ and set $\hat{\boldsymbol{x}}_0 := \bar{\boldsymbol{x}}$. This lets us write

$$p(\boldsymbol{x})\tau(\boldsymbol{x}) = \sum_{i=0,1,\cdots} T_i N(\boldsymbol{x}|\hat{\boldsymbol{x}}_i, \boldsymbol{K}_i) \tag{20}$$

We use this expression and the moments of normal distributions to calculate $\hat{\Lambda}$, $\hat{\boldsymbol{x}}$, and $\hat{\boldsymbol{\Sigma}}$:

$$\hat{\Lambda} := \mathbb{E}\left[\lambda(\boldsymbol{x})\right] = \mathbb{E}\left[\theta\tau(\boldsymbol{x})\right] = \sum_{i=0,1,\cdots} T_i\hat{\theta}_i \tag{21}$$

$$\hat{\boldsymbol{x}} := \frac{1}{\hat{\Lambda}}\mathbb{E}\left[\boldsymbol{x}\lambda(\boldsymbol{x})\right] = \frac{1}{\hat{\Lambda}}\mathbb{E}_p\left[\boldsymbol{x}\theta\tau(\boldsymbol{x})\right] = \frac{1}{\hat{\Lambda}}\int_{\boldsymbol{x}} \begin{pmatrix} \phi\theta \\ \theta^2 \end{pmatrix} p(\boldsymbol{x})\tau(\boldsymbol{x})d\boldsymbol{x}$$

This expression picks out non-central second moment terms of each normal distributions in 20, 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 $\boldsymbol{K}_i$, we can write

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

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

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}} :=& \frac{1}{\hat{\Lambda}} \mathbb{E}_p \left[[[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 \lambda(\boldsymbol{x}) \right] = \frac{1}{\hat{\Lambda}} \mathbb{E}_p \left[[[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 \theta \tau(\boldsymbol{x}) \right] \\
=& \frac{1}{\hat{\Lambda}} \sum_{i=0,1,\cdots} T_i \mathbb{E}_p \left[\theta [[\boldsymbol{x} - \bar{\boldsymbol{x}}_{t+}]]^2 N(\boldsymbol{x} | \hat{\boldsymbol{x}}_i, \boldsymbol{K}_i) \right] \\
=& \frac{1}{\hat{\Lambda}} \sum_{i=0,1,\cdots} T_i \hat{\theta}_i \mathbb{E}_p [[[\boldsymbol{x} - \hat{\boldsymbol{x}}_i]]^2 N(\boldsymbol{x} | \hat{\boldsymbol{x}}_i, \boldsymbol{K}_i)] \\
& + T_i \hat{\theta}_i [[\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+}]]^2 \\
& + T_i (\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+}) \mathbb{E}_p [(\boldsymbol{x} - \hat{\boldsymbol{x}}_i)^T (\theta - \hat{\theta}_i) N(\boldsymbol{x} | \hat{\boldsymbol{x}}_i, \boldsymbol{K}_i)] \\
& + T_i \mathbb{E}_p [(\boldsymbol{x} - \hat{\boldsymbol{x}}_i)(\theta - \hat{\theta}_i) N(\boldsymbol{x} | \hat{\boldsymbol{x}}_i, \boldsymbol{K}_i)] (\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+})^T] \qquad (23) \\
=& \frac{1}{\hat{\Lambda}} \sum_{i=0,1,\cdots} T_i [\hat{\theta}_i \boldsymbol{K}_i + \hat{\theta}_i [[\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+}]]^2 \\
& + (\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+}) \begin{pmatrix} K_i^{21} & K_i^{22} \end{pmatrix} + \begin{pmatrix} K_i^{12} \\ K_i^{22} \end{pmatrix} (\hat{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{t+})^T] \qquad (24) \\
& \qquad (25)
\end{aligned}$$

These expressions coupled with (28) 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 $\mathbf{\Sigma}$ to zero except V. However, this requires finessing various degeneracies, e.g. wherever $\mathbf{\Sigma}$ is inverted. More straightforward is to follow the same process as above, starting from the PIPPET generative model (3). Either way ultimately yields the PIPPET filter (3).

For multiple event streams j,:

$$dp_t(\boldsymbol{x}) = \mathcal{L}[p_t(\boldsymbol{x})]dt + p_t(\boldsymbol{x}) \sum_j (\lambda_j(\phi) - \mathbb{E}_p[\lambda_j(\phi)]) \cdot (\mathbb{E}_p[\lambda_j(\phi)]^{-1} dN_j - dt) \quad (26)$$

This follows directly from application of the derivation above to equation (5) in [74] with a discrete spatial dimension. By the methods above, it yields the multi-PIPPET filter (8) and the multi-PATIPPET filter:

$$\begin{cases} d\mu = dt - \sum_j (\mu^{*j} - \mu)(dN_t^j - \hat{\Lambda}^j dt) \\ d\Sigma = \sigma^2 dt - \sum_j (\Sigma^{*j} - \Sigma)(dN_t^j - \hat{\Lambda}^j dt) \end{cases} \quad (27)$$

and the multi-PATIPPET filter:

$$\begin{cases} d\bar{\boldsymbol{x}}_t = \begin{pmatrix} \bar{\theta} \\ 0 \end{pmatrix} dt + \sum_j \left(\hat{\boldsymbol{x}}^j - \bar{\boldsymbol{x}}_{t-} \right) \cdot (dN_t^j - \hat{\Lambda}^j dt) \\ d\mathbf{\Sigma} = \begin{pmatrix} 2\Sigma^{12} + \sigma_\phi^2 & \Sigma^{22} \\ \Sigma^{22} & \sigma_\theta^2 \end{pmatrix} dt + \sum_j \left(\hat{\mathbf{\Sigma}}^j - \mathbf{\Sigma}_{t-} \right) \cdot (dN_t^j - \hat{\Lambda}^j dt) \end{cases} \quad (28)$$

6.3 Simulation parameters.

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

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

Figure 1: $\phi_i = t_i = \{0.5, 1, 1.5\}$, $v_i = 0.0001$, $\tau_i = 0.02$, $\tau_0 = 0.01$, $\sigma = 0.05$

Figure 2A: $\phi_i = t_i = \{0.25, 0.5, 0.75, 1\}$, $v_i = 0.0001$, $\tau_i = 2$, $\tau_0 = 0.01$, $\sigma = 0.05$.

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

Figure 3A:

$$
\begin{aligned}
t_i &= \{0, 0.150, 0.25, 0.5, 0.65, 0.9, 1\} \\
\phi_i &= \{0, 0.15, 0.25, 0.4, 0.5, 0.65, 0.75, 0.9, 1, 1.15\} \\
v_i &= \{.0001, .0005, .0001, .0005, .0001, .0005, .0001, .0005\} \\
\tau_i &= \{.05, .01, .05, .01, .05, .01, .05, .01\} \\
\tau_0 &= 0.01 \\
\sigma &= 0.05
\end{aligned}
$$

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

Figure 4: Same as Figure 3A, but with $t_i = \{0, 0.15, .65, .9, 1.15, 1.25\}$.

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

$$
\begin{aligned}
\phi_i^j &= 0.25i \text{ for } j = bass, snare, hihat \\
v_i^{bass} &= .0001, v_i^{snare} = .0003, v_i^{hihat} = .001 \\
\tau_i^{bass} &= \{.05, .005, .005, .005, \dots\} \\
\tau_i^{snare} &= \{.005, .005, .05, .005, \dots\} \\
\tau_i^{hihat} &= \{.05, .05, .05, .05, \dots\} \\
\tau_0 &= 0.01
\end{aligned}
$$

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

Figure 6:

$$
\begin{aligned}
t_i &= \frac{i}{1.15} \\
\phi_i &= i \\
v_i &= \{.0001, .0003, .0001, .0003, .0001, .0003, .0001, .0003\} \\
\tau_i &= \{.02, .01, .02, .01, .02, .01, .02, .01\} \\
\tau_0 &= 10^{-4} \\
\sigma &= 0.05 \\
\sigma_\theta &= 0.05 \\
\boldsymbol{\mu_0} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\boldsymbol{\Sigma_0} &= \begin{pmatrix} .001 & 0 \\ 0 & .04 \end{pmatrix}
\end{aligned}
$$

Figure 7: In four simulations, we set the inter-onset interval Δ to $0.4s$, $0,7s$,

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

$$
\begin{aligned}
t_i &= \{\Delta, 2\Delta, 3\Delta, 4\Delta + \delta\} \\
\phi_i &= i \\
v_i &= 0.0002 \\
\tau_i &= \{.02, .01, .02, .01, .02, .01, .02, .01\} \\
\tau_0 &= 10^{-5} \\
\sigma &= 0.01 \\
\sigma_\theta &= 0.01 \\
\boldsymbol{\mu_0} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
\boldsymbol{\Sigma_0} &= \begin{pmatrix} 10^{-4} & 0 \\ 0 & 10^{-4} \end{pmatrix}
\end{aligned}
$$

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. 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

14. 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

15. 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

16. 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

17. Friston K. Does predictive coding have a future? Nature Neuroscience 2018; 21:1019–21. DOI: 10.1038/s41593-018-0200-7

18. 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

19. 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

20. 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

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

22. 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

23. 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

24. 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

25. 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

26. 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

27. 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

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

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

30. Eden UT and Brown EN. CONTINUOUS-TIME FILTERS FOR STATE ESTIMATION FROM POINT PROCESS MODELS OF NEURAL DATA. Statistica Sinica 2008; 18:1293–310

31. 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

32. 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

33. 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

34. 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

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

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

37. 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

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

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

40. 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

41. 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

42. 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

911 43. Repp BH. Tapping in synchrony with a perturbed metronome: The phase
912 correction response to small and large phase shifts as a function of tempo.
913 Journal of Motor Behavior 2011; 43:213–27. DOI: 10.1080/00222895.
914 2011.561377

915 44. Repp BH, Keller PE, and Jacoby N. Quantifying phase correction in sen-
916 sorimotor synchronization: Empirical comparison of three paradigms. Acta
917 Psychologica 2012; 139:281–90. DOI: 10.1016/j.actpsy.2011.11.002.
918 Available from: http://dx.doi.org/10.1016/j.actpsy.2011.11.002

919 45. Witek MA, Clarke EF, Kringelbach ML, and Vuust P. Effects of Poly-
920 phonic Context, Instrumentation, and Metrical Location on Syncopation
921 in Music. Music Perception 2014; 32:201–17

922 46. Hove MJ, Marie C, Bruce IC, and Trainor LJ. Superior time perception for
923 lower musical pitch explains why bass-ranged instruments lay down musical
924 rhythms. Proceedings of the National Academy of Sciences of the United
925 States of America 2014; 111:10383–8. DOI: 10.1073/pnas.1402039111

926 47. Repp BH. Phase Correction , Phase Resetting , and Phase Shifts After Sub-
927 liminal Timing Perturbations in Sensorimotor Synchronization. Journal
928 of Experimental Psychology: Human Perception and Performance 2001;
929 27:600–21. DOI: 10.1037//0096-1523.27.3.600

930 48. Heggli OA, Cabral J, Konvalinka I, Vuust P, and Kringelbach ML. A Ku-
931 ramoto model of self-other integration across interpersonal synchronization
932 strategies. PLoS Computational Biology 2019; 15:1–17. DOI: 10.1371/
933 journal.pcbi.1007422

934 49. Koban L, Ramamoorthy A, and Konvalinka I. Why do we fall into sync
935 with others? Interpersonal synchronization and the brain's optimization
936 principle. Social Neuroscience 2019; 14:1–9

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

51. 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

52. 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

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

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

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

56. 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

57. 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

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

59. 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

60. 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

61. 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

62. 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

63. 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

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

65. 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

66. 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

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. 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

69. 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

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

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

72. 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

73. 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

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