Journal of Mathematical Psychology 122 (2024) 102869

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Inferred representations behave like oscillators in dynamic Bayesian models of beat perception

Jonathan Cannon [a,*,1], Thomas Kaplan [b,c,2]

[a] Department of Psychology, Neuroscience & Behavior, McMaster University, Hamilton, Ontario, Canada
[b] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, UK
[c] William Harvey Research Institute, Barts and the London Faculty of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK

ARTICLE INFO

Keywords:
Rhythm
Beat
Bayesian brain
Predictive processing
Oscillators
Entrainment
Model

ABSTRACT

The human's capacity to perceptually entrain to an auditory rhythm has been repeatedly modeled as a dynamical system consisting of one or more forced oscillators. However, a more recent perspective, closely related to the popular theory of Predictive Processing, treats auditory entrainment as an inference process in which the observer infers the phase, tempo, and/or metrical structure of an auditory stimulus based on event timing. Here, we propose a close relationship between these two perspectives. We show for the first time that a system performing variational Bayesian inference about the circular phase underlying a rhythmic stimulus takes the form of a forced, damped oscillator with a specific nonlinear phase response function corresponding to the internal metrical model of the underlying rhythm. This algorithm can be extended to simultaneous inference on both phase and tempo using one of two possible approximations that closely align with the two most prominent models of auditory entrainment: one yields a single oscillator with an adapting period, and the other yields a networked bank of oscillators. We conclude that an inference perspective on rhythm perception can offer similar descriptive power and flexibility to a dynamical systems perspective while also plugging into the fertile unifying framework of Bayesian Predictive Processing.

1. Introduction

Humans show a strong capacity to perceptually identify recurring temporal patterns underlying an auditory stimulus and use them to contextualize the stimulus as it unfolds (henceforth referred to as "auditory entrainment"). In music listening, this capacity gives rise to the experience of the "beat", a subjective impression of a (usually) periodic pulse at a walking or running tempo (Merchant et al., 2015). This pulse is consistent with, but may not be unambiguously specified by, the objective temporal structure of the music (the "rhythmic surface"). Auditory entrainment occurs alongside entrainment of neural activity, and entrained neural signals seem to emphasize the level of periodicity associated with the experienced beat (Nozaradan, 2013, 2014).

In recent decades, interest has grown in understanding the cognitive and neurophysiological substrates of the experience of rhythmic structure, with particular emphasis on the beat. An expression of this interest has been a series of efforts to build mathematical models of rhythm/beat perception. One fruitful modeling thread posits that physical oscillations in the brain synchronize to external rhythms, giving rise to beat perception as well as other rhythmic aspects of perception. These oscillations are modeled as dynamical systems made up of one or more oscillatory components. This thread has birthed a series of models that explain a range of facets of human rhythm perception in terms of generic properties of such dynamical systems.

A more recent modeling thread posits that perception of beat and metre in rhythm represents an "inference" about underlying structure based on evidence presented by the rhythmic surface. Although this perspective has not yet achieved the same descriptive range as oscillator models, it is appealing in that its mathematical and conceptual language naturally connects rhythm perception with other perceptual and motor processes within the same modeling framework.

Below, we briefly review these two modeling threads. Although it is difficult to fully separate models of rhythm perception from models of physical entrainment to rhythm, we attempt to narrow our scope to the former (but see Palmer and Demos (2022) for a similar review including the latter).

* Corresponding author.
E-mail address: cannoj9@mcmaster.ca (J. Cannon).

[1] JC was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
[2] TK was supported by a doctoral studentship from the Engineering and Physical Sciences Research Council, United Kingdom and Arts and Humanities Research Council, United Kingdom Centre for Doctoral Training in Media and Arts Technology [EP/L01632X/1].

https://doi.org/10.1016/j.jmp.2024.102869
Received 31 December 2023; Received in revised form 22 May 2024; Accepted 2 July 2024
Available online 23 July 2024

1.1. Oscillator models of auditory entrainment

Early models of auditory entrainment as oscillation were employed to describe the perception of musical meter (Large & Kolen, 1994; Large & Palmer, 2002), and more generally to describe the entrainment of periodic fluctuations of attention and time perception (Large & Jones, 1999; McAuley, 1995; McAuley & Jones, 2003). In these models, the state of the entrained perceptual process (e.g., attention) was described by a single phase variable (a "phase oscillator") that was assumed to advance steadily around the circle with time, and rhythmic sensory events were assumed to perturb this phase differently depending on the current phase at the time of the event (i.e., according to a "phase response curve"). In some formulations, the phase response curve was defined as the gradient of an abstract "expectancy function" that specified how much an event was expected at each phase: the phase advanced when events occurred slightly earlier than expected and regressed when they occurred slightly later. This basic model allowed the perceptual process to entrain to a range of rhythm periods near its characteristic unforced frequency. The model is then extended to adapt its tempo/period according to a similar response function, allowing it to entrain at a range of tempi, and sometimes with an adaptable phase response curve that makes event prediction more precise when predictions of temporal regularity are repeatedly satisfied. We shall refer to this class of models as "adaptive oscillators".

A separate set of models of auditory entrainment used "banks" of linear (Todd et al., 2002) or nonlinear (Large et al., 2010) oscillators with a gradient of intrinsic frequencies. Oscillators in these models have amplitudes as well as phases — metrical structure is identified by the amplitudes of groups of oscillators in response to the forcing input, where *meter* can be defined as a perceived and hierarchically embedded collection of recurring temporal periods within a stimulus (the *beat* generally being the most salient) (London, 2004). Damping terms that cause amplitudes to decay over time allow them to account for the loss of entrainment after the offset of a rhythmic stimulus. Learnable coupling strengths between oscillators arranged in multiple layers give these models formidable capacity to pick out underlying pulses from complex rhythms (Kim & Large, 2021; Tichko & Large, 2019) and continue them in the absence of rhythmic input if appropriately modulated (Large et al., 2015). We shall refer to this class of models as "gradient frequency neural networks".

These models fit more or less comfortably into the perspective that properties of behavior and perception emanate first and foremost from the generic properties of dynamical systems. This perspective can account for many aspects of rhythm perception and production with appropriate parameter and feature choices. However, it offers us little insight into the cognitive and evolutionary facets of human rhythmicity. Can our ability and tendency to entrain movement to complex rhythms within a specific tempo range rightly be described as an example of a broader class of perceptual processes in which we use sensation to identify underlying structure? Even a novice musician can flexibly interact with a beat to create a wide range of rhythms — it may be possible to account for this phenomenon with elaborate networks of oscillators, but it becomes increasingly appealing to break out of the strict dynamical systems perspective and allow the brain some type of "internal representation" of a rhythm that it can flexibly utilize.

1.2. Entrainment by inference

A distinct approach to modeling auditory entrainment treats it as a process of inference — identifying and making sense of what is happening in the world. Where oscillator approaches to human rhythmicity use math to describe the dynamics of hypothesized physiological processes, inference approaches use math to describe the process of integrating sensory evidence into representations of rhythmic stimuli. Sensory information can inform a representation of something in the world via some understanding of how things in the world produce sensation; we will call such an understanding a "generative model". Based on this model, the listener can back-track from sensation to cause (i.e., "invert" the model). Unlike the modeling language of oscillators, this language is essentially cognitive and representational — the dynamic states in such a model are "representations" that carry meaning about something in the world. Importantly, our generative models are thought to be inherently stochastic, and therefore our representations include not only estimates of world states, but also of uncertainty regarding those states. Since the mathematics of inverting stochastic causal models draws on Bayes Rule, this perspective on the brain is often referred to as the "Baysian Brain" hypothesis.

The Bayesian Brain is the groundwork of the theory of Predictive Processing (at least in its most recent incarnations: Friston, 2005, 2010), which proposes that the updating of representations of the world based on sensory data is performed by adjusting representations to minimize error between the sensory data predicted by the model and actual sensory data. Note that Predictive Processing has been extended to a hypothesized set of neurophysiological mechanisms that implement the Bayesian computations, which should be considered distinct from the essentially cognitive foundation of the theory. See Friston (2005, 2010) for discussion of the value of the Predictive Processing/Bayesian Brain perspective, and Koelsch et al. (2019) for its value in understanding music cognition.

In the first work connecting rhythm perception directly to Predictive Processing, Vuust and Witek (2014), the authors proposed that listeners carry a generative model of when sound events should occur within the metrical structure of a rhythm, and interpret all incoming rhythmic information in light of this model, using it to continuously infer and update a representation of the underlying meter. (This proposal was fleshed out more fully in Vuust et al. (2018)). Other modeling work has proposed that listeners may continuously infer whether a rhythmic input stream is caused by a single agent or two (Elliott et al., 2014). A recent model (Heggli et al., 2021) suggests that even self-produced sounds may be inferred to come from the same source as external sounds; such self/other integration may then influence the dynamics of synchronization (also see Koban et al., 2019).

In our recent work, we have modeled entrainment to a rhythm as a process of approximate Bayesian inference about the progress of a stimulus through an expected metrical structure, which may also include inference about the speed (or tempo) of the stimulus (Cannon, 2021). We further developed this model to include inference over multiple possible metrical templates (Kaplan et al., 2022), as hypothesized by Vuust and Witek (2014). However, both of these models were formulated in a manner agnostic to periodicity – what was inferred was linear progress through a structure that might or might not be periodic – and therefore the oscillator dynamics that provide such an elegant description of entrainment to a beat were not present.

Several authors have proposed that Bayes-like inference about the phase and/or tempo of a stimulus may be performed by oscillators. Heggli et al. (2019) model dyadic synchronization by treating each partner as a pair of coupled oscillators that can be loosely interpreted as models of the self and other. Doelling et al. (2023) show that a forced oscillator with adaptive frequency reproduces Bayesian aspects of human judgements about rhythmic timing. In both these cases, oscillators are shown to have convenient inference-like characteristics, but in neither case is the state of the oscillator mapped onto a meaningful representation in an actual formal inference problem.

1.3. Objective

Here we aim to demonstrate that an inference perspective on rhythm perception naturally gives rise to the dynamics and mathematics of forced and coupled oscillators. In other words, if we adopt the view that our perception of rhythmic structure can be usefully described as a representation of a dynamic process underlying the rhythmic surface, it leads us to models of rhythm perception that

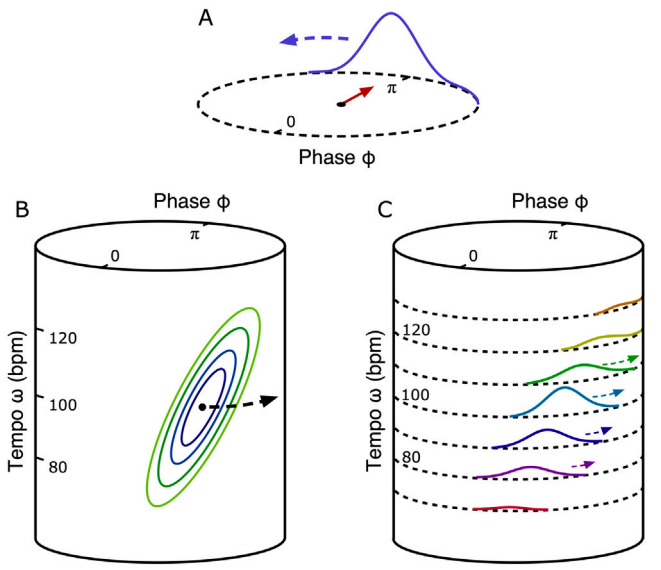

Fig. 1. Three proposed models of rhythm perception as inference. (A) c-PIPPET estimates the instantaneous circular phase underlying a rhythmically patterned stimulus, tracking a Gaussian-like distribution over possible phases that progresses around the circle based on a known underlying tempo (while at the same time spreading out as phase uncertainty increases) and that resets at each sound event. A vector representation of this distribution (red arrow) acts like a damped linear oscillator with nonlinear forcing. (B) Variational c-PATIPPET estimates the instantaneous phase and tempo of the stimulus, tracking a multivariate Gaussian-like distribution over phase and tempo that progresses around the phase/tempo cylinder (while at the same time tilting and spreading) and that resets at each sound event. (C) Gradient c-PATIPPET also estimates the instantaneous phase and tempo of the stimulus, but instead tracks a distribution over phase and tempo in the form of a collection of phase distributions conditioned on a gradient of possible tempi, each progressing around the circle at its own rate and resetting at sound events.

mathematically resemble existing oscillator models and show a similar descriptive repertoire reproducing key properties of human rhythm perception.

Our demonstration takes the form of three models, illustrated in Fig. 1. The first, c-PIPPET (Cycle Phase Inference from Point Process Event Timing, Fig. 1A), describes a formal inference process in which the listener continuously makes their best possible probabilistic estimate of the momentary phase of a rhythmically patterned stimulus given a probabilistic internal model of the stimulus/partner's behavior. In short, we find that the dynamics of a properly designed oscillator solve the formal "phase inference" problem necessary to make metrical sense of a complex rhythm.

The second and third models extend c-PIPPET to include inference about tempo as well as phase; both are therefore called c-PATIPPET (Cycle Phase And Tempo Inference from Point Process Event Timing). Each uses a different approximation to reduce a dynamic distribution over phase and tempo to a finite collection of dynamic variables. "Variational c-PATIPPET" (Fig. 1B) tracks a fully variational approximation of a distribution over possible phases an tempos: the distribution is parameterized by a mean phase, a mean tempo, and a covariance structure relating phase and tempo. By contrast, "gradient c-PATIPPET" (Fig. 1C) instead breaks tempo space into a range of discrete tempo bins such that the distribution over phase and tempo is parameterized by a circular Gaussian distribution over phase in each tempo bin. These models closely resemble existing oscillator models ("adaptive oscillator" models and "gradient frequency neural networks", respectively), but all of the variables carry specific representational meanings in the context of ongoing inference of underlying rhythmic structure.

Below, we present mathematical formulations of each model and the inference problems they solve. For each, we show simulation results demonstrating the model's behaviors that correspond to properties of human rhythm perception. In the Discussion, we show that by linking the dynamics of rhythm perception to problems of inference, these models raise interesting theoretical questions which suggest new directions for empirical research.

2. c-PIPPET: Cycle phase inference from point process event timing

Here we will build on recent work that began the process of modeling entrainment as inference (Cannon, 2021). In this earlier work, the listener's generative model was stated in a general form that did not assume that the stimulus was periodic. Here, we show that if the listener's generative model assumes some type of periodicity (e.g., a beat) underlying the stimulus, then the solution to this inference process is naturally approximated by a damped linear oscillator with nonlinear pulse forcing, where the forcing function is derived directly from a representation of the listener's metrical expectations.

2.1. Generative model

Suppose that the observer believes that the rhythmic stream of events they are listening to is generated by a rule that repeats periodically. One highly general form of such a rule is to assume that the stimulus possesses an underlying phase ϕ that advances steadily on the circle, and that events occur probabilistically at specific phases. For example, when listening to a rhythm with a beat that is sometimes split into duple subdivisions (i.e., with sound events halfway between the beats), one might strongly expect sound events to occur when the underlying beat cycle reaches phase zero and weakly expect them at the opposite phase. Phase-based event expectancies operationalize the assumption underlying the metrical inference hypothesis of Witek & Vuust: sound events at specific points in a metrical cycle. See Fram and Berger (2023) for evidence that this probabilistic formulation is an appropriate way to describe the subjective experience of metre.

The observer's observations are, however, corrupted by noise. We may suppose that the observer suffers from three sources of "noise":

1. random delays in processing the timing of auditory events
2. noisiness in measuring elapsed time
3. random auditory events unrelated to the stimulus.

Noise source 1 will cause the observer to experience events as randomly perturbed from the expected event phases. We can describe the resulting sequence of events in terms of a point process, where events occur at rate $\lambda(\phi)$ (a function of the stimulus phase). $\lambda(\phi)$ will consist of peaks near expected event phases with some spread on either side:

$$\lambda(\phi) = \sum_j \lambda_j \varphi_{wr}(\phi|\phi_j, \upsilon_j) \tag{1}$$

where $\varphi_{wr}(\phi|\phi_j, \upsilon_j)$ is a Gaussian function wrapped over the ϕ circle with mean ϕ_j and variance υ_j, and λ_j is a scaling factor corresponding to the overall likelihood of events associated with peak j. We can think of $\lambda(\phi)$ as a "metrical expectation template" over phase, representing the points in the metrical cycle when events are expected (ϕ_j) and how precisely these events specify the underlying phase (υ_j, with large values indicating less precision). See Fig. 2 for illustration.

Noise source 2 would most properly be modeled by assuming that the inference process somehow takes into account noise in its own execution. However, since this source of noise would be experienced by the observer as noise disrupting the steady advance of stimulus phase, we operationalize it from the observer's perspective by incorporating Brownian noise into the dynamics of stimulus phase:

$$\dot{\phi} = \omega + \sigma_\phi B_t \tag{2}$$

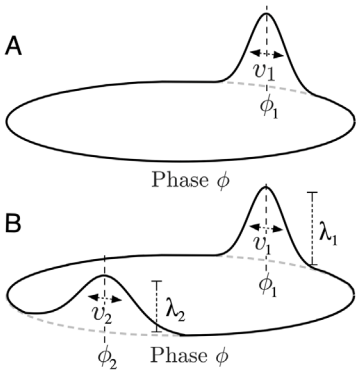

Fig. 2. Illustration of two possible metrical expectation templates. (A) The listener expects to perceive sound events in a close vicinity (quantified by variance υ_1) to an ongoing periodic beat. As a result, each sound event will cause them to infer that the phase of the ongoing pulse is near the beat phase ϕ_1. (B) The listener expects events on the beat, and expects events halfway between beats less strongly (quantified by scale λ_2 relative to the scale of on-beat expectations λ_1) and with greater temporal uncertainty (quantified as variance υ_2).

where ω is the frequency of the repetitive rule underlying the stimulus, B_t is a Wiener process (Brownian noise) with unit variance, and σ_ϕ is the amplitude of the Brownian noise.

The third source of noise, random events unrelated to phase, can be modeled by simply incorporating a small constant λ_0 into the sum (1) defining the expected event probability $\lambda(\phi)$, representing the fact that an event can happen at any time with a certain low probability.

2.2. Phase inference

Given these noise sources, the ideal observer should track not just the stimulus phase but a distribution over possible phases. In Cannon (2021), we let stimulus phase live on the real line and let the observer represent inferred phase with a Gaussian distribution. Here, we let it live on the circle instead, representing a distribution over possible phases within a metrical cycle. Tracking a distribution over stimulus phase (rather than a single most-likely stimulus phase) is particularly important at the onset of a stimulus: immediately preceding the first event, the observer should have no expectations for the phase of the stimulus, and should therefore start with a uniform distribution over phase that is updated at the first event. The simplest non-trivial form of a distribution over possible stimulus phases is a wrapped Gaussian:

$$p(\phi) = \varphi_{wr}(\phi|\mu_t, V_t)$$

with two parameters: a mean μ_t representing a most likely stimulus phase at time t and variance V_t representing the spread of possible stimulus phases around μ_t (i.e., the level of uncertainty about stimulus phase). See Cannon (2021) for a discussion of experimental evidence that listeners are indeed tracking both a most likely stimulus phase and their uncertainty about stimulus phase as a rhythm unfolds. Both of these parameters evolve in time by applying Bayes' rule at each dt time step to incorporate the observation of presence or absence of a sound event in that time step into a prior distribution over stimulus phase, while taking into account the anticipated forward progress of stimulus phase. We will see that at sound events, the inferred distribution over phase resets discontinuously, whereas between events, it evolves continuously, with mean progressing around the circle and variance gradually increasing (Fig. 3). As pointed out in Cannon (2021), this is a Kalman–Bucy filter for phase based on point process observations.

If the circular distribution $p(\phi)$ is conceived of as a distribution on the unit circle $e^{i\pi\phi}$ in the complex plane, its center of mass (or first moment) is

$$Z := \int_\phi e^{i\pi\phi} p(\phi) d\phi$$

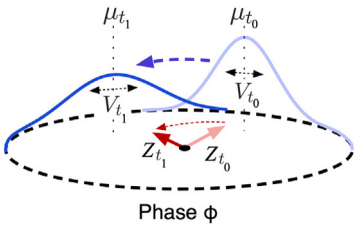

Fig. 3. The inferred distribution over stimulus phase evolves continuously between sound events, with mean μ_t progressing around the circle and variance V_t gradually increasing from time t_0 (lighter blue) to time t_1 (darker blue). This dynamic distribution can be represented as a vector Z_t with angle μ_t and amplitude $e^{-\frac{V_t}{2}}$. This vector's angle moves around the circle from t_0 (lighter red) to t_1 (darker red) as its amplitude decays, following the pattern of a damped oscillator.

A wrapped normal distribution is fully specified by its first moment Z. It is well established that this moment can be written in terms of the wrapped normal's mean μ and variance V:

$$Z = e^{\mu i - \frac{V}{2}} \tag{3}$$

Thus, if the dynamic distribution over stimulus phase is assumed to take the simplified form of a wrapped normal, then the dynamic parameters μ_t and V_t of this distribution can be recovered from dynamic first moment Z_t via $\mu_t = arg(Z_t)$ and $V_t = -2log(|Z_t|)$. When Z_t is on the unit circle ($|Z_t| = 1$), then we have $V_t = 0$: the observer is totally certain about their estimate of stimulus phase. Conversely, when $Z_t = 0$, then we have $V = \infty$: the observer is totally uncertain about stimulus phase, and their wrapped normal distribution takes the limiting form of a uniform distribution over the circle.

In the Derivation section, we show that at each event, Z_t resets to another point $\hat{Z} = F(Z_t)$ in the complex plane. The resetting function F is defined in the Derivations section in terms of the parameters of the expectation $\lambda(\phi)$. Here we will only note the key properties of this function:

- When $|Z_t| = 1$, i.e. $V_t = 0$, Z_t does not change at an event: $F(Z_t) = Z_t$. (When the observer is totally confident about stimulus phase, events have no influence.)
- For smaller $|Z_t|$, $F(Z_t)$ generally moves in the direction of the nearest peak of $\lambda(\phi)$, as illustrated in Fig. 4. (When the observer is uncertain about the stimulus phase, an event pulls their phase estimate toward the nearby phase at which events are expected.)
- If $|Z_t| = 0$, i.e., $V_t = \infty$, then the phase of $F(Z_t)$ is generally the phase at which events are most strongly expected. (When the observer knows nothing about stimulus phase, the first event is assumed to correspond to the strongest metrical position.)

We further show in the Derivations section that the continuous evolution of Z_t between events is governed by a differential equation of the form

$$\dot{Z} = \left(i\omega - \frac{(\sigma_\phi)^2}{2}\right) Z_t + \Lambda_t(Z_t - \hat{Z}) \tag{4}$$

where Λ_t represents the degree to which an event is expected at time t given the inferred distribution on stimulus phase (or the "subjective hazard function"). The second term above allows strong expectations to manipulate the ongoing estimate of phase, even in the absence of events. See Cannon (2021) for evidence that the absence of expected events does subtly manipulate our rhythmic anticipation in this way. In this presentation, we will ignore these effects for simplicity by assuming that event expectancy is relatively weak, i.e., Λ_t is small. As a result, the expression reduces to

$$\dot{Z} = \left(i\omega - \frac{(\sigma_\phi)^2}{2}\right) Z_t \tag{5}$$

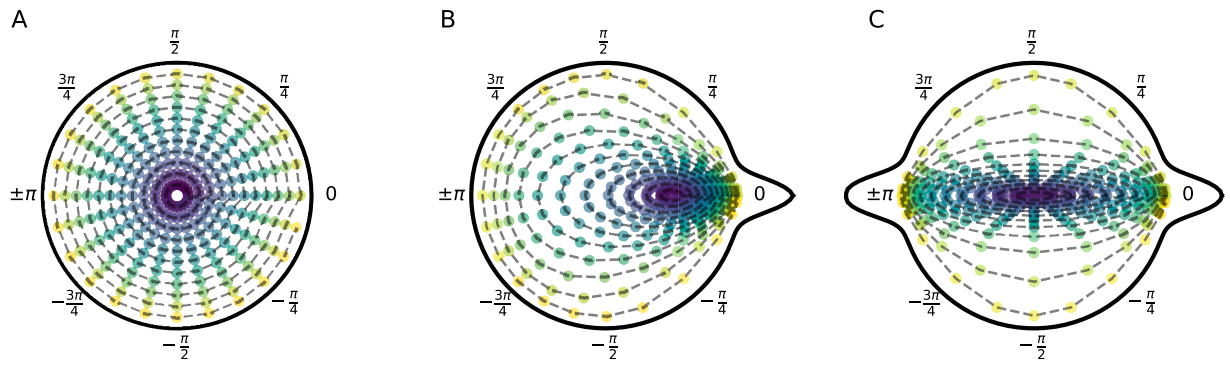

Fig. 4. Phase and amplitude resetting at events in c-PIPPET models. (A) A polar grid of possible oscillator phases and amplitudes, corresponding to a full range of possible estimated stimulus phases and degrees of phase certainty. Phase and amplitude resetting at a sound event will be illustrated in (B) and (C) by displacing these dots to their new values. (B) If the metrical expectation template (black line) consists of a single phase $\mu_j = \{0.0\}$ at which events are expected (plus a small constant λ_0), then an event resets the estimated phase toward 0. If the phase is already near 0, the amplitude (phase certainty) increases. (C) If the metrical expectation template consists of two phases $\mu_j = \{0.0, \pi\}$ at which events are equally expected (plus a constant λ_0), then an event resets the phase toward the nearest one.

This differential equation that describes an under-damped harmonic oscillator with frequency ω that decays exponentially back to a stable fixed point at 0 at rate $\frac{(\sigma_\phi)^2}{2}$. This is the generic behavior of a dynamical system near a stable fixed point with a pair of complex eigenvalues.

Uncertainty

Multiple variables in this model relate to different types of uncertainty, so for clarity we list and disambiguate them here:

- $V_t = -2\log(|Z_t|)$ is a dynamic variable quantifying the observer's uncertainty about the stimulus phase at time t in the form of the variance of their posterior distribution on phase.
- σ_ϕ is a constant quantifying the magnitude of phase noise assumed by the observer's generative model; as a result, this parameter determines the rate of growth of phase uncertainty V_t (i.e. the rate of decay of $|Z_t|$) between sound events.
- υ_j is a constant quantifying the (im)precision with which sound events are associated with a particular stimulus phase ϕ_j in the observer's generative model. When a sound event occurs at a time that associates it unambiguously with phase ϕ_j, phase uncertainty V_t is reduced a lot if υ_j is small and less if υ_j is large.

2.3. Phase inference simulations

Please see Appendix B for a list of parameters used in the simulations below.

The behavior of this model is particularly interesting in response to *syncopated* rhythms, i.e. rhythms in which sound events are absent at metrical positions associated with strong expectations (e.g., on certain beats) but present at weaker metrical points nearby (e.g., the subdivisions between beats). In Fig. 4, we illustrate the function F that describes the resetting of the c-PIPPET oscillator at events. Note that this is not an ordinary phase resetting function in that it depends on and resets not only phase but also radius. In Fig. 5, we simulate c-PIPPET inferring the phase of a perturbed metronome over several clicks, and illustrate the phase and radius dynamics of the oscillator that implements this inference.

Fitch and Rosenfeld (2007) note that for sufficiently syncopated rhythms, the listener may choose an interpretation in which the even timing of beats is violated in order to place the beat at a position that better matches the rhythmic surface. By simulating the phase inference process on the circle rather than the line (as in Cannon (2021)), c-PIPPET can more precisely describe and visualize the process by which perceived beat phase may or may not realign over the course of a syncopated rhythm. In Fig. 6, we illustrate a rhythm that can be perceived in two ways: either as a highly syncopated rhythm with the majority of sound events taking place off the beat, or as a simple rhythm with most events taking place on the beat. Two c-PIPPET simulations are initialized with a metronome count-in such that the subjective beat is initially placed in the highly syncopated alignment.

In the first simulation, σ_ϕ is set to a lower value, giving the model more confidence in its estimated phase over time. As a result, perceived beat phase continues to advance steadily over the course of the syncopated rhythm, and the rhythm continues to be perceived as syncopated. In the second simulation, σ_ϕ is set to a higher value, giving the model less confidence in its phase estimate. In this simulation, phase uncertainty accumulates over the course of three off-beat events, and by the third the distribution over phase is broad enough that the phase is realigned to place this event on the beat. In effect, the model is so uncertain about phase that it pays no attention to its own estimate and instead aligns the event with the phase at which events are most likely to occur, i.e., the beat.

c-PIPPET gives a clear, intuitive account of phase realignment during a syncopated rhythm: it is a result of the accumulation of phase uncertainty that eventually leads the model to discount its phase estimate and instead align events with whatever part of the metrical expectation template fits them best. When the model does *not* reset during a syncopated rhythm, it generally means that the model is capable of sustaining two different phase alignments with the rhythm; in the language of dynamical systems, the oscillator/forcing system is bistable. This bistability gives way with changing parameters, leading to monostability (and thus mandatory phase realignment) under the forcing rhythm.

3. c-PATIPPET: Cycle phase and tempo inference

We now incorporate the process of inferring stimulus tempo into an inference model. To do so, we propose that the observer aims to maintain a representation of a distribution over stimulus phase and tempo, and that they do so in light of a generative model of rhythm in which tempo may vary and drift over time.

3.1. Generative model

The generative model is largely identical to the c-PIPPET generative model except that phase is assumed to advance at a variable rate (tempo) ω:

$$\dot{\phi} = \omega + \sigma_\phi B_t^\phi$$

And ω is assumed to gradually vary up and down at random as a Wiener process with drift rate σ_ω, while drifting toward a preferred tempo ω_p at rate k:

$$\dot{\omega} = k(\omega_p - \omega) + \sigma_\omega B_t^\omega$$

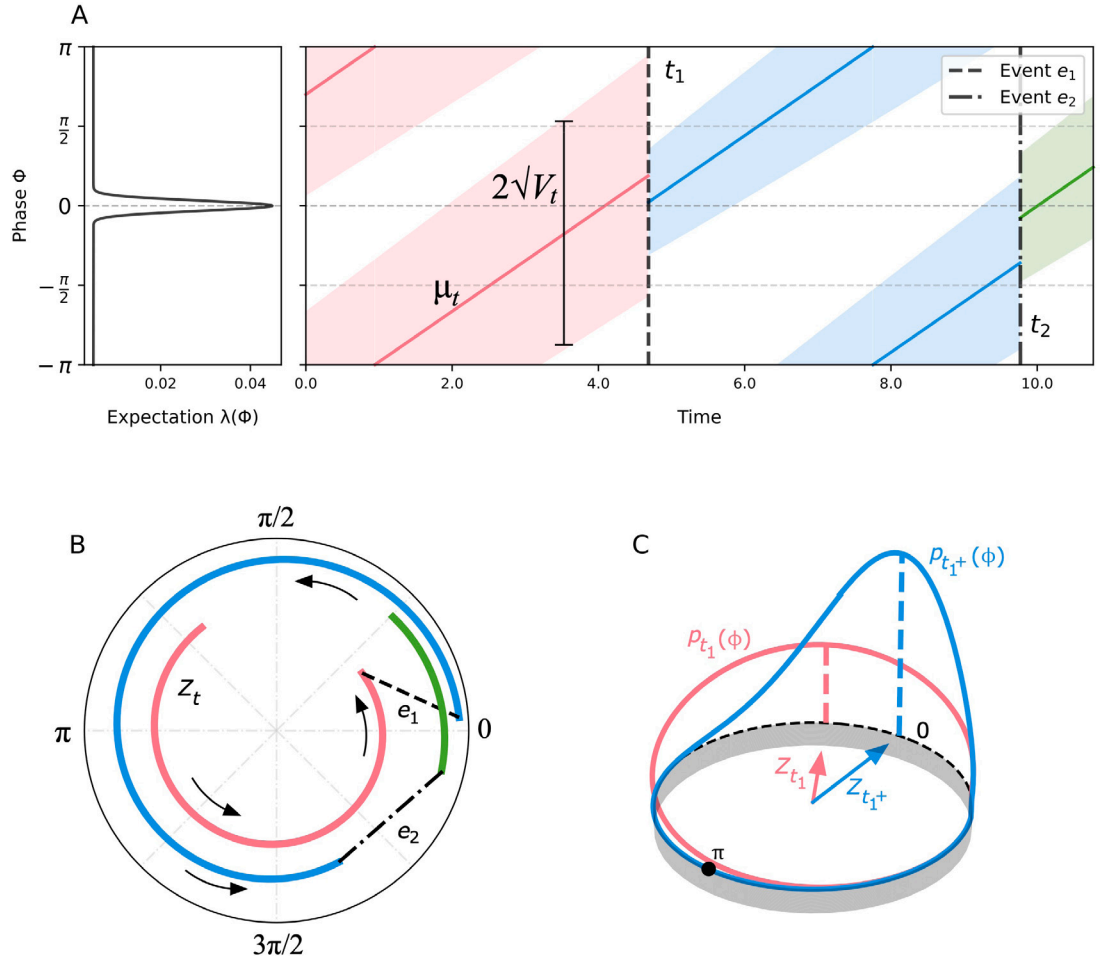

Fig. 5. Phase tracking of a two-event sequence, with expectation for events at intervals of 2π. (A) Tracking cycle phase over time, estimated phase μ_t resets, and phase uncertainty V_t contracts, at each of the two sound events e_1 and e_2 (marked by dashed lines) at times t_1 and t_2 respectively. Phase uncertainty V_t grows between events. (B) Phase and amplitude of the oscillator Z_t corresponding to this c-PIPPET solution, plotted over time, with dashed lines representing the instantaneous resets at sound events. (C) Distributions over phase and their corresponding oscillator vectors just before and just after the event at time t_1.

In simulations below, k is assumed to be zero for simplicity. Finally, the expectation function is made a function of both phase and tempo in order for v_j to reflect a constant precision of event expectation *with respect to time* rather than phase:

$$\lambda(\phi,\omega) = \sum_j \lambda_j \varphi_{wr}\left(\phi|\phi_j, v_j\omega^2\right) \tag{6}$$

A full Bayesian solution to this inference problem would maintain a full (infinite-dimensional) posterior distribution over phase and tempo at all times. There are several ways the brain might reduce the dimensionality of the posterior distribution to make the problem tractable.

3.2. Variational c-PATIPPET

Just as the variational solution to the phase inference problem maintains a wrapped Gaussian posterior, a complete variational solution the phase and tempo inference problem would maintain a bivariate Gaussian distribution over phase and tempo, wrapped around the circle in the phase direction but not the tempo direction (i.e., on the cylinder). A non-wrapped bivariate distribution is most straightforwardly parameterized by a two-dimensional mean and a covariance matrix with three degrees of freedom: the variance of the distribution in each of the two directions and the covariance between the two variables. The situation is not as straightforward on the cylinder, but one analogous way to parameterize a bivariate Gaussian posterior distribution $p(\phi,\omega)$ is in terms the five dynamic variables μ_t, Ω_t, V_t^ϕ, V_t^ω, and a scalar S_t similar to covariance.

$$Z_t := \mathbb{E}[e^{i\pi\phi}] = \int_{\omega,\phi} e^{i\pi\phi} p_t(\phi,\omega)$$

is the complex oscillator-like variable that yields μ_t and V_t^ϕ;

$\mu_t := arg(Z_t)$ represents the best estimate of the phase of the stimulus;

$V_t^\phi := -2\ln(Z_t)$ represents the uncertainty about that phase estimate;

$$\Omega_t := \mathbb{E}[\omega] = \int_{\omega,\phi} \omega p_t(\phi,\omega)$$

is the best estimate of the tempo of the stimulus;

$$V_t^\omega := \mathbb{E}[(\omega-\Omega_t)^2] = \int_{\omega,\phi} (\omega-\Omega_t)^2 p_t(\phi,\omega)$$

is the uncertainty about that tempo estimate; and a final variable S_t quantifying the dependency between the estimates of ϕ and ω (much like a covariance, but better suited to the cylinder) is produced using the ansatz

$$p_t(\phi,\omega) = \varphi(\omega|\Omega_t, V_t^\omega)\varphi_{wr}(\phi|\mu_t + S_t(\omega-\Omega_t), V_t^\phi)$$

which allows the joint distribution to take the form of a partially-wrapped 2D Gaussian with a variable slant S_t over the face of the cylinder. This ansatz distribution is illustrated in Fig. 7. In the Derivations section, these variables are discussed further, and equations are

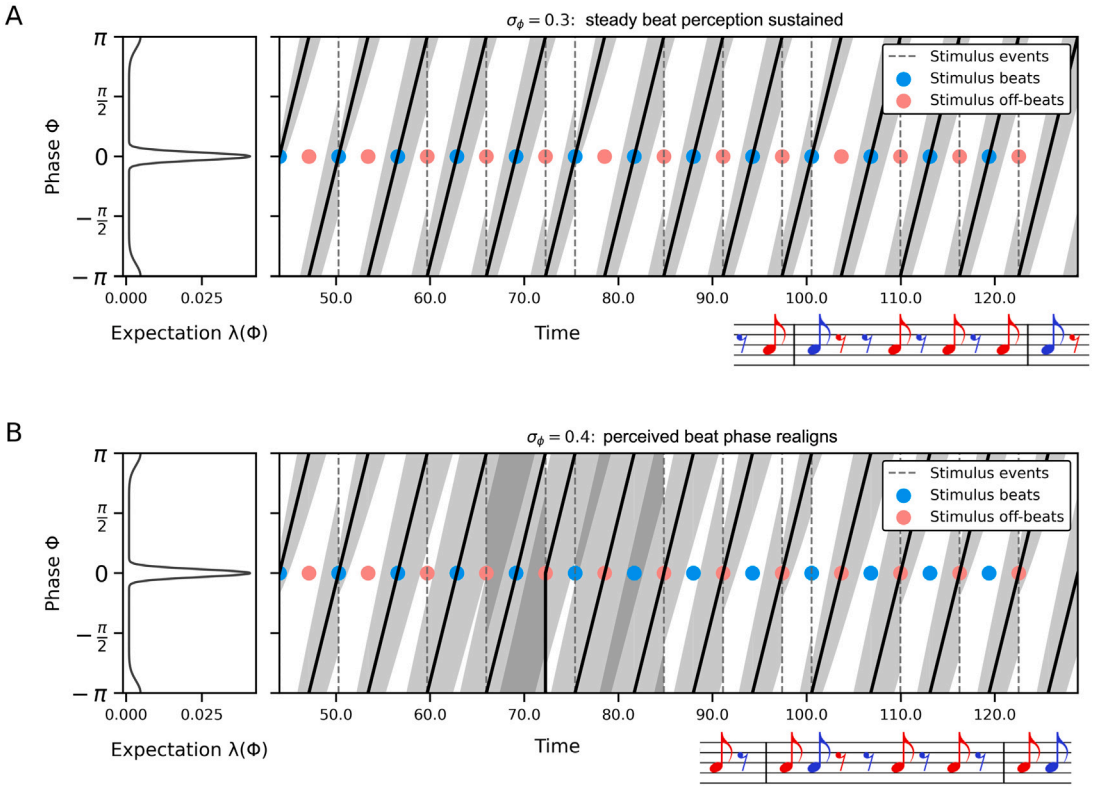

Fig. 6. Tracking phase of a syncopated rhythm with more stimulus events in metrically weak positions (red dots and notes) than metrically strong positions (blue dots and notes). μ_t and V_t are displayed over time as in Fig. 5. (A) When σ_ϕ (the phase noise assumed by the generative model, which determines the rate of accumulation of phase uncertainty) is set to the lower value of 0.3, the c-PIPPET model treats the first event as the on-beat and maintains this alignment loyally, as illustrated by the placement of the bar line in the music notation below. Beats are perceived (estimated phase crosses zero) on the intended beat (blue dots). (B) When σ_ϕ is set to the higher value of 0.4, the accumulation of phase uncertainty leads the model to align the metrically weak events with the perceived beat, as illustrated by the placement of the bar line in the music notation below. By the end, beats are perceived (estimated phase crosses zero) at points in the stimulus that were originally aligned to be off the beat (red dots).

derived describing their evolution over time and their resetting at events for a particular metrical expectation template $\lambda(\phi, \omega)$.

This solution corresponds roughly to the adaptive oscillator models discussed above, in which a single oscillator adapts both its phase (in this case, μ_t) and its intrinsic frequency (in this case, Ω_t) in response to the temporal structure of its forcing. This correspondence is mapped out more fully in Section 4.1.

Variational c-PATIPPET behavior

Although the formula describing the evolution of the posterior is complex, the general dynamics in most cases are relatively straightforward. Between sound events:

- The phase mean of the posterior μ_t advances steadily at rate Ω_t.
- The tempo mean of the posterior Ω_t drifts toward the preferred tempo ω_p at rate k.
- Tempo uncertainty V_t^ω grows at a rate determined by the tempo noise σ_ω assumed in the generative model, but its growth is gradually reigned in if centralized tempo drift rate k is nonzero.
- The growth of phase uncertainty V_t^ϕ is effectively the sum of uncertainty accumulation due to phase noise σ_ϕ in the generative model, and uncertainty accumulation due to tempo uncertainty V_t^ω. High tempo uncertainty increases the rate of growth of phase uncertainty.
- S_t grows proportionate to time elapsed since the last precisely predicted event, as future judgements about tempo and phase become more codependent — e.g., if a sound event later implies

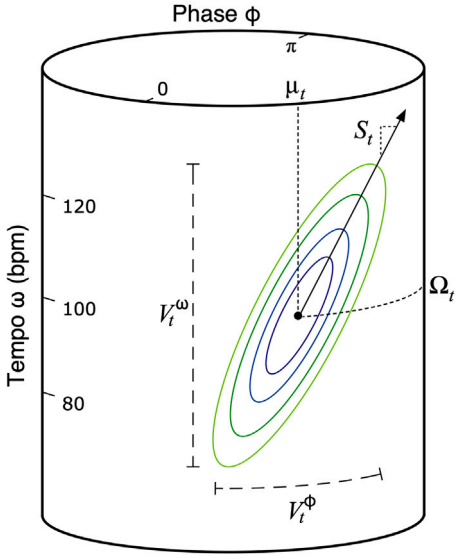

Fig. 7. Illustration of a distribution over phase and tempo inferred by the variational c-PATIPPET model. A gaussian-like ansatz distribution on the phase/tempo cylinder is parametrized by the expected value of phase on the circle μ_t, the expected tempo Ω_t, its phase variance V_t^ϕ, its tempo variance V_t^ω, and the slope S_t describing the dependency of phase on tempo (similar to a covariance).

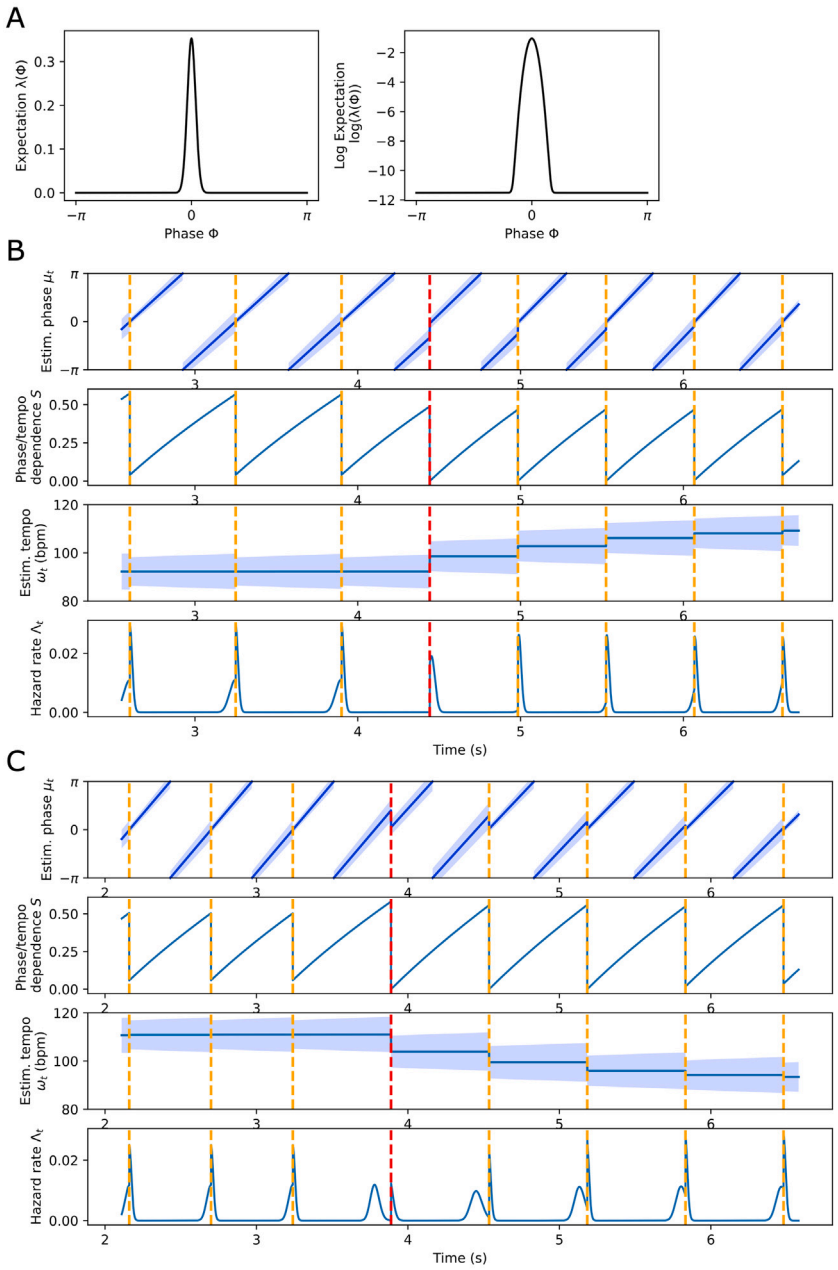

Fig. 8. A variational c-PATIPPET model responds to an increase and then to a decrease in metronome tempo. (A) Plots of the expectation template for events over stimulus phase $\lambda(\phi)$ and its log. (Note that the template $\lambda(\phi)$ is also a function of tempo, as per Eq. (6) — here it is plotted at an intermediate tempo of 108 beats per minute (bpm).) (B) is the example of an increase in tempo (from 93 bpm to 111 bpm), and (C) the decrease in tempo (from 11 bpm to 93 bpm). In both (B) and (C), the tempo change starts with an event marked in red, estimated phase μ_t is shaded with a radius of uncertainty $2\sqrt{V_t^{\phi}}$, and estimated tempo Ω_t is shaded with a radius of uncertainty $2\sqrt{V_t^{\omega}}$. The slope S_t of the dependence between phase and tempo increases between sound events and resets to near zero at events. Increases in the hazard rate Λ_t (the probability with which events are expected) realign with the metronome as the tempo estimate is adjusted.

that the stimulus phase is further advanced than expected, it also implies that the tempo is faster than expected.

At each sound event that occurs when μ_t is close to the phase of an expected event, μ_t resets to place the estimated stimulus phase closer to the expected event phase; Ω_t increases if the event is earlier than expected and decreases if it is later than expected; phase uncertainty V_t^{ϕ} and tempo uncertainty V_t^{ω} decrease; and S_t resets to near zero.

In Fig. 8, we plot the dynamics of these variables as variational c-PATIPPET infers phase and tempo over the course of a tempo-changing metronomic sequence.

As a demonstration that this model is capable of inferring phase and tempo from complex rhythms, we equipped the model with a metrical template representing a subdivided beat and presented it with a "missing pulse" rhythm – a rhythm with no spectral power at the frequency of the beat due to equal numbers of events at strong and weak metrical positions (see Tal et al. (2017)) – at two different tempi. In Fig. 9, we show that the tempo and beat phase are successfully inferred at both tempi, and that as a result, the model begins to anticipate beats at the appropriate moments shortly after the rhythm starts.

3.3. Gradient c-PATIPPET

In variational c-PATIPPET, we hypothesized that the observer performed approximate phase and tempo inference by approximating the

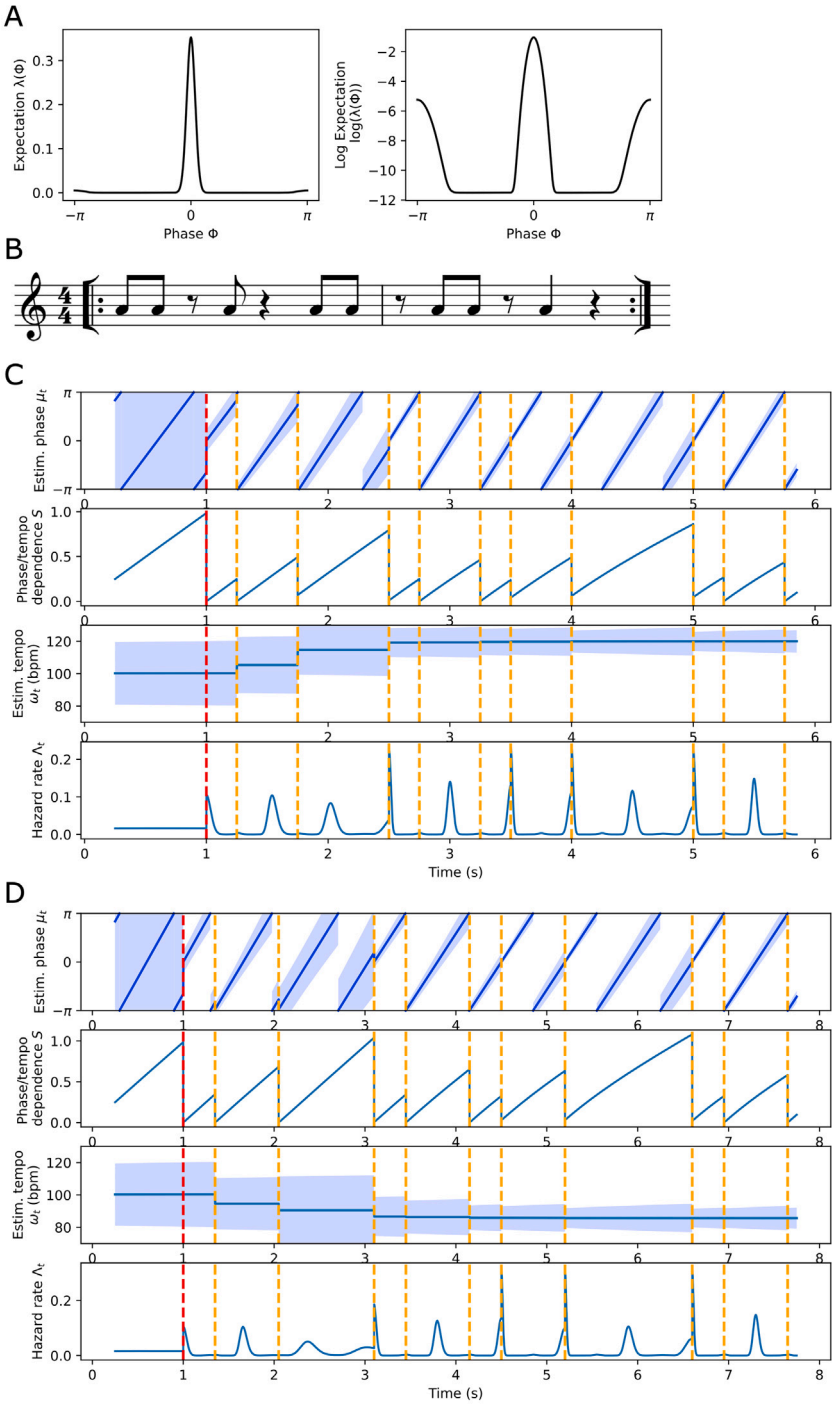

Fig. 9. A variational c-PATIPPET model infers the tempo for a complex "missing pulse" rhythm, where equally many events occur at metrically weak and strong positions, at a fast tempo (120 bpm) and at a slow tempo (86 bpm). (A) The metrical expectation template includes weakly expected even subdivisions of the beat. Its log is included to emphasize the presence of subdivision expectations at $\pm\pi$. (B) The stimulus rhythm in music notation. (C) and (D) The model responds to the faster and then the slower stimulus, which start at the event marked in red. In both cases, it quickly adjusts its estimated tempo appropriately and aligns its pulses of event expectancy (hazard rate Λ_t) to the beat. Rows are the same as Fig. 8.

2D distribution over phase and tempo with a Gaussian distribution whose means and covariances are continuously updated. Here we propose another plausible hypothesis closely related to gradient frequency neural network models discussed above, in which a collection of coupled phase/amplitude oscillators with a range of intrinsic frequencies respond to rhythmic forcing, and underlying stimulus periodicities are picked out by the oscillators responding with high amplitude. As illustrated in Fig. 10, we assume that the observer maintains a distribution over stimulus phase and tempo by slicing phase/tempo space into N discrete tempo bins $\omega^{[i]}$ of width $\Delta\omega$. The distribution $p(\phi, \omega)$ is approximated by a discrete set of N functions $p(\phi, \omega^{[i]})$ over phase ϕ. Each of these functions is a product of a distribution $p(\phi|\omega^{[i]})$ conditioned on tempo bin $\omega^{[i]}$ and a probability density $P_t^{[i]}$ where $P_t^{[i]}\delta\omega$ is the probability that the tempo is indeed in bin i:

$$p(\phi, \omega^{[i]}) = P_t^{[i]} p(\phi|\omega^{[i]})$$

The conditional distribution for each bin i is approximated as a wrapped normal:

$$p(\phi|\omega^{[i]}) = \varphi_{wr}(\phi|\mu_t^{[i]}, V_t^{[i]})$$

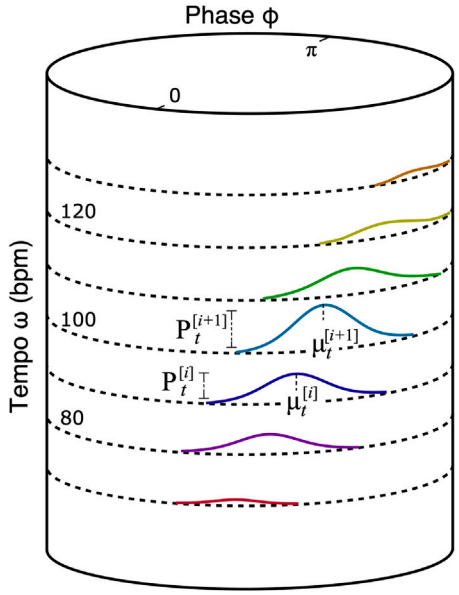

Fig. 10. Illustration of a distribution over phase and tempo inferred by the gradient c-PATIPPET model. A distribution on the cylinder is discretized into a collection of functions $p(\phi, \omega^{[i]})$ over phase ϕ. These functions are described by the product of Gaussian conditional distributions $p(\phi|\omega^{[i]})$ parametrized by their means $\mu_t^{[i]}$ and variances $V_t^{[i]}$ (not shown), and multipliers $P_t^{[i]}$ representing the posterior marginalized over phase (i.e., $P_t^{[i]}\Delta\omega$ is the probability that $\omega = \omega^{[i]}$ at time t).

For each tempo bin i, the variables $\mu_t^{[i]}$ and $V_t^{[i]}$ represent the estimated phase of the stimulus and its accompanying uncertainty, conditioned on the stimulus's instantaneous tempo being within tempo bin $\omega^{[i]}$. Thus, the dynamic distribution over continuous phase and discrete tempo is fully characterized by a set of $3N$ dynamic variables: $\mu_t^{[i]}$, $V_t^{[i]}$, and $P_t^{[i]}$, where $\mu_t^{[i]}$, $V_t^{[i]}$ can be recovered from the state of a c-PIPPET oscillator-like variable $Z_t^{[i]}$. Thus, like a gradient frequency neural network, the gradient c-PATIPPET model consists of a bank of oscillators with a range of intrinsic frequencies (though with the addition of a corresponding bank of scalars $P_t^{[i]}$). These variables evolve over time according to a repeated application of Bayes' rule at each dt time step that integrates the observation of presence or absence of a sound event into the estimated distribution over phase and tempo.

In the Derivations section, we show that the evolution of these variables is described by a set of coupled differential equations with instantaneous resetting at events. The dynamic variables associated with tempo bin $\omega^{[i]}$ are coupled with those of the bins above and below. These couplings exist because of the possibility of the stimulus tempo in the generative model drifting from one bin into an adjacent bin. The effect of the coupling is to keep the phases of adjacent oscillators $Z_t^{[i]}$ close together and to allow the probability mass represented by $P_t^{[i]}$ to diffuse outward.

Gradient c-PATIPPET behavior

In general, between sound events:

- The phase mean $\mu_t^{[i]}$ of the posterior conditioned on tempo $\omega^{[i]}$ advances steadily at rate $\omega^{[i]}$.
- $P_t^{[i]}$, the marginalized distribution over tempos $\omega^{[i]}$, diffuses outward at a rate that increases with increasing σ_ω (the magnitude of tempo noise in the generative model), approaching a steady-state normal distribution centered on ω_p if $k > 0$.
- The variance $V_t^{[i]}$ of each conditional distribution increases at a rate that increases with increasing σ_ϕ (the magnitude of phase noise in the generative model), and further increases due to diffusion from adjacent conditional distributions.

At each sound event, conditional estimated stimulus phase $\mu_t^{[i]}$ resets closer to the expected event phase and $V_t^{[i]}$ decreases. $P_t^{[i]}$ increases if $\mu_t^{[i]}$ was close to an expected event phase and decreases if it was far, reflecting the inference that whichever possible tempo corresponded to the most accurate prediction is the most likely current tempo.

In Fig. 11 (using the same generative model and stimuli as Fig. 8), we simulate and plot the evolution of these variables as gradient c-PATIPPET infers phase and tempo over the course of tempo-changing metronomic sequences. In Fig. 12 (using the same generative model and stimuli as Fig. 9), we simulate and plot the response of gradient c-PATIPPET to the same complex "missing pulse" rhythm at two different tempi.

3.4. Parameter fitting

To reconcile this modeling framework more thoroughly with human behavior, we fit a set of variational c-PATIPPET parameters to an existing data set. In this work, Repp (2007), participants listened to a steady click at two different tempi, with and without subdivisions, and tapped once on a specified beat. Immediately preceding the tap, some combination of beats and subdivisions were shifted in time, and the effect of these shifts was observed in the timing of the tap (Fig. 13A). This data set was ideal for our purposes because it eliminated any perceptual effects of ongoing finger tapping, allowing the single finger tap to serve as something like a direct readout of perceived rhythm phase. A c-PATIPPET model was assumed to produce a tap when its phase passed zero on the appropriate cycle. The model was given a template in which events were expected at the quadruplet subdivision level, with expectations at the second and fourth event in each quadruplet assumed to be identical and λ_0 set to zero. Parameters were fit by hand in an effort to match the tap timing responses for trials with no subdivisions, duplet subdivisions, and quadruplet subdivisions.

The model had nine parameters, but the fit was relatively insensitive to the values of λ_j, leaving five key parameters: υ_1, υ_3, $\upsilon_{2/4}$, σ_ω, and σ_ϕ. The best fit was achieved with $\upsilon_1 = 0.0001$, $\upsilon_3 = 0.0003$, $\upsilon_{2/4} = 0.0025$, $\sigma_\phi = 0.3$, and $\sigma_\omega = 0.15$. The resulting template is shown in Fig. 13B, and the quality of the fit to experimental data is shown in Fig. 13C. Most noteworthy is that the fits reproduce the relationships between tap timing responses at the 540 ms period and at the 720 ms period. When no events intervene between a timing shift and the tap, tap timing shifts more at the slower tempo; when unshifted events do intervene, tap timing shifts more at the faster tempo. Both of these effects arise naturally from the steady accumulation of phase and tempo uncertainty during the silences between sound events. To the best of our knowledge, no existing oscillator models account for differences in the phase resetting function at different tempi.

4. Discussion

We have shown through derivation and simulation that modeling rhythm perception as inference leads naturally to oscillators, and thus that the inference approach to modeling rhythm perception possesses much of the descriptive power and flexibility of the oscillator approach. The process of maintaining an up-to-date posterior distribution on the circular phase of a cyclically patterned stimulus (c-PIPPET) is described by the dynamics of a damped oscillator, where its radius represents the momentary precision of the phase estimate and its resetting function at each sound event is determined by the pattern and precision of event expectations over phase. And the process of maintaining an up-to-date posterior on the phase and tempo of such a stimulus (c-PATIPPET) is described either by the dynamics of a set of five dynamic variables, two of which act like the phase and amplitude of an oscillator and one of which controls that oscillator's frequency (variational approximation), or by the dynamics of a coupled row of damped oscillators with a corresponding coupled row of scalars (gradient approximation).

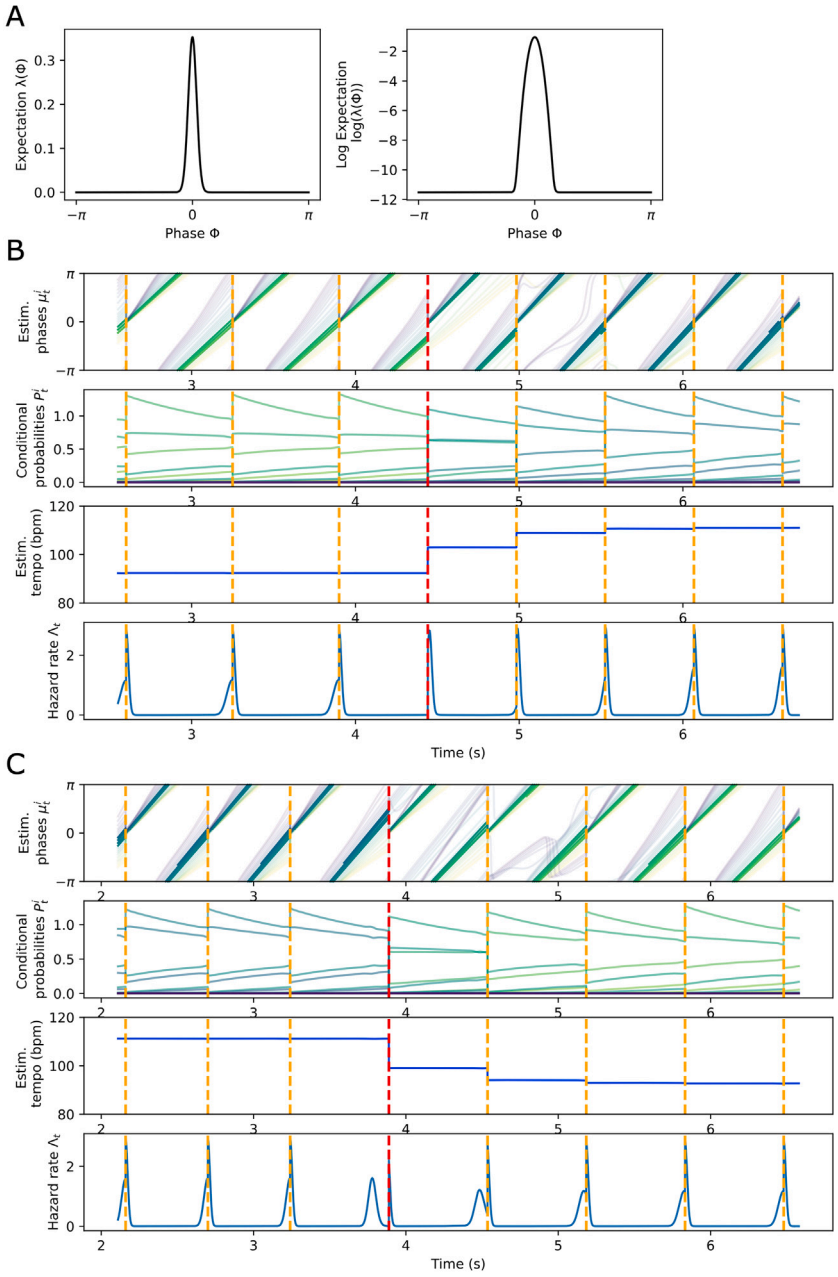

Fig. 11. A gradient c-PATIPPET model with the same parameters as the variational model in Fig. 8 responds to the same changes in metronome tempo. (A) The expectation template (and its log) are the same as in Fig. 8. (B) is the example of an increase in tempo, and (C) the decrease in tempo. In both, the first event at the new tempo is marked in red. The collection of oscillators are represented by a spectrum of colors, with darker colors associated with faster tempo. The phases $\mu_t^{[i]}$ of these oscillators are plotted opaquely only if their associated probability densities $P_t^{[i]}$ are greater than an arbitrary threshold of 0.4, highlighting the phases of the oscillators representing the most likely range of tempi. At the tempo changes, a different (faster or slower) set of oscillators assumes higher probabilities. We integrate over the phase/tempo distribution to calculate an expected value of tempo $\mathbb{E}[\omega]$ over time. Increases in the hazard rate Λ_t (the probability with which events are expected) realign with the metronome as the tempo estimate is adjusted.

These mappings from inference problems onto dynamical systems that approximate their solutions require assumptions about the parametrization of the observer's posterior distribution, which are effectively assumptions about what variables are represented in the brain during auditory entrainment. c-PIPPET makes the "variational" assumption that the observer's distribution on phase is represented by just two parameters describing the mean and variance of a Gaussian distribution wrapped on the circle. Variational c-PATIPPET assumes that phase and tempo estimates are represented by the parameters of a wrapped multivariate Gaussian. Gradient c-PATIPPET makes the variational assumption for phase, and posits that the tempo axis is chopped into discrete bins, allowing a distribution over phase and tempo to be represented by a few variables for at each bin: a mean phase, a phase variance, and the probability that the stimulus tempo is the tempo associated with that bin. By making different assumptions about these distributions, we could derive other valid algorithms for phase and tempo inference; using these algorithms as models of rhythm cognition would be positing that different parameters of the distributions are represented in the observer's brain. We chose these sets of assumptions because they highlight routes by which the inference perspective on rhythm perception can acquire key attributes of prominent oscillator models, and because, as proponents of oscillator models

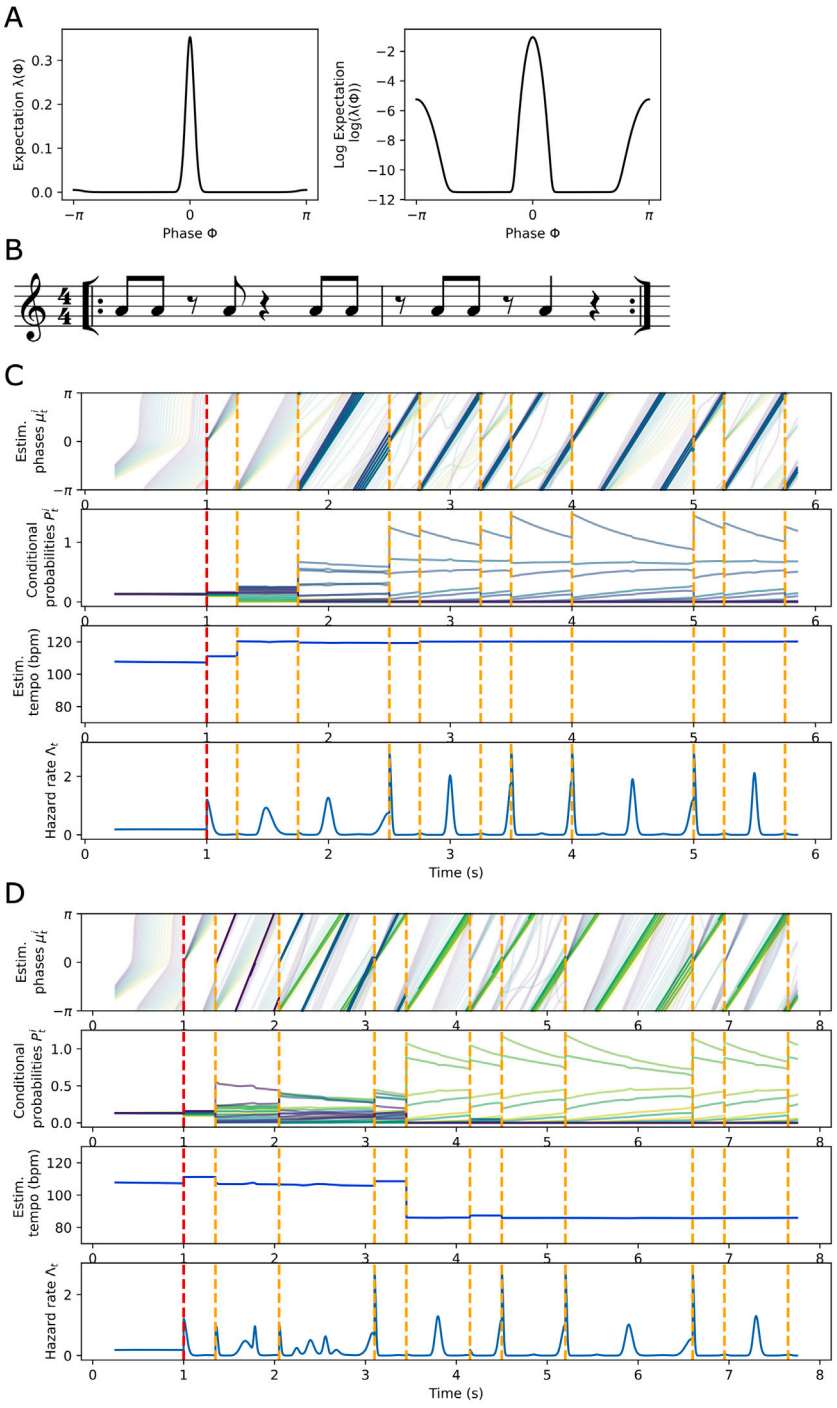

Fig. 12. A gradient c-PATIPPET model infers a phase and tempo distribution from a complex "missing pulse" rhythm at a fast and a slow tempo (same stimuli and generative model as Fig. 9). (A) The metrical expectation template includes weakly expected subdivisions of the beat, as in Fig. 9. (B) The stimulus rhythm in music notation. (C) and (D) The model responds to the faster and then the slower stimulus, and in both cases it successfully adjusts its estimated tempo and aligns it pulses of event expectancy (hazard rate Λ_t) with the beat. Rows are the same as Fig. 11.

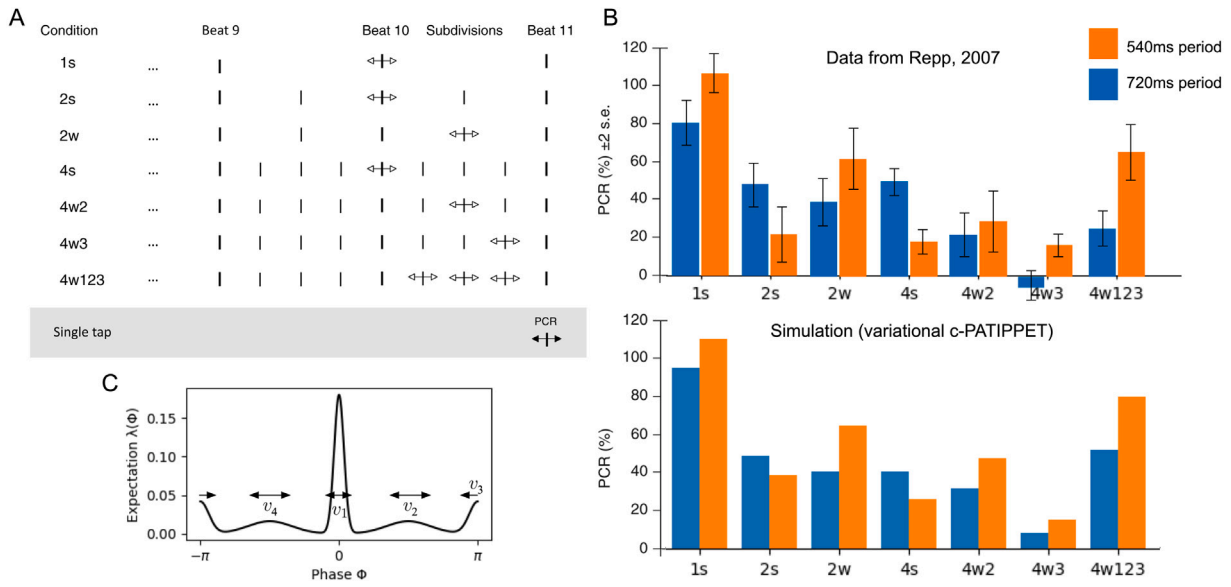

Fig. 13. (A) Illustration of the experiment performed in Repp (2007), reproduced with permission. In this experiment, the participant signals the position of a specific anticipated beat with a single finger tap. Leading up to the tap, the participant listens to a metronome with or without subdivisions. In each of the seven conditions illustrated here, a different set of the sound events immediately preceding the tap is time-shifted, and the effect of this shift on tap timing is measured. (B) Above, the results of this experiment are reproduced with permission. Bar heights represent the relationship between the sound event timing shift magnitude and resulting tap time shift magnitude, expressed as a percentage ratio of tap time shift to event time shift. Below, a variational c-PATIPPET model has been fit by hand to these results by tuning five parameters. Note that the relationships between the tap timing shift at the faster and the slower tempo are reproduced by the model for each condition. (C) The metrical expectation template corresponding to the identified model fit is plotted over phase. (Note that the template $\lambda(\phi)$ is also a function of tempo, as per Eq. (6) — here it is plotted at the faster of the two tempi.).

have argued, such dynamics could plausibly be implemented by neural circuits.

4.1. Oscillators with cognitive interpretations

As dynamical systems, our oscillatory inference models consist of oscillators, and as a result they attain the dynamical range and descriptive repertoire of oscillator models of rhythm perception. However, their variables maintain comprehensible cognitive/representationalist interpretations with theoretical grounding. We highlight the correspondence with oscillator models and the meanings of the static and dynamic variables below.

Formally, the basic c-PIPPET model is simply a pulse-forced damped linear oscillator with a special nonlinear phase/amplitude resetting function.[3] Its variables can be interpreted in terms of an inference process based on a generative model:

- The phase of a damped oscillator corresponds to the mean of a distribution over possible stimulus phase.
- The radius represents the precision of this distribution.
- The damping of the oscillator represents the noisiness of stimulus phase in a generative model, which paces the progressive loss of phase precision over time in the absence of rhythmic events, and determines the flexibility of the inference process in response to phase shifts.
- The phase (and radius) resetting function at events is derived from the pattern of event expectancy over stimulus phase, and tends to reset estimated phase to align with phases at which events are strongly expected.

In gradient c-PATIPPET, we have a network of variables very similar to those used in (oscillator-based) gradient frequency neural network models, but that can be fully interpreted as an inference algorithm:

- The activity of each oscillator in the gradient frequency bank represents a distribution over phase conditioned on the possibility that the current stimulus tempo corresponds to the natural tempo of that oscillator.
- The coupling strength between oscillators of adjacent frequencies represents the rate and direction of tempo drift in a generative model, which paces the progressive loss of tempo precision over time in the absence of rhythmic events and determines the flexibility of the inference process in response to tempo shifts.

Gradient c-PATIPPET also includes a set of variables $P_t^{[i]}$ that are not included in canonical Gradient Frequency Neural Networks. Without the additional information provided by these variables, the amplitudes of the bank of oscillators conflate the precision of a phase estimate with the distribution over tempi. However, it is possible that correlations between $P_t^{[i]}$ and $V_t^{[i]}$ could allow this algorithm to be readily approximated without this additional set of dynamic variables.

The basic mechanic of underlying phase and tempo inference in variational c-PATIPPET looks very much like previous "adaptive oscillator" models of auditory entrainment to rhythmic stimulation that consist of a single phase oscillator which corrects its phase and adapts both its period and the precision of its expectations according to the earliness or lateness of expected events (Large & Jones, 1999; Large & Kolen, 1994; Large & Palmer, 2002; McAuley, 1995). In variational c-PATIPPET,

- Basic phase correction is accomplished by dynamic μ_t.
- Period correction is made possible by a combination of dynamic Ω_t and S_t: S_t acts as a timer from one event to the next, and the correction to estimated tempo Ω_t at each event is proportionate to the accumulated S_t. This mechanic allows for period corrections that are roughly linear with respect to current period, but makes different predictions about how period correction should change

[3] This equivalence is conditional on the assumption of a weak expectation template that allowed us to go from Eq. (4) to Eq. (5) and neglect the influence of unfulfilled expectations on estimated phase — without this assumption, the oscillator's ODE acquires a nonlinear term.

with the presence of subdividing events leading up to an event timing perturbation.
- Event prediction precision increases over the course of extended regular sequences through a combination of dynamic phase and tempo certainty, which makes the peaks in the hazard function signifying expectations sharper over the course of entrainment to a steady rhythm.

4.2. Oscillatory inference models raise new questions

Although this family of models is not formally very different from existing oscillator-based dynamical systems models of beat perception, the cognitive/representational interpretations of the parameters and moving parts can help frame the generation of new questions, hypotheses, and empirical experiments. For example:

- In c-PIPPET-based models, the accumulation of phase and tempo uncertainty – i.e., the tendency to lose a sense of the beat during silent intervals – accounts for differences in the response to event timing shifts across different tempi, as demonstrated in Section 3.4. Specifically, the models make larger phase corrections when there has been more time for uncertainty to accumulate since the last event. Accumulating uncertainty also accounts for the tendency toward phase realignment of ambiguous rhythms simulated in Section 2.3. Are these two effects really attributable to the same underlying dynamic of uncertainty? If so, measures of the two processes should correlate across individuals, and may covary with other individual differences. For example, individuals with ADHD, who seem to lose track of the beat at low tempi (Gilden & Marusich, 2009), may show more rapid accumulation of uncertainty.
- In these models, the accumulation of uncertainty is directly linked to the phase and tempo noise σ_ϕ and σ_ω in the observer's generative model. Do humans learn to base their phase/tempo uncertainty accumulation on their brain's actual levels of phase/tempo noise? Uncertainty accumulation could be measured either through the scaling of the phase-resetting function over tempi or through the tendency toward phase realignment during complex rhythms. Phase/tempo noise could be quantified through performance precision in timing and tempo discrimination tasks.
- Similarly, the precision of event expectations and the strength of resulting phase corrections at events is a function of perceived event timing noise v_j in the generative model. Do humans show the same relationship between perceptual phase error correction (as measured by Repp (2007), described in Section 3.4) and auditory timing noise (as measured perhaps by the variance in delays in auditory evoked EEG responses)?
- Both c-PATIPPET models are initialized with a prior over tempi, and (for $k > 0$) return to this prior given sufficient time. This prior seems closely related to the various measures of "preferred tempo" and "spontaneous motor tempo" in the rhythm psychophysics literature. Does this prior differ by individual? Do these differences reflect differences in exposure to different tempi? Such systematic differences might arise from differences in the tempo of the beats we hear most often: the cadence of our own gait (Dahl et al., 2014) or our parents' gait (Rocha et al., 2021).

4.3. Modeling philosophy

We have shown that dynamic inference of phase and tempo can be performed by a bank of oscillators or a frequency adapting oscillator, partially reconciling the perspectives of the two prominent threads of rhythm perception modeling. However, an inference perspective is inconsistent with philosophical commitments that often come along with dynamical systems modeling. Proponents of dynamical systems models of behavior argue that if the existence of internal "representations" of hidden world states is not necessary to explain a system, then there is no reason to invoke them (Stepp & Turvey, 2010). Indeed, many aspects of sensorimotor synchronization to rhythm can be explained purely in terms of interacting oscillators (Tognoli et al., 2020). In response to this point, we would like to point out that, in theory, *all* physical phenomena should be explainable in terms of dynamical systems underpinned by physical laws. That does not, however, mean that introducing an intermediate level of explanation is pointless or meaningless: just as chemistry is built on top of physics to provide more compact explanations of observed phenomena, so a quantitative science of "representations" and "generative models" can be built on top of dynamical systems to more compactly explain our subjective perceptual experiences and our resulting behaviors. The theoretical foundations of this construction have been laid by work deriving the Free Energy Principle and the resulting dynamics of probabilistic representations from underlying physics (Ramstead et al., 2023). At a more practical level, Poldrack (Poldrack, 2021) notes that neurophysiologists are increasingly turning to a "fluid combination of representationalist and dynamicist thinking" that treats network states as representations of the world while acknowledging that the formation and transformation of these states is orchestrated according to dynamical rules.

4.4. Model comparison and validation

The different models presented here can be compared with each other and with other models by two metrics: neurophysiological and behavioral. From the neurophysiological standpoint, if the phases of multiple oscillators of different intrinsic frequencies can be identified in neural recordings, or if the underlying physiology is shown to support multiple oscillators of different frequencies, these observations would support gradient c-PATIPPET as the more plausible model of human rhythm; conversely, if signals representing tempo and/or tempo uncertainty can be identified, or if a neural mechanism is found for an adjustable-frequency neural oscillator, this would favor variational c-PATIPPET as more plausible. From the behavioral standpoint, there are likely to be specific situations in which variational and gradient c-PATIPPET behave differently due to the more stringent assumptions on the variational model. For example, gradient c-PATIPPET should allow the observer to simultaneously keep two beats at different tempi, at least transiently, whereas variational c-PATIPPET precludes this possibility by design. By comparing model performance to human performance in these situations, we could determine which model is more behaviorally accurate.

4.5. Limitations and future directions

This work should be considered a preliminary proof-of-concept showing that dynamical systems that act like oscillators can perform formal inference about rhythmic stimuli. We have demonstrated that it is possible to match model parameters to the results of a psychophysics experiment, but we have not validated or fit the models with targeted experimentation. Future work will aim to refine these models and tune them to human behavior across a range of perceptual tasks.

One potentially fruitful direction for model development would be to introduce noise into the inference process itself (as opposed to just the generative model) which might account for the element of randomness inherent in human judgements of rhythmicity and timeliness. Further model development might also require refinement of the generative model underlying the inference process. For example, the assumption that tempo drifts continuously may not allow the observer much leeway to recover a sense of the beat following large shifts in tempo as quickly as a human listener; if this is the case, a generative model that assumes that tempo can change via jump discontinuities may prove a better description of the logic underlying human rhythm perception. Finally, context-dependent changes in human perception

could be modeled by introducing additional dynamic variables into the generative model, which would be estimated dynamically during the inference process. For example, possible effects of a history of stimulus jitter or tempo change on the observer's phase and tempo flexibility could be accounted for by introducing dynamic estimates of phase and tempo noise similar to the estimates of "volatility" commonly assumed in hierarchical predictive processing models (Mathys et al., 2011, 2014). Similarly, expectations that an accelerating stimulus would continue to accelerate (as modeled by the "anticipation" module of the ADAM model (Van der Steen & Keller, 2013)) could be represented by introducing a dynamic estimate of the current rate of tempo change. And listeners' rapid adaptation to changes in the pattern of subdivisions of the beat (e.g., duplets vs. triplets) could be modeled as dynamic selection of an metrical expectation template, as in Kaplan et al. (2022).

Metrical structure in music generally involves multiple hierarchical levels of periodicity, but here we have only modeled a single beat level. This is at least partially justified by the perceptual primacy of the beat level of periodicity (London, 2004), and allows the model to describe entrainment by non-isochronous subdivisions. However, it does not explain how the metrical expectation templates describing beat subdivisions arise, does not account for biases toward integer ratios at levels below the beat (Jacoby & McDermott, 2017), and does not account for structure above the level of the beat. Extending it to describe multiple levels of periodicity would first require augmenting the generative model, perhaps by including faster or slower phase variables that couple tightly to the phase and tempo of the beat. Dynamic inference with this generative model would likely take the form of a hierarchy of coupled circular variables, making the dynamics of inference even more closely resemble the dynamics of a gradient frequency neural network model.

If the generative model is developed further, it is likely to become analytically intractable. The theory of Predictive Processing poses prediction error minimization as the canonical scheme for approximating intractable Bayesian solutions to dynamic inference problems, in part due to its neural plausibility. Thus, a future direction may be to replace the analytical solutions used here with a more powerful prediction-error-minimizing algorithm that also might be more easily mapped onto specific brain mechanisms.

The phase and tempo inference problems formulated here describe the perception of rhythmic structure, but are insufficient to describe motor entrainment to rhythm. The canonical Predictive Processing approach to including movement in this picture would be to draw on the theory of Active Inference (Adams et al., 2013; Friston, 2010), which posits that physical movement serves to minimize surprisal by creating the sensations that the organism expects. A PIPPET model with entrained tapping might encode a tapping goal by as a template of expectations for the timing of motor feedback relative to the rhythm. This avenue will be explored in future work.

Funding

This research was funded in part by the Natural Sciences and Engineering Research Council of Canada (Award Number RGPIN-2022-05027).

CRediT authorship contribution statement

Jonathan Cannon: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Thomas Kaplan:** Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Conceptualization.

Data availability

Data will be made available on request.

Appendix A. Derivations

Here we derive equations describing the variational Bayesian estimation of stimulus phase in c-PIPPET and of phase and tempo in c-PATIPPET.

A.1. c-PIPPET

The wrapped Gaussian φ_{wr} takes the form:

$$\varphi_{wr}(\phi|\mu,V) := \frac{1}{\sqrt{2\pi V}} \sum_{q=-\infty}^{\infty} e^{-\frac{(\phi-\mu-2\pi q)^2}{2V}}$$

It can also be written as a sum of Fourier components:

$$\varphi_{wr}(\phi|\mu,V) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-\frac{n^2V}{2}+in(\phi-\mu)}$$

If we place the stimulus phase on the unit circle in the complex plane, $e^{i\phi}$, both parameters of the wrapped normal are uniquely specified by the complex first moment of the wrapped normal distribution:

$$\begin{aligned} Z =& \mathbb{E}[e^{i\phi}] = \int_\phi e^{i\phi}\varphi_{wr}(\phi|\mu,V) \\ =& \int_\phi e^{i\phi}\frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{-\frac{n^2V}{2}+in(\phi-\mu)}d\phi \\ =& \int_\phi \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{-\frac{n^2V}{2}+i(n+1)\phi-in\mu}d\phi \end{aligned}$$

All terms in this sum integrate to zero except the $n=-1$ term, so

$$Z = e^{-\frac{V}{2}+i\mu} \tag{7}$$

From (Snyder, 1972), we have:

$$dp_t(\phi) = \mathcal{L}[p_t(\phi)]dt + p_t(\phi)\left(\frac{\lambda(\phi)}{\Lambda_t}-1\right)\cdot(dN_t - \Lambda_t dt) \tag{8}$$

where $\Lambda_t := \mathbb{E}[\lambda(\phi)]$ (with $\mathbb{E}$ denoting expectation under distribution $p_t(\phi)$), 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 (2):

$$\mathcal{L}[p(\phi)] = -\frac{\partial}{\partial\phi}p(\phi) + \frac{(\sigma_\phi)^2}{2}\frac{\partial^2}{\partial\phi^2}p(\phi) \tag{9}$$

As the observer's distribution over stimulus phase evolves, we will make the "variational" assumption that it maintains the simple form of a wrapped normal distribution. We operationalize this assumption by continuously (i.e., at each dt time step) replacing the distribution with the "nearest" wrapped normal distribution. Measuring "nearest" by KL divergence, the standard measure of distance between distributions, this means continuously replacing the distribution with the wrapped normal that has the same first moment in the complex plane. Thus, we need only describe how the first moment Z_t of the observer's distribution evolves in time, and the full wrapped normal distribution is specified.

To describe the evolution of Z_t:

$$\begin{aligned} dZ :=& d\mathbb{E}[e^{i\phi}] \\ =& d\left[\int_\phi e^{i\phi}p_t(\phi)d\phi\right] \\ =& \int_\phi e^{i\phi}dp_t(\phi)d\phi \end{aligned}$$

From (8),

$$\begin{aligned} =& \int_\phi e^{i\phi}\left[\mathcal{L}[p_t(\phi)]dt + p_t(\phi)\left(\frac{\lambda(\phi)}{\Lambda_t}-1\right)\cdot(dN_t-\Lambda_t dt)\right]d\phi \\ =& \left[\int_\phi -e^{i\phi}\frac{\partial}{\partial\phi}p_t(\phi)d\phi + \int_\phi \frac{(\sigma_\phi)^2}{2}e^{i\phi}\frac{\partial^2}{\partial\phi^2}p_t(\phi)d\phi\right]dt \end{aligned}$$

$$+\int_{\phi} p_t(\phi)\left(\frac{\lambda(\phi)e^{i\phi}}{\Lambda_t}-e^{i\phi}\right)\cdot(dN_t-\Lambda_t dt)d\phi$$

Integrating by parts:

$$\begin{aligned}dZ=&\left[\int_{\phi} ie^{i\phi}p_t(\phi)d\phi-\int_{\phi}\frac{(\sigma_\phi)^2}{2}e^{i\phi}p_t(\phi)\right]dt\\&+\int_{\phi} p_t(\phi)\left(\frac{\lambda(\phi)e^{i\phi}}{\Lambda_t}-e^{i\phi}\right)\cdot(dN_t-\Lambda_t dt)d\phi\end{aligned}$$

Rewriting integrals against p_t as expected values:

$$\begin{aligned}dZ=&i\mathbb{E}[e^{i\phi}]dt-\frac{(\sigma_\phi)^2}{2}\mathbb{E}[e^{i\phi}]dt+\left(\frac{\mathbb{E}[\lambda(\phi)e^{i\phi}]}{\Lambda_t}-\mathbb{E}[e^{i\phi}]\right)\cdot(dN_t-\Lambda_t dt)\\=&\left(i-\frac{(\sigma_\phi)^2}{2}\right)Z_t dt+\left(\hat{Z}-Z_t\right)\cdot(dN_t-\Lambda_t dt)\end{aligned}\tag{10}$$

where we have set $\hat{Z}:=\frac{\mathbb{E}[\lambda(\phi)e^{i\phi}]}{\Lambda_t}$.

To calculate Λ_t and $\hat{Z}$, we first write a reduced expression for $p_t(\phi)\lambda(\phi)$ using the Fourier sum expression of the wrapped normal:

$$\begin{aligned}p_t(\phi)\lambda(\phi)=&\varphi_{wr}(\phi|\mu,V)\sum_j\lambda_j\varphi_{wr}(\phi|\phi_j,v_j)\\=&\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}e^{-\frac{n^2V}{2}+in(\phi-\mu)}\sum_j\lambda_j\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}e^{-\frac{m^2v_j}{2}+im(\phi-\phi_j)}\\=&\frac{1}{(2\pi)^2}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_j\lambda_j e^{-\frac{n^2V}{2}-\frac{m^2v_j}{2}+in(\phi-\mu)+im(\phi-\phi_j)}\end{aligned}$$

Setting $\ell:=n+m$, we have

$$p_t(\phi)\lambda(\phi)=\frac{1}{(2\pi)^2}\sum_{\ell=-\infty}^{\infty}e^{\ell\phi}\left(\sum_{m=-\infty}^{\infty}\sum_j\lambda_j e^{-\frac{(\ell-m)^2V}{2}-\frac{m^2v_j}{2}-i(\ell-m)\mu-im\phi_j}\right)$$

All terms in the initial sum integrate to zero except $\ell=0$, so we have

$$\begin{aligned}\Lambda_t=&\int_{\phi}p_t(\phi)\lambda(\phi)d\phi\\=&\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\sum_j\lambda_j e^{-\frac{m^2V}{2}-\frac{m^2v_j}{2}+im\mu-im\phi_j}\\=&\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\sum_j\lambda_j e^{-m^2\frac{V+v_j}{2}+im(\mu-\phi_j)}\end{aligned}\tag{11}$$

When $p_t(\phi)\lambda(\phi)$ is integrated against $e^{i\phi}$, all terms integrate to zero except $\ell=-1$, so we have

$$\begin{aligned}\hat{Z}=&\frac{1}{\Lambda_t}\int_{\phi}p_t(\phi)\lambda(\phi)e^{i\phi}d\phi\\=&\frac{1}{2\pi\Lambda_t}\sum_{m=-\infty}^{\infty}\sum_j\lambda_j e^{-\frac{(m+1)^2V}{2}-\frac{m^2v_j}{2}+i(m+1)\mu-im\phi_j}\end{aligned}\tag{12}$$

Thus, as the rhythm unfolds, the observer's distribution over stimulus phases is specified by $p_t(\phi):=\varphi(\phi|\mu_t,V_t)$, where μ_t and V_t are determined by the evolving complex variable Z_t. From (7) we have:

$$\begin{aligned}\mu_t:=&arg(Z_t)\\V_t:=&-2log(|Z_t|)\end{aligned}$$

and, from (10), Z_t evolves according to a differential equation of the form

$$\dot{Z}=\left(\omega i-\frac{\sigma^2}{2}\right)Z_t-\Lambda_t(\hat{Z}-Z_t)$$

with Z_t instantaneously resetting to $\hat{Z}$ at any event time.

To simulate variational Bayesian inference on phase, we simulate the dynamics of Z_t according to (10), approximating sums over m by summing from $m=-40$ to 40.

A.2. Variational c-PATIPPET

Generative model:

$$\begin{aligned}\dot{\phi}=&\omega+\sigma_\phi B_t\\\dot{\omega}=&k(\omega_p-\omega)+\sigma_\omega B_t\end{aligned}$$

with events occurring at Poisson rate

$$\lambda(\phi,\omega)=\sum_j\lambda_j\varphi_{wr}\left(\phi|\phi_j,v_j\omega^2\right)$$

We assume that the evolving posterior distribution over phase and tempo takes the form of a continuous family $p_t(\phi,\omega)$ of wrapped Gaussians over phase ϕ, parametrized by tempo ω, with identical variance V_t. These are scaled by a second Gaussian distribution over tempo, and the means μ_t^ω of these wrapped Gaussians are assumed to increase (or decrease) linearly with tempo.

This distribution can be parameterized in terms of the center of mass within the cylinder (Z_t,Ω_t), the variance V_t^ω of the Gaussian over tempo, and the slope S of the dependence of μ_t^ω on tempo: $\mu_t^\omega=\mu_t+S(\omega-\Omega_t)$, where $\mu_t:=arg(Z_t)$.

We can calculate dp_t, the change in the posterior over phase and tempo in a dt time step, using (8), where the Kolmogorov forward operator is

$$\mathcal{L}p_t=-\frac{\partial}{\partial\phi}(\omega p_t)-\frac{\partial}{\partial\omega}(k(\omega_p-\omega)p_t)+\frac{(\sigma_\phi)^2}{2}\frac{\partial^2p_t}{\partial^2\phi}+\frac{\sigma_\phi\sigma_\omega}{2}\frac{\partial^2p_t}{\partial\phi\partial\omega}+\frac{(\sigma_\omega)^2}{2}\frac{\partial^2p_t}{\partial^2\omega}\tag{13}$$

Similarly to the derivation above, we calculate the differentials of expected values over time by taking integrals over the cylinder, substituting from (13), and then calculating the integrals by parts. We will calculate S at each time step based on a related term, $V^{\phi\omega}:=\mathbb{E}[\omega e^{i\phi}]$. The terms $\hat{Z}$, $\hat{\omega}$, etc. are defined below.

$$\begin{aligned}dZ=&d\mathbb{E}[e^{i\phi}]=\int_{\phi,\omega}e^{i\phi}dp_t d\phi d\omega\\=&-\frac{(\sigma_\phi)^2}{2}Z_t dt+(V_t^{\phi\omega}+\Omega_t z)idt+(\hat{Z}-Z_t)(dN_t-\Lambda_t dt)\\d\Omega=&d\mathbb{E}[\omega]=\int_{\phi,\omega}\omega dp_t d\phi d\omega\\=&k(\omega_p-\Omega_t)dt+(\hat{\Omega}-\Omega_t)(dN_t-\Lambda_t dt)\\dV^{\phi\omega}=&d\mathbb{E}[\omega e^{i\phi}]=\int_{\phi,\omega}\omega e^{i\phi}dp_t d\phi d\omega\\=&(Z_tV^\omega i+(\Omega_t-k)V_t^{\phi\omega}+kZ_t(\omega_p-\Omega_t))dt\\&+(V^{\hat{\phi}\omega}-V_t^{\phi\omega})(dN_t-\Lambda_t dt)\\dV^\omega=&d\mathbb{E}[(\omega-\Omega_t)^2]=\int_{\phi,\omega}(\omega-\Omega_t)dp_t d\phi d\omega\\=&(-2kV_t^\omega+\frac{(\sigma_\omega)^2}{2})dt+(\hat{V^\omega}-V_t^\omega)(dN_t-\Lambda_t dt)\end{aligned}$$

To calculate S_t at any time, we note that

$$V_t^{\phi\omega}=\int_\omega\left(\int_\phi e^{i\phi}\varphi_{wr}(\phi|\mu_t^\omega,V_t)d\phi\right)(\omega-\Omega_t)\varphi(\omega|\Omega_t,V_t^\omega)d\omega$$

Setting $u=\omega-\Omega_t$,

$$V_t^{\phi\omega}=\int_u\left(\int_\phi e^{i\phi}\varphi_{wr}(\phi|S_tu+\mu_t,V_t)d\phi\right)u\varphi(u|0,V_t^\omega)du$$

From (3),

$$\begin{aligned}V_t^{\phi\omega}=&\int_u e^{S_tu+\mu_t-\frac{V_t}{2}}u\varphi(u|0,V_t^\omega)du\\=&e^{\mu_t-\frac{V_t}{2}}\int_u e^{S_tu}u\frac{1}{\sqrt{2\pi V_t^\omega}}e^{\frac{-u^2}{2V_t^\omega}}du\\=&e^{\mu_t-\frac{V_t}{2}}\int_u u\frac{1}{\sqrt{2\pi V_t^\omega}}e^{-\frac{1}{2V_t^\omega}(u-S_tiV_t^\omega)^2-\frac{S_t^2V_t^\omega}{2}}du\end{aligned}$$

$$= e^{\mu_t - \frac{V_t}{2} - \frac{S_t^2 V_t^\omega}{2}} \int_u u \frac{1}{\sqrt{2\pi V_t^\omega}} e^{-\frac{1}{2V_t^\omega}(u - S_t i V_t^\omega)^2} du$$

The integral term expresses the expected value of u over a Gaussian distribution with mean $S_t i V_t^\omega$, and therefore equals that mean:

$$V_t^{\phi\omega} = e^{\mu_t - \frac{V_t}{2}} e^{-\frac{S_t^2 V_t^\omega}{2}} S_t i V_t^\omega \tag{14}$$

Next, we calculate:

$$\begin{aligned} Z_t &= \int_{\phi,\omega} e^{i\phi} p_t(\phi,\omega) d\phi d\omega \\ &= \int_\omega \varphi(\omega|\Omega_t, V_t^\omega) \int_\phi e^{i\phi} \varphi_{wr}(\phi|\mu_t^\omega, V_t) d\phi d\omega \end{aligned}$$

Calculating the integral as in (14), we have

$$Z_t = e^{i\mu_t^\omega - \frac{V_t}{2} - \frac{S_t^2 V_t^\omega}{2}}$$

Substituting into (14):

$$V_t^{\phi\omega} = Z_t S_t i V_t^\omega$$

So we can calculate S_t at any time using

$$S_t = \frac{V_t^{\phi\omega}}{Z_t i V_t^\omega} \tag{15}$$

(S_t should be a real number, but may not be due to numerical issues; we therefore only take the real part of this expression in the code.)

We calculate expectations using Fourier expansions as in the previous section, using the expected tempo Ω_t to scale event expectation precision v_j:

$$\begin{aligned} \Lambda_t &:= \mathbb{E}[\lambda(\phi,\omega)] \\ &= \sum_j \frac{\lambda_j}{2\pi} \sum_{m=-\infty}^{\infty} exp\left(-m^2 \frac{V_t + v_j \Omega_t^2 + V_t^\omega S^2}{2} - im(\mu_t - \phi_j)\right) \\ \hat{Z} &:= \frac{1}{\Lambda_t} \mathbb{E}[e^{i\phi} \lambda(\phi,\omega)] \\ &= \frac{1}{\Lambda_t} \sum_j \frac{\lambda_j}{2\pi} \\ &\quad \times \sum_{m=-\infty}^{\infty} exp\left(-\frac{m^2(V_t + V_t^\omega S^2)}{2} - \frac{v_j \Omega_t^2 (m+1)^2}{2} - im(\mu_t - \phi_j) + i\phi_j\right) \\ \hat{\Omega} &:= \frac{1}{\Lambda_t} \mathbb{E}[\omega \lambda(\phi,\omega)] \\ &= \frac{1}{\Lambda_t} \sum_j \frac{\lambda_j}{2\pi} \\ &\quad \times \sum_{m=-\infty}^{\infty} (\Omega_t - imV_t^\omega) exp\left(-m^2 \frac{V_t + v_j \Omega_t^2 + V_t^\omega S^2}{2} - im(\mu_t - \phi_j)\right) \\ V^{\hat{\phi}\omega} &:= \frac{1}{\Lambda_t} \mathbb{E}[e^{i\phi}(\omega - \Omega_{t+}) \lambda(\phi,\omega)] \\ &= \frac{1}{\Lambda_t} \sum_j \frac{\lambda_j}{2\pi} \sum_{m=-\infty}^{\infty} (\Omega_t - imV^\omega) \\ &\quad \times exp\left(-\frac{m^2(V + V_t^\omega S^2)}{2} - \frac{v_j \Omega_t^2 (m+1)^2}{2} - xim(\mu - \phi_j) + i\phi_j\right) \\ &\quad - \Omega_{t+} \hat{Z} \\ \hat{V^\omega} &:= \frac{1}{\Lambda_t} \mathbb{E}[(\omega - \Omega_{t+})^2 \lambda(\phi,\omega)] \\ &= \frac{1}{\Lambda_t} \sum_j \frac{\lambda_j}{2\pi} \\ &\quad \times \sum_{m=-\infty}^{\infty} Re\left(V_t^\omega exp\left(-m^2 \frac{V_t + v_j \Omega_t^2 + V_t^\omega S^2}{2} - im(\mu_t - \phi_j)\right)\right) \\ &\quad + (\Omega_{t+} + \Omega_t - 2\hat{\omega})(\Omega_{t+} - \Omega_t) \end{aligned}$$

A.3. Gradient c-PATIPPET

We define

$$\begin{aligned} Y_t^\omega &:= \int_\phi e^{i\phi} p_t(\phi,\omega) d\phi \\ P_t^\omega &:= \int_\phi p_t(\phi,\omega) d\phi \end{aligned}$$

As in the previous section, we calculate the differentials using (13) and then calculate the integrals by parts:

$$\begin{aligned} dY^\omega &= \int_\phi e^{i\phi} d p_t(\phi,\omega) d\phi \\ &= Y_t \left(i\omega - \frac{(\sigma_\phi)^2}{2}\right) dt - k(\omega - \omega_p) \frac{\partial Y}{\partial \omega} dt + \frac{(\sigma_\omega)^2}{2} \frac{\partial^2 Y}{\partial \omega^2} dt \\ &\quad + (\hat{Y} - Y_t^\omega)(dN_t - dt) \\ dP^\omega &= \int_\phi d p_t(\phi,\omega) d\phi \\ &= -k(\omega - \omega_p) \frac{\partial P}{\partial \omega} dt + \frac{(\sigma_\omega)^2}{2} \frac{\partial^2 P}{\partial \omega^2} dt + (\hat{P} - P_t^\omega)(dN_t - dt) \end{aligned}$$

where $\hat{Y}$ and $\hat{P}$ are defined below. Breaking tempo into discrete $\Delta\omega$ bins at evenly-spaced values $\omega^{[i]}$, we can rewrite the partial derivatives as discrete approximations:

$$\begin{aligned} \frac{\partial Y^{[i]}}{\partial \omega} &\approx \frac{Y^{[i+1]} - Y^{[i-1]}}{2\Delta\omega} = \frac{Y^{[i+1]} - Y^{[i]}}{2\Delta\omega} + \frac{Y^{[i]} - Y^{[i-1]}}{2\Delta\omega} \\ \frac{\partial^2 Y^{[i]}}{\partial \omega^2} &\approx \frac{Y^{[i+1]} - Y^{[i]}}{\Delta\omega^2} - \frac{Y^{[i]} - Y^{[i-1]}}{\Delta\omega^2} \end{aligned}$$

and similarly for P. Thus, $Y^{[i]}$ behaves like a chain of coupled oscillators, and $P^{[i]}$ behaves like a chain of coupled scalars. At the edges of the chain, we implement closed boundary conditions by setting the appropriate terms to zero.

Recall that $p_t(\phi,\omega^{[i]}) = P_t^{[i]} \varphi_{wr}(\phi|\mu_t^{[i]}, V_t^{[i]})$. Drawing on the calculations in Appendix A.1, we write expressions for $\hat{Y}$ and $\hat{P}$:

$$\begin{aligned} \Lambda_t^{[i]} &:= \int_\phi p_t(\phi|\omega^{[i]}) \lambda(\phi,\omega) d\phi = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sum_j \lambda_j e^{-m^2 \frac{V_t^{[i]} + v_j(\omega^{[i]})^2}{2} + im(\mu_t^{[i]} - \phi_j)} \\ \Lambda_t &:= \int_{\phi,\omega} p_t(\phi,\omega) \lambda(\phi,\omega) d\phi d\omega \approx \sum_i P_t^{[i]} \Lambda_t^{[i]} \Delta\omega \\ \hat{P}^{[i]} &= \frac{1}{\Lambda_t} \int_\phi p_t(\phi,\omega^{[i]}) \lambda(\phi,\omega^{[i]}) d\phi = \frac{P_t^{[i]} \Lambda_t^{[i]}}{\Lambda_t} \\ \hat{Y}^{[i]} &= \frac{1}{\Lambda_t} \int_\phi p_t(\phi,\omega^{[i]}) e^{i\phi} \lambda(\phi,\omega) d\phi \\ &= \frac{P_t^{[i]}}{2\pi\Lambda_t} \sum_{m=-\infty}^{\infty} \sum_j \lambda_j e^{-\frac{(m+1)^2 V_t^{[i]}}{2} - \frac{m^2 v_j(\omega^{[i]})^2}{2} + i(m+1)\mu_t^{[i]} - im\phi_j} \end{aligned}$$

We define $Z_t^{[i]}$ as the center of mass of the distribution over phase, conditioned on tempo $\omega^{[i]}$:

$$Z_t^{[i]} := \int_\phi e^{i\phi} p_t(\phi|\omega^{[i]}) d\phi = \frac{Y_t^{[i]}}{P_t^{[i]}}$$

As in A.1, we can define estimated phase $\mu_t^{[i]}$ and uncertainty/variance $V_t^{[i]}$, this time conditioned on tempo $\omega^{[i]}$:

$$\begin{aligned} \mu_t^{[i]} &:= arg(Z_t^{[i]}) \\ V_t^{[i]} &:= -2\ln(|Z_t^{[i]}|) \end{aligned}$$

Appendix B. Simulation parameters

All code will be made available at https://github.com/Kappers/cpippet-JMathPsych.

Configuration for c-PIPPET simulations:

$$dt = 0.005$$
$$\sigma^{\phi} = 0.2$$
$$\mu_0 = 0.0$$
$$V_0 = 0.1$$
$$\lambda_0 = 0.001$$
$$\mu_j = \{0.0, \}$$
$$\upsilon_j = \{0.01, \}$$
$$\lambda_j = \{0.01, \}$$

Configuration specific to gradient c-PATIPPET:

$$N = 21$$
$$\Delta\omega = \frac{72 \text{ bpm}}{N - 1}$$
$$\omega^{[i]} = \{72 \text{ bpm}, 72 \text{ bpm} + \Delta\omega, \ldots, 144 \text{ bpm}\}$$

Configuration for c-PATIPPET tempo change simulations (Figs. 8 and 11):

$$dt = 0.005$$
$$\sigma^{\phi} = 0.3$$
$$\sigma^{\omega} = 0.4$$
$$\phi_j = \{0, \}$$
$$\upsilon_j = \{0.0001, \}$$
$$\lambda_j = \{0.01, \}$$
$$\lambda_0 = 0.00001$$

Configuration for c-PATIPPET complex rhythms (Figs. 9 and 12):

$$\sigma^{\phi} = 0.3$$
$$\sigma^{\omega} = 0.4$$
$$\phi_j = \{0, \pi\}$$
$$\upsilon_j = \{0.0001, 0.0004\}$$
$$\lambda_j = \{0.1, 0.003\}$$
$$\lambda_0 = 0.00001$$

Initial conditions for variational c-PATIPPET complex rhythms (Fig. 9):

$$\mu_0 = 0$$
$$V_0 = 1$$
$$\omega_0 = 10.5$$
$$V_0^{\omega} = 1$$
$$S_0 = 0$$

Initial conditions for gradient c-PATIPPET complex rhythms (Fig. 12):

$$\mu_0^{[i]} = 0$$
$$V_0^{[i]} = 1$$
$$P_0^{[i]} = \frac{1}{N\Delta\omega} \text{ (uniform distribution)}$$

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. http://dx.doi.org/10.1007/s00429-012-0475-5.

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

Dahl, S., Huron, D., Brod, G., & Altenmüller, E. (2014). Preferred dance tempo: Does sex or body morphology influence how we groove? *Journal of New Music Research*, *43*(2), 214–223. http://dx.doi.org/10.1080/09298215.2014.884144.

Doelling, K. B., Arnal, L. H., & Assaneo, M. F. (2023). Adaptive oscillators support Bayesian prediction in temporal processing. *PLoS Computational Biology*, *19*(11), Article e1011669. http://dx.doi.org/10.1371/journal.pcbi.1011669.

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2014). Moving in time: Bayesian causal inference explains movement coordination to auditory beats. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1786), http://dx.doi.org/10.1098/rspb.2014.0751.

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception*, *25*(1), 43–58. http://dx.doi.org/10.1525/mp.2007.25.1.43.

Fram, N. R., & Berger, J. (2023). Syncopation as probabilistic expectation: Conceptual, computational, and experimental evidence. *Cognitive Science*, *47*(12), Article e13390. http://dx.doi.org/10.1111/cogs.13390.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *360*(1456), 815–836. http://dx.doi.org/10.1098/rstb.2005.1622.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. http://dx.doi.org/10.1038/nrn2787.

Gilden, D. L., & Marusich, L. R. (2009). Contraction of time in attention-deficit hyperactivity disorder, 23 (2). (pp. 265–269). http://dx.doi.org/10.1037/a0014553.

Heggli, O. A., Cabral, J., Konvalinka, I., Vuust, P., & Kringelbach, M. L. (2019). A Kuramoto model of self-other integration across interpersonal synchronization strategies. *PLoS Computational Biology*, *15*(10), 1–17. http://dx.doi.org/10.1371/journal.pcbi.1007422.

Heggli, O. A., Konvalinka, I., Kringelbach, M. L., & Vuust, P. (2021). A metastable attractor model of self–other integration (MEAMSO) in rhythmic synchronization. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *376*(1835), Article 20200332. http://dx.doi.org/10.1098/rstb.2020.0332.

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, *27*(3), 359–370. http://dx.doi.org/10.1016/j.cub.2016.12.031.

Kaplan, T., Cannon, J., Jamone, L., & Pearce, M. (2022). Modeling enculturated bias in entrainment to rhythmic patterns. *PLoS Computational Biology*, *18*(9), Article e1010579. http://dx.doi.org/10.1371/journal.pcbi.1010579.

Kim, J. C., & Large, E. W. (2021). Multifrequency Hebbian plasticity in coupled neural oscillators. *Biological Cybernetics*, *115*(1), 43–57. http://dx.doi.org/10.1007/s00422-020-00854-6.

Koban, L., Ramamoorthy, A., & Konvalinka, I. (2019). Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. *Social Neuroscience*, *14*(1), 1–9. http://dx.doi.org/10.1080/17470919.2017.1400463.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, *23*(1), 63–77. http://dx.doi.org/10.1016/j.tics.2018.10.006.

Large, E. W., Almonte, F. V., & Velasco, M. J. (2010). A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*, *239*(12), 905–911. http://dx.doi.org/10.1016/j.physd.2009.11.015.

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

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

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*(2–3), 177–208. http://dx.doi.org/10.1080/09540099408915723.

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

London, J. (2004). Research on temporal perception and its relevance for theories of musical meter. In *Hearing in time: Psychological aspects of musical meter*. Oxford University Press, http://dx.doi.org/10.1093/acprof:oso/9780195160819.003.0003.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, http://dx.doi.org/10.3389/fnhum.2011.00039.

Mathys, C., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the hierarchical Gaussian filter. *Frontiers in Human Neuroscience*, *8*, 1–24. http://dx.doi.org/10.3389/fnhum.2014.00825.

McAuley, J. D. (1995). *Perception of time as phase: Toward an adaptive-oscillator model of rhythmic pattern processing* (Ph.D. thesis), Indiana University.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1102–1125. http://dx.doi.org/10.1037/0096-1523.29.6.1102.

Merchant, H., Grahn, J., Trainor, L. J., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: A neural perspective across humans and non-human primates.. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *370*(1664), http://dx.doi.org/10.1098/rstb.2014.0093.

Nozaradan, S. (2013). *Exploring the neural entrainment to musical rhythms and meter: A steady-state evoked potential approach* (Ph.D. thesis), Montréal, Quebec, Canada: Université de Montréal.

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *369*, Article 20130393.

Palmer, C., & Demos, A. P. (2022). Are we in time? How predictive coding and dynamical systems explain musical synchrony. *Current Directions in Psychological Science*, *31*(2), 147–153. http://dx.doi.org/10.1177/09637214211053635.

Poldrack, R. A. (2021). The physics of representation. *Synthese*, *199*(1), 1307–1325. http://dx.doi.org/10.1007/s11229-020-02793-y.

Ramstead, M. J. D., Sakthivadivel, D. A. R., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., Klein, B., & Friston, K. J. (2023). On Bayesian mechanics: a physics of and by beliefs. *Interface Focus*, *13*(3), Article 20220029. http://dx.doi.org/10.1098/rsfs.2022.0029.

Repp, B. H. (2007). Multiple temporal references in sensorimotor synchronization with metrical auditory sequences. *Psychological Research*, *72*(1), 79–98. http://dx.doi.org/10.1007/s00426-006-0067-1.

Rocha, S., Southgate, V., & Mareschal, D. (2021). Infant spontaneous motor tempo. *Developmental Science*, *24*(2), Article e13032. http://dx.doi.org/10.1111/desc.13032.

Snyder, D. L. (1972). Filtering and detection for doubly stochastic Poisson processes. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, *18*(1), 91–102. http://dx.doi.org/10.1109/TIT.1972.1054756.

Stepp, N., & Turvey, M. T. (2010). On strong anticipation. *Cognitive Systems Research*, *11*(2), 148–164. http://dx.doi.org/10.1016/j.cogsys.2009.03.003.

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Golumbic, E. Z. (2017). Neural entrainment to the beat: The "missing-pulse " phenomenon. *Neuroscience*, *37*(26), 6331–6341. http://dx.doi.org/10.1523/JNEUROSCI.2500-16.2017.

Tichko, P., & Large, E. W. (2019). Modeling infants' perceptual narrowing to musical rhythms: Neural oscillation and Hebbian plasticity. *Annals of the New York Academy of Sciences*, http://dx.doi.org/10.1111/nyas.14050.

Todd, N. P. A., Lee, C. S., & O'Boyle, D. J. (2002). A sensorimotor theory of temporal tracking and beat induction. *Psychological Research*, *66*(1), 26–39. http://dx.doi.org/10.1007/s004260100071.

Tognoli, E., Zhang, M., Fuchs, A., Beetle, C., & Kelso, J. A. (2020). Coordination dynamics: A foundation for understanding social behavior. *Frontiers in Human Neuroscience*, *14*, http://dx.doi.org/10.3389/fnhum.2020.00317.

Van der Steen, M., & Keller, P. E. (2013). The adaptation and anticipation model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, *7*, http://dx.doi.org/10.3389/fnhum.2013.00253.

Vuust, P., Dietz, M. J., Witek, M., & Kringelbach, M. L. (2018). Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*, *1423*(1), 19–29. http://dx.doi.org/10.1111/nyas.13622.

Vuust, P., & Witek, M. A. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*, http://dx.doi.org/10.3389/fpsyg.2014.01111.