A Bayesian framework for generalized entrainment to stochastic rhythms

2021 CoSyNe submission

November 2020

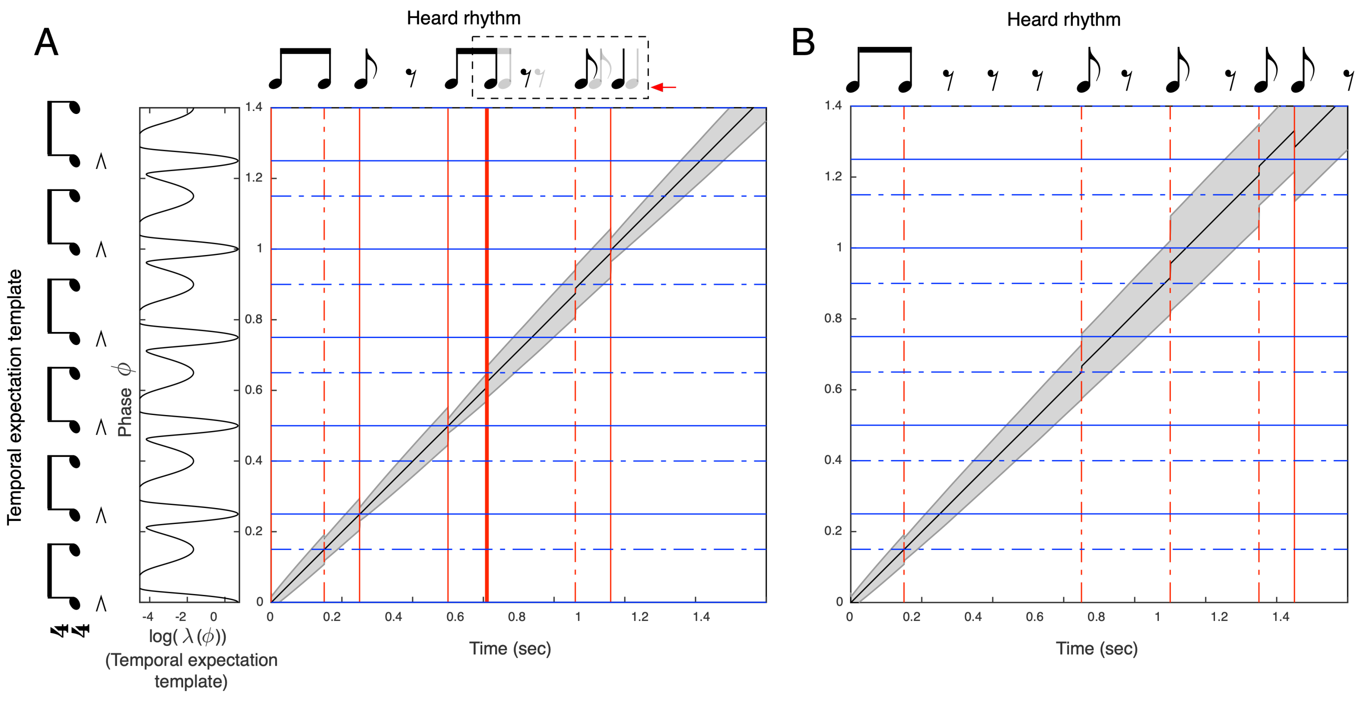
Summary

When presented with complex rhythmic auditory stimuli, most humans are able to track underlying temporal structure (e.g., a "beat") with overt movement or covertly while automatically and unconsciously adjusting for timing irregularities. We propose that the problem of overt and covert rhythm tracking is most naturally characterized as a problem of point process filtering: continuously estimating a hidden underlying phase and tempo based on information provided by the precise timing of temporally localized events. Event timing is informative because events are expected to occur at certain phases with certain probabilities and degrees of temporal precision, as specified by a flexible periodic or aperiodic "temporal expectation template." We demonstrate that approximate (variational Bayesian) solutions to this inference problem reproduce characteristics of overt and covert human rhythm tracking that have not been well addressed by other models, including interval-dependent phase correction, failure to track overly syncopated rhythms, and the distortion of perceived timing by disappointed expectations.

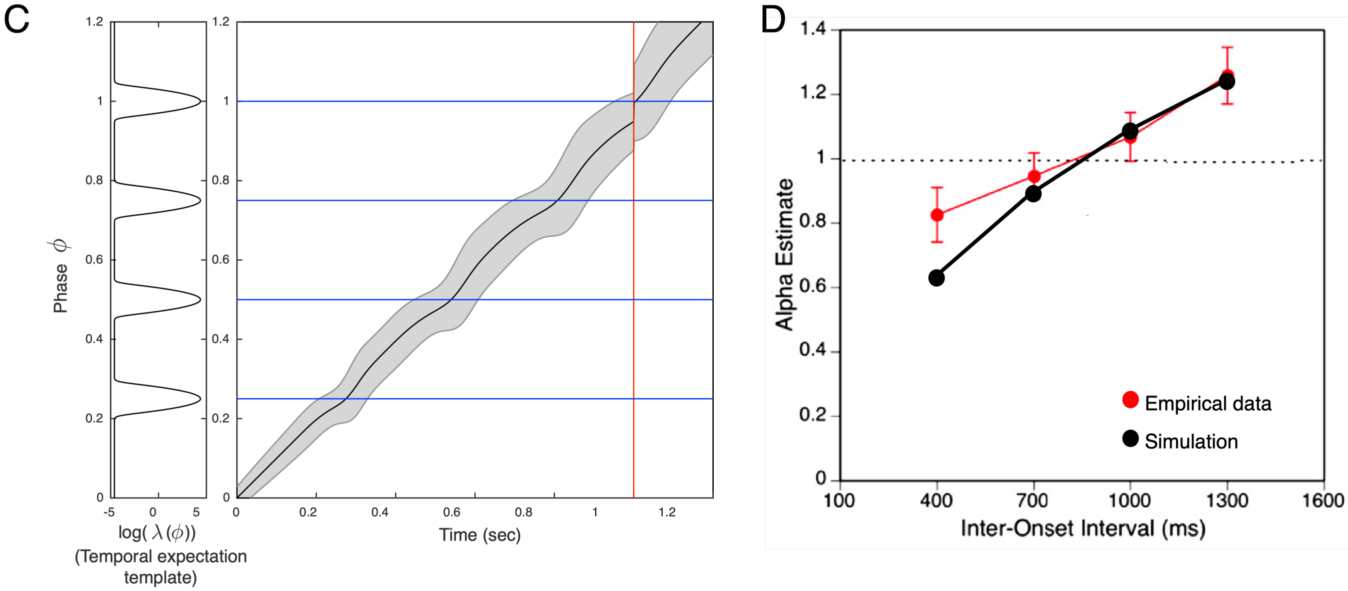
This framework can treat rhythms of arbitrary complexity, and can therefore be used model the behavioral results of rich musical psychophysics experiments. Further, its continuous time dynamics serve as a plausible model of brain activity during entrainment. This work is motivated by a neurophysiologically detailed hypotheses of the brain dynamics of entrainment, and a search for brain activity similar to these inference dynamics could build a link between the algorithmic and physiological levels of entrainment. Finally, as a variational Bayesian inference method in continuous time, this approach fits comfortably into the "predictive processing" framework, allowing for future extension into models of hierarchical inference based on event timing that could describe even more complex sensorimotor processes.

Additional Information

Results



Black line represents estimated phase μ_ϕ; grey shading represents phase uncertainty $\sqrt{\Sigma_{11}}$. A) The model is given a temporal expectation template corresponding to a nonisochronous swing grid. It is exposed to a rhythm based on this grid, but with a phase shift introduced part of the way through (heavy red line). The filter tracks phase and uncertainty, correcting for the phase shift. B) As in humans, when many events in the rhythm correspond to phases with weak event expectations, this can lead to a failure to track phase accurately even if the rhythm is perfectly timed.



C) The disappointment of strong expectations for interval subdivision leads the phase estimate to lag, creating the "filled duration" illusion – empty durations seem to terminate unexpectedly early. D) A listener tapping to a metronome makes more substantial corrections to timing perturbations for longer metronome inter-onset intervals, as measured by "alpha," the fraction of a phase shift corrected on the next tap. Like humans, a variational Bayesian filter produces larger alpha for longer intervals, and over-corrects ($\alpha > 1$) for very long intervals.

Equations

The generative model for hidden phase ϕ and tempo θ dynamics is:

$$d\mathbf{\Phi} = \begin{pmatrix} d\phi \\ d\theta \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} dt + \begin{pmatrix} \sigma_\phi dW_t^\phi \\ \sigma_\theta dW_t^\theta \end{pmatrix}$$

Observable point process events are generated at rate

$$\lambda(\phi) = \lambda_0 + \sum_i \lambda_i N(\phi|\phi_i, v_i) \tag{1}$$

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

The variational Bayesian filter for Gaussian approximation of the joint distribution on phase and tempo tracks a mean $\boldsymbol{\mu} = \begin{pmatrix} d\mu_\phi \\ d\mu_\theta \end{pmatrix}$ and a covariance matrix $\mathbf{\Sigma}$ in continuous time according to the rule:

$$\begin{cases} d\boldsymbol{\mu} = \begin{pmatrix} d\mu_\phi \\ d\mu_\theta \end{pmatrix} = \begin{pmatrix} \mu_\theta \\ 0 \end{pmatrix} dt - (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})(dN_t - \bar{\Lambda} dt) \\ d\mathbf{\Sigma} = \left(\begin{pmatrix} {\sigma_\phi}^2 & 0 \\ 0 & {\sigma_\theta}^2 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \mathbf{\Sigma} + \mathbf{\Sigma} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right) dt - (\bar{\mathbf{\Sigma}} - \mathbf{\Sigma})(dN_t - \bar{\Lambda} dt) \end{cases}$$

where dN_t denote increments in the event-counting process N_t, and we define

$$\begin{cases} \bar{\Lambda} := & \lambda_0 + \sum_i \Lambda_i \\ \bar{\boldsymbol{\mu}} := & \frac{\lambda_0}{\bar{\Lambda}} \boldsymbol{\mu} + \sum_i \frac{\Lambda_i}{\bar{\Lambda}} \bar{\boldsymbol{\mu}}_i \\ \bar{\mathbf{\Sigma}} := & \frac{\lambda_0}{\bar{\Lambda}} \mathbf{\Sigma} + \sum_i \frac{\Lambda_i}{\bar{\Lambda}} \left(\boldsymbol{K}_i + (\bar{\boldsymbol{\mu}}_i - \boldsymbol{\mu})(\bar{\boldsymbol{\mu}}_i - \boldsymbol{\mu})^T \right) \end{cases} \begin{cases} \Lambda_i := & \lambda_i N(\phi_i|\mu_\phi, v_i + \mathbf{\Sigma}_{11}) \\ \bar{\boldsymbol{\mu}}_i := & \boldsymbol{K}_i \left(\begin{pmatrix} v_i^{-1} \phi_i \\ 0 \end{pmatrix} + \mathbf{\Sigma}^{-1} \boldsymbol{\mu} \right) \\ \boldsymbol{K}_i := & \left(\begin{pmatrix} v_i^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \mathbf{\Sigma}^{-1} \right)^{-1} \end{cases} \tag{2}$$