Contents lists available at ScienceDirect

Cognition

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



Original articles

Probabilistic modelling of microtiming perception

Thomas Kaplan^{a,*}, Lorenzo Jamone^b, Marcus Pearce^{a,c}

^a School of Electronic Engineering & Computer Science, Queen Mary University of London, London, United Kingdom
^b School of Engineering & Materials Science, Queen Mary University of London, London, United Kingdom

^c Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

Department of Culture Medicine, Aurilias Oniversity, Aurilias, Denina

ARTICLE INFO

ABSTRACT

Dataset link: http://dx.doi.org/10.17605/OSF.I O/XHZ3K Keywords: Music cognition Pluther persention

Music cognition Rhythm perception Microtiming Bayesian inference Music performances are rich in systematic temporal irregularities called "microtiming", too fine-grained to be notated in a musical score but important for musical expression and communication. Several studies have examined listeners' preference for rhythms varying in microtiming, but few have addressed precisely how microtiming is perceived, especially in terms of cognitive mechanisms, making the empirical evidence difficult to interpret. Here we provide evidence that microtiming perception can be simulated as a process of probabilistic prediction. Participants performed an XAB discrimination test, in which an archetypal popular drum rhythm was presented with different microtiming. The results indicate that listeners could implicitly discriminate the mean and variance of stimulus microtiming. Furthermore, their responses were effectively simulated by a Bayesian model of entrainment, using a distance function derived from its dynamic posterior estimate over phase. Wide individual differences in participant sensitivity to microtiming were predicted by a model parameter likened to noisy timekeeping processes in the brain. Overall, this suggests that the cognitive mechanisms underlying perception of microtiming reflect a continuous inferential process, potentially driving qualitative judgements of rhythmic feel.

1. Introduction

Humans can detect remarkably precise timing perturbations in auditory rhythms, in the order of tens of milliseconds (Repp, 2005; Repp & Su, 2013). Such perturbations from strictly metronomic (quantised) stimuli are commonplace in expressive music performance, often called *microtiming*, which we define as sub-syntactic timing deviations that are too small for formal musical notation. In the performance of written music, expressive microtiming corresponds to a departure from the nominal note durations encoded in the score, which are quantised to a time grid of fixed granularity. Expressive microtiming has also been described as the "remainder" of timing information that is lost following categorical perception of a rhythm (Clarke, 1987), given the well-documented tendency to perceive rhythms performed with expressive microtiming in terms of integer-ratio interval categories, e.g. 1:1, 1:2 and 1:3 (Desain & Honing, 2003; Jacoby & McDermott, 2017; Ravignani et al., 2018; Schulze, 1989).

Microtiming is often considered essential in establishing a rhythmic "feel" that encourages motor engagement with music. Most notably, Keil (1987, 1995) proposed the theory of "participatory discrepancies", whereby variation within a musician's expressive performance of a rhythmic pattern, and phase-shifting between musicians in an ensemble, brings about the pleasurable sensation of "groove" (Senn et al.,

2019) for listeners. Such expressive microtiming has been widely studied, particularly for African-American groove-based music (e.g. Collier & Wright, 1995; Ellis, 1991; Friberg & Sundström, 2002; Hosken, 2021), with authors often relating systematic microtiming patterns to qualitative feelings of a rhythm being "pushed" or "laid-back" (alternatively "on top" or "relaxed", see Iyer, 2002; Prögler, 1995). In spite of this literature demonstrating the importance of expressive timing for skilled musicians, it is unclear whether this is the case for an average listener. Surprisingly, several studies have reported similar or greater subjective "groove" ratings for rhythms without microtiming (Datseris et al., 2019; Frühauf et al., 2013; Madison, 2006; Senn et al., 2016; Skaansar et al., 2019), with listeners struggling to discern leading instruments for moderate timing deviations (~30 ms, Butterfield, 2010; Matsushita & Nomura, 2016). It is possible this surprising finding relates to the implicit, unconscious nature of phase corrections made in response to subtle timing variation of auditory stimuli in sensorimotor synchronisation studies (Repp, 2005). However, it is challenging to interpret these results, as there are currently no formal models of the cognitive processes underlying microtiming perception.

Several authors have invoked high-level concepts from the framework of Dynamic Attending Theory (DAT, Jones & Boltz, 1989) in

https://doi.org/10.1016/j.cognition.2023.105532

Received 15 December 2022; Received in revised form 11 June 2023; Accepted 21 June 2023 Available online 11 July 2023 0010-0277/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



^{*} Correspondence to: School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, United Kingdom *E-mail address:* t.m.kaplan@qmul.ac.uk (T. Kaplan).

describing microtiming perception. DAT characterises temporal expectations as "attentional pulses" stemming from the entrainment of neural oscillations. Each pulse reflects an anticipated event at a point in time (at approximately the mode of the pulse), and the pulse "width" determines how focused an expectation is. While concepts such as "attentional pulses" are difficult to define in terms of neural mechanisms, Large and Jones (1999) defined them computationally in terms of periodic probability density functions within an adaptive oscillator model, which adapts to the level of temporal regularity in a stimulus rhythm. Large and Palmer (2002) successfully applied this model to simulate the perception of large temporal fluctuations (rubato), and temporal asynchronies within chords (melody leads), but it has not been applied directly to expressive microtiming and associated qualitative rhythmic feel. Nevertheless, several authors have made the theoretical proposal that microtiming causes phase perturbations which determine the temporal width of attentional pulses, ultimately influencing the certainty in event timing (Butterfield, 2010; Danielsen, 2018; Skaansar et al., 2019). Similarly, Hosken (2021) characterised systematic microtiming patterns in drum rhythms from popular music in terms of probabilistic timing distributions, likening the sufficient statistics of these distributions to qualitative accounts of rhythmic feel, e.g. "pushed" or "laid-back" timing, as illustrated in Fig. 1.

Given that existing accounts of microtiming perception remain somewhat vague as to the underlying cognitive mechanisms, we attempt to formalise them using a probabilistic description of entrainment-defined here as flexible temporal alignment of a biological or behavioural process with the regularities in an exogenous stimulus. Specifically, we consider the "Phase Inference from Point Process Event Timing" (PIPPET) framework, where entrainment is considered a process of continuous Bayesian inference of stimulus phase (Cannon, 2021), inspired by neural theories of the motor systems' predictive role in human beat perception (Cannon, 2021; Proksch et al., 2020; Ross et al., 2016). PIPPET explicitly tracks phase as a Gaussian distribution, as per variational approaches to modelling perceptual inference (Buckley et al., 2017; Friston et al., 2010), dynamically estimating both phase and uncertainty. While PIPPET does not address periodicity (tempo) estimation, a variant called PATIPPET estimates both phase and periodicity (Cannon, 2021, p. 7), and could be evaluated in future work. PIPPET's posterior estimate on phase adjusts in response to timing deviations, with respect to an individual listener's prior expectations for timing variance and noise sources, plausibly driving rhythmic feel. For two differently microtimed performances of the same (notated) rhythm, we expect differences in PIPPET's posterior distribution over time-stemming from microtiming-to predict listeners' qualitative judgements of rhythmic feel. Although we formulate and test PIPPET as a model of musical timing, it could equally be applied to perception of event timing in other sensory domains (e.g., speech). It is worth noting that predictive coding schemes are largely in-keeping with DAT, as noted by Skaansar et al. (2019), but PIPPET differs from DAT in its explicit dynamic estimation of phase uncertainty, and how it moderates the continuous effects of prior expectations (see Cannon, 2021, pp. 1-2).

Here, we perform an experimental test of microtiming perception, and simulate the behavioural data using PIPPET to validate its suitability as a model for explaining microtiming perception. We assess participants' ability to discriminate drum rhythms with systematically manipulated microtiming, then use PIPPET to predict the trial-by-trial responses of individual participants, in order to explain inter-individual differences in task performance.

2. Methods

2.1. Participants

Fifty-six participants (27 F) were recruited and paid via the Prolific recruitment platform (prolific.co [Oct 2022]). All participants were at

least 18 years old (M = 38, SD = 11), based in the UK, and self-reported normal or corrected-to-normal vision and hearing and no cognitive impairments. Participants were not pre-screened for musical experience. Experimental procedures were approved by the Queen Mary Ethics of Research Committee (reference QMERC22.103) and informed consent was obtained.

2.2. Stimuli

Stimuli were synthesised using the same archetypal drum rhythm, reported as the predominant drum pattern in Western popular music (e.g. pop, rock, funk and hip-hop, Mauch & Dixon, 2012). It consists of regular (eighth-note) beats on the hi-hat (HH), with alternating (quarter-note) beats on the bass drum (BD) and snare drum (SD). Each stimulus had four measures (or bars), with each measure containing four beats (i.e. 4/4 meter), resulting in eight pattern cycles (considering the BD and SD back-beat pattern has a periodicity of two beats).

Each stimulus was first constructed from this common pattern using metronomic (i.e. perfectly quantised) timing. Microtiming deviations were then introduced for BD and SD events using noise sampled randomly (with replacement) from a normal distribution characterising a specific "timing profile" (Table 1). Each timing profile is associated with a mean onset time, either early ("pushed", PU) or late ("laidback", LA) relative to a metronome; and variance in onset times, either low ("tight", TI) or high ("loose", LO); refer to Fig. 1. These timing profiles have proven descriptive for comparing performances of popular drummers (Hosken, 2021, Ch. 3). We arrived at the task parameters through a combination of pilot experiments and preliminary simulations. First, we used the parameters tested by Hosken (2021, Ch. 4) in an earlier version of this task (M = ± 16.36 ms, SD = {10.91, 21.82} ms), but were unable to replicate their results. Therefore we widened the parameter ranges (M = ± 27.27 ms, SD = {4.09, 30.0} ms) in accordance with simulated performance improvements using a preliminary configuration of the model introduced in the Simulations section. While the timing ranges chosen exceed that of average performances (M = 0 ms, SD = 16.1 ms, at the average tempo of 103.3 bpm, Hosken et al., 2022, p. 129), they fall within the range of perceptible differences reported by other empirical studies using this archetypal rhythm (~ 16-80 ms, Frühauf et al., 2013; Matsushita & Nomura, 2016; Skaansar et al., 2019)—we revisit this in the Discussion.

Microtiming deviations were sampled separately for each BD and SD event from the same timing profiles, and once for each stimulus. Timing was fixed upon creation, such that all participants heard the same stimuli—it was not practical to generate stimuli dynamically for every stimulus presentation. 144 stimuli were created overall (excluding metronomic baseline stimuli), corresponding to 36 stimuli in each timing profile (PU_TI, PU_LO, LA_TI, LA_LO).

Stimuli were represented as MIDI and synthesised in Ogg format ($f_s = 48$ kHz) using high-quality drum kit samples from the Loop Loft commercial sample library (Matt Chamberlain's "Vintage Gretsch" drum kit, thelooploft.com [Oct 2022]). Each drum was synthesised using a single (identical) sample. The tempo was 110 beats per minute (BPM, period of ~ 545 ms), so each stimulus lasted roughly 9 s.

2.3. Procedure

jsPsych (De Leeuw & Motz, 2016, version 7.1.2) was used to construct an online listening experiment, which was hosted on Pavlovia (pavlovia.org [Oct 2022]). Participants were informed that headphones were required, and had to complete and pass a headphone screening test to ensure reasonable listening equipment and low background noise (Milne et al., 2020). The headphone screening test was performed before the experiment itself, consisted of six trials, and a participant was unable to proceed (and excluded) if they provided any incorrect responses.



Fig. 1. Archetypal drum kit rhythm from popular music, with example timing profiles annotated for events on the snare drum (SD) and bass drum (BD). In probabilistic terms, as described in Hosken (2021), "pushed" and "laid-back" rhythms correspond to slightly early or late timing on average, respectively; whereas "tight" and "loose" timing reflect low or high variance, respectively.

Table 1

Timing "profiles" used to construct stimuli, each a normal distribution characterising SD and BD event timing deviations with respect to a metronomic pulse (in milliseconds, at 110BPM). Metronomic timing was only used for stimuli in the baseline trials. Note that the task parameters were initially calculated in terms of beats, $M = \pm .05$, SD = {.055, .0075}, and then converted into milliseconds for stimulus synthesis.

Timing profile	Task		Baseline trials	
	Mean (ms)	SD (ms)	Mean (ms)	SD (ms)
Metronomic	-	-	0.0	10.0
Pushed & Tight (PU_TI)	-27.27	4.09	-100.0	10.0
Pushed & Loose (PU_LO)	-27.27	30.0	-100.0	100.0
Laidback & Tight (LA_TI)	27.27	4.09	100.0	10.0
Laidback & Loose (LA_LO)	27.27	30.0	100.0	100.0

The experiment consisted of an XAB test, modified from Hosken (2021, Ch. 4), where participants listened to three stimuli per trial (X, A and B) and chose whether A or B sounded most similar to X. Either A or B shared a timing profile with X, but did not match exactly in terms of microtiming. The rhythms were referred to as the "Reference Drummer" (X), "Drummer A" and "Drummer B". Participants were simply asked "Who sounds most like the reference drummer?"; with a note that the drummers were always different, "Note that they are new drummers each time!", to encourage independent responses in each trial. Task instructions and a screenshot of the task interface are provided in the Supplementary Materials. Participants listened to the rhythms sequentially in the order XAB, and responses were self-paced. Stimuli were played only once, i.e. could not be repeated, to ensure consistent playback across participants. In addition to their choice (Drummer A or B), participants rated their confidence along a fivepoint Likert scale (Not at all confident, Not very confident, Somewhat confident, Fairly confident, Completely confident).

Each participant performed 48 trials, presented in random order. This included 24 timing profile combinations, where X = A for 12 stimuli and X = B the other 12 stimuli. Each timing profile combination was presented twice, for a total of 48 trials—using all 144 stimuli that were generated. For each combination, the assignment of timing profiles to Drummer A and B was balanced—taking PU_TI vs PU_LO as an example, in two trials PU_TI was assigned to A and PU_LO to B; and in two trials PU_LO was assigned to A and PU_TI to B. Trials were presented in 6 blocks (8 trials each), and participants could take a brief self-paced break between blocks.

To assess participant attention without a prompt, a single "baseline" trial was added to each block in a random position (for a total of

9 trials per block), comparing a near-metronomic stimulus with a highly exaggerated stimulus of a random timing profile (see Table 1). Participants were informed about these trials before the experiment, which was not terminated early based on their performance. Following the experiment, participants reported their musical engagement and training (Gold-MSI, Müllensiefen et al., 2014).

2.4. Simulations

Model overview. The experiment was simulated using PIPPET (Cannon, 2021), a computational model which provides an explanation for how listeners anticipate the timing of a sequence of events (e.g. beats in a song, ticking of a clock) using Bayesian inference; suggesting that the brain constructs a statistical model of the temporal structure of events, which it then uses (and adjusts) in real-time to generate predictions of precisely when an event should occur next. PIPPET is one of an increasing number of dynamic models of musical rhythm perception which applies ideas from Predictive Processing theory (Large et al., 2023, p. 11, also see Vuust et al., 2022).

More precisely, PIPPET describes how an observer tracks a rhythm by continuously estimating its phase, using a generative model describing (prior) expectations for the temporal structure of events. The generative model of rhythmic events is composed of two processes: (1) a drift-diffusion process ϕ representing the continuously progressing state of phase; and (2) an inhomogeneous point process $\lambda(\phi)$ representing the observer's prior expectations for event structure. The latter is referred to as an "expectation template", and is composed by a summation of Gaussian distributions, each corresponding to an expectation for an event at a specific phase value ϕ_i , with an associated expectation precision and strength. The expectation template flexibly describes a stochastic sequence of events, which might not be periodic in nature, such that phase ϕ should be thought to advance along the real number line as opposed to a unit circle—phase can be thought of as an internal sense of time, which might deviate from actual time.

In order to optimally track the stimulus, the generative model of event timing in a rhythm given $\lambda(\phi)$ is inverted (used in reverse) through variational Bayesian inference (Friston, 2008; Friston et al., 2008) to dynamically estimate the phase ϕ based on the timing of observations. The phase estimate at any time *t* is represented as a Gaussian posterior with a mean μ_t and variance V_t (inverse of precision).¹ The

¹ Note that it is a PIPPET convention to denote the variance of its Gaussian posterior with *V* instead of σ^2 (Cannon, 2021; Kaplan et al., 2022), and σ will be introduced later to denote a different parameter.



Fig. 2. mPIPPET expectation template for one measure of the drum kit rhythm. An expectation template in mPIPPET represents the instantaneous rate of events $\lambda^{s}(\phi)$ occurring at phase ϕ for an event-stream *s*: HH, SD or BD events. In other words, the degree of expectation for events to occur at phase ϕ . Expectations are represented by Gaussians centred at a specific phases with associated variances and expectation strengths.

optimal solution to rhythm tracking is approximated by a filter which draws upon the expectation template, adjusting the estimate based on the presence or absence of events at any time *t*. In the absence of events, the mean μ_t increases at a steady rate towards the next expected event, and variance v_t (i.e. uncertainty) accumulates. Upon observing an event, the estimate is pushed towards expected event phases and variance is adjusted, by a precision-weighted sum of the prior and posterior.

Here we use a multi-stream variant of PIPPET, called mPIPPET (Cannon, 2021, p. 8), which explicitly maintains separate expectation templates for different event types, i.e. $\lambda^{s}(\phi)$ where *s* is a stream of events with a specific type. For the present paper, there are three event streams, corresponding to the different types of drum kit voice: HH, SD and BD. Upon observations, the respective expectation template is used to inform a single posterior estimate for phase. The expectation template used here is shown in Fig. 2. Expected phases are centred around metronomic timing, expectation strengths are held constant, and expectation variances are homogeneous but scaled according to the model configuration—this is covered in the final paragraph of this section, which addresses model configuration.

Please refer to Cannon (2021) for further background on PIPPET, namely the variational filtering equations, their derivation, and illustrative examples (also see Kaplan et al., 2022 for simulation of empirical data on rhythm perception and synchronisation).

Simulating discrimination. To simulate discrimination in an XAB paradigm, a distance function was implemented that directly compares the posterior distributions between two stimuli-which here have a shared rhythmic pattern and differ solely in microtiming. For two stimuli s_1 and s_2 , with a shared duration t_{max} , the posterior distribution at any time t can be compared using a statistical distance measure which quantifies the divergence between the two samples (the overlap between probability density functions). Here we use Bhattacharyya distance, which can be calculated for many common distributions within the exponential family. For the purpose of the present paper, we use a form suitable for comparing univariate Gaussian distributions (Kailath, 1967, Eq. 60), D_{BC}^{t} in Eq. (1) (see Fig. 3 for an example). Note that this is just one of several measures for comparing statistical samples (Cha, 2007), and an optimal measure should be identified in future work. The total distance D_{BC} is calculated by aggregating this value over time:

$$D_{BC}^{t}(s_{1}, s_{2}) = \frac{1}{4} \log \left(\frac{1}{4} \left(\frac{V_{t}^{s_{1}}}{V_{t}^{s_{2}}} + \frac{V_{t}^{s_{2}}}{V_{t}^{s_{1}}} + 2 \right) \right) + \frac{1}{4} \left(\frac{(\mu_{t}^{s_{1}} - \mu_{t}^{s_{2}})^{2}}{V_{t}^{s_{1}} + V_{t}^{s_{2}}} \right)$$

$$D_{BC}(s_{1}, s_{2}) = \sum_{t} D_{BC}^{t}(s_{1}, s_{2})$$
(1)

The distance D_{BC} as specified above relies on a perfect recall of the posterior phase estimate (at every time *t*) for the two stimuli s_1 and s_2

being compared, which might be implausible for human listeners. In the Discussion we revisit the realism of this implementation and consider improvements which could simulate perception more accurately.

The stimulus (A or B) deemed most similar by the model to the reference stimulus (X) is referred to as the *match*, and the other the *mismatch*. But please note again that whilst one of A or B will share a timing profile with X, the microtiming deviations in each stimulus are unique. The accuracy (or confidence) of PIPPET's prediction of a listener response can be computed by the relative difference, Δ , of the distance between the reference and mismatch, and the reference and match (similarly to Millet et al., 2021). In other words, how different the reference is to the poorly-matching stimulus, compared to the difference between the reference and the closely-matching stimulus. If this value is very small, then the reference X is similar to both stimuli A and B. If this value is large, then the reference X is much more similar to one stimulus than the other. This can be calculated as:

$$\Delta = D_{BC}(s_{\text{mismatch}}, X) - D_{BC}(s_{\text{match}}, X)$$
(2)

Configurations. Model discrimination was tested using a variety of configurations (N = 400), generated by a sweep across three parameters. Given the number of free parameters in PIPPET, this was necessary to determine the configuration(s) that are able to accurately simulate participant performance. It is also possible that variation in participant performance is reflected in different model configurations (this prediction is outlined later). Three parameters were selected which broadly span different aspects of phase inference, and they were co-varied. The parameters are:

- Expectation variance, v_i: this is the expected variance (inverse of precision) for each event's onset. Expectation variance is used to construct the expectation template, part of the generative model responsible for representing expected event structure, so serves as a prior in the model's inference of posterior phase uncertainty V;
- Expected phase noise, σ: this regulates the rate by which phase uncertainty V_t grows in the absence of expected events, whilst phase continuously progresses between observations;
- 3. Phase-tracking noise, η_{μ} : this noise term continuously perturbs the phase estimate mean μ_t , reflecting noisiness of timekeeping processes in the brain. It is not considered part of the observer's generative model, but a separate consequence of how the model is implemented physiologically (Kaplan et al., 2022, p. 23).

Discrete ranges for the above parameters were chosen manually for a balance between coverage and tractability, i.e. to achieve ample variation in model behaviour given a practicable number of simulations. Other PIPPET parameters were held constant, but might be explored in future work. When taking all combinations of values for the above three parameters, there were a total of 400 different model configurations. All parameters are listed in Supplemental Materials.

Given that the phase-tracking noise parameter η_{μ} perturbs the phase estimate μ_t using noise drawn from a standard Normal distribution (see Kaplan et al., 2022, p. 23), it introduces randomness (in the model's phase estimate) at the level of single trials. Therefore each trial was simulated ten times by a given model configuration, and the output distance measures were averaged.

2.5. Statistical analysis

Task performance was analysed in Python (version 3.8.1), using the pysdt module of the psychocaoustics package (version 0.4.6, github.com/sam81/pychoacoustics [Oct 2022]) to calculate performance in the XAB task, in terms of the detection-theoretic sensitivity measure, d', a standardised estimate of the distance between distributions characterising the two possible responses—match and mismatch (or more generally hits and false alarms, see Herzog et al., 2019,



Fig. 3. Example of distance function D'_{BC} (Eq. (1)), for a PIPPET instance tracking three stimuli: *X*, *A* and *B*. For the sake of illustration, each stimulus consists of a single metronomic hi-hat event at an expected time, e_{HH} , and a bass-drum event which is either pushed, e_{BD}^X and e_{BD}^B , or laid-back, e_{AD}^A . (A) shows the expected event timing, i.e. the prior, which is shared across all stimuli, and centred around metronomic timing—note that this is a rotated and truncated version of Fig. 2. (B) compares the distance between *X* and *A* at each time step. (C) shows the posterior phase estimates for *X* and *A* over time, which are adjusted based on event timing (dashed lines) with respect to prior expectations, and used to derive the distance. For example, e_{BD}^X arrives early, causing the phase estimates μ_i^X to be adjusted forwards, and increasing the distance between *X* and *A*. (D) and (E) do the same for *X* and *B*, showing lower distance and aligned posterior phase estimates, given the similar microtiming. Note that PIPPET's 'phase' estimate and expectations are not wrapped around a unit circle (in radian units), but instead correspond to perceived progress along a line (the expectation template), so 'phase' has a similar range to time but is not equivalent.

for an introduction). *d'* is calculated using the "difference" decision strategy, which assumes participant responses are driven by relative stimulus differences (between X and A, then X and B), as opposed to an "independent-observations" decision strategy in which participants are able to independently categorise stimuli. This is in accordance with Hautus and Meng (2002)'s recommendation to choose a difference strategy when the decision strategy adopted by observers is unknown (noting that participants were not informed of the timing profiles used to construct stimuli). Standard deviations of *d'* values were calculated as per Eq. 13.4 from Macmillan and Creelman (2004, p. 325).

Generalised linear mixed-effects models (GLMMs) were fit in R (version 4.2.1) using the lme4 package (Bates et al., 2015). Effect sizes in terms of Cohen's *d* were additionally calculated using the effectsize (Ben-Shachar et al., 2020) package, and interpreted following Gignac and Szodorai (2016)'s recommendations—these *d* statistics should not get confused with the separate *d'* statistics introduced above. The proportion of variance explained by a model was calculated using the MuMIn (Bartoń, 2022) package, which outputs two values: (1) marginal R^2 , the proportion of variance explained by the fixed effects alone, R_m^2 ; and (2) conditional R^2 , the proportion of variance explained by both the fixed and random effects, R_c^2 .

Classification performance was analysed using sklearn.metrics (version 1.1.1, scikit-learn.org/stable/modules/model_evaluation.html #classification-metrics [Oct 2022]), to determine consistency between model and participant responses, in terms of the timing profile associated with the response (PU_TI, PU_LO, LA_TI, LA_LO). We report three common metrics, given the model and participant responses: (1) precision, the proportion of model responses for a specific timing profile that agree with participants; (2) recall, the proportion of participant responses for a specific timing profile that are correctly matched by the model; and (3) F1-score, a measure which balances precision and recall using a harmonic mean. An F1 score of one is optimal, and only achieved when precision and recall both equal one (100%).

2.6. Predictions

In absence of a strong empirical basis to motivate hypotheses for improved discrimination of specific timing profiles, we expect our behavioural data to corroborate preliminary findings by Hosken (2021, Ch. 4) in an earlier version of this XAB task. Hosken reported improved discrimination when both the mean and variance of timing are different (LA_LO vs PU_TI and LA_TI vs PU_LO). These orthogonal stimuli vary along both parameters (late/pushed and tight/loose), so presumably maximise the information participants can implicitly draw upon when discriminating between stimuli.

We expect to observe varying levels of participant sensitivity in microtiming discrimination, as per Hosken's findings, and expect these individual differences will be captured by PIPPET models configured with different parameters. Given that Hosken did not observe effects of musical sophistication, albeit with a less comprehensive measure than the Gold-MSI used here, we do not expect that model parameters will correlate with musical sophistication.

3. Results

3.1. Variability in participant sensitivity to microtiming

Most participants could discriminate timing profiles with sensitivity above chance (d' > 0), with just one participant achieving at-chance sensitivity (d' = 0), see Fig. 4. However, there were wide performance differences across participants, with d' values ranging between 0.0 and 3.9, M = 1.7, SD = 0.8.

A binomial generalised linear mixed model (GLMM) evaluated discrimination performance between all pairs of timing profiles forming a trial (i.e. combinations of timing profiles for stimuli A and B), see Table 2. Seven participants were excluded due to poor performance on the baseline tests (two or more answers incorrect). The GLMM had relatively weak explanatory power, accounting for a small proportion of variance ($R_c^2 = .06$, $R_m^2 = .12$), but exhibited several significant



Fig. 4. Participant task performance, as: (A) d' values, error-bars denoting 95% CI; and (B) hit- and false-alarm rates. Note that the median split considers the lower CI of d' scores, which resulted in an even split of participants (28 per group), and is used in subsequent analysis.

Table	2
-------	---

Binomial generalised linear mixed model fit to predict participant correctness from timing profile comparison type, with a fixed effect controlling for trial order, and random participant effects ($N_{Participant} = 49$, df = 2344, $R_c^2 = .06$, $R_m^2 = .12$). Effect sizes are reported as contrasts with respect to the intercept, i.e. discrimination LA_TI - PU_TI is not significantly different to that of the intercept LA_LO - LA_TI which has a high estimate (reasonable discrimination). Note that the Estimate and d values are calculated using separate R packages (see Methods), and have small differences for two of the reported effects.

see methods), and have small differences for two of the reported circles.					
Predictor	Estimate	SE	z	р	d [95% CI]
Trial order	-0.00	0.00	-0.30	0.766	-0.01 [-0.11, 0.08]
LA_LO - LA_TI (Intercept)	0.72	0.15	4.83	< .001	0.70 [0.45, 0.95]
LA_LO - PU_LO	-0.43	0.15	-2.84	0.004	-0.43 [-0.72, -0.13]
LA_LO - PU_TI	0.66	0.17	3.98	< .001	0.66 [0.33, 0.98]
LA_TI - PU_LO	0.95	0.17	5.44	< .001	0.95 [0.61, 1.29]
LA_TI - PU_TI	-0.07	0.15	-0.45	0.651	-0.07 [-0.37, 0.23]
PU_LO - PU_TI	0.49	0.16	3.04	0.002	0.49 [0.18, 0.81]

effects. The intercept model used to report contrasts was discrimination of timing variance when the mean was laidback (LA_LO vs LA_TI), as assigned automatically–and for our purposes arbitrarily–by the GLMM software used (refer back to Methods). Participants successfully discriminated rhythms differing in timing variance (LA_LO vs LA_TI and PU_LO vs PU_TI); and discrimination was moderately better when the mean timing was pushed (PU_LO vs PU_TI). Discrimination was much better for stimuli differing in both mean and variance of timing (LA_LO vs PU_TI and LA_TI vs PU_LO). The only discrimination that did not significantly differ from the intercept model (LA_LO vs LA_TI) was discrimination of mean timing with low timing variance (PU_TI vs LA_TI). Finally, discrimination was moderately worse when the timing variance was high (LA_LO vs PU_LO).

The Gold-MSI subscales were tested step-wise as fixed effects in the GLMM, but did not improve the model's Bayes Information Criterion (BIC). Participants' mean self-reported general sophistication was relatively low, M = 69.4, SD = 23.5, 95% CI [63.3, 75.6], corresponding to the 28th percentile for a very large sample (Müllensiefen et al., 2014, Table S3). The mean score for general sophistication in the present study was significantly lower than that of Müllensiefen et al., t(147687) = -4.408, p < .001 (unpaired t-test). The Supplementary Materials contains score distributions for each survey sub-scale.

Participant confidence ratings were significantly greater in trials where participants provided the correct response (Mann–Whitney U = 7.21e5, p < .001; Mdn_{Incorrect} = 2, Somewhat confident; Mdn_{Correct} = 3, Fairly confident), suggesting that participants had some explicit awareness of their perceived differences in stimulus timing profiles.

Binomial tests were performed to assess biases in participant responses – note that the order of presentation for stimulus A and B in every XAB combination was balanced across trials (refer back to Procedure). Two-sided tests were performed and are reported here with respect to response A. Overall, the proportion of A responses of 50.2% did not differ significantly from the expected 50.0% (p = .772). The proportion of responses for the stimulus presented first after the reference was 52.4%, narrowly greater than the expected 50.0% but statistically significant (p = .014). This indicates a mild primacy bias in participants' responses.

3.2. Individual differences explained by the probabilistic model

Across the parameter ranges tested, most model configurations performed above chance (d' > 0), see Fig. 5A&B. Model d' values spanned a larger range than participants, from -0.8 to 3.2, M = 1.3, SD = 0.7. The performance floor of models was lower than participants, with several configurations performing below chance (left side of Fig. 5A), unlike participants. Additionally, models appeared to suffer from a performance ceiling, with few models achieving hit rates exceeding 85% (top left of Fig. 5B)-this is revisited in the Discussion. Five numericallyunstable model configurations were excluded, as they could not track the stimulus of at least one trial without excessive accumulated error in the phase estimate. For the remaining model configurations, phase-tracking noise η_{μ} and expected phase noise σ parameters were significant predictors of a specific configuration's performance d', $R^2 =$.81, F(2, 392) = 550.7, p < .001 (ordinary least squares regression, OLS); whereas the expected timing variance v_i parameter did not improve the model's BIC. Better performance of a model configuration was associated with lower phase-tracking noise, $t_{\eta_u} = -36.4$, and higher expected phase noise, $t_{\sigma} = 19.0$.

Next we addressed the question of individual variance between participants by identifying the model configuration(s) whose responses best matched those of individual participants. Configurations were assigned to individual participants by maximising agreement across trials, measured by Cohen's kappa (κ). It was possible to assign the same configuration to multiple participants, such that only 33 unique

model configurations were assigned to the 56 participants. Participants with poor baseline performance were included in this analysis, as lower-performing models could feasibly predict their performance. As shown in Fig. 5D, agreement between participants and assigned models generally fell in the ranges of moderate (0.41–0.60) to substantial (0.61–0.80). The assigned models' *d'* values were significant (if moderate) predictors of the respective participant's *d'* values, Fig. 5C, $R^2 = .35$, F(1,54) = 28.85, p < .001 (OLS); confirming that participants and their individually-matched models shared similar levels of overall performance.

The phase-tracking noise (η_{μ}) parameter of assigned models was a weak but significant predictor of participant d', t = -3.11, $R^2 = .15$, F(1, 54) = 9.64, p = 0.003, see Fig. 5E. The other model parameters were not significant predictors, not improving the Akaike information criterion (AIC) when included and excluded stepwise in the OLS regression. This suggests that higher levels of phase-tracking noise in participants' rhythm tracking contribute negatively to overall performance, which is unsurprising, given perturbations to the posterior phase estimate originating from noise might offset perturbations originating from stimulus microtiming. It is also therefore unsurprising that agreement κ of assigned models also predicted participant's overall performance d', $R^2 = .37$, F(1, 54) = 32.07, p < .001; given noisy discrimination in both models and participants would reduce the likelihood of agreement, whilst also reducing performance.

Next, in order to validate the assigned models' ability to predict their participant's responses at the level of specific trials, we separate participants with lower performance, which (as described above) reflects noisy discrimination across trials. Participants were partitioned for subsequent analysis using a median split on the lower CI of participant d' (Mdn_{d'-CI} = 1.02, see Fig. 4A), to separate participants with closer-to-chance (noisy) performance from those with stronger performance. The use of the lower limit of the 95% CI allowed a perfectly even split, where each group had N = 28 participants. The shorthand < Mdn is used for the below-median group, and \geq Mdn the above-median group.</sub>

Model prediction of participant responses was first analysed in terms of classification performance—the agreement between participant and model responses, given the timing profile associated with the response. In other words, when participants' response (A or B) was associated with a specific timing profile (e.g. PU_TI), whether the model responded the same way. Table 3 reports the performance using classification measures, for both the above-median and below-median groups. Precision and recall clearly improved for participants in the above-median group, as reflected by overall classification performance: $F1_{\geq Mdn} = .82$, $F1_{<Mdn} = .66$. These overall *F*1 scores indicate good model performance for participants in the above-median split, compared with merely acceptable performance for participants in the below-median split. The F1 scores were relatively homogeneous across timing profiles, suggesting the model was consistent in predicting participant responses.

Binomial GLMMs were also created to assess whether model responses (A or B) predicted participant responses, alongside the absolute confidence (Δ) of model predictions, which corresponds to the difference between distances X-A and X-B. We hypothesised an effect of both model response and confidence, as well as an interaction between the two. For low levels of model confidence, i.e. where the X-A and X-B distances are similar, the model would predict lower participant accuracy, hence the model response would be a less effective predictor of participant responses. GLMM fit for the above-median participant group is shown in Table 4, see Supplementary Materials for the belowmedian participant group. For above-median participants, the GLMM accounted for a reasonable proportion of variance ($R_{-}^2 = .55$, $R_{-}^2 = .68$). The model response (match/mismatch) had the largest effect, strongly predicting participant responses. There was a moderate effect of confidence, and as expected a large interaction between predictors, such that the predictive power of the response decreased for low confidence values. Together, this suggests the continuous model-derived distance

Table 3

Overall classification of participant responses by individually assigned models, i.e. timing profile of response rhythm, for all participants ($N_{Support} = 336$ per timing profile). Performance is reported separately for the below- and above-median participant groups (< Mdn and \geq Mdn respectively).

Timing profile	Precision		Recall		F1	
	< Mdn	\geq Mdn	< Mdn	\geq Mdn	< Mdn	\geq Mdn
Pushed & Tight (PU_TI)	.65	.82	.74	.89	.69	.85
Pushed & Loose (PU_LO)	.69	.90	.64	.74	.66	.81
Laidback & Tight (LA_TI)	.75	.96	.50	.70	.60	.81
Laidback & Loose (LA_LO)	.61	.72	.78	.97	.68	.83
Accuracy	-	-	-	-	.66	.83
Mean	.67	.85	.66	.83	.66	.82

Table 4

Binomial generalised linear mixed model fit to predict participant responses from interacting effects of model response (A or B) and Δ (absolute confidence), for participants in the above-median split, with random effects of timing profile comparison type and participant ($N_{Participant} = 28$, df = 1338, $R_c^2 = .55$, $R_m^2 = .68$). Results for the below-median split are reported in the Supplementary Materials.

	-			-	
Predictor	Estimate	SE	z	р	d [95% CI]
(Intercept)	-2.68	0.50	-5.34	< .001	-2.56 [-3.55, -1.58]
Model response	4.88	0.25	19.59	< .001	4.69 [4.20, 5.18]
Δ	0.00	0.00	5.27	< .001	0.53 [0.17, 0.88]
$\Delta \times$ Model response	-0.00	0.00	-8.15	< .001	-0.88 [-1.27, -0.49]

explains some additional variance in participant responses, albeit a smaller portion than the (discretised) response derived from this value.

There were no significant correlations between the assigned models' parameters and musical sophistication, only a marginally significant relationship between phase-tracking noise (η_{μ}) and participant's self-reported general levels of musical sophistication; with noise decreasing for higher general Gold-MSI scores, $b_1 = -81.55$, $R^2 = .07$, F(1, 54) = 3.81, p = 0.056.

4. Discussion

The results suggest that listeners could detect systematic patterns of auditory microtiming (\sim 30 ms), characterised by specific probabilistic distributions of onset timing, and this could be simulated in terms of a continuous inferential probabilistic process tracking stimulus phase. This research makes two key contributions.

First, we found that many participants were sensitive to both the mean and variance of microtiming, without any instructed listening strategy or guidance about stimulus properties. Discrimination was most sensitive for stimuli with contrasting means and variances; as per earlier findings by Hosken (2021) using the same method but stimuli with lower levels of microtiming. This result is not necessarily intuitive, since the most sensitive discrimination might be expected where the mean timing was different and the timing variance was consistently low (PU_TI vs LA_TI)-one stimulus would be clearly pushed, and the other clearly laid-back. It is possible that such an effect would depend on smaller (and more naturalistic) pushed and laid-back mean offsets, as timing variance would introduce greater noise in interpreting mean differences-instead of being perceived as a separate parameter. However, even with our stimuli, participants were less sensitive to mean onset timing under high variance (LA_LO vs PU_LO). Interestingly, participants were more sensitive to timing variance when the mean onset timing was early rather than late (PU_LO vs PU_TI). Similar results were reported by Matsushita and Nomura (2016), where a bass guitar preceding a hi-hat (by a similar asynchrony to that of the present work, ± 31.25 ms) was more perceptible than the opposite, and participants also rated these stimuli with higher 'pleasantness'. The reason for this asymmetry requires further investigation, but one promising explanation is auditory masking (Oxenham & Wojtczak, 2010) in which earlier sounds mask later ones if the hi-hat, being higher in frequency, is a stronger simultaneous mask then the other



Fig. 5. Model task performance, and fit to individual participants: (A) model d' values across configurations (N = 400), represented as a continuous line due to the large number of data-points, with models assigned to participants annotated, and error-bars (shading) denoting 95% CI; (B) model hit- and false-alarm rates; (C) model d' versus participant d'; and (D) the distribution of κ values between participants and assigned models, with histogram area normalised with respect to a kernel density estimate (KDE, which uses the bandwidth method from Scott, 1992); and (E) model phase-tracking noise η_{μ} versus participant d'.

drum samples. However, this might not be a widespread phenomenon, given a study measuring pupil dilation in response to microtiming asynchronies ($\pm 40 \text{ ms}$ and $\pm 80 \text{ ms}$) between a double bass and a drumkit found that pupil responses did not differ significantly between early and late microtiming (Skaansar et al., 2019).

Second, participant responses were effectively predicted by a Bayesian model of entrainment (PIPPET, Cannon, 2021) fit to individual participants. Our results support a plausible mechanism underlying the perception of microtiming: an observer continuously estimates phase and tracks phase uncertainty using an approximation to optimal (Bayesian) inference about the rhythmic stimulus, and this estimate serves as a basis to compare the similarity of expressive microtiming in different performances. Additionally, we found that wide inter-individual differences in participant sensitivity were partially accounted for by the model's configured level of phasetracking noise-how much the dynamically estimated phase mean is perturbed by hypothesised noise in neural timekeeping processes. Future work might consider a larger PIPPET parameter space, and assess whether this results in more unique assignments of model configurations to participants, given several participants in the present work were matched to the same model configurations; but this might simply reflect high-agreement between participants. While there was no prior empirical basis for the continuous (i.e. not event-based) distance function developed here for PIPPET (D_{BC}) , related models involving continuous error accumulation have been proposed for discrimination of irregular rhythms (Espinoza-Monroy & de Lafuente, 2021) and tempo drift (Madison, 2004, ritardandi in music). Further, listeners have been shown to implicitly collect higher-order statistics describing auditory stimuli, which can be used to detect changes and inform perceptual discrimination through Bayesian inference (Skerritt-Davis & Elhilali, 2018, 2021).

The way in which PIPPET models microtiming perception is similar to that of the "timing-function" (TIF) model by Honing (2001), which characterises musical timing as a function mapping between score and performance time, where timing asynchronies (early/late) depend on tempo dynamics. The local time shifts represented in a TIF resemble updates to the mean (posterior) phase estimate in PIPPET, but PIPPET differs in some important ways: it operates dynamically, hence serves as a satisfying model of real-time behaviour; prior expectations do not require a one-to-one mapping onto the score-based representation of a stimulus, unlike TIFs; it takes direct inspiration from motor neurophysiology (Cannon & Patel, 2021); and phase estimates include an explicit uncertainty term, V_t . In some contexts, phase uncertainty in PIPPET might increase so much in response to repeating stimulus asynchronies that the mean phase is not subsequently shifted; which more closely resembles wide attentional pulses in DAT (Danielsen, 2018, also see Danielsen, 2010 for consideration of TIFs).

We found that PIPPET configurations hit a performance ceiling, where few configurations achieved a hit-rate exceeding 85% (Fig. 5B), unlike a handful of participants who achieved hit rates above 90% (Fig. 4B). Possibly, this stems from the distance function used to simulate participant discrimination-comparing the distance between posterior estimates at every time-step, for two stimuli, in parallel. This limited the model to a linear comparison of two stimuli, whereas participants did not listen to multiple stimuli simultaneously and are unlikely to (actively or unconsciously) recall the posterior estimates for both stimuli perfectly-recalling just one stimulus (~ 9 s) would stretch the capacity of temporal working memory (Repp, 2005, p. 972). Participants might have instead compared memorable and similar asynchronies at different times within each stimulus (e.g. BD greatly preceded the HH at the start of stimulus X, and similarly at the end of stimulus B). A similar approach might be implemented for PIPPET using a non-linear comparison of posterior estimates (e.g. using dynamic time warping, Müller, 2007), which may better-represent participant discrimination. It is also worth bearing in mind that some discrepancies between simulations and participant responses might stem from procedural biases; for example, we noticed a mild but nevertheless significant bias towards the first stimulus (of A and B) presented after the reference.

There are other important considerations for developing our modelling in future work. We configured models with mean expectations for events with metronomic timing (ϕ_i) , which is appropriate (on average) for drum rhythms from popular music (Hosken et al., 2022), but in certain music styles expressive timing might not be considered a 'deviation' from quantised timing but instead define prior expectations (Kaplan et al., 2022, p. 19) and aesthetic standards (Bengtsson, 1987; Repp, 1997). For example, expressive timing patterns systematically interact with syntactic processes in the phrasing of off-beats by samba percussionists (Gerischer, 2006), backbeats in rock and jazz rhythms (Butterfield, 2006; Iyer, 2002), or subdivision patterns in Malian jembe music (Polak, 2010). It is also possible that some of the observed participant sensitivity to specific microtiming profiles can be explained by low-level auditory features of the specific samples used for synthesis, such as the amplitude envelope (including rise time), sound frequency or duration (Danielsen et al., 2019; London et al., 2019), which deserves consideration in terms of the model's prior expectations. One other avenue for future work might be considering how the fixed rate of the model's posterior update (i.e. PIPPET's time step parameter dt) relates to sensory sampling schemes (Morillon et al., 2019), and implications for low level temporal integration or bounding of auditory events (Simon et al., 2019).

The results reported here indicate that some listeners struggled to explicitly distinguish between different performances, despite using stimuli with exaggerated microtiming compared to realistic human performances in Western popular music (Hosken et al., 2022). This raises the important question of what the precise perceptual thresholds are for specific microtiming profiles. Future iterations of the XAB task presented might explore this, in tandem with corpus studies that quantify typical performance ranges for specific patterns of microtiming deviations (e.g. for the patterns of drum stroke asynchronies identified by Câmara et al., 2022). For realistic performance ranges, participants might not reliably perceive systematic microtiming patterns–or at least draw upon the effects for explicit discrimination–without prompts of what to listen to (such as the "assertiveness" of a given instrument, Butterfield, 2010); whereas participants in the present study distinguished between microtiming patterns without an advised listening strategy.

Finally, this study might be repeated with musical experts. The lack of expertise effects in our results was consistent with Hosken (2021, Ch. 4)'s findings. Participant differences were predicted here by a model parameter attributed to noisy timekeeping processes in the brainas opposed to a parameter inside the generative model of temporal expectancy, which would presumably depend on musical experience. But our participants had relatively low levels of musical sophistication (measured by the Gold-MSI index, Müllensiefen et al., 2014), and several perceptual studies have reported more precise perception of timing deviations for musicians (e.g. Jones & Yee, 1997; Matthews et al., 2016; Repp, 2010) alongside perceptual centres dependent on genre-specific musical experience (Danielsen et al., 2021). Experience effects might also depend on more-naturalistic microtiming patterns. For drum rhythms this might include expressive timing in articulation of the hi-hat (Câmara et al., 2022), whereas it was metronomic in the stimuli used for the present paper. Expressive timing profiles might also require more detailed probabilistic descriptions that include, for example, distributional asymmetries (Hosken, 2021, Ch. 3).

5. Conclusion

Our results reveal wide individual differences in auditory microtiming sensitivity, with some listeners showing excellent performance and others struggling to explicitly distinguish microtiming between different performances of a popular drum rhythm. The cognitive mechanisms underlying sensitivity to microtiming (characterised by probabilistic distributions of timing) can be simulated as a Bayesian process of predictive entrainment. The PIPPET model embodying these mechanisms provides a robust basis for future hypothesis-driven research using actual performed rhythms and comparisons across modalities, including timing in language.

CRediT authorship contribution statement

Thomas Kaplan: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualisation, Writing – original draft, Writing – review & editing. **Lorenzo Jamone:** Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Visualization, Writing – review & editing. **Marcus Pearce:** Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

Data availability

Data and code used for running the experiment and analysis are available on an OSF repository (http://dx.doi.org/10.17605/OSF.IO/XHZ3K).

Acknowledgements

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]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Ana Clemente for advice on statistical analysis, and Anne Danielsen for comments on an earlier version of the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105532. Supplementary material: contains simulation parameters and further results.

References

- Bartoń, Kamil (2022). MuMIn: Multi-Model Inference. R package version 1.47.1 https: //CRAN.R-project.org/package=MuMIn.
- Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. http: //dx.doi.org/10.18637/jss.v067.i01.
- Ben-Shachar, Mattan S., Lüdecke, Daniel, & Makowski, Dominique (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56), 2815. http://dx.doi.org/10.21105/joss.02815.
- Bengtsson, Ingmar (1987). Notation, motion and perception: Some aspects of musical rhythm. In A. Gabrielsson (Ed.), Action and Perception in Rhythm and Music: Vol. 55, (pp. 69–80). Stockholm: The Royal Swedish Academy of Music.
- Buckley, Christopher L., Kim, Chang Sub, McGregor, Simon, & Seth, Anil K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81, 55–79. http://dx.doi.org/10.1016/j.jmp.2017.09.004.
- Butterfield, Matthew (2006). The Power of Anacrusis. Music Theory Online, 12(4), http://dx.doi.org/10.30535/mto.12.4.2.
- Butterfield, Matthew (2010). Participatory discrepancies and the perception of beats in jazz. *Music Perception*, 27(3), 157–176. http://dx.doi.org/10.1525/mp.2010.27. 3.157.
- Câmara, Guilherme Schmidt, Sioros, George, & Danielsen, Anne (2022). Mapping timing and intensity strategies in drum-kit performance of a simple back-beat pattern. *Journal of New Music Research*, 1–24. http://dx.doi.org/10.1080/09298215.2022. 2150649.
- Cannon, Jonathan (2021). Expectancy-based rhythmic entrainment as continuous Bayesian inference. *PLoS Computational Biology*, 17(6), 1–29. http://dx.doi.org/10. 1371/journal.pcbi.1009025.
- Cannon, Jonathan, & Patel, Aniruddh D. (2021). How beat perception co-opts motor neurophysiology. Trends in Cognitive Sciences, 25(2), 137–150. http://dx.doi.org/10. 1016/j.tics.2020.11.002.

T. Kaplan et al.

- Cha, Sung-Hyuk (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models* and Methods in Applied Sciences, 1(4), 300–307.
- Clarke, Eric F. (1987). Categorical rhythm perception: An ecological perspective. In Alf Ed Gabrielsson (Ed.), *International Conference on Event Perception and Action* (3rd ed.). (pp. 10–33). Royal Swedish Academy of Music.
- Collier, Geoffrey L., & Wright, Charles E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 602–627. http://dx.doi.org/10.1037/0096-1523. 21.3.602.
- Danielsen, Anne (2010). Here, There, and Everywhere: Three accounts of pulse in D'Angelo's 'Left and Right'. In Musical Rhythm in the Age of Digital Reproduction (pp. 19–36). http://dx.doi.org/10.4324/9781315596983-2.
- Danielsen, Anne (2018). The Routledge companion to popular music analysis. In Ciro Scotto, Kenneth M. Smith, & John Brackett (Eds.), *Routledge*. http://dx.doi. org/10.4324/9781315544700.
- Danielsen, Anne, Nymoen, Kristian, Anderson, Evan, Câmara, Guilherme Schmidt, Langerød, Martin Torvik, & Thompson, Marc R. (2019). Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. Journal of Experimental Psychology: Human Perception and Performance, 45(3), 402–418. http://dx.doi.org/10.1037/xhp0000611.
- Danielsen, Anne, Nymoen, Kristian, Langerød, Martin Torvik, Jacobsen, Eirik, Johansson, Mats, & London, Justin (2021). Sounds familiar(?): Expertise with specific musical genres modulates timing perception and micro-level synchronization to auditory stimuli. Attention, Perception & Psychophysics, 84(2), 599–615. http://dx. doi.org/10.3758/s13414-021-02393-z.
- Datseris, George, Ziereis, Annika, Albrecht, Thorsten, Hagmayer, York, Priesemann, Viola, & Geisel, Theo (2019). Microtiming deviations and swing feel in jazz. *Scientific Reports*, 9(1), http://dx.doi.org/10.1038/s41598-019-55981-3.
- De Leeuw, Joshua R., & Motz, Benjamin A. (2016). Psychophysics in a web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48(1), 1–12.
- Desain, Peter, & Honing, Henkjan (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3), 341–365. http://dx.doi.org/10.1068/p3370.
- Ellis, Mark C. (1991). An analysis of "swing" subdivision and asynchronization in three jazz saxophonists. *Perceptual and Motor Skills*, 73(3), 707–713. http://dx.doi.org/10. 2466/pms.1991.73.3.707.
- Espinoza-Monroy, Marisol, & de Lafuente, Victor (2021). Discrimination of regular and irregular rhythms explained by a time difference accumulation model. *Neuroscience*, 459, 16–26. http://dx.doi.org/10.1016/j.neuroscience.2021.01.035.
- Friberg, Anders, & Sundström, Andreas (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19(3), 333–349. http://dx.doi.org/10.1525/mp.2002.19.3.333.
- Friston, K. (2008). Variational filtering. NeuroImage, 41(3), 747–766. http://dx.doi.org/ 10.1016/j.neuroimage.2008.03.017.
- Friston, Karl, Stephan, Klaas, Li, Baojuan, & Daunizeau, Jean (2010). Generalised filtering. Mathematical Problems in Engineering, 2010, http://dx.doi.org/10.1155/ 2010/621670.
- Friston, K., Trujillo-Barreto, N., & Daunizeau, J. (2008). DEM: A variational treatment of dynamic systems. *NeuroImage*, 41(3), 849–885. http://dx.doi.org/10.1016/j. neuroimage.2008.02.054.
- Frühauf, Jan, Kopiez, Reinhard, & Platz, Friedrich (2013). Music on the timing grid: The influence of microtiming on the perceived groove quality of a simple drum pattern performance. *Musicae Scientiae*, 17(2), 246–260. http://dx.doi.org/10.1177/ 1029864913486793.
- Gerischer, Christiane (2006). Io Suingue Baiano:/i rhythmic feeling and microrhythmic phenomena in Brazilian percussion. *Ethnomusicology*, 50(1), 99–119. http://dx.doi. org/10.2307/20174425.
- Gignac, Gilles E., & Szodorai, Eva T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. http: //dx.doi.org/10.1016/j.paid.2016.06.069.
- Hautus, Michael J., & Meng, Xiandong (2002). Decision strategies in the ABX (matchingto-sample) psychophysical task. *Perception & Psychophysics*, 64(1), 89–106. http: //dx.doi.org/10.3758/bf03194559.
- Herzog, Michael H., Francis, Gregory, & Clarke, Aaron (2019). Experimental design and the basics of statistics: Signal detection theory (SDT). In Understanding Statistics and Experimental Design (pp. 13–22). Springer International Publishing, http://dx.doi. org/10.1007/978-3-030-03499-3_2.
- Honing, Henkjan (2001). From time to time: The representation of timing and tempo. Computer Music Journal, 25(3), 50–61. http://dx.doi.org/10.1162/ 014892601753189538.
- Hosken, F. (2021). The pocket: A theory of beats as domains (Ph.D. thesis), Northwestern University.
- Hosken, F., Bechtold, T., Hoesl, F., Kilchenmann, L., & Senn, O. (2022). Drum groove corpora. *Empirical Musicology Review*, 16(1), http://dx.doi.org/10.17605/OSF.IO/ 3SEJT.
- Iyer, Vijay (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception*, 19(3), 387–414. http://dx.doi.org/10. 1525/mp.2002.19.3.387.

- Jacoby, Nori, & McDermott, Josh 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.
- Jones, Mari R., & Boltz, Marilyn (1989). Dynamic attending and responses to time. Psychological Review, 96(3), 459–491. http://dx.doi.org/10.1037/0033-295X.96.3. 459.
- Jones, Mari Riess, & Yee, William (1997). Sensitivity to time change: The role of context and skill. Journal of Experimental Psychology: Human Perception and Performance, 23(3), 693–709. http://dx.doi.org/10.1037/0096-1523.23.3.693.

Kailath, Thomas (1967). The divergence and Bhattacharyya distance measures in signal Selection. *IEEE Transactions on Communications*, 15, 52–60.

- Kaplan, Thomas, Cannon, Jonathan, Jamone, Lorenzo, & Pearce, Marcus (2022). Modeling enculturated bias in entrainment to rhythmic patterns. In Samuel J. Gershman (Ed.), *PLoS Computational Biology*, 18(9), Article e1010579. http://dx.doi.org/10. 1371/journal.pcbi.1010579.
- Keil, Charles (1987). Participatory discrepancies and the power of music: Vol. 2, (3), (pp. 275–283). Wiley, http://dx.doi.org/10.1525/can.1987.2.3.02a00010.
- Keil, Charles (1995). The theory of participatory discrepancies: A progress report. Ethnomusicology, 39(1), 1. http://dx.doi.org/10.2307/852198.
- Large, Edward, & Jones, Mari Riess (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119–159. http://dx.doi.org/ 10.1037/0033-295X.106.1.119.
- Large, Edward, & Palmer, Caroline (2002). Perceiving temporal regularity in music: Vol. 26, (1), (pp. 1–37). Wiley, http://dx.doi.org/10.1207/s15516709cog2601_1.
- Large, Edward, Roman, Iran, Kim, Ji Chul, Cannon, Jonathan, Pazdera, Jesse K., & Trainor, Laurel J. (2023). Dynamic models for musical rhythm perception and coordination. Frontiers in Computational Neuroscience, 17, http://dx.doi.org/10. 3389/fncom.2023.1151895.
- London, Justin, Nymoen, Kristian, Langerød, Martin Torvik, Thompson, Marc Richard, Code, David Loberg, & Danielsen, Anne (2019). A comparison of methods for investigating the perceptual center of musical sounds. Attention, Perception, & Psychophysics, 81(6), 2088–2101. http://dx.doi.org/10.3758/s13414-019-01747-y.
- Macmillan, Neil, & Creelman, Douglas (2004). Detection theory: A user's guide: Vol. xix, (2nd ed.). http://dx.doi.org/10.4324/9781410611147.
- Madison, Guy (2004). Detection of linear temporal drift in sound sequences: Empirical data and modelling principles. Acta Psychologica, 117(1), 95–118. http://dx.doi. org/10.1016/j.actpsy.2004.05.004.
- Madison, Guy (2006). Experiencing groove induced by music: Consistency and phenomenology. *Music Perception*, 24(2), 201–208. http://dx.doi.org/10.1525/mp. 2006.24.2.201.
- Matsushita, Soyogu, & Nomura, Shingo (2016). The asymmetrical influence of timing asynchrony of bass guitar and drum sounds on groove. *Music Perception*, 34(2), 123–131. http://dx.doi.org/10.1525/mp.2016.34.2.123.
- Matthews, Tomas E., Thibodeau, Joseph N. L., Gunther, Brian P., & Penhune, Virginia B. (2016). The impact of instrument-specific musical training on rhythm perception and production. *Frontiers in Psychology*, 7, http://dx.doi.org/10.3389/fpsyg.2016. 00069.
- Mauch, Matthias, & Dixon, Simon (2012). A corpus-based study of rhythm patterns. In 13th International Conference on Music Information Retrieval (pp. 163–168). Porto, Portugal.
- Millet, Juliette, Chitoran, Ioana, & Dunbar, Ewan (2021). Predicting non-native speech perception using the perceptual assimilation model and state-of-the-art acoustic models. In Proceedings of the 25th Conference on Computational Natural Language Learning. Association for Computational Linguistics, http://dx.doi.org/10.18653/ v1/2021.conll-1.51.
- Milne, Alice E., Bianco, Roberta, Poole, Katarina C., Zhao, Sijia, Oxenham, Andrew J., & Billig, Alexander J. (2020). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, 53(4), 1551–1562. http://dx.doi.org/10.3758/ s13428-020-01514-0.
- Morillon, Benjamin, Arnal, Luc H., Schroeder, Charles E., & Keitel, Anne (2019). Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neuroscience & Biobehavioral Reviews*, 107, 136–142. http://www.sciencedirect.com/science/article/pii/S0149763419300922.
- Müllensiefen, Daniel, Gingras, Bruno, Musil, Jason, & Stewart, Lauren (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. In Joel Snyder (Ed.), *PLoS ONE*, 9(2), Article e89642. http: //dx.doi.org/10.1371/journal.pone.0089642.
- Müller, Meinard (2007). Dynamic time warping. Information Retrieval for Music and Motion, 69–84.
- Oxenham, Andrew J., & Wojtczak, Magdalena (2010). Frequency Selectivity and Masking. Oxford University Press, http://dx.doi.org/10.1093/oxfordhb/9780199233557.013. 0002.
- Polak, Rainer (2010). Rhythmic feel as meter: Non-isochronous beat subdivision in jembe music from Mali. *Music Theory Online*, 16(4), http://dx.doi.org/10.30535/ mto.16.4.4.
- Prögler, J. A. (1995). Searching for swing: Participatory discrepancies in the jazz rhythm section. *Ethnomusicology*, 39(1), 21. http://dx.doi.org/10.2307/852199.
- Proksch, Shannon, Comstock, Daniel C., Médé, Butovens, Pabst, Alexandria, & Balasubramaniam, Ramesh (2020). Motor and predictive processes in auditory beat and rhythm perception. *Frontiers in Human Neuroscience*, 14, http://dx.doi.org/10.3389/ fnhum.2020.578546.

T. Kaplan et al.

- Ravignani, Andrea, Thompson, Bill, Lumaca, Massimo, & Grube, Manon (2018). Why do durations in musical rhythms conform to small integer ratios? Frontiers in Computational Neuroscience, 12, http://dx.doi.org/10.3389/fncom.2018.00086.
- Repp, Bruno H. (1997). Expressive timing in a Debussy prelude: A comparison of student and expert pianists. *Musicae Scientiae*, 1(2), 257–268. http://dx.doi.org/ 10.1177/102986499700100206.
- Repp, Bruno H. (2005). Sensorimotor synchronization: A review of the tapping literature. Psychonomic Bulletin & Review, 12(6), 969–992. http://dx.doi.org/10. 3758/BF03206433.
- Repp, Bruno H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science*, 29(2), 200–213. http://dx.doi.org/10.1016/j.humov.2009.08.002.
- Repp, Bruno H., & Su, Yi-Huang (2013). Sensorimotor synchronization: A review of recent research (2006–2012). Psychonomic Bulletin & Review, 20(3), 403–452. http://dx.doi.org/10.3758/s13423-012-0371-2.
- Ross, Jessica M., Iversen, John R., & Balasubramaniam, Ramesh (2016). Motor simulation theories of musical beat perception. *Neurocase*, 22(6), 558–565. http: //dx.doi.org/10.1080/13554794.2016.1242756.
- Schulze, Hans-Henning (1989). Categorical perception of rhythmic patterns. Psychological Research, 51(1), 10–15. http://dx.doi.org/10.1007/bf00309270.
- Scott, David W. (1992). Multivariate density estimation. Wiley, http://dx.doi.org/10. 1002/9780470316849.

- Senn, Olivier, Kilchenmann, Lorenz, von Georgi, Richard, & Bullerjahn, Claudia (2016). The effect of expert performance microtiming on listeners' experience of groove in swing or funk music. *Frontiers in Psychology*, 7, http://dx.doi.org/10.3389/fpsyg. 2016.01487.
- Senn, Olivier, Rose, Dawn, Bechtold, Toni, Kilchenmann, Lorenz, Hoesl, Florian, & Jerjen, Rafael (2019). Preliminaries to a psychological model of musical groove. *Frontiers in Psychology*, 10, http://dx.doi.org/10.3389/fpsyg.2019.01228.
- Simon, Júlia, Balla, Viktória, & Winkler, István (2019). Temporal boundary of auditory event formation: An electrophysiological marker. *International Journal of Psychophysiology*, 140, 53–61. http://dx.doi.org/10.1016/j.ijpsycho.2019.04.006.
- Skaansar, Jo Fougner, Laeng, Bruno, & Danielsen, Anne (2019). Microtiming and mental effort. *Music Perception*, 37(2), 111–133. http://dx.doi.org/10.1525/mp.2019.37.2. 111.
- Skerritt-Davis, Benjamin, & Elhilali, Mounya (2018). Detecting change in stochastic sound sequences. In Wolfgang Einhüser (Ed.), PLoS Computational Biology, 14(5), Article e1006162. http://dx.doi.org/10.1371/journal.pcbi.1006162.
- Skerritt-Davis, Benjamin, & Elhilali, Mounya (2021). Computational framework for investigating predictive processing in auditory perception. *Journal of Neuroscience Methods*, 360, Article 109177. http://dx.doi.org/10.1016/j.jneumeth.2021.109177.
- Vuust, Peter, Heggli, Ole A., Friston, Karl J., & Kringelbach, Morten L. (2022). Music in the brain. Nature Reviews Neuroscience, 23(5), 287–305. http://dx.doi.org/10.1038/ s41583-022-00578-5.