# MODELING BEAT UNCERTAINTY AS A 2D DISTRIBUTION OF PERIOD AND PHASE: A MIR TASK PROPOSAL

**Martin A. Miguel[1, 2] and Diego Fernández Slezak[1, 2]**
[1]Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Departamento de Computación. Buenos Aires, Argentina.
[2]CONICET-Universidad de Buenos Aires. Instituto de Investigación en Ciencias de la Computación (ICC). Buenos Aires, Argentina.
mmiguel@dc.uba.ar

## ABSTRACT

This work proposes modeling the beat percept as a 2d probability distribution and its inference from musical stimulus as a new MIR task. We present a methodology for collecting a 2d beat distribution of period and phase from free beat-tapping data from multiple participants. The methodology allows capturing beat-tapping variability both within (e.g.: mid-track beat change) and between annotators (e.g.: participants tap at different phases). The data analysis methodology was tested with simulated beat tracks assessing robustness to tapping variability, mid-tapping beat change and disagreement between annotators. It was also tested on experimental tapping data where the entropy of the estimated beat distributions correlated with tapping difficulty reported by the participants. For the MIR task, we propose using optimal transport as an evaluation criterion for models that estimate the beat distribution from musical stimuli. This criterion provides better scores to beat estimations closer in phase or period to distributions obtained from data. Finally, we present baseline models for the task of estimating the beat distribution. The methodology is presented with aims to enhance the exploration of ambiguity in the beat percept. For example, it exposes if beat uncertainty is related to a pulse that is hard to produce or conflicting interpretations of the beat.

## 1. INTRODUCTION

The beat is a cognitive percept fundamental for the human musical experience. It is often conceived as an isochronous pattern expressed by tapping with a hand or foot. Mentally, the location and duration of musical events are defined with respect to it. It is also subjective as different listeners may select different tapping speeds or locations when producing the beat [1–3]. Also, how easily the beat percept is perceived – its pulse clarity – may vary according to the musical stimulus [4].



The uncertainty on whether there is a perceived beat and which is finally perceived is also part of the music experience. For example, pulse clarity has been related with higher valence [5] and arousal [6]. Pulse clarity has also been shown to correlate with the degree and variability of movement [7, 8]. Overall, uncertainty on the musical structure is considered relevant for the analysis of emotion in music. Theories on the mechanisms through which music produces affective responses point towards unfulfilled expectations. We listen to music and predict how it will develop. If such predictions are defied, an affective response emerges [9–11]. Predictions on the musical structure include how its events organize in time, which is arranged based on the periodicity of the beat. On top of the beat, other hierarchical structures are conceived, such as the downbeat and the meter. These, in turn, organize the length and location of repetitions and sections of a song [12, 13]. With this in mind, both the frequency and the location of the beat are of relevance.

Moreover, the certainty of our predictions is important as well. It is considered that prediction error is what causes affective response [14]. If no structure is predicted with high probability, no outcome is considered a prediction error [15]. In order to build models to analyze expectations with respect to the timing of musical events, we need to model which beat percepts emerge in listeners, if there are conflicting interpretations, and what is the certainty about them.

Ambiguity in the interpretation of the beat has been analyzed experimentally only by looking into how listeners select different tapping speeds. [16] analyzed subjective tempo on various musical stimuli, concluding that, on top of the music's base tempo, its structure and accents can also influence the selected tapping tempo. [1] observed that listeners may have different strategies to select a tapping speed. Some can comfortably tap at the fastest consistent pulse of the music, while others had a tendency towards a slower subdivided pulse.

From a Music Information Retrieval perspective, the tempo estimation task aims to estimate the most likely tapping speeds. For example, in the MIREX tempo estimation challenge, algorithms must provide the two most salient tempos together with their saliency [17] (for a review see [18–20]). Another related MIR task is beat track-

ing. The task aims to estimate, from a musical surface, the time points where a listener would tap. Several beat tracking models have been introduced, many within the beat tracking task in MIREX (for a review, see [21, 22]). These models find the best beat track, but rarely produce information on other possible interpretations.

Considering these limitations where only tempo is analyzed or only one possible beat is estimated, this work proposes modeling the beat as a 2D probability distribution of tempo and phase. The probability modeling allows capturing both tempo and location of the beat, whether more than one interpretation is likely, and overall beat certainty. Furthermore, we propose to construct the inference of the beat distribution as a MIR task. This constitutes an extension of the tempo estimation task, as it yields more information about the possible interpretations of the beat and their saliences. The extended information is relevant to understand which metrical interpretations are likely for a listener and how certain she might be, which is deemed important for the analysis of affect in music. We present the pipeline to construct this task, including how to obtain these distributions from the listener's tapping data and a proposal evaluation metric for models estimating the distributions from musical stimuli.

In the next sections, we present the formalisms of the proposed task. First, we define the 2d distribution and present the algorithm for obtaining it from tapping data. Next, we assert that the definition and the algorithm capture relevant beat situations from simulated tapping data. The algorithm is also applied on experimental tapping data where participants were required to tap to a self-selected beat on rhythms of varying complexity. We describe the experiment and show how different beat uncertainty scenarios are present from the data and that a more spread distribution correlates with reported tapping difficulty. Finally, we present an evaluation metric for MIR models producing these distributions from musical stimuli. We also provide 3 baseline models and their scores on the experimental dataset.

## 2. BEAT AS A 2D DISTRIBUTION

We propose modeling the beat percept as a 2d probability distribution of period (or tempo) $\delta$ and phase $\rho$. Such distribution allows illustrating the beat variability commonly expressed in beat-tapping data. We consider $\delta$ to be in some time unit in a bounded range reasonable for beat perception. A proposed range is 250 ms (240 bpm) to 1800 ms (33 bpm) [1]. We consider $\rho \in [0, 1]$ as a relative location within the period. The methodology estimates a discretized probability distribution $p(\delta, \rho|\text{stimulus})$ from the beat-tapping data of multiple participants while listening to the stimulus. For example, in the simulated data (section 2.1.1), we discretize the support every 25 ms for the period $\delta$ and 0.05 for the phase $\rho$. Examples of the 2d distribution for different rhythmic stimuli are presented in Figure 1.

Let us consider the tapping data as a set of tap times series $\{t_i^j\}$, with $j$ indicating the participant ($j = 1, \ldots, J$), and $i$ indicating the tap index ($i = 1, \ldots, N_j$), where $J$ is

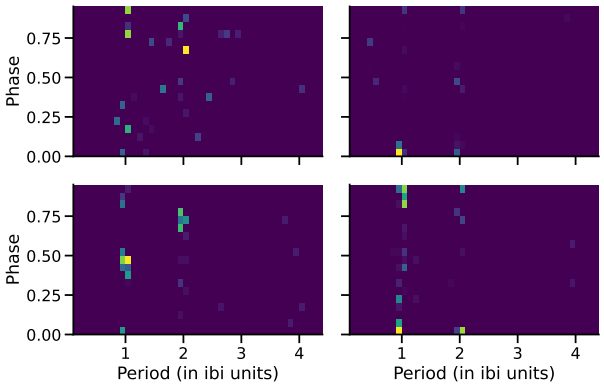


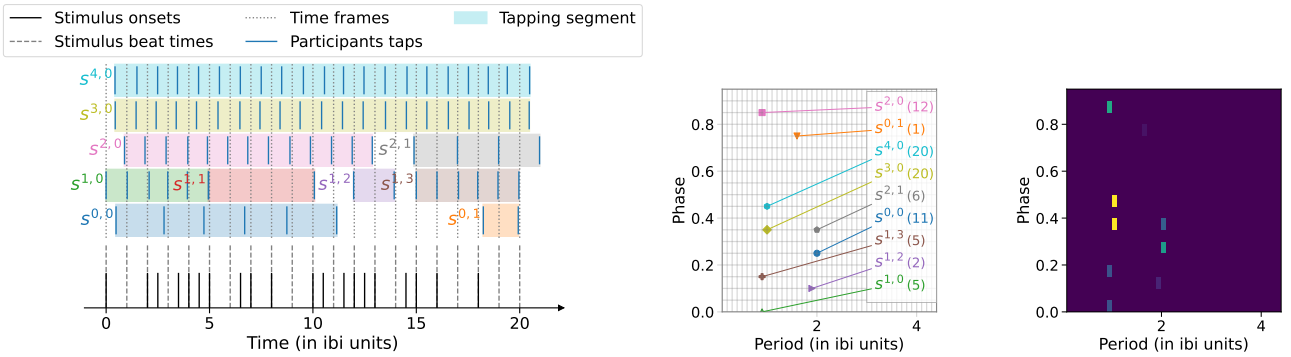

**Figure 1**: Example beat distributions from experimental tapping data for various rhythmic stimuli. The distributions depict different beat perception situations, such as high and low variability (top left and right, respectively), one period with multiple phases (bottom left) or one main phase but multiple periods (bottom right).

the number of participants and $N_j$ the number of taps by participant $j$. Our aim is to calculate the evidence for each beat bin present in the data. The process is illustrated in Figure 2. First, taps are split into segments $\{s_i^{j,k}\}_{i=1}^{S^{j,k}}$ of similar inter-tap intervals (Figure 2a). From each segment, a tactus, with inter-beat interval (IBI) $\delta$ and phase relative to the beginning of the stimuli $\rho$, is estimated with a linear regression. The period and phase are obtained using only the tapping data, without requiring a true beat level or a reference beat sequence. Next, the segment is assigned to a bin in the beat distribution (Figure 2b). To quantify the evidence provided by a segment, the stimulus is segmented into time frames which are assigned to overlapping segments (Figure 2a). The final distribution histogram is generated by adding the frame counts assigned to each beat bin by each segment and then normalizing (Figure 2c).

Taps are segmented by iterating them in order. The first two tap times constitute the first segment. Then, the segment is extended by inspecting if the distance between the last tap in the segment and the following tap time ($\Delta_*$) is similar to the average inter-tap interval of the segment ($\Delta\bar{s}^{j,k}$) within a threshold $\Delta_{th}$. The criterion is expressed in equation 1. If the equation holds, the segment is extended with the new tap time. Otherwise, a new segment is initiated with the next two tap times. $\Delta_{th}$ is defined as 0.175 based on the Continuity metric used as evaluation in the beat tracking task [23].

$$\frac{|\Delta_* - \Delta\bar{s}^{j,k}|}{\Delta\bar{s}^{j,k}} \leq \Delta_{th} \tag{1}$$

With each tap segment $\{s_i^{j,k}\}$ defined, we estimate which beat is being produced during the segment. To do so, we perform a linear regression with parameters $\alpha$ and $\beta$ minimizing the mean squared difference between $s_x^{j,k}$ and $(\alpha + \beta \times (x - 1))$ (with $x = 1, \ldots, S_i^{j,k}$). The beat selected for the segment corresponds to the beat bin of the discretized distribution containing $(\delta = \beta, \rho =$




(a) Segmented participants taps.

(b) Segments assigned to tactus beats (with assigned frame count).

(c) Final beat distribution for the section stimulus.

**Figure 2**: Illustration of the process to obtain a beat distribution from tapping data. (a) presents tapping data from 5 participants (one per row) on a section of a stimulus. Participants taps are clustered in segments $\{s_i^{j,k}\}$ of similar inter-tap intervals. From each segment, a tactus ($\rho$, $\delta$) is obtained with a linear regression. The stimulus is divided into time frames, which are assigned to the overlapping tapping segment (if any). (b) the tactus of each segment is mapped onto the distribution support. Segments contribute to their tactus bins according to the number of frames assigned to the segment (in parenthesis). (c) shows the distribution generated from the process.

$(\alpha \bmod \beta)/\beta)$. The estimated values correspond to the inter-tap interval and location with respect to the beginning of the rhythm. The segmentation process allows having more tapping information to inform the parameter estimation in spite of tapping variability.

Finally, the duration of the stimulus is split into time-frames. Each time-frame is assigned the beat bin of the segment that overlaps with it. If no segment overlaps with the frame, the frame is ignored. The beat distribution $p(\delta, \rho|\text{stimulus})$ is defined as the normalized histogram of frame counts for each beat. Figure 1 presents examples of this procedure applied to data from a tapping experiment (the experiment and the adaptations to the procedure are described in section 2.1.2).

## 2.1 Evaluation

### *2.1.1 Simulated data*

We evaluated the methodology by producing simulated tap time series from a selected beat and then calculating the beat distribution with the procedure presented in the previous section. First, we asserted that distributions obtained were robust to tapping variability. Second, considering a listener may change the beat they are tapping, we examined whether the probability of two different beats is proportional to the time each beat is produced. Third, considering data from various listeners, we assessed that the probability of two beats is proportional to the number of listeners tapping to it. For the experiments we used beat bins with period $\delta \in [250ms, 1800ms]$ with $25ms$ increments, and phase $\rho \in [0, 1]$ with steps of $0.05$.

For the first evaluation, we created simulated beat-tapping series with tapping variability and evaluated whether the beat distribution obtained had most of the probability mass in the expected beat. To produce each simulated tapping series, a beat was defined by first randomly selecting a beat bin and then drawing a specific period $\delta$ and phase $\rho$ values within the bin. With the selected period and phase, tap times were generated starting at $\rho \times \delta$ and then adding $\delta$ time for each successive tap until reaching 30 seconds. Tapping variability was controlled with parameter $\sigma$ and was incorporated into the tap series by adding Gaussian noise N(0, $\sigma \times \delta$) to each tap time. We tested 20 tapping variability magnitudes, with $\sigma \in [0, 0.08]$. Tapping variability has been reported to range from 2% to 4%, depending on inter-beat interval duration and musical training of the listener [24, 25]. It may be as high as 5% for very slow tempos (below 60 bpm) [26]. We selected 100 random beats and used each to produce 20 tapping series, one for each value of $\sigma$.

Figure 3 presents the average probability for the original beat bin at each $\sigma$. Since the variability in the tapping times might yield a different beat bin, we evaluated whether the probability mass was captured by neighboring phase and period bins. We see the probability of the originally selected beat bin decays with tapping variability, but up to 5% it is mostly captured by the immediate neighboring phase bin. The probability of the original bin together with the neighboring period bins is practically the same as the original bin, meaning that probability mass is rarely transferred to a different tempo.

The methodology is expected to work in free-tapping situations where the listener may stop and even change the beat she is tapping during the musical excerpt. The segmentation and framing procedure is used to capture these changes. We evaluate whether this behavior is captured by simulating tapping series where the beat is changed mid-tapping. Each series is generated by selecting two random period and phase values as before, selecting the proportion of time the tapping series will produce each period

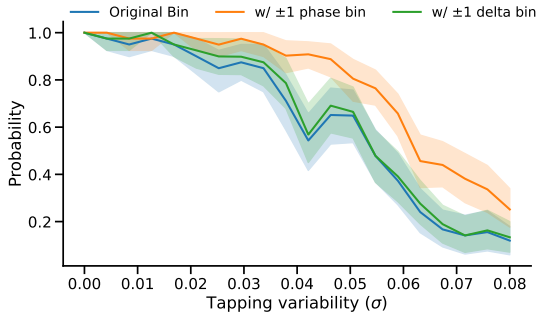



**Figure 3**: Evaluation on whether a beat distribution obtained from simulated tapping data assigns the probability to the original beat bin, with respect to tapping variability ($\sigma$). Most probability mass is given to the original bin, although it decays with variability. Up to 5% variability, most of the probability is assigned to either the original bin or its neighboring phase bins. Little probability is transferred to nearby period. Values are presented as mean probability for 100 simulations per sigma value, along with the 90% confidence interval.

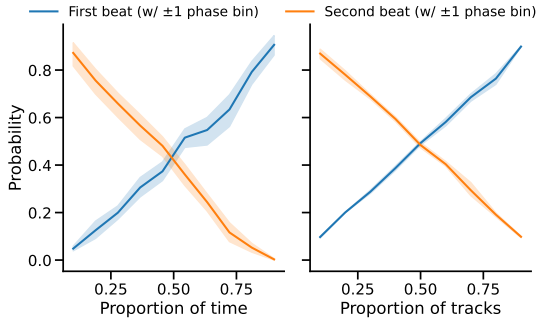



**Figure 4**: Probability of the first and second selected beats with respect to the proportion of time (left) or series (right) each beat is produced. Probability for both beats are shown to be linearly related to the amount of time or series expressing this beat. Values are presented as mean probability for 50 simulations per proportion value, together with 90% confidence intervals.

and phase, and then generating the tapping for the first and second beat for the defined proportion of the time. From these tapping series, the beat distribution is obtained and the probability assigned to the two selected beat bins is observed. Here, tapping series are generated with 3% tapping variability and are also 30 seconds long.

Similarly, the methodology is designed to capture the proportion in which different beats are selected by a set of listeners. In the last simulation, we produced a set of tapping series where a proportion was generated based on one beat and the rest on a second one. For each simulation, the beat distribution is obtained by counting the number of frames assigned to each beat bin from the complete set of tapping series. Then, the probability of each beat, with respect to the number of simulated participants producing that beat, is observed. Tapping series were also generated with 3% tapping variability and 30-second duration.

In Figure 4 we present the results of the simulations. In both cases, the probability of the first selected beat decays linearly with the proportion of time or series the beat is produced. Equally, as the proportion of time or series expresses the second beat, its probability increases.

### *2.1.2 Experimental data*

Finally, the proposed methodology was used to obtain 2d beat distributions from free-tapping data. In Figure 1 we present examples of the beat distribution for four stimuli. Data was collected on an on-site experiment where participants were asked to tap on a sensing surface to a self-selected beat while listening to rhythmic stimuli of varying rhythmic complexity [27]. The experiment was designed to capture subjective beat. Participants were instructed to tap to any self-selected beat and were allowed to stop or change their tapping throughout the stimulus. After each trial, participants rated how difficult the tapping task was with values between 1 (easy) and 5 (hard). Stimuli consisted of repeating rhythms produced using identical click sounds. A total of 33 rhythmic passages were presented to each participant. To use rhythms with validated complexity, 11 of the rhythms were taken from [3] and 7 from [28]. 5 were isochronous beats at 150, 200, 250, 500, 800 ms inter-beat intervals. 10 new patterns were created from four beat patterns (in contrast with 8 from [3] and [28]) to have participants familiarize with the task. 7 of these were always presented at the beginning of the experiment. All other stimuli presentations were randomized. With the exception of the isochronous stimuli, pattern-based stimuli were presented varying the notated inter-beat interval between 450 and 550 ms. IBIs were pseudo-randomized avoiding using the same one in two consecutive trials. Each stimulus consisted of repeating the rhythmic pattern to last a minimum of 24 seconds (and up to 31 seconds). From 35 total participants, 30 remained after filtering participants that were deemed to not understand the concept of beat. They were selected as participants who replicated the stimulus instead of defining a beat in more than three trials. 6 participants were female, and 24 are male. The overall average age was 28.27 (sd = 7.94) and overall mean musical training was 5.43 years (sd = 4.62).

From the data collected, we gathered a training subset to be used in the exploration performed in this work. 15 participants were selected to uniformly represent the range of training years. The rhythmic stimuli subset contains all the isochronous excerpts, 5 from [3], 3 from [28] and 4 from the newly proposed rhythms. Rhythms subsets were selected to uniformly represent the reported tapping difficulty range. In the experimental data, tap times are normalized to the inter-beat interval used during the experiment. This allows comparing presentations of the same stimulus at different IBIs. For the analysis in Figure 1, the beat distribution was obtained with period $\delta \in [0.1, 4.5]$ with increments of $0.01$. The figure shows how inferred beats concentrate on specific areas of the distribution. For example, we can observe that the tapping period is most often the stimulus' inter-beat interval and only in some rhythms

double-period tapping is also present. The distributions also show that the tapping phase is commonly near 0 (or 1), indicating synchrony with the stimulus' beat. This is not always the case, as in the lower left subfigure, anti-phase tapping (phase nearby 0.5) is more prevalent, indicating that participants considered this beat more likely than the one originally defined in the stimulus.

We also assessed whether the spread of the probability mass was related to the tapping difficulty reported by the participants. We estimated tapping difficulty from the distribution by calculating its entropy. Reported tapping difficulty for each stimulus was obtained as the mean of the per-participant z-standardized difficulty scores. Spearman rank's correlation was calculated on the training subset, ignoring the isochronous stimuli as they were not intended to express rhythmic complexity. The correlation yielded $r = 0.88$ and $p < 0.001$.

## 3. EVALUATION METRIC

We propose estimating the probability of different beats being perceived as the pulse of a musical stimulus as a new MIR task. An empirical discrete beat distribution (considered as period and phase) is obtained from tapping data produced by listeners. A model for this task would be required to produce estimates of the probability of each beat from the musical stimulus. To evaluate the model, both distributions must be compared. We propose using the Earth Mover's distance (or Wasserstein distance) for this comparison [29]. This distance evaluates how much probability mass must be moved in order to convert one distribution into the other. It considers two distributions with probability mass nearby to be less distant than if their mass is further apart. In contrast to the more commonly used Kullback-Leiber divergence, the Earth Mover's distance does not require both distributions to have non-zero mass on the entire support. It also takes into account the topology of the support as it allows defining a distance between the bins. We consider the distance between beat bins $(\delta_i, \rho_i)$ and $(\delta_j, \rho_j)$ in the distribution's support as the Manhattan distance between the bins. We propose adding a multiplier $M_\delta$ to the distance between periods to penalize the difference in the period more than in the phase. Here, we use $M_\delta = 5$. We also modify the distance calculation to allow a phase value close to 1 to be also close to 0, given the circularity of the phase [2, 30]. We present the used topology in equation 2.

$$d((\delta_i, \rho_i), (\delta_j, \rho_j)) = |\delta_j - \delta_i| \times M_\delta + min(|\rho_j - \rho_i|, 1 - |\rho_j - \rho_i|) \quad (2)$$

To test the proposed distance we generated pairs of distributions from increasingly distanced phase bins. Figure 5 presents the mean Earth Mover's distance and Kullback-Leiber divergence calculations for 20 simulations at each possible phase distance. We observe that the proposed distance increases linearly with the distance in phase between the distributions and then decreases when the distance in phase is reduced by the corresponding circularity. Contrastingly, the Kullback-Leiber divergence fails to capture the proximity of the distributions.

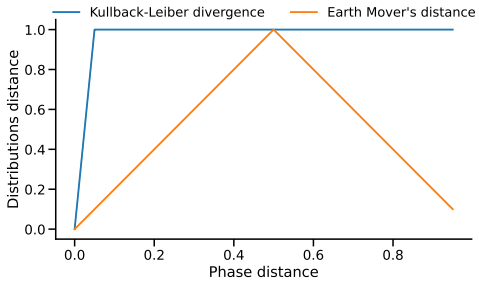



**Figure 5**: Distance metrics for two distributions only shifted in phase. The Earth Mover's distance is proportional to the shift in phase and responds to the circularity of phase. When the shift is greater than 0.5, the distance decreases.

### 3.1 Baseline models

We now present three reference models to provide an overview of the expected distance values for the task, as well as exemplify a first approach. The models are designed to take as an input a series of onset times, equivalent to the stimuli used in the experiment. The models are also expected to provide a discretized beat distribution, considering the support described in section 2.1.2. We describe the models and present their scores on the training subset.

The first reference model is the uniform distribution on any beat, i.e.: $p(\delta, \rho|\text{stimulus}) \propto 1$. Although a uniform distribution might not have the largest distance to the target distribution, we will use its distance to express the scores of the models with respect to it. In case the estimated distribution is more distant than a uniform distribution, the score will be higher than 1. The second reference model is fixed on phase 0 and only provides non-zero probability for periods $\delta_{1..4} = \{0.5, 1, 2, 4\}$. The model, named *Phase Zero*, is expressed in equation 3. The probability $D_i$ provided to each period was calibrated by fitting a Gamma distribution to the distribution of tempos selected by the participants in the isochronous stimuli of the training subset.

$$p(\,\delta, \rho \mid \text{stimulus}) \propto \begin{cases} D_i & \text{if } \delta = \delta_i \text{ and } \rho = 0 \\ 0 & otherwise \end{cases} \quad (3)$$

The third reference model, named *Beat*, assigns probability proportional to a beat fitness score multiplied with a prior distribution on the period. The fitness score projects the given period and phase throughout the length of the stimulus and evaluates whether the projected beat times coincide with musical onsets. Coincidence is measured as the density of a Gaussian window centered on the expected beat time with variance $\sigma_w = 0.1 \times \delta$ [31]. To avoid favoring faster or slower tempos, the number of correctly predicted beat times is multiplied by the number of correctly predicted onset times [22]. The model is described

| Model | Distance | Rel. Dist. |
|---|---|---|
| Beat | 6.740872 | 0.398315 |
| Phase Zero | 7.307728 | 0.460842 |
| Uniform | 17.895050 | 1.000000 |

**Table 1**: Evaluation scores on training subset for reference and baseline models. Distance is the mean Earth Mover's distance of each estimated to each empirical beat distribution. Relative Distance is the mean distance, relative to the distance of the uniform distribution, per stimulus. Lower values mean a closer estimation to the target distribution.

in equation 4. The prior for the period is given by the same Gamma described above. The distribution provided by this model is later filtered by turning to zero all beat bins where probability is below the 95th percentile, as only the most salient beats are expressed by a listener's tapping.

$$p(\,\delta, \rho \mid \text{stimulus}) \propto \text{score}(\delta, \rho, \text{stimulus}) \times \text{prior}(\delta) \quad (4)$$

$$\text{score}(\delta, \rho, s) = \frac{[\sum_j \max_b W((\gamma_b - p_j)/(0.1\delta))]^2}{|p| \times |\gamma|}$$

with $\gamma_b$ the onset times and $p_j$ the projected beat series.

In Table 1 we present the average Earth Mover's distance for each model's beat distribution to the tapping data. The table also presents the mean relative distance when compared with the uniform model for each stimulus. In this evaluation, the isochronous stimuli are excluded from the evaluation since they were used to calibrate the period prior. The table shows how assigning most of the distribution to phase zero reduces the relative distance from the uniform distribution by half. The *Beat* model adds a 6% improvement. In Figure 6 we present a sample of estimated and empirical beat distributions from the training dataset for the *Beat* model. We present the two closest and two most distant estimations in the dataset.

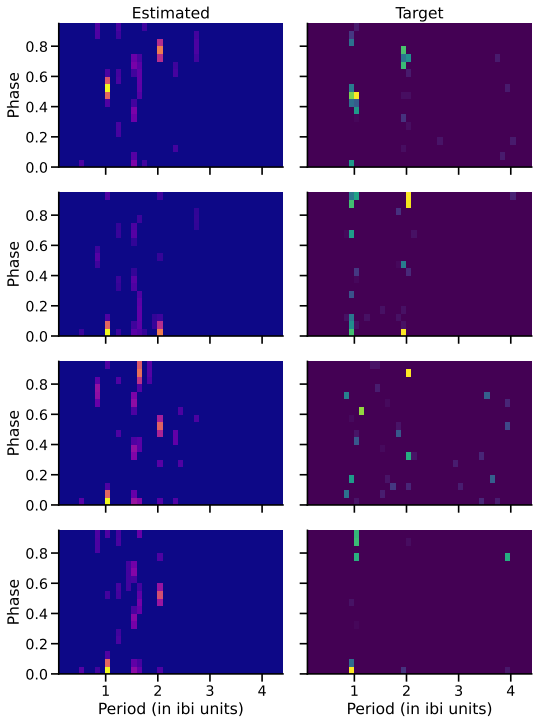



**Figure 6**: Sample of the two best (top rows) and two worst (bottom rows) beat distribution estimations by the *beat* model (left column).

## 4. DISCUSSION

We propose a new MIR task consisting of estimating the probability distribution of which beats are perceived by listeners for a musical stimulus. For this task, the beat is modeled as a 2d distribution of period and phase. The proposal includes a methodology for obtaining the distribution from tapping data of listeners and an evaluation metric for comparing estimated and empirical distributions.

The entire pipeline behavior was assessed on simulated tapping data. It was also tested on experimental data from listeners tapping to a self-selected beat while exposed to rhythms of different complexity. We also propose a set of reference models that estimate the beat distribution from the stimulus.

Modeling the beat considering period and phase extends previous analyses of beat ambiguity that mainly focused on tempo. The phase adds another dimension, as some beat ambiguity comes from where to tap, instead of at which speed. The modeling also allows a more detailed analysis of the concept of pulse clarity. For example, low pulse clarity may be due to multiple competing beat interpretations or to a single beat that is hard to follow. In the 2d distribution, the first scenario would be portrayed as multiple separated bins with equal probability and the second one as having all probability mass concentrated, but with no clear mode. Furthermore, the pipeline can be applied to tapping data from single or multiple listeners, exhibiting individual beat ambiguity or group disagreement.

This proposal can be further developed into modeling the distribution of beat series. In the beat tracking task, models are required to produce one beat track and therefore cannot capture situations where annotators disagree because more than one beat is reasonable. Another limitation to this approach is that it does not allow modeling non-isochronous beats. This would require a richer description of the beat which may not be as simple to visualize.

Finally, the focus on both dimensions of the beat, period and phase, is required for models of the meter. The added focus on uncertainty can be carried onto meter, yielding uncertainty on the whole rhythmic interpretation. This, in turn, can be used for the analysis of expectation in music as a mechanism driving affective responses. Most recent theories on this mechanism assign a key role to prediction error, which takes into account the certainty with which predictions of future events are made [11, 14]. Estimating the certainty of different beat estimations constitutes an initial step in this direction.

## 5. REFERENCES


[1] P. A. Martens, "The Ambiguous Tactus: Tempo, Subdivision Benefit, And Three Listener Strategies," *Music Perception*, vol. 28, no. 5, pp. 433–448, 06 2011. [Online]. Available: https://doi.org/10.1525/mp.2011.28.5.433

[2] E. W. Large, J. A. Herrera, and M. J. Velasco, "Neural networks for beat perception in musical rhythm," *Frontiers in Systems Neuroscience*, vol. 9, p. 159, 2015. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnsys.2015.00159

[3] W. T. Fitch and A. J. Rosenfeld, "Perception and Production of Syncopated Rhythms," *Music Perception*, vol. 25, no. 1, pp. 43–58, 09 2007. [Online]. Available: https://doi.org/10.1525/mp.2007.25.1.43

[4] O. Lartillot, P. Toiviainen, and T. Eerola, "A matlab toolbox for music information retrieval," in *Data Analysis, Machine Learning and Applications*, C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 261–268.

[5] W. Trost, S. Frühholz, T. Cochrane, Y. Cojan, and P. Vuilleumier, "Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity," *Social Cognitive and Affective Neuroscience*, vol. 10, no. 12, pp. 1705–1721, 05 2015. [Online]. Available: https://doi.org/10.1093/scan/nsv060

[6] G. Luck, P. Toiviainen, J. Erkkilä, O. Lartillot, K. Riikkilä, A. Mäkelä, K. Pyhäluoto, H. Raine, L. Varkila, and J. Värri, "Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations," *Psychology of Music*, vol. 36, no. 1, pp. 25–45, 2008. [Online]. Available: https://doi.org/10.1177/0305735607079714

[7] V. E. Gonzalez-Sanchez, A. Zelechowska, and A. R. Jensenius, "Correspondences between music and involuntary human micromotion during standstill," *Frontiers in Psychology*, vol. 9, p. 1382, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.01382

[8] B. Burger, M. R. Thompson, G. Luck, S. Saarikallio, and P. Toiviainen, "Music Moves Us: Beat-Related Musical Features Influence Regularity of Music-Induced Movement," no. July, 2012, pp. 183–187. [Online]. Available: http://icmpc-escom2012.web.auth.gr/sites/default/files/papers/183_Proc.pdf

[9] L. B. Meyer, *Emotion and meaning in music*. Chicago University Press, 1956.

[10] D. B. Huron, *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2006.

[11] P. Vuust and M. A. G. Witek, "Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music," *Frontiers in Psychology*, vol. 5, p. 1111, 2014. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2014.01111

[12] D. Temperley, "Computational models of music cognition," *The psychology of music*, pp. 327–368, 2012.

[13] C. Palmer and C. L. Krumhansl, "Mental representations for musical meter." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 16, no. 4, p. 728, 1990.

[14] R. J. Zatorre, "Why do we love music?" in *Cerebrum: the Dana forum on brain science*, vol. 2018. Dana Foundation, 2018.

[15] P. Vuust, M. J. Dietz, M. Witek, and M. L. Kringelbach, "Now you hear it: a predictive coding model for understanding rhythmic incongruity," *Annals of the New York Academy of Sciences*, vol. 1423, no. 1, pp. 19–29, 2018. [Online]. Available: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.13622

[16] D. Moelants and M. McKinney, "Tempo perception and musical content: What makes a piece fast, slow or temporally ambiguous?" in *Proceedings of the 8th International Conference on Music Perception and Cognition*, 2004, pp. 558–562.

[17] M. McKinney and D. Moelants, "Deviations from the resonance theory of tempo induction," in *Conference on Interdisciplinary Musicology*, R. Parncutt, A. Kessler, and F. Zimmer, Eds. Department of Musicology, University of Graz, 2004, pp. 124–125.

[18] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano, "An experimental comparison of audio tempo induction algorithms," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1832–1844, 2006.

[19] H. Schreiber, F. Zalkow, and M. Müller, "Modeling and estimating local tempo: A case study on chopin's mazurkas," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Montreal, Quebec, Canada*, 2020.

[20] S. Böck, F. Krebs, and G. Widmer, "Accurate tempo estimation based on recurrent neural networks and resonating comb filters." in *ISMIR*, 2015, pp. 625–631.

[21] S. Böck, F. Krebs, and G. Widmer, "Joint beat and downbeat tracking with recurrent neural networks." in *ISMIR*, 2016, pp. 255–261.

[22] M. A. Miguel, M. Sigman, and D. Fernandez Slezak, "From beat tracking to beat expectation: Cognitive-based beat tracking for capturing pulse clarity through time," *PLOS ONE*, vol. 15, no. 11, pp. 1–22, 11 2020. [Online]. Available: https://doi.org/10.1371/journal.pone.0242207

[23] A. Klapuri, A. Eronen, and J. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 342–355, 2006.

[24] S. Fujii, M. Hirashima, K. Kudo, T. Ohtsuki, Y. Nakamura, and S. Oda, "Synchronization error of drum kit playing with a metronome at different tempi by professional drummers," *Music Perception: An Interdisciplinary Journal*, vol. 28, no. 5, pp. 491–503, 2011.

[25] B. H. Repp and Y.-H. Su, "Sensorimotor synchronization: a review of recent research (2006–2012)," *Psychonomic bulletin & review*, vol. 20, no. 3, pp. 403–452, 2013.

[26] B. H. Repp and R. Doggett, "Tapping to a very slow beat: a comparison of musicians and nonmusicians," *Music Perception*, vol. 24, no. 4, pp. 367–376, 2007.

[27] M. A. Miguel, P. Riera, and D. Fernández Slezak, "A simple and cheap setup for timing tapping responses synchronized to auditory stimuli," *Behavior Research Methods (in press)*, 2021. [Online]. Available: https://doi.org/10.3758/s13428-021-01653-y

[28] D.-J. Povel and P. Essens, "Perception of temporal patterns," *Music Perception: An Interdisciplinary Journal*, vol. 2, no. 4, pp. 411–440, 1985.

[29] G. Peyré and M. Cuturi, "Computational optimal transport," 2020.

[30] N. I. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press, 1993.

[31] A. T. Cemgil, B. Kappen, P. Desain, and H. Honing, "On tempo tracking: Tempogram representation and kalman filtering," *Journal of New Music Research*, vol. 29, no. 4, pp. 259–273, 2000.