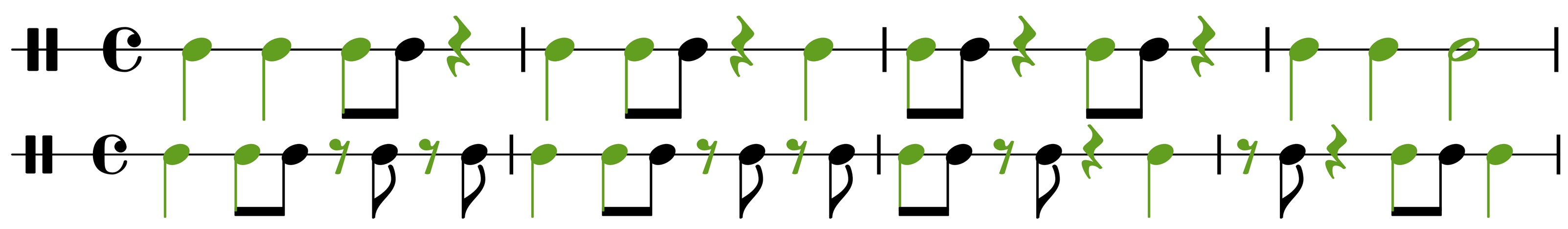
Beat tracking model for cognitive evaluation

Martin A. Miguel[1], Diego Fernandez Slezak[1], Mariano Sigman[2]

[1]Laboratorio de Inteligencia Artifical Aplicada, Departamento de Computación, Universidad de Buenos Aires, Buenos Aires, Argentina; CONICET-Universidad de Buenos Aires. Instituto de Investigación en Ciencias de la Computación (ICC). Buenos Aires, Argentina.
[2]Laboratorio de Neurociencia Integrativa, Universidad Torcuato Di Tella, Buenos Aires, Argentina
mmiguel@dc.uba.ar

Motivation

Music is known for its capability to produce a wide range of emotions; yet the mechanics by which emotions emerge are still misterious [2]. Several theories have been proposed and a particularly interesting one is the concept that emotions emerge when our expectations are unfulfilled [4].
Music consists of a series of events that unfold in time, and so do our expectations. *When* will things happen next is just as important as *what* [3]. The most basic construct to expectation in time is the **beat** or **tactus** [5]. The tactus is an internal periodic clock that a listener creates to track the timing of the events. It is generally expressed by tapping with the hand or foot while listening to a song.



Rhythm examples that abide and challenge beat expectations (top and bottom, respectively). Notes and rests that coincide with the beat are colored.

We would like to have a model of timing expectations over a rhythmic pattern to measure how expectations are (un)realized. The model should manage the following features:

- Track the beat throughout a rhythmic pattern.
- The model should be able to track the beat in expressive performances of the rhythmic pattern.
- The model should be able to track the beat without looking into the future.
- The model should provide a score to describe whether a listener would be likely to track the beat that way.

We are interested in measuring the congruency of the expectations over time. By inspecting how the congruency of the expectations fluctuate over time we might have a window into emotions evoked by music.

How does the beat come to be?

When listening to a song, we may keep track of the beat. The beat is a steady inner clock that emerges to help us organize the timing of events. A song might be harder or easier to track. Different interpretations might be possible and it must be able to adapt throughout the song.

Here we will only focus on the timing of the events. We will describe a song as a sequence of events (r_i) in ms.

The *tactus* should allow the listener to understand the timing of events. How well it does so will be the **congruency** of the beat with the timing of the events.

We can model the beat or tactus as an isochronous pulse that adapts over time. At a given point in time it can be described as a phase ρ and a period δ (both in ms).

We will consider varying possible beat interpretations as a subset of:

$$\{(\rho, \delta) \mid \rho = r_i \wedge \ \delta = r_j - r_i, \ \forall \ i < j < d\}$$

Music is often expressive in its timing, and as such the underlying beat used to organize the events may shift or even change suddenly. We will account for this by adapting the tactus hypotheses and measuring a beat's **congruency** over time.

In the beat we trust

$$congruency((\rho, \delta), (r_i)) = \frac{\text{hits of the tactus hypothesis}}{\text{number of beat events}} \times \frac{\text{hits of the tactus hypothesis}}{\text{number of song events}}$$

Where:

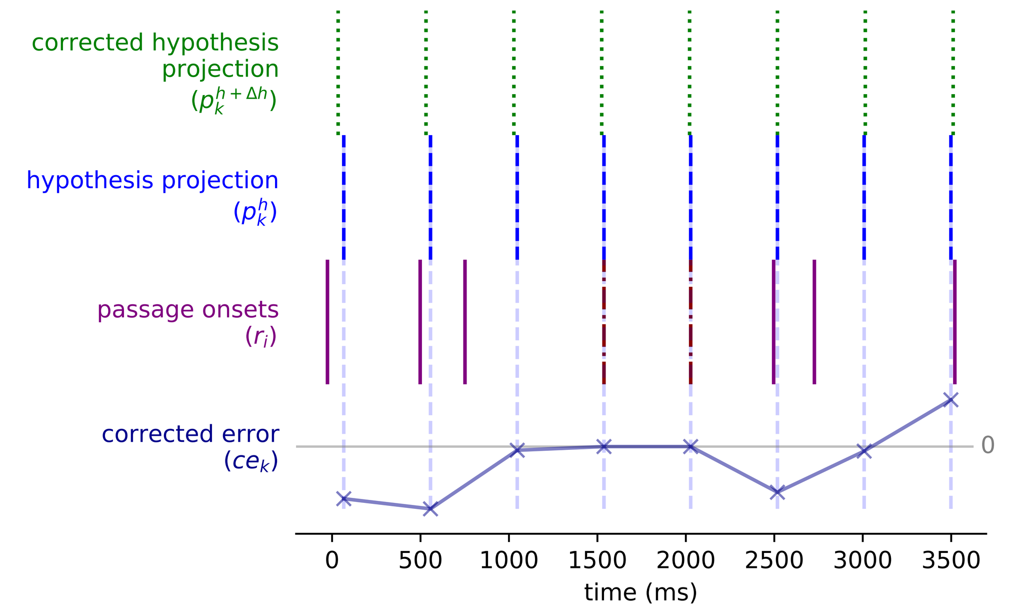
hits of the tactus hypothesis
: Number of times a the beat times predicted by the hypothesis are concurrent with an event in the song

number of beat events
: Given the ρ and δ parameters of a beat hypothesis, the number of beat events produced by this clock.

number of song events
: That is: $|(r_i)|$

Given a beat hypothesis (ρ, δ) we will have a sequence of expected beat times we will call the hypothesis projection (p_k^h). As the song unfolds, the timing of the events might shift and we will need for the hypothesis to adapt.
Moreover, in complex rhythmic patterns, not every beat time will concurr with a music event and we will need to detect such cases. We will calculate a *corrected prediction error* ce_k that will provide a correction factor for timing mismatch between beat events and music events and dismiss beat events with no corresponding music event from the correction.
m and d are error multiplication and decay factors and r_k is the closest music event to the p_k^h time event.
Since the beat times are calculated as a linear function of the hypothesis we can us a linear regression to correct the ρ and δ parameters. Because we want to dismiss beat times with no events, we will use ce_k as the error of the regression.
Corrections to the hypotheses are performed with each new music event.

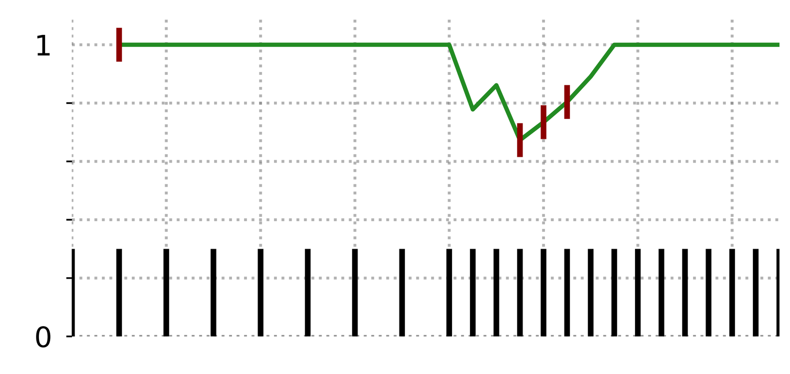
$$ce_k = m \times (r_k - p_k^h) \times d^{\frac{|p_k^h - r_k|}{\delta}}$$



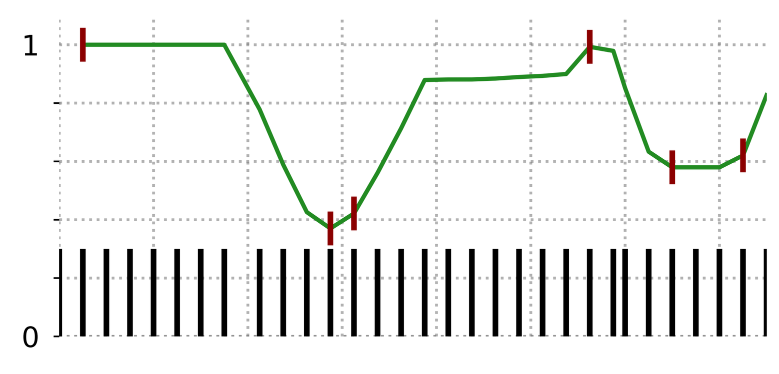
Equation for corrected error ce_k (left). Example of corrected hypothesis (right).

Evaluation

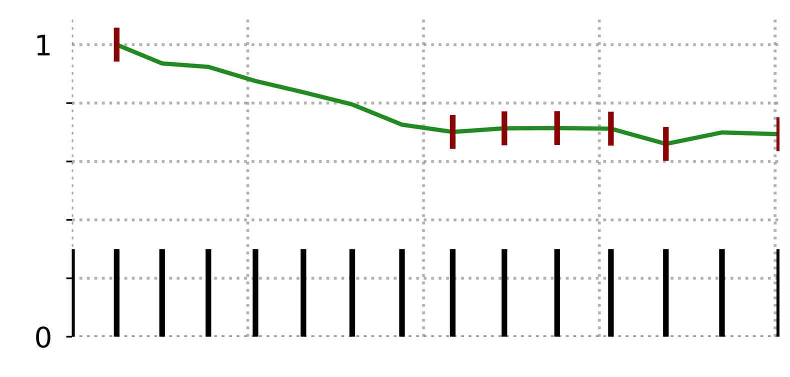
cognitive evaluation. The cognitive section of the model is expressed in certain beat tracking behaviour and development of the congruency score over-time. Examples to test specific behaviour were developed. In (a) and (b), changes in the beat reflect with a lowering of the score until a new fitting tracking is selected. In (c), the tracking slowly adapts to the rallentando without an explicit change of tracking. In (d) we show that on non trivial rhythms the score is lower than one but tries to adapt to the changes.



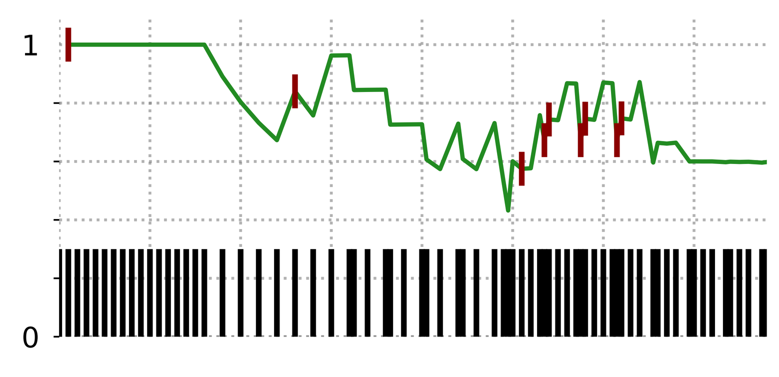
(a) period changing beat



(b) phase changing beat



(c) rallentando



(d) mixed patterns

Examples of specific passages for cognitive analysis. The plots show the confidence of the top hypothesis over time. The X axis is time. The Y axis is the congruency score of the best hypothesis at a given time. Markers are placed each time the main hypothesis changes. At the bottom of the plot, the musical events are displayed over time.

beat-tracking. We evaluated our model on the beat tracking task by using the training dataset for the MIREX Audio Beat Tracking competition. We compared our performance against Dixon's *beat root* model [1]. Performance was comparable in most metrics.

Bibliografía

[1] Simon Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58, 2001.

[2] David Huron and Elizabeth Hellmuth Margulis. Musical expectancy and thrills. 2010.

[3] Mari R Jones. Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological review*, 83(5):323, 1976.

[4] Leonard B Meyer. Emotion and meaning in music. 1956. *for an important attempt to distinguish image processes, connotations, moods, and affective experience in the apprehension of musical phenomena*, pages 256–272, 1956.

[5] Dirk-Jan Povel and Peter Essens. Perception of temporal patterns. *Music Perception: An Interdisciplinary Journal*, 2(4):411–440, 1985.