Co-regulated timing in music ensembles: a Bayesian listener perspective

Marc Leman

Ghent University, Department of Musicology, IPEM@Krook, Miriam Makebaplein 1, 9000 Gent, Belgium. Email: marc.leman@ugent.be

Abstract:
Co-regulated timing in a music ensemble can be understood as a dynamic system whose components (i.e., musicians) establish an overall state (of timing) through coordinated action (=co-regulation). An algorithm is proposed in which co-regulated timing is modelled from the perspective of a musician-listener, who uses latent processes based on Bayesian inference, simulating timing constancy. Global features of the latent processes such as their fluctuation, stability and joint constancy can be extracted. The algorithm is demonstrated and benchmarked on the off-line re-analysis of a database containing performances of duet singers (Dell'Anna et al., 2020), showing how global features correlate with perceived performance quality and agency (feeling of control). The algorithm could be upscaled to real-time applications in interactive technological environments.

1. Introduction and background

Co-regulated timing in a music ensemble can be understood as a dynamic system whose components (i.e., musicians) establish an overall state (of timing) through coordinated action (=co-regulation). From the viewpoint of a music ensemble as a whole, the capacity for timing rests on different modalities of information exchange (Bishop et al., 2019; Bishop and Goebl, 2020) and corporeal expression (Keller and Appel, 2010, Chang et al. 2019), and it is characterized by typical emergent behavior, such as resilience in response to perturbation (Glowinski et al., 2016, 2017; Hilt et al., 2019). Emergent behavior requires a balancing of underlying components and processes, so that the system can maintain a state of constancy despite fluctuations (Davies, 2016; Changeux, 1999, p. 57). In timing, these regulatory forces are directed towards a particular constancy, otherwise joint music performance would become uncertain and unpredictable, and the performance would perhaps be experienced as a "badly timed" performance The capacity for timing constancy, obviously, depends on musical skills, rehearsal time, musical complexity, and other confounds such as expression (Cochrane et al., 2013; Fabian et al., 2014, Keller, 2014). Ultimately, the quality of the timing will co-affect emotional experiences (Levitin et al. 2018, Schiavio et al., 2017) and embodied synchronizations (Witek et al., 2014; Burger et al., 2017).

From the viewpoint of a musician within a music ensemble, the capacity for timing rests on each musician's capacity to synchronize, entrain, and socially co-regulate the sound-producing actions in view of joint timing intentions (D'Ausilio et al. 2015; Volpe et al., 2016; Bishop, 2018). For example, when an onset from a musician of my ensemble is perceived earlier than expected, I can counterbalance this unexpected event. Accordingly, my next onset would be timed in such a way that an assumed joint latent hypothesis about the ensemble's tempo, meter, or expressive rhythmical structuring, can be maintained. Note that the assumed constancy also applies to expressive timing, such as increase and decrease of

tempo, because what matters is how a musician's observations of timing fit with constancy assumptions about the ensemble's joint intentional timing.

Co-regulated timing has been assumed to rest on the human sensitivity for isochrony (Ravignani and Madison, 2017). While isochrony is a basis for the emergence of timing patterns at higher temporal levels, known as meter and rhythms (Repp and Su, 2013, Kotz et al. 2018, Scheurich et al., 2020), isochrony may be embedded in rather complex temporal patterns and endowed with non-isochronous pulse structures. Yet, even in those circumstances, timing can afford precise and stable rhythmic performance and entrainment (Frühauf et al., 2013; Polak et al., 2016), suggesting that timing is based on latent processes capable of handling timing fluctuations. We thereby assume that, in these latent processes, listeners are able to "construct" the parameters of (joint) timing constancy, using perceived traces of timing (e.g., onsets) as produced by the music ensemble.

Latent perception processes can be handled by Bayesian inferences on parameters extracted from observations. As known, the Bayesian viewpoint involves a likelihood (how likely it is that a new temporal event stems from our assumed timing constancy) and a prior (the expected timing constancy in absence of observations), from which a posterior (expected timing constancy given the new temporal event) is inferred. In view of a next onset, the prior is updated by replacing the old prior with the posterior. Bayesian inference (Aitchison and Lengyel, 2017) can be conceived of in terms of neural circuits (Friston et al. 2017), hierarchical circuits (Kanai et al., 2015), homeostatic regulation (Pezzulo et al., 2015), or predictive coding (Koelsch et al., 2019). Vuust and Witek (2014) show how rhythm perception can be conceptualized as an interaction between what is heard ("rhythm") and the brain's anticipatory structuring of music ("meter"), using the predictive coding model and its implied Bayesian inference.

As far as we know, understanding the dynamics of co-regulated timing from the viewpoint of an active listener is novel. Thereby, we model the listener as a dynamic linear system equipped with latent processes that capture constancy in timing, in view of the prediction that is necessary for co-acting. Pioneering work in cognitive dynamics (e.g. Haken, 1991; Kelso,1995; Port and Van Gelder, 1995) has been inspiring for using dynamic systems in musical applications, such as gesture analysis (Demos et al., 2014), simulations of musical expectations (Agres et al., 2018), emotion research (Grimaud and Eerola, 2020), and adaptive digital music synthesis and control (Van Nort and Depalle, 2017), but not for co-regulated timing and listening as far as we know.

The goal of the present paper is to propose and benchmark an algorithm which we call the Bayesian listener algorithm, or BListener for short. BListener can be considered a local component of a global dynamic system for co-regulated timing. It is focused on perception rather than on action. However, perception is modelled from the perspective of acting and therefore, a link with operating components will be rather straightforward, provided that the algorithm is implemented in real-time.

The structure of the paper is as follows. In section 2, we present the BListener algorithm and in section 3 its basic diagnostics. In section 4, we benchmark the algorithm with different datasets, and finally in section 5 the limitations and future perspectives are discussed. We encourage the reader to experiment with the code and examples that come along with this paper. At github.ugent.be/mleman/BListener/, we provide an R-package with analysis

functions and a Supplementary Material section with scripts that generate all figures and tables.

2. The Bayesian listener algorithm

In this section, we define the basic timing concepts, and global analysis measures. Then we describe how the Bayesian listening algorithm works at sample level.

2.1. Basic concepts of co-regulated timing

Our modelling will be framed in terms of discrete events. This framing applies to musical contexts in which onsets can be clearly discerned. Using manual annotations, onset-detection algorithms, or a combination of both, we obtain a one-dimensional array of onset times (in milliseconds) that is used as input to the Bayesian listener algorithm. All other timing concepts, at least of Western music, such as inter-onset-intervals, meter and the tempo, can be based on these onsets (Clarke, 1999; London, 2012).

The recording of a music ensemble using a single microphone might suffice for extracting the onsets. Accordingly, all timing objects are based on representations of successive onsets produced by the music ensemble, regardless of the number of musicians and the musician who produces the onset. Successive onsets define **inter-onset-intervals** (IOIs), or durations of the ensemble's onset-output, and these are conceived as the timing objects from which other timing objects are derived. The algorithm aims at tracking timing objects that express timing constancy.

The timing objects can be categorized as inter-onset observations, inter-onset classes (of constancy), and meter and tempo. We use the prefix IOI to stress the fact that all these objects have a duration that can be traced back to inter-onset-intervals. The **IOI-observations**, or **IOI-observed objects**, result from onset-detection and/or onset-annotation of the music ensemble's recorded musical signal. Given two successive onsets extracted from the signal, the IOI-observed object (with its defined duration) becomes available after the second onset. The **IOI-classes** (of constancy but this explicit reference to constancy will be deleted from now on) stand for IOIs that are inferred from the IOI-observations, and they are used to anticipate future IOI-observations. IOI-classes are thus conceived in terms of latent timing parameters. They exist only as the result of a Bayesian inference about the observed co-regulated timing of the music ensemble. As latent parameter values of a (stochastic) process, IOI-classes will be defined by (Gaussian) distributions, with mean and variance. The mean will define the duration of the IOI-class and the variance will define the uncertainty about this duration. **Multiple IOI-classes** can be tracked in parallel, and as musical time proceeds, this IOI-classes result is a multivariate time series of constancy. Moreover, the Bayesian inference about the IOI-classes can be regularized by a meter, which operates as a hyper-parameter that constraints the variance in the ratio among IOI-classes. That meter is called the **out-of-time IOI-meter**, as it defines an assumed ratio among IOI-classes, disregarding how time evolves. For example, an IOI-meter with ratio {3, 2, 1} represents IOIs having an assumed ideal duration, for example, of 1500, 1000, 500 milliseconds, or, of 600, 400, 200 milliseconds. The IOI-meter can be extracted from a musical score or it can be an assumption about the relationship among IOI-classes, possibly derived from an inspection of data. For example, in an off-line analysis, one could use k-means clustering of the IOIs to obtain an estimate of the out-of-time IOI-meter. The **time-related IOI-meter** defines ratios among the multivariate IOI-class time series, representing the IOI-class ratio at each sample of the

multivariate time series. As shown below, the time-related IOI-meter can be regularized (constrained) by the out-of-time IOI-meter, and this regularization can in turn affect the Bayesian inference about the IOI-classes. Finally, the **IOI-tempo** at a particular moment in time is defined as a linear combination of the IOI-classes at any moment in time, using the out-of-time IOI-meter ratios to calculate a mean of the IOI-class time series. Examples are given below.

Blistener obeys dynamic laws which (for reasons of computational efficiency) will be updated at a sampling rate of 100 samples per second, unless specified otherwise (up to 1000 samples per second). However, inside this dynamic processing, we handle IOI objects in **log2dur scale**. The latter is a log2 transformation of the millisecond scale, and "dur" stands for duration. For example, the relation between 880 and 800 milliseconds is expressed in log2dur as a difference between log2(880) and log2(800), which is 0.1375 log2dur, and this is the same as the difference between log2(440) and log2(400), log2(440/400) or log2(1.1). Many people, however, are used to work with durations in milliseconds rather than in log2dur. Accordingly, if we speak about IOI-meter, we will specify whether the ratio is determined in milliseconds or in log2dur. What counts here is that inside Blistener, IOI objects are all handled in log2dur while the dynamic updating of Blistener proceeds in milliseconds.

2.2. Global features of co-regulated timing

Global features of the analysis of co-regulated timing are prediction error, fluctuation, stability, narration, and collapse. All these measures somehow involve the IOI-classes. The **prediction error** is defined as the difference, in log2dur, between the duration of an IOI-observed object and the duration of the IOI-class object at the time when it was assigned to this IOI-class. The mean of the absolute values of all those differences for an IOI-class is a measure of **fluctuation**, and it is calculated for each IOI-class separately. Alternatively, we can take the standard deviation of those differences, which we call **fluctuation2**. A measure of **stability** is obtained by calculating the IOI-class' standard deviation over time.

As IOI-observations get assigned to the IOI-class whose duration is closest, it is possible to label each IOI-observation assignment with a code that represents an IOI-class. Given three IOI-classes, for example, one could thus get a sequence of respective assignments, such as: 1, 2, 1, 3, 2, 2, 3, etc... This sequence is called a **narration**, and a measure of entropy or structure can be inferred from it. Without going in much detail, we will apply a recurrence analysis (e.g., Nakayama et al., 2020; Tolston et al., 2020) and calculate the recurrence ratio (RR), which amounts to a percentage representing structure in terms of the number of points that appear closely together in the phase space versus all points in the phase space, including those that don't appear closely together. Finally, **collapse** is a measure of outliers, defined as the sum of all IOI-observed durations greater than a specified threshold. Sometimes, outliers can be very small values due to onset mistakes. However, they can also be large, due to a performance breakdown of a few seconds. When detected, outliers are neglected and the system navigates on its own, based on the system's dynamics.

2.3. The Bayesian listener algorithm

The Bayesian listener algorithm is described in terms of a state-space model (Petris et al. 2009; Shumway and Stoffer, 2017), a concept that can be related to a multivariate version of the Kalman filter (cf. Meinhold and Singpurwalla, 1983).

2.3.1. The univariate approach

Figure 1 offers a graphical view of the algorithm at sample level, for the univariate case. The horizontal axis represents discrete time, with black circles as samples at t-2, t-1, and t. Associated with this time line are instances of one IOI-class (c_{t-2}, c_{t-1}, c_t). The IOI-observed object occurs at o_t. The four small arrows labelled 1 to 4, in Figure 1 show the different steps of a Bayesian inference. While in the multivariate version of the algorithm, Bayesian updating may apply to another IOI-class that runs in parallel, the IOI-observed objects appear in sequence and so, at sample level, Bayesian update is applied to only one IOI-class while the other IOI-classes will be updated according to the system equation, as explained below.

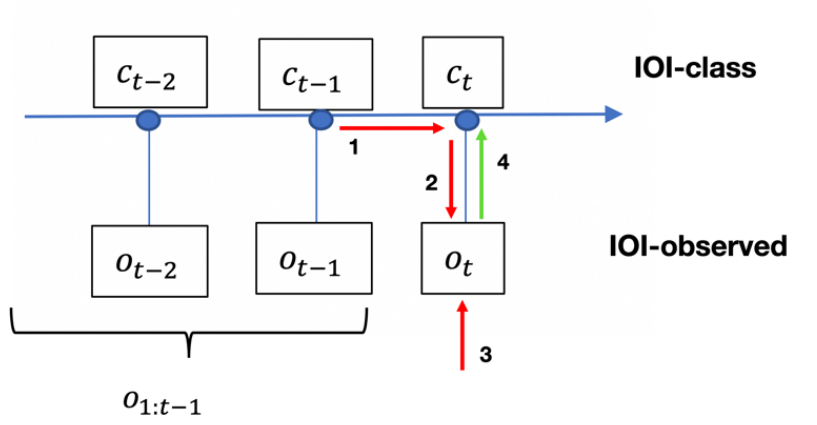


Figure 1. State-space model for Bayesian inference on a single IOI-class

The transition of the IOI-class from one sample to another sample is described by the state equation (Equation1), while the connection of an IOI-class to the IOI-observation is described by the observation equation (Equation 2). Both equations represent stochastic processes. The label *c* is used for the IOI-class, and *o* is used for the IOI-observation. The labels w_t and v_t represent independent noise whose normal distribution is characterized by zero mean and a fixed variance W_t and V_t.

$$c_t = c_{t-1} + v_t \sim N(0, V_t) \quad \text{(Equation 1)}$$

$$o_t = F_t c_t + w_t \sim N(0, W_t) \quad \text{(Equation 2)}$$

We adopt a simple rule for the state equation (Equation 1), which is that the IOI-class at time t is the same as the IOI-class at time $t-1$, except that system noise is added. Equation 1 therefore characterizes the stochastic process as *random walk*. Obviously, as the IOI-class is a latent process, its value cannot be directly observed, and thus the goal is to estimate it using IOI-observations that become available.

Provided that the IOI-observation is not an outlier, it is assigned to an IOI-class on the basis of a metric, here: the smallest difference between the duration of the IOI-observed object and the duration of any of the IOI-class objects. After the assignment, the IOI-observed object is handled as an assigned IOI, and Equation 2 applies in order to update the IOI-class using Bayes rule. The noise introduced by Equation 2 accounts for the uncertainty in the observation, such as measurement mistakes in onset detection. While this IOI-observed object is processed (i.e. involving Bayesian inference), the process of the other non-assigned IOI-classes get updated according to Equation 1.

The state equation (Equation 1) and the observation equation (Equation 2) drive stochastic processes that obey Gaussian probability laws. Hence, using the graph of Figure 1, it is possible to specify the steps from one sample to another in terms conditionals:

$$c_t|c_{t-1} \sim N(c_{t-1}, V_t)$$
$$o_t|c_t \sim N(F_t c_t, W_t)$$

The step from $c_{t-1}$to c_t obeys a Gaussian probability distribution with c_{t-1} as mean and V_t as variance. The step from $c_t$to o_t obeys a Gaussian probability distribution with $F_t c_t$ as mean and W_t as variance. Both mean and variance follow directly from Equation 1 and Equation 2. The arrows in Figure 1 indicate an updating of the IOI-class, given an IOI-observed object that becomes available. This updating can be understood as a Bayesian inference in three steps.

Step 1 (Prior). Assume that we are standing in position c_{t-1} where we have not yet encountered the IOI-observed object. Given the previously IOI-observed objects up to now (from 1 to $t-1$), we have a known mean and a known variance of the IOI-class at our disposal, which we denote by: $c_{t-1}|\, o_{1:t-1} \sim N(\hat{c}_{t-1}, \check{c}_{t-1})$. Note that at the start of the latent process, we just have to make a guess about this mean and variance, denoted $\hat{c}_{t-1}$ and $\check{c}_{t-1}$. This can be done manually, or by other methods, for example, by k-means clustering on the beginning part of a data-set. Next, we use the state equation (Equation 1) to make a hypothesis (= the prior) about our next IOI-class c_t (see arrow 1): $c_t|\, o_{1:t-1} = c_{t-1}|\, o_{1:t-1} \sim N(\hat{c}_{t-1}, \check{c}_{t-1} + V_t)$. Given the conservative approach, our hypothesis about the mean for the IOI-class c_t does not change from our current position and therefore, it has the same mean as the IOI-class c_{t-1}, namely: $\hat{c}_{t-1}$. The system variance, however, will be the proper variance of the IOI-class that we had, plus the new variance due to system noise, thus $\check{c}_{t-1} + V_t = R_t$.

Step 2 (Likelihood). While we still await the IOI-observed object (being ourselves at position c_{t-1}), we can already make a forecast about this object. That forecast will be conditioned by our previous forecast about c_t , and the observation equation that allows us to go from c_t to a forecast of o_t . Using Equation 2, we thus get (arrow 2): $o_t|c_t, o_{1:t-1} \sim N(F_t\hat{c}_{t-1}, R_t + W_t)$. The obtained mean denoted $F_t\hat{c}_{t-1}$ is our forecast about the IOI-observed o_t . The total variance Q_t will include the system variance, R_t, plus the variance due to observation, W_t, thus $R_t + W_t = Q_t$.

Next, our IOI-observed object becomes available (arrow 3) and now we can calculate the prediction error e_t as the difference between the forecast $F_t\hat{c}_{t-1}$, and the IOI-observation o_t. The prediction error is therefore $e_t = o_t\, F_t\hat{c}_{t-1}$.

Step 3 (Posterior). We now have our observation and so we can look back (in the graph) to adapt the IOI-class on which we based our forecast and prediction error measurement. This step involves the core of the Bayesian inference (arrow 4), which we formulate here in terms of probabilities: $P(\ c_t|o_{1:t}) \propto P(\ o_t|c_t{}_, o_{1:t-1})P(c_t{}_, o_{1:t-1})$. In terms of the graphism, it results in a step where we go from the IOI-observed object back to the IOI-class: $c_t|\ o_{1:t} \sim N(\widehat{c}_t, \check{c}_t)$, where $\widehat{c}_t = \widehat{c}_{t-1} + K_t e_t$ and $\check{c}_t = R_t - K_t R_t$. The mean of c_t, denoted $\widehat{c}_t$ is equal to the previous mean $\widehat{c}_{t-1}$ plus the observation error e_t, weighted. The weighting factor K_t is sometimes called the Kalman gain. It is a ratio of the system variance R_t with respect to the total variance Q_t (= system variance + observation variance). If the observation variance is small, then the effect of the error on the IOI-class is large. If the observation variance is large, the effect is small, and that will imply that a large difference between the expected IOI-observed object and the real IOI-observed object will have a small impact on the IOI-class. The variance of c_t, denoted $\check{c}_t$ is based on the previous variance, weighted by the Kalman gain. The formulas for $\widehat{c}_t$ and $\check{c}_t$ are standard for the evaluation of conditional Gaussian distributions (Bishop, 2006).

Once these three steps (four arrows in Figure 1) are taken, our updating of the IOI-class is finalized. We use this result about $\widehat{c}_t$ and $\check{c}_t$ as the assumption for our next sample, to further continue the updating As such, the Bayesian inference is made dynamic.

2.3.2. The multivariate approach, with regularization

The aforementioned state-space model applies to a multivariate model, with several IOI-class processes running at the same time. Thereby, the IOI-classes can be forced to obey the IOI-meter by adding a regularizing term to the state equation (Equation 1). That term will drive the time-related IOI-meter (or the ratio among IOI-class time series) towards the out-of-time IOI-meter. The regularization is defined as $x_t = (m - m_t) - mean(m - m_t)$, where m stands for the out-of-time IOI-meter, and m_t stands for the time-related IOI-meter at time t. The latter is obtained by taking the duration of the IOI-class objects at time t and subtracting the lowest value from these values (in log2dur !). In fact, a fraction of x_t, defined by the parameter g, is added to the IOI-class so that it gradually drives the IOI-class to the ideal out-of-time IOI-meter.

For example, assume that the three IOI-class durations at t are 8.5, 7, and 5.7 log2dur, respectively, and m is {3, 1, 0} (ratios in log2dur). Then m_t is {8.5, 7, 5.7} 5.7 = {2.8, 1.3, 0}, and $m - m_t$ is {3, 1, 0}- {2.8, 1.3, 0} = {.2,-.3,0}, whose mean is -.03. So, x_t becomes {.17, -.33, -.03}. If, in our state equation, we would add this term to the IOI-class durations at t we would drive our IOI-class to: {8.5, 7, 5.7} + {.17, -.33, -.03} = {8.67, 6.67, 5.67}. The corresponding time-related IOI-meter has {3, 1, 0}, which we wanted. However, in our state equation, we will drive the IOI-class process incrementally to this goal, hence, we multiply with a parameter that depends on the sampling rate of the process. Given the fact that new IOI-observations constantly come in occasionally, the regularization will be constantly at work, slightly driving the IOI-class processes to the out-of-time IOI-meter.

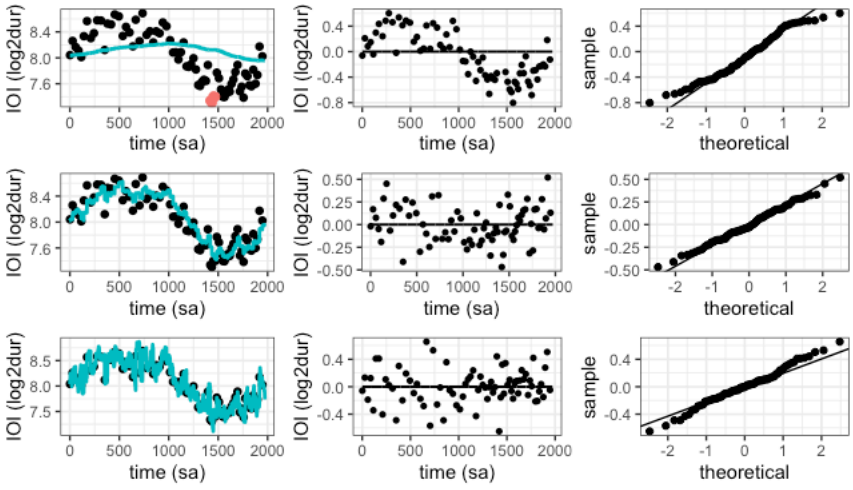
3. Diagnostic

In this section, artificial data are created and used for testing the algorithm's behavior. First, we test smoothness (in a univariate setting), then the regularization of the IOI-classes (in a

multivariate setting). See Supplementary Material for more details on parameters. Using the scripts all figures can be replicated and parameters can be changed.

3.1. Smoothness effects

Smoothness is a feature of the IOI-classes' constancy. Smoothness is co-defined by the system variance V_t and the observation variance W_t, and their ratio $r = V_t / W_t$ is overall indicative of smoothing. To test smoothing, we generated a sinusoidal signal consisting of 75 data points. Then we generated 75 random numbers using a normal distribution with zero mean and standard deviation of .15. These numbers were then added to the sinusoidal signal and the resulting values served as duration values of IOI-observed objects. We transformed these values to milliseconds so that they could be used as input to the Bayesian listener algorithm (dots in Figure 2, left panels). The goal is to retrieve a sinusoid-like IOI-class time series from the IOI-observed objects, using different values for V_t and W_t.



	Fluctuation1	Fluctuation2	Stability
r = 1e-05	0.32	0.37	0.08
r = 0.01	0.17	0.21	0.38
r = 100	0.2	0.26	0.43

Figure 2. Smoothness effects. Different values for r=V/W are shown in three rows of plots. The columns show the data and the retrieved IOI-class (log2dur over time in samples, 100 sa/sec); the residuals (log2dur over time in samples), and the distribution (samples versus theoretical normal distribution, with -1 and 1 as standard deviation). The standard deviation of the residuals can be seen in the plot of the left column, on the y-value that corresponds to -1 or 1 on the x-axis. The rows of plots correspond with the rows in the table with r=1e-4, r=0.01, and r = 100.

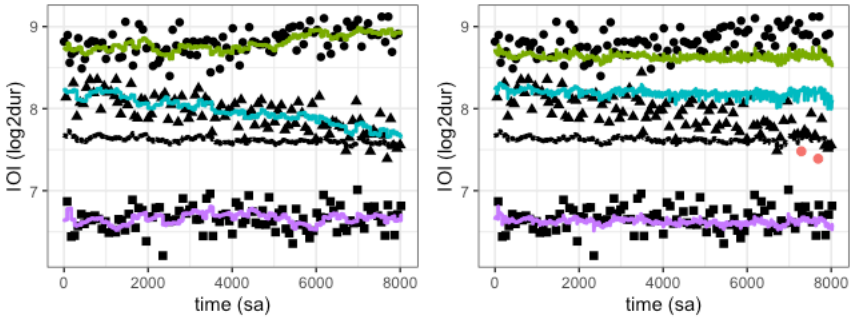
Figure 2 shows the effect for different ratios *r* in plots and a table. The first row of plots reveals a smooth curve (see the left panel) but the residuals lack a uniform distribution over time (see the middle panel) due to the fact that the curve has a delay, despite its normal distribution out-of-time (see the right panel). The second row of plots reveals a relatively smooth curve with improved uniform distribution. The third row of plots reveals severe overfitting as the IOI-class jumps to each IOI-observed object with overshooting. The table shows the measured values for fluctuation1, fluctuation2, and stability. The fluctuation1 measure is based on the prediction error before Bayesian updating. After updating, the IOI-

class will move to the IOI-observation but that brings it in a position that is away from the true mean and so the distance with another random IOI-observation is likely to be slightly larger than when the IOI-class would be positioned at the true mean. In addition, the algorithm is stochastic which means that noise is added to each observation. The second row with $r = 0.01$ and $W = 0.001$ and $V = 0.00001$ gives the best results for fluctuation1 and fluctuation2. Stability is just the standard deviation of the smoothed curve.

3.2. Regularization effects

In this section, we demonstrate the regularization effect with an artificial stimulus. As before we first define the IOI-classes and then we build IOI-observed objects that fluctuate around those IOI-classes. Here we work with three IOIs. IOI1 is going from 400 to 500 milliseconds, IOI2 goes from 300 to 200 milliseconds, and IOI3 remains constant at 100 milliseconds. Each have 100 data points, which we then translate to the log2dur scale. Then, random numbers are generated at each data point, using a normal distribution with zero mean and standard deviation of .15 log2dur. When these random numbers are added to the IOI-classes, one obtains the IOI-observed objects, shown as dots in Figure 3. The next step, then, is the creation of a one-dimensional array of IOI-observed objects, by repeatedly taking values from the three IOI-classes. Finally, all values are translated to the millisecond scale. Accordingly, the first 6 first IOI-observed values are: 354, 321, 96, 472, 312, 100 milliseconds, and so on. The dataset thus simulates the situation that an ensemble plays in a meter which, while maintaining tempo, gradually shifts to another meter.

The task is to reconstruct the IOI-classes from the IOI-observed objects. The parameter g thereby defines the strength of the regularization of the IOI-classes. Initial values are given as starting values for the IOI-classes, using an out-of-time meter indication of {4, 3,1} in milliseconds. The expected smallest IOI is 100 milliseconds. The left plot (Figure 3) shows that the IOI-classes capture the drift in the upper two IOIs. The right plot shows that the IOI-classes maintain the meter, despite the drift in the data. Due to regularization, the IOI-classes are forced to have a similar time-related IOI-meter and therefore, fluctuation1 and 2 should should increase for IOI1 and IOI2, as the prediction error becomes larger. The horizontal line around 7.6 log2dur is the IOI-tempo, which is derived from a linear combination of the IOI-classes, using the time-related IOI-meter as parameters. Often the goal is to plot tempo in the vicinity of about 2 Hz, which equals an IOI of 500 milliseconds, or about 9 log2dur (Van Noorden and Moelants, 1999). Given the relationship among the IOI-classes, the estimated tempo is very similar in both the non-regularized and the regularized analysis. We plotted the IOI-tempo one log2dur unit lower in order to have a clear picture.



g=0

	F	F2	S
IOI1	0.13	0.16	0.09
IOI2	0.13	0.17	0.15
IOI3	0.13	0.17	0.06

g=0.1

	F	F2	S
IOI1	0.22	0.18	0.03
IOI2	0.28	0.21	0.04
IOI3	0.13	0.16	0.03

Figure 3. Regularization effect in plots and table. Left plot: no regularization (g = 0). Right plot: with regularization (g=0.1). The vertical axis is log2dur over time samples (100 sa/sec). Dot-shapes (squares, triangles, circles) are IOI-observed objects that got assigned to the IOI-classes. The dotted line at about 7.6 log2dur indicates the tempo. F1 = fluctuation1, F2 = fluctuation2, and S = Stability.

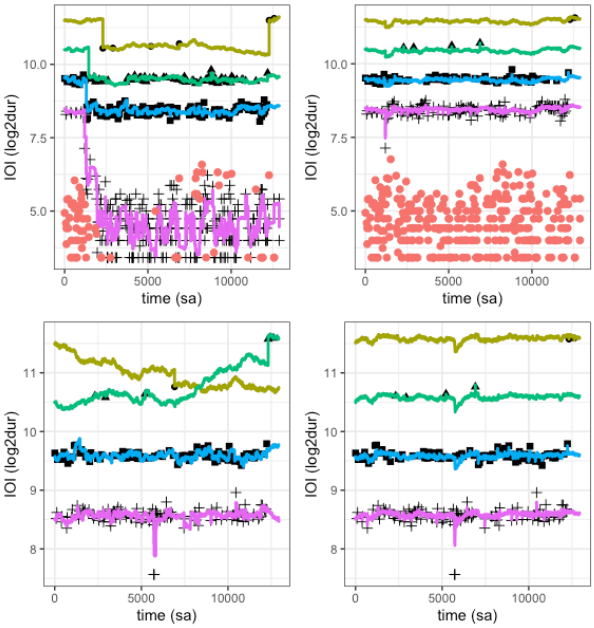
4. Applications

In this section, we apply BListener to data from real music ensembles. The first example is a recording of student choir consisting of four singers. The second example is a re-analysis of a dataset consisting of 14 duet singers each performing 8 times the same song.

4.1. The MIT2019 dataset: trial 19

Figure 4 shows data from a choir of four singers (recorded in 2019 at Gent University). Here, we analyze one single performance, called "trial 19". The top plots of Figure 4 are based on a merging of individual recordings of each singer. These onset times were then concatenated, sorted and differentiated so that one single IOI sequence was obtained, which was then given as input to the BListener. The bottom plots of figure4 are based on an omni-microphone recording of the same performance. The analysis parameters are: outt=.7, tg=360, meter = c(8,4,2,1), V= .00001, W = .001.

Due to asynchronization among singers, the merged recording shows many short IOIs. In the top left plot the analysis goes wrong because the lowest IOI-class drifts away and starts capturing those very short durations. In the top right plot, regularization is applied so that this drift doesn't happen and a rather decent pattern is obtained. In the bottom left plot, we see a random walk phenomenon in the upper two IOI-classes due to the fact that only few data are available. In the bottom right plot, regularization is applied and this reduces the variance among IOI-class time series. The corresponding tables show global measures for each IOI-class.



1

	F1	F2	S
IOI1	0.42	0.69	0.4
IOI2	0.11	0.18	0.32
IOI3	0.12	0.19	0.31
IOI4	0.65	0.76	1.23

2

	F1	F2	S
IOI1	0.05	0.08	0.06
IOI2	0.2	0.06	0.06
IOI3	0.09	0.11	0.06
IOI4	0.12	0.18	0.07

3

	F1	F2	S
IOI1	0.28	NA	0.22
IOI2	0.15	0.22	0.33
IOI3	0.07	0.1	0.07
IOI4	0.11	0.17	0.09

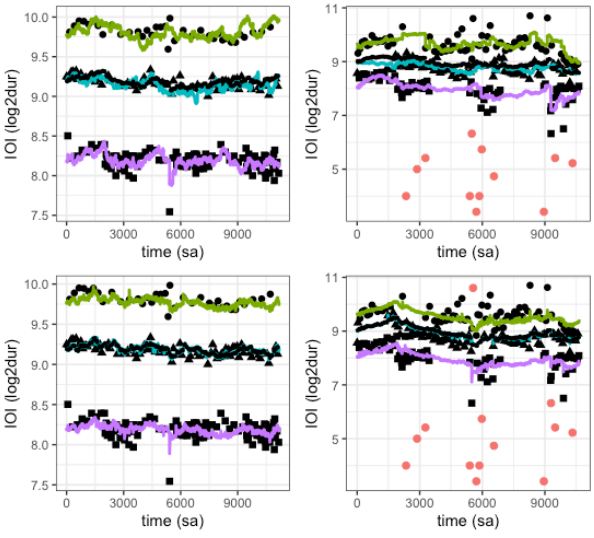
4

	F1	F2	S
IOI1	0.06	0	0.04
IOI2	0.07	0.1	0.04
IOI3	0.07	0.09	0.04
IOI4	0.09	0.15	0.05

Figure 4. BListener applied to choir singing. The left plots have no regularization (g=0), the right plots have regularization (g=.1). The top plots have onset detection based on a merge of four microphones. The bottom plots have onset detection based on a single omni-microphone recording. The tables show fluctuation and stability measures of IOI-classes. The labels "IOI1", "IOI2" etc. refer to the IOI-classes shown in the panels, starting from top to bottom. F1= fluctuation1, F2 = fluctuation2, S= Stability, and Fluctuation2.

4.2. The JustHockIt dataset

The Bayesian listener algorithm is here applied to the JustHockIt dataset of duet singers (Dell' Anna et al., 2020). Our goal was to compare the BListener results with previous reported findings. We used 14 music ensembles plus 1 reference from the JustHockIt dataset, each performing eight times the same song in two conditions, with movement (four trials) and without movement (four trials). The music ensembles consisted of two singers (duets), who alternatingly sang a note except for some short note repetitions. The parameters are: meter = {3,2,1}, outt = 1.5, V = 1e-05, W = 1e-03, and either g = 0 or g = 0.1.



1

	F	F2	S
IOI1	0.11	0.13	0.09
IOI2	0.11	0.13	0.07
IOI3	0.11	0.15	0.08

2

	F1	F2	S
IOI1	0.27	0.36	0.23
IOI2	0.18	0.22	0.13
IOI3	0.28	0.39	0.27

3

	F1	F2	S
IOI1	0.11	0.14	0.05
IOI2	0.1	0.12	0.05
IOI3	0.11	0.15	0.05

4

	F1	F2	S
IOI1	0.27	0.34	0.24
IOI2	0.23	0.22	0.24
IOI3	0.32	0.42	0.24

Figure 5. Analysis of duets 16 and 4 from the JustHockIt dataset, in plots and tables. Left plots show duet 16 and right plots show duet 4. Same for tables. The top row is without regularizatio. The bottom row is with regularization. The horizontal axis in each plot shows the time in samples (100 sa/sec), the vertical axis shows duration in log2dur scale. Each plot shows three horizontal lines representing the IOI-classes. The dotted horizontal line (not always very visible due to overlap) represents the IOI-tempo. Big dots are outliers but they appear only in the plots of duet 4. In tables, F1 is fluctuation1, F2 is fluctuation2, S is stability.

Figure 5 shows an analysis of duet 16 and duet 4 without and with regularization. When the music ensemble has its timing in agreement with the score, then the IOI-tempo (represented by dotted lines) will fully overlap with the IOI-class (as in the plots of duet 16). However, for expressive reasons, or for reasons that have to do with musical capabilities, it may happen that a music ensemble's co-regulated timing slightly deviates from the prescribed meter in the score, for example when short notes are performed shorter and long notes are performed longer. Understanding how this timing elasticity relates to musical expressivity is a topic of ongoing research (e.g., Coorevits et al., 2019). Here we see that duet 4 generates an IOI-meter of about {3, 1.5, 1} (in milliseconds ratio) right from the beginning of the performance, while the meter is in fact {3,2,1}. When regularization is applied, the IOI-classes are forced to stay within the constraints of the out-of-time IOI-meter.

Figure 6 shows how narration reflects the structure of the music (as ABAB... scheme) in a recurrence plot. Sequences of four labels holding IOI-class assignments (e.g., 1, 2, 1, 3) are compared in a four-dimensional phase space, and at each time point this is done for past as well as future time points. The comparison is either 1 or 0, depending on a threshold, and the 1s are represented as a dot. The recurrence ratio (RR) can be used as an additional mark of homeostatic stability in co-regulated timing.

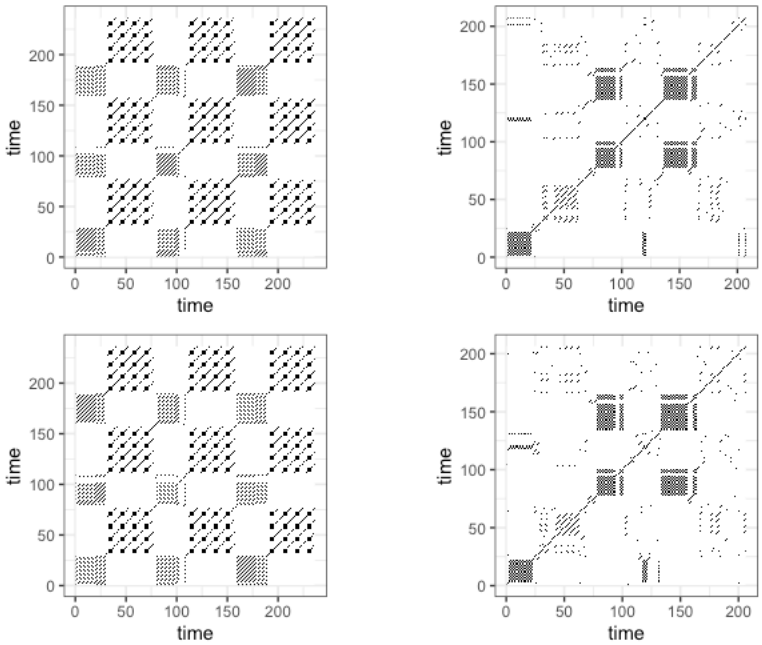


Figure 6. Recurrence plots of the plots in Figure 5. Time starts at the lower left corner and can be interpreted as going up to the upper right corner.

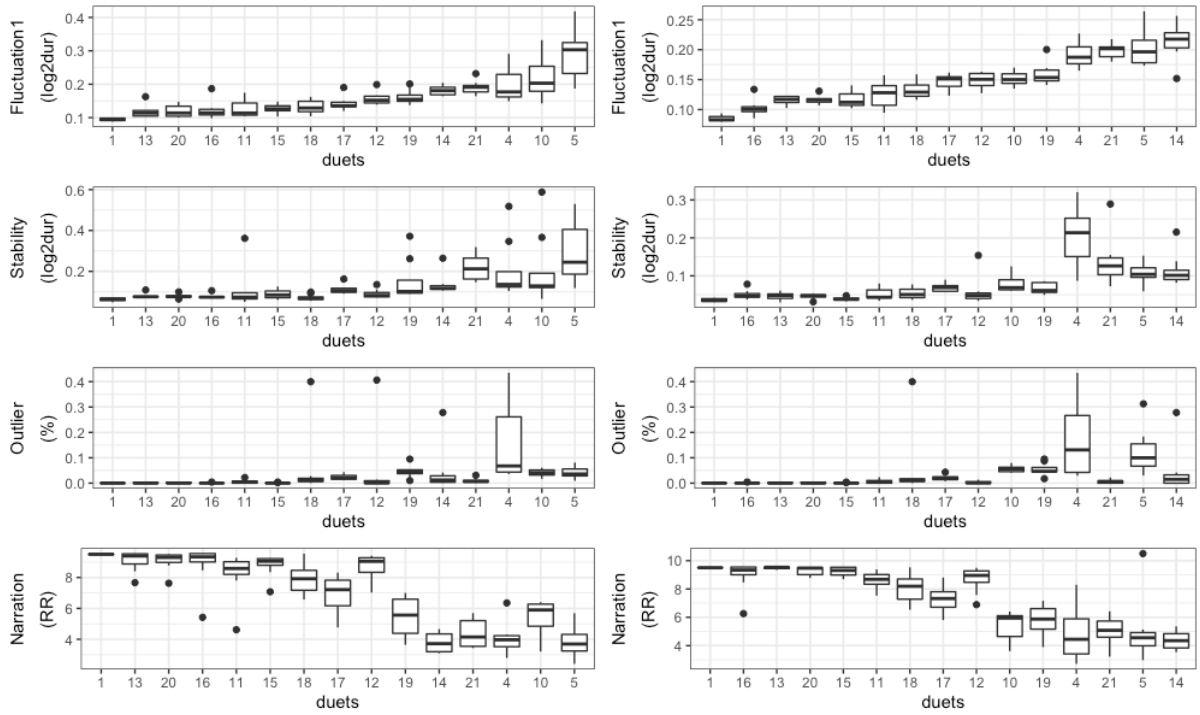


Figure 7. Comparison of global measures for the non-regularized (g=0) (left column) and regularized (g=.1) (right column) analysis of the JustHockIt dataset. The horizontal axis in each panel shows the duets ordered according to the fluctuation1 per duet, as in the top plots[1]. The bars show the mean and standard deviations for each measure. Conditions (movement, non-movement) are mixed as tests showed no distinction between moving and non-moving.

Figure 7 provides an overview of the measures of the JustHockIt dataset. The duet numbers are labels, 15 in total. Duet 1 is a midi-based performance used here as control. Overall, the graph suggests that the measures are correlated, as summarized in Figure 8. Narration is reflecting the fact that low fluctuation implies a structure in the assignation of IOI-observed objects to IOI-classes. The difference between no regularization (left) and regularization (right) is small, despite a different ordering of duets.

Dell'Anna et al. (2020) also provides a subjective assessment of performance quality and agency. Here we focus on the data of the performers. Figure 7 gives an overview of the correlations of the measures (using Kendall's tau), as well as the correlations of the measures with quality and agency. The results are similar to the results of Dell'Anna et al. (2020).

[1] Fluctuation1 calculates one value for each IOI-class. Rather than taking the mean over IOI-classes, we use only the fluctuation1 value of the second IOI-class.

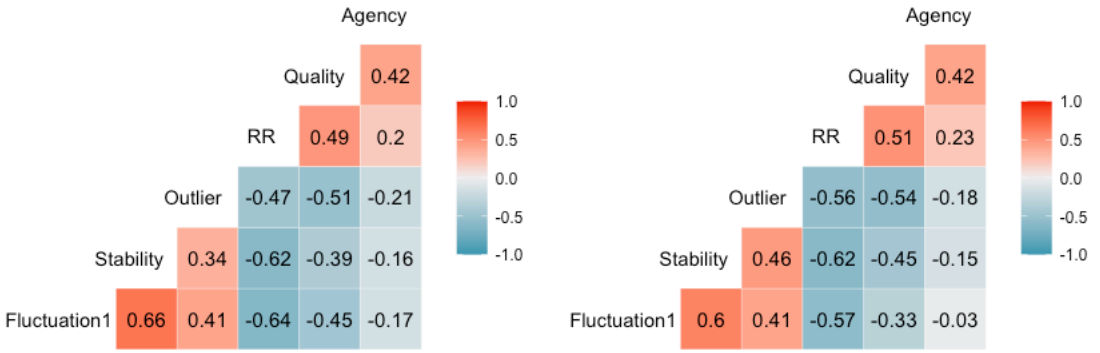


Figure 8. Correlation analysis of data shown in Figure 7 with subjective data. Without (top) and with (bottom) regularization.

The question whether regularization should be applied or not deserves caution, especially in relation to subjective self-assessments such as annotation and agency. Regularization checks for the correct meter, and it discounts deviations from the assumed correct meter. However, it is possible that performers have a good impression of their own performance, as reflected in the annotation task and in the responses to agency questionnaires, despite the fact that the prescribed meter was not followed.

Discussion and conclusion

From the viewpoint of the ensemble as a whole, the dynamics of co-regulated timing can be understood in terms of collaborative actions that bring about emergent patterns related to meter, rhythms, and tempo. From the viewpoint of a participating musician, the dynamics can be understood in terms of actions informed by hypotheses about the ensemble's overall timing. BListener is an attempt to model hypotheses formation using Bayesian inferencing. The outcome of BListener is a multivariate time series representing perceived timing constancy, given the music ensemble's co-regulated timing events (as inter-onset-intervals) as input. From these time series it is possible to extract global features about the latent processes.

BListener has the option to regularize hypotheses about timing constancy by binding their variance. Thus, rather than assuming that the IOI-classes evolve independently from each other, we make them dependent using a hyper-parameter, which is the out-of-time IOI-meter. When regularization is turned on, BListener "believes" (weakly or strongly, depending on the control) that observed IOIs should be processed using the meter as prior. This regularization can be useful in contexts where obedience to the meter is regarded as a feature of the global timing intention. For example, the co-regulated timing of duet 10 in the JustHockIt database shows constancy but no adherence to the meter. A basic idea is that regularization also prevents drift of IOI-classes.

Constancy in timing correlates with the subjective assessment of timing quality and even with the subjective assessment of agency (Dell'Anna et al., 2020). While statistical models can be built for predicting subjective experiences such as performance quality and agency, we restricted ourselves here to a simple correlation analysis which suggests that when timing becomes more predictive, the feeling of control becomes more pronounced. However, this finding needs further investigation since there are many cofounding variables that play a role, such as level of musicianship and rehearsal time. Overall, it can be assumed that a music ensemble's capacity to establish timing constancy contributes to the resulting affective power on listeners, such as feelings of control or embodied synchronization. BListener thereby offers measures for timing constancy in non-stationary timing conditions such as expressive timing with rubato. However, further work is also needed here.

BListener focuses on perception rather than on action. However, by turning the BListener into a real-time algorithm equipped with musical synthesis tools it would be rather straightforward to set up computer simulations of co-regulated timing actions. A proof of concept was developed by Laghetto (2019) in a setting of three percussionists who jointly played a musical piece on percussion instruments. The algorithm's task was to track the timing constancy and provide feedback so that the musicians could adapt their timing in view of improving their co-regulation. The real-time application was based on the onset detector of the Python madmon package (Böck et al., 2012) and it used shifting averages of IOIs along the IOI-classes, as described in Dell'Anna et al. (2020). Colored lightening was used as feedback to measured fluctuation (which is used as marker of timing constancy, and for interaction quality), using variance levels as thresholds for different colors. The proof of concept showed that co-regulated timing can be traced and used as an indicator of interaction quality, which can be used in the feedback that drives the social interaction towards homeostatic regulation.

The current approach and implementation also have some weak points. BListener is currently based on a discrete approach to timing, as it takes onsets as input. However, some music is based on amplitude modulations without energetic bursts, which make it hard to perform onset detection. A next point concerns the particular dependency of BListener on priors, which is a strength but also a weakness of the Bayesian approach. In particular, the IOI-classes can be set to be highly adaptive but their trajectory can go terribly wrong when certain parameters are not properly set, as illustrated in Figure 4. Especially in view of real-world applications, more work is needed to prepare for unexpected situations via proper priors. Another possible point of improvement is the rather simple assignment rule of IOI-observations to IOI-classes, which is currently based on a difference between durations. This rule is blind for musical structure, since IOI-classes have no clue about narrative expectations.

Despite these and other limitations, BListener may be useful in interaction research. Previous work in the domain of music-based biofeedback (for example in Moens et al., 2014; Van den Berghe et al., 2020; Lorenzoni et al., 2019; Moumddjan et al., 2019; Buhmann et al., 2018) suggested that beneficial outcomes of human-machine synchronization may be conditioned by co-regulated timing of human and machine. However, if the interaction is of low quality, then beneficial effects will probably be poor or even neglectable, suggesting a dependency of effect on interaction quality. Interaction can thus be reinforced, provided that a human-machine interaction is of good quality. At this point, BListener can provide measures of homeostatic co-regulation in social groups, which are useful in interactive multimedia systems that train synchronized social interaction skills in view of affective outcomes. In an

application with fitness-machines (Fritz et al., 2015), it was shown that participants can co-regulate an ongoing audio stream using physical effort and concentration. However, this co-regulation could establish a social interaction state that affected agency in participants. Similarly, research on individual-oriented music-based biofeedback systems shows that reinforcement learning can be used to steer users towards particular behaviors (e.g., Van den Berghe et al., 2020; Lorenzoni et al., 2019). The proposed model for homeostatic co-regulation of timing could be a component of a biofeedback system that uses reinforcement learning to steer users towards particular co-regulation behavior in view of attaining particular levels of performance quality.

Finally, it is important to realize that co-regulated timing in a music ensemble is more than Bayesian-inferencing. Co-regulated timing in a music ensemble is obviously intended, even before the music ensemble starts playing. Moreover, co-regulated timing is likely to involve states of affect and emotion such as feelings of agency and arousal, suggesting that Bayesian inferencing forms part of a more encompassing story. Earlier references to the concept of homeostasis pointed to this more encompassing dynamic of reward-based regulation and emotional states (Leman, 2016, pp. 188; Damasio, 2017). The metaphors reveal that humans are intrigued by extraordinary precarious states, especially when they are difficult to self-realize.

References

Aitchison, L. and Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. Current opinion in neurobiology, 46, 219-227.

Agres, K., Abdallah, S. and Pearce, M. (2018). Information-theoretic properties of auditory sequences dynamically influence expectation and memory. Cognitive science, 42(1), 43-76.

Bishop, C. (2006). Pattern recognition and machine learning. Berlin: Springer.

Bishop, L. (2018). Collaborative musical creativity: How ensembles coordinate spontaneity. Frontiers in psychology, 9, 1285.

Bishop, L., Cancino-Chacón, C. and Goebl, W. (2019). Moving to communicate, moving to interact: Patterns of body motion in musical duo performance. Music perception, 37(1), 1-25.

Bishop, L. and Goebl, W. (2020). Negotiating a shared interpretation during piano duo performance. Music and science, 3, 2059204319896152.

Böck, S., Arzt, A., Krebs, F. and Schedl, M. (2012, September). Online real-time onset detection with recurrent neural networks. In Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12), York, UK.

Buhmann, J., Moens, B., Van Dyck, E., Dotov, D. and Leman, M. (2018). Optimizing beat synchronized running to music. PLOS ONE, 13(12).

Burger, B., London, J., Thompson, M. R. and Toiviainen, P. (2018). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. Psychological research, 82(6), 1195-1211.

628 Chang, A., Kragness, H., Livingstone, S., Bosnyak, D. and Trainor, L. (2019). Body sway
629 reflects joint emotional expression in music ensemble performance. Scientific reports, 9(1),
630 1-11.

631 Changeux, J-P. (1999). Leçon inaugurale, 16 janvier 1976, p.57. In A. Berthoz (Ed.) Leçons
632 sur le corps, le cerveau et l'esprit. Paris: Editions Odile Jacob.

633 Clarke, E. (1999). Rhythm and timing in music. In D. Deutsch (Ed.) The psychology of
634 music (pp. 473-500). Academic Press.

635 Cochrane, T., Fantini, B. and Scherer, K. (Eds.). (2013). The emotional power of music:
636 Multidisciplinary perspectives on musical arousal, expression, and social control. Oxford:
637 OUP.

638 Coorevits, E., Moelants, D., Maes, P-J. and Leman, M. (2019). Exploring the effect of tempo
639 changes on violinists' body movements. Musicae scientiae, 23(1), 87-110.

640 Damasio, A. (2017). L'ordre étrange des choses: la vie, les sentiments et la fabrique de la
641 culture. Paris: Odile Jacob.

642 D'Ausilio, A., Novembre, G., Fadiga, L. and Keller, P. (2015). What can music tell us about
643 social interaction? Trends in cognitive science, 19, 111-114.

644 Davies, K. (2016). Adaptive homeostasis. Molecular aspects of medicine, 49, 1-7.

645 Dell'Anna, A., Buhmann, J., Six, J., Maes, P. J. and Leman, M. (2020). Timing markers of
646 interaction quality during semi-hocket singing. Frontiers in neuroscience, 14.

647 Demos, A., Chaffin, R. and Kant, V. (2014). Toward a dynamical theory of body movement
648 in musical performance. Frontiers in psychology, 5, 477.

649 Fabian, D., Timmers, R. and Schubert, E. (Eds.). (2014). Expressiveness in music
650 performance: Empirical approaches across styles and cultures. Oxford: OUP.

651 Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. and Pezzulo, G. (2017). Active
652 inference: a process theory. Neural computation, 29(1), 1-49.

653 Fritz, T., Hardikar, S., Demoucron, M., Niessen, M., Demey, M., Giot, O., Li, Y., Haynes, J.-
654 D., Villringer, A. and Leman, M. (2013). Musical agency reduces perceived exertion during
655 strenuous physical performance. PNAS, 110(44), 17784–17789.

656 Frühauf, J., Kopiez, R. and Platz, F. (2013). Music on the timing grid: The influence of
657 micro-timing on the perceived groove quality of a simple drum pattern performance. Musicae
658 scientiae, 17(2), 246-260.

659 Glowinski, D., Bracco, F., Chiorri, C. and Grandjean, D. (2016). Music ensemble as a
660 resilient system. Managing the unexpected through group interaction. Frontiers in
661 psychology, 7, 1548.

662 Glowinski, D., Bracco, F., Chiorri, C. and Grandjean, D. (2017). The resilience approach to
663 studying group interaction in music ensemble. In Lesaffre et al. (2017). The Routledge
664 companion to embodied music interaction (pp. 96-104). New York: Routledge.

Grimaud, A. M. and Eerola, T. (2020). EmoteControl: an interactive system for real-time control of emotional expression in music. Personal and ubiquitous computing. doi.org/10.1007/s00779-020-01390-7

Haken, H. (1990). Synergetics as a tool for the conceptulization and mathematization of cognition and behaviour. How far can we go? In H. Haken and M. Stadler (Eds.) Synergetics of cognition, (pp. 2-31) Berlin, Heidelberg: Springer.

Hilt, P.M., Badino, L., D'Ausilio, A., Volpe, G., Tokay, S., Fadiga, L. and Camurri, A. (2019). Multi-layer adaptation of group coordination in musical ensembles. Scientific reports, 9(1), 1-10.

Kanai, R., Komura, Y., Shipp, S. and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. Philosophical transactions of the royal society B: biological sciences, 370(1668), 20140169.

Keller, P. and Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. Music perception, 28(1), 27-46.

Keller P. (2014). Ensemble performance: interpersonal alignment of musical expression. In D Fabian et al. (Eds). Expressiveness in music performance: Empirical approaches across styles and cultures, pp. 260 282. Oxford: OUP.

Kelso, J. S. (1995). Dynamic patterns: The self-organization of brain and behavior. Cambridge, MA: The MIT press.

Kotz, S. A., Ravignani, A. and Fitch, W. T. (2018). The evolution of rhythm processing. Trends in cognitive sciences, 22(10), 896-910.

Laghetto, P. (2019). Time-evaluation model for live interaction with multiple performers. MA-thesis at Padova University and Ghent University.

Leman, M. (2007). Embodied music cognition and mediation technology. Cambridge, MA: The MIT press.

Leman, M. (2016). The expressive moment: How interaction (with music) shapes human empowerment. Cambridge, MA: The MIT press.

Levitin, D. J., Grahn, J. A. and London, J. (2018). The psychology of music: Rhythm and movement. Annual review of psychology, 69, 51-75.

London, J. (2012). Hearing in time: Psychological aspects of musical meter. Oxford: OUP.

Lorenzoni, V., Staley, J., Marchant, T., Onderdijk, K., Maes, P.-J. and Leman, M. (2019). The sonic instructor: A music-based biofeedback system for improving weightlifting technique. PLOS ONE, 14(8).

Moens, B., Muller, C., van Noorden, L., Franěk, M., Celie, B., Boone, J., Bourgois, J. and Leman, M. (2014). Encouraging spontaneous synchronisation with D-Jogger, an adaptive music player that aligns movement and music. PLOS ONE, 9(12).

Nakayama, S., R. Soman, V. and Porfiri, M. (2020). Musical collaboration in rhythmic improvisation. Entropy 2020, 22, 233.

Meinhold, R. and Singpurwalla, N. (1983). Understanding the Kalman filter. The American statistician, vol.37, no.2, 123-127.

Moumddjian, L., Moens, B., Vanzeir, E., De Klerck, B., Feys, P. and Leman, M. (2019). A model of different cognitive processes during spontaneous and intentional coupling to music in multiple sclerosis. Annals of the New York academy of sciences, 1445(1), 27–38.

Petris, G., Petrone, S. and Campagnoli, P. (2009). Dynamic linear models with R. New York: Springer.

Pezzulo, G., Rigoli, F. and Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. Progress in neurobiology, 134, 17-35.

Polak, R., London, J. and Jacoby, N. (2016). Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of Malian jembe drumming. Frontiers in neuroscience, 10, 285

Port, R. and Van Gelder, T. (1995). Mind as motion: Explorations in the dynamics of cognition. Cambridge, MA: The MIT press.

Ravignani, A. and Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. Frontiers in psychology, 8, 1820.

Repp, B. and Su, Y. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). Psychonomic bulletin & review, 20(3), 403-452.

Schiavio, A., van der Schyff, D., Cespedes-Guevara, J. and Reybrouck, M. (2017). Enacting musical emotions. Sense-making, dynamic systems, and the embodied mind. Phenomenology and the cognitive sciences, 16(5), 785-809.

Scheurich, R., Pfordresher, P. and Palmer, C. (2020). Musical training enhances temporal adaptation of auditory-motor synchronization. Experimental brain research 238, 81–92 (2020).

Shumway, R. and Stoffer, D. (2017). Time series analysis and its applications: with R examples. New York: Springer.

Tolston, M. T., Funke, G. J., and Shockley, K. (2020). Comparison of cross-correlation and joint-recurrence quantification analysis based methods for estimating coupling strength in non-linear Systems. Dynamics, 29(31), 32.

Van den Berghe, P., Gosseries, M., Gerlo, J., Lenoir, M., Leman, M. and De Clercq, D. (2020). Change-point detection of peak tibial acceleration in overground running retraining. Sensors, 20, 17.

Van Nort, D. and Depalle, P. (2017). Adaptive musical control of time-frequency representations. In R. Bader (Ed.). Springer handbook of systematic musicology. Berlin: Springer, pp. 313-328.

739 Volpe, G., D'Ausilio, A., Badino, L., Camurri, A. and Fadiga, L. (2016). Measuring social
740 interaction in music ensembles. Philosophical transactions of the royal society B: biological
741 sciences, 371(1693), 20150377.

742 Vuust, P. and Witek, M. (2014). Rhythmic complexity and predictive coding: a novel
743 approach to modeling rhythm and meter perception in music. Frontiers in psychology, 5,
744 1111.

745 Witek, M., Clarke, E., Wallentin, M., Kringelbach, M. and Vuust, P. (2014) Syncopation,
746 body-movement and pleasure in groove music. PLOS ONE 9, e94446.