A KALMAN FILTER MODEL FOR SYNCHRONIZATION IN MUSICAL ENSEMBLES

Hugo T. Carvalho[1*] **Min S. Li**[2] **Massimiliano di Luca**[3] **Alan M. Wing**[3]
[1] Department of Statistical Methods, Federal University of Rio de Janeiro, Brazil
[2] Bristol Interaction Group, School of Computer Science, University of Bristol, United Kingdom
[3] Virtual Reality Lab, School of Psychology, University of Birmingham, United Kingdom
[*] Corresponding author: hugo@dme.ufrj.br

ABSTRACT

The synchronization of motor responses to rhythmic auditory cues is a fundamental biological phenomenon observed across various species. While the importance of temporal alignment varies across different contexts, achieving precise temporal synchronization is a prominent goal in musical performances. Musicians often incorporate expressive timing variations, which require precise control over timing and synchronization, particularly in ensemble performance. This is crucial because both deliberate expressive nuances and accidental timing deviations can affect the overall timing of a performance. This discussion prompts the question of how musicians adjust their temporal dynamics to achieve synchronization within an ensemble. This paper introduces a novel feedback correction model based on the Kalman Filter, aimed at improving the understanding of interpersonal timing in ensemble music performances. The proposed model performs similarly to other linear correction models in the literature, with the advantage of low computational cost and good performance even in scenarios where the underlying tempo varies.

1. INTRODUCTION

Synchronization of motor responses to rhythmic auditory cues represents a biological phenomenon found across various species [1], and social collectives often engage in activities necessitating precise temporal coordination among members, a crucial factor for successful group endeavors. For example, in scenarios such as rowing eights, temporal alignment may not be the primary focus, but individual timing remains tied to collective timing dynamics [2]. In domains like musical performances, achieving precise temporal synchronization serves as a prominent goal [3].

Typically, musicians do not adhere strictly to the exact timing of note onsets as indicated in the musical score: due to expressiveness, they often introduce deviations from the prescribed timing [4]. These fluctuations require a high level of control over relative timing, where the phase of notes produced by the musician deviating from the timing aligns differently with the phases of other musicians. Rehearsals often involve reaching a consensus on expressive variations, ensuring that timing deviations are synchronized among players while maintaining relative timing [5]. Nevertheless, even with a unified understanding of the musical interpretation, individual musicians may opt to vary the timing of note onsets in specific passages between different performances [5, 6]. Musical performance timing is also susceptible to inadvertent variations due to factors such as rhythmic intricacies, technical demands beyond timing (e.g., pitch, volume), lapses in concentration, and the inherent variability of biological timing [7]. While extensive individual practice can mitigate some of these unintended variations, complete elimination is unlikely.

The previous discussion raises the inquiry: how do musicians within an ensemble modulate their temporal dynamics to achieve synchronization with one another? In this paper a novel feedback correction model is presented, based on the Kalman Filter and aimed at improving timing accuracy in ensemble music performances. The proposed model generalizes the linear autoregressive model in [8] with the improvement of allowing two important quantities, the *phase* and *period correction gains*, to vary along time, since it makes the model suitable to describe synchronization in scenarios where the underlying tempo greatly varies (a realistic case in ensemble performance).

The paper is organized as follows: Section 2 recalls some linear models for synchronization, and the dynamic generalization of the model in [8] is presented, which is formulated as a Kalman Filter in Section 4; the fundamentals of the Kalman Filter are briefly recalled in Section 3; the computational experiments are shown and discussed in Section 5; conclusions are presented in Section 6. Directions for future work are identified throughout the paper.

2. LINEAR MODELS FOR ENSEMBLE SYNCHRONIZATION

The starting point for contextualizing the proposed model is [9], where a phase-correction model is presented as a method for an individual performer to achieve synchrony with a periodic metronome click or with another performer

(see also [10]). The fundamental concept revolves around utilizing the asynchrony, termed as a *phase error*, between a tone onset and the metronome click (or between two tone onsets produced by different performers) in a feedback mechanism, that guides the performer to adjust the time interval preceding the next tone onset. Consequently, the performer either decreases or increases the interval leading up to the subsequent tone onset proportionally to the preceding asynchrony. This process aims to achieve greater synchrony ("in phase") between the next tone onset and the metronome click (or pair of tone onsets). This synchronization scheme can be represented by Equation (1):

$$t_n = t_{n-1} + T_n - \alpha A_{n-1} + \varepsilon_n, \tag{1}$$

where t_n and t_{n-1} represent the current and previous observed tone onset event times respectively, T_n denotes the time interval generated by an internal timekeeping mechanism, α is the *phase correction strength* or *phase correction gain*, A_{n-1} refers to the asynchrony of the previous onset event, and ε_n represents a random error term, which includes the internal timekeeper noise. The complete reduction of asynchrony to zero hinges on the value assigned to the gain, α, since this parameter determines the proportion of the preceding phase error that the performer endeavors to eliminate.

Following [11], *period correction* can also be incorporated in the model in Equation (1), by imposing that

$$T_n = T_{n-1} - \beta A_{n-1}, \tag{2}$$

where β is the *period correction strength* or *period correction gain*. Phase correction involves a local, within-cycle adjustment to the timing, while period correction entails a more enduring alteration to the underlying tempo, influencing subsequent cycles as well. Phase correction typically occurs automatically, without the need for conscious awareness of synchronization discrepancies. However, period correction appears to be more cognitively demanding, relying on the conscious detection of tempo variations in the external rhythm [12, 13].

As previously mentioned, Equations (1) and (2) model the asynchrony correction of an individual tapping according to a periodic metronome, or between two individuals tapping together. In [8] it is argued that the same modeling framework is also suited to describe synchronization in music ensemble performance, where a specific musician now tries to reduce asynchrony between him/her and every other performer. Therefore, Equations (1) and (2) can be jointly generalized to an ensemble of K performers as:

$$t_{i,n} = t_{i,n-1} + T_{i,n} - \sum_{\substack{i=1 \\ j\neq i}}^{K} \alpha_{ij} A_{ij,n-1} + \varepsilon_{i,n} \tag{3}$$

$$T_{i,n} = T_{i,n-1} - \sum_{\substack{i=1 \\ j\neq i}}^{K} \beta_{ij} A_{ij,n-1}, \tag{4}$$

where $i = 1, \ldots, K$ indicate a specific performer, $t_{i,n}$ and $t_{i,n-1}$ are respectively the current and previous observed tone onset event times for player i, $T_{i,n}$ is the timekeeper interval for player i at time instant n, $A_{ij,n-1} = (t_{i,n-1} - t_{j,n-1})$ is the asynchrony at the time instant $n-1$ between players i and j, α_{ij} and β_{ij} are respectively the phase and period correction gain applied by player i to compensate for $A_{ij,n-1}$, and $\varepsilon_{i,n}$ is a noise term identified with the internal timekeeper. Estimation of the values of α_{ij} and β_{ij} can be performed using the *bounded Generalized Least Squares* method (bGLS) [14, 15].

In [8], the model in Equation (3) is implemented and largely investigated for the case of a string quartet ensemble playing a homophonic section from the string quartet Op. 74 no. 1 by Joseph Haydn (fourth movement, bars 13–24), as this part has a steady tempo and all player's quarter notes are aligned. In [14] the coupling of Equations (3) and (4) is investigated, with a simulated string quartet data with mild tempo changes, and the bGLS algorithm is shown to be capable of recovering the values of α and β. However, due to the nature of the bGLS algorithm, the authors point out that many data points are necessary for robust estimation of these variables, which may not be available or is an unrealistic aim in the case of a real-time implementation of the correction model (eg. for a virtual reality musical ensemble). In [16] the ADAM model (ADaptation and Anticipation Model) is proposed, including not only correction terms but also anticipatory ones, and in [17] this model is tested with tempo-changing tapping data, but since there is no adaptation of the bGLS algorithm to this new set of equations, the parameter estimation is done by exhaustive search, which is infeasible for real-life applications. Moreover, due to the nature of its parameters, the ADAM model is non-identifiable, meaning that more than one configuration of the parameters leads to the same estimate.

In order to circumvent the aforementioned issues, an alternative is to consider not a single value of α and β for each pair of performers through time, but *time-dependent correction gains*. Developing this intuition, a dynamic α_{ij} allows that a performer changes the phase correction at each onset, and a dynamic β_{ij} would allow him/her to correct differently for tempo variations during the performance of an excerpt. To model a dynamic variable, a good balance between simplicity and accuracy is a random walk, and in this case phase and period correction occur according to Equations (3) and (4), respectively, but with additional equations to allow the evolution of both correction gains. This new model is summarized in Equations (5), (6), (7), and (8):

$$t_{i,n} = t_{i,n-1} + T_{i,n} - \sum_{\substack{i=1 \\ j\neq i}}^{K} \alpha_{ij,n} A_{ij,n-1} + \varepsilon_{i,n} \tag{5}$$

$$T_{i,n} = T_{i,n-1} - \sum_{\substack{i=1 \\ j\neq i}}^{K} \beta_{ij,n} A_{ij,n-1} \tag{6}$$

$$\alpha_{ij,n} = \alpha_{ij,n-1} + w_{ij,n}^{(\alpha)} \tag{7}$$

$$\beta_{ij,n} = \beta_{ij,n-1} + w_{ij,n}^{(\beta)}, \tag{8}$$

where $w_{ij,n}^{(\alpha)}$ and $w_{ij,n}^{(\beta)}$ are independent zero-mean Gaussian random variables, allowing the evolution of $\alpha_{ij,n}$ and $\beta_{ij,n}$ through time, respectively (notice the novel subscript "n" in both α_{ij} and β_{ij}).

However, in the model proposed in Equations (5), (6), (7), and (8), it is not clear how to employ the bGLS method to obtain estimate of the variables of interest, and two distinct paths can now be followed: generalize the bGLS algorithm to this new situation, or resort to estimation techniques within the theory of dynamic models [18]. This work follows the latter, adopting the Kalman Filter as a framework to analyze Equations (5), (6), (7), and (8), due to its balance between flexibility and simplicity, as well as its simple and highly interpretable update equations. Section 3 recalls the basics of the Kalman Filter and Section 4 formulates the proposed model in this scenario.

3. A BRIEF RECALL ON THE KALMAN FILTER

The Kalman Filter (KF) was developed in the 1960's, and served originally as a way to produce accurate estimates of variables of interest (eg. position of an object) by reaching a consensus between physical models and noisy measurements [19]. More generally, the KF can be seen as a state-space dynamic model, employed to describe more general time-series as a dynamic linear regression model as function of an underlying Markov model [18].

The main contribution of this paper is to propose the model in Equations (5), (6), (7), and (8), and formulate it as a KF, employing its filtering and smoothing equations to estimate the phase and period correction gains through time. The choice of a KF to achieve this goal are: linear nature of the model in Equations (5), (6), (7), and (8), high interpretability of the KF and its update equations, and potential low computational cost of its implementation.

The notation and basic equations of the KF are now briefly recalled, following [18]. In what follows, the index n ranges from 1 to N. Let $\mathbf{y}_n \in \mathbb{R}^m$ be a sequence of *observed variables* (or *measurements*), and $\boldsymbol{\theta}_n \in \mathbb{R}^p$ be a sequence of unobserved vectors, which are called the *hidden* (or *state*) variables. The KF model assumes that these two entities are related by Equations (9) and (10):

$$\mathbf{y}_n = \mathbf{F}_n \boldsymbol{\theta}_n + \mathbf{v}_n \tag{9}$$

$$\boldsymbol{\theta}_n = \mathbf{G}_n \boldsymbol{\theta}_{n-1} + \mathbf{w}_n, \tag{10}$$

where $\mathbf{F}_n \in \mathbb{R}^{m\times p}$ and $\mathbf{G}_n \in \mathbb{R}^{p\times p}$ are sequences of known matrices (*observation model* and the *state-transition model*, respectively). Vectors $\mathbf{v}_n \in \mathbb{R}^m$ and $\mathbf{w}_n \in \mathbb{R}^p$ are independent *observation* and *process* noise terms, respectively, and it is assumed that they follow Gaussian probability distributions, that is, $\mathbf{v}_n \sim N(\mathbf{0}, \mathbf{V}_n)$ and $\mathbf{w}_n \sim N(\mathbf{0}, \mathbf{W}_n)$,[1] where $\mathbf{V}_n \in \mathbb{R}^{m\times m}$ and $\mathbf{W}_n \in \mathbb{R}^{p\times p}$ are sequences of known covariance matrices of the observation and process noise terms respectively.

[1] The symbol $\sim$ means "follows the probability distribution", and $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The dimension of the support of the random vector is omitted, and compatibility between dimensions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is always assumed.

The KF dynamically estimates variables $\boldsymbol{\theta}_n$ and $\mathbf{y}_n$ based on observations up to time $n-1$, and updates the estimate of $\boldsymbol{\theta}_n$ when the observation at time n is available. This process is done accordingly to Equations (11), (12) and (13), called the *filtering equations*:[2]

Prediction step for hidden variables:

$$\boldsymbol{\theta}_n | \mathbf{y}_{1:n-1} \sim N(\mathbf{a}_n, \mathbf{R}_n) \tag{11}$$

Prediction step for observed variables:

$$\mathbf{y}_n | \mathbf{y}_{1:n-1} \sim N(\mathbf{f}_n, \mathbf{Q}_n) \tag{12}$$

Update step (compare predictions to measurements):

$$\boldsymbol{\theta}_n | \mathbf{y}_{1:n} \sim N(\mathbf{k}_n, \mathbf{C}_n), \tag{13}$$

where[3]

$$\mathbf{a}_n = \mathbf{G}_n \mathbf{k}_{n-1} \tag{14}$$

$$\mathbf{R}_n = \mathbf{G}_n \mathbf{C}_{n-1} \mathbf{G}_n^T + \mathbf{W}_n \tag{15}$$

$$\mathbf{f}_n = \mathbf{F}_n \mathbf{a}_n \tag{16}$$

$$\mathbf{Q}_n = \mathbf{F}_n \mathbf{R}_n \mathbf{F}_n^T + \mathbf{V}_n \tag{17}$$

$$\mathbf{k}_n = \mathbf{a}_n + \left[\mathbf{R}_n \mathbf{F}_n^T \mathbf{Q}_n^{-1}\right] \mathbf{e}_n \tag{18}$$

$$\mathbf{e}_n = \mathbf{y}_n - \mathbf{f}_n \tag{19}$$

$$\mathbf{C}_n = \mathbf{R}_n - \left[\mathbf{R}_n \mathbf{F}_n^T \mathbf{Q}_n^{-1}\right] \mathbf{F}_n \mathbf{R}_n, \tag{20}$$

assuming that the initial state is chosen according to a normal distribution, that is, $\boldsymbol{\theta}_0 \sim N(\mathbf{k}_0, \mathbf{C}_0)$. For more details on the KF, see [18, 19].

One of the appealing aspects of the KF is its ability to perform estimation and forecasting sequentially, as new data emerge. However, if observations $\mathbf{y}_n$ for $n = 1, \ldots, N$ are available beforehand, one is also able to retrospectively reconstruct the system's states, in order to analyze its behavior given all the observations. For this purpose, a backward-recursive algorithm can be employed to compute the conditional distributions of $\boldsymbol{\theta}_n$ given $\mathbf{y}_{1:N}$, for any $n < N$ [18, 19]. The main ingredient of this algorithm is the *smoothing equation* (21):

$$\boldsymbol{\theta}_n | \mathbf{y}_{1:N} \sim N(\mathbf{s}_n, \mathbf{S}_n), \tag{21}$$

where

$$\mathbf{s}_n = \mathbf{k}_n + \mathbf{C}_n \mathbf{G}_{n+1}^T \mathbf{R}_{n+1}^{-1} \left[\mathbf{s}_{n+1} - \mathbf{a}_{n+1}\right] \tag{22}$$

$$\mathbf{S}_n = \mathbf{C}_n - \mathbf{C}_n \mathbf{G}_{n+1}^T \mathbf{R}_{n+1}^{-1} \times \left[\mathbf{R}_{n+1} - \mathbf{S}_{n+1}\right] \mathbf{R}_{n+1}^{-1} \mathbf{G}_{n+1} \mathbf{C}_n, \tag{23}$$

assuming that $\boldsymbol{\theta}_{n+1} | \mathbf{y}_{1:N} \sim N(\mathbf{s}_{n+1}, \mathbf{S}_{n+1})$. Notice that since the smoothing is performed backwards, it is necessarily to previously filter the set of observations to gain access to vectors $\mathbf{k}_n$ and $\mathbf{a}_n$, and matrices $\mathbf{C}_n$ and $\mathbf{R}_n$.

4. KALMAN FILTER MODEL FOR ENSEMBLE SYNCHRONIZATION

Equations (5), (6), (7), and (8) can be written as a KF by considering proper choices for the observed and hidden

[2] The conditional distribution of $\mathbf{u}$ given $\mathbf{z}$ is denoted by $\mathbf{u}|\mathbf{z}$, and $i : j$ means "observations from time instants i to j", both extremes included.

[3] The superscript ^{T} after a vector or matrix denotes its transpose; the superscript $^{-1}$ after a matrix denotes its inverse.

variables, as well as the observation and state-transition matrices. The main goal of this section is to construct a sequence of matrices $\mathbf{F}_n$ and $\mathbf{G}_n$, as well as vectors $\mathbf{y}_n$ and $\boldsymbol{\theta}_n$ of observed and hidden variables respectively, such that Equations (9) and (10) recover the model proposed in Equations (5), (6), (7), and (8). Firstly, to simplify the formulation of the model, the observed variables are not the tone onset times for each player, but rather the *inter-onset-intervals* (IOIs), denoted by $r_{i,k} = t_{i,n} - t_{i,n-1}$, for $i = 1, \ldots, K$. These values are assembled as in Equation (24):

$$\mathbf{y}_n = [r_{1,n} \ \ldots \ r_{K,n}] \in \mathbb{R}^K. \tag{24}$$

The hidden variable $\boldsymbol{\theta}_n$ can be written as in Equation (25):

$$\boldsymbol{\theta}_n = \left[\mathbf{T}_n^T \mid \mathbf{r}_n^T \mid \boldsymbol{\alpha}_n^T \mid \boldsymbol{\beta}_n^T\right]^T \in \mathbb{R}^{2K^2}, \tag{25}$$

where

$$\mathbf{T}_n = [T_{1,n} \ \ldots \ t_{K,n}]^T \in \mathbb{R}^K \tag{26}$$

$$\mathbf{r}_n = [r_{1,n} \ \ldots \ r_{K,n}]^T \in \mathbb{R}^K \tag{27}$$

$$\boldsymbol{\alpha}_n = [\alpha_{ij,n} \text{ in the lexicographical order on } ij, \text{ for } 1 \leq i, j \leq K, i \neq j] \in \mathbb{R}^{K(K-1)} \tag{28}$$

$$\boldsymbol{\beta}_n = [\beta_{ij,n} \text{ in the lexicographical order on } ij, \text{ for } 1 \leq i, j \leq K, i \neq j] \in \mathbb{R}^{K(K-1)}. \tag{29}$$

The relation between $\boldsymbol{\theta}_n$ and $\mathbf{y}_n$ is described by the observation matrix in Equation (30): [4]

$$\mathbf{F}_n = \left[\mathbf{0}_K \mid \mathbf{I}_K \mid \mathbf{0}_{K\times K(K-1)} \mid \mathbf{0}_{K\times K(K-1)}\right]. \tag{30}$$

Notice that matrices $\mathbf{F}_n \in \mathbb{R}^{K\times 2K^2}$ are constant through time. The evolution of the hidden variables in $\boldsymbol{\theta}_n$ is modelled by a sequence of state-transition matrices $\mathbf{G}_n \in \mathbb{R}^{2K^2\times 2K^2}$, described in Equation (31):

$$\begin{bmatrix} \mathbf{I}_K & \mathbf{0}_K & \mathbf{0}_{K\times K(K-1)} & \mathbf{G}_n^{T\beta} \\ \mathbf{I}_K & \mathbf{0}_K & \mathbf{G}_n^{r\alpha} & \mathbf{G}_n^{r\beta} \\ \mathbf{0}_{K(K-1)\times K} & \mathbf{0}_{K(K-1)\times K} & \mathbf{I}_{K(K-1)} & \mathbf{0}_{K(K-1)} \\ \mathbf{0}_{K(K-1)\times K} & \mathbf{0}_{K(K-1)\times K} & \mathbf{0}_{K(K-1)} & \mathbf{I}_{K(K-1)} \end{bmatrix}, \tag{31}$$

where matrices $\mathbf{G}_n^{T\beta}$, $\mathbf{G}_n^{r\alpha}$, and $\mathbf{G}_n^{r\beta}$ (of dimensions $K \times K(K-1)$ each) describe the interaction between variables in their respective superscripts. These three matrices are equal to the matrix in Equation (32):

$$\begin{bmatrix} -\mathbf{A}_{1:,n-1}^T & \mathbf{0}_{1\times(K-1)} & \cdots & \mathbf{0}_{1\times(K-1)} \\ \mathbf{0}_{1\times(K-1)} & -\mathbf{A}_{2:,n-1}^T & \cdots & \mathbf{0}_{1\times(K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times(K-1)} & \mathbf{0}_{1\times(K-1)} & \cdots & -\mathbf{A}_{K:,n-1}^T \end{bmatrix}, \tag{32}$$

where each $\mathbf{A}_{i:,n-1} \in \mathbb{R}^{K-1}$ contain the asynchronies $A_{ij,n-1}$ of player i to all players j, for $j \neq i$, at time $n-1$. Vector $\mathbf{A}_{i:,n-1}$ is made explicit in Equation (33):

$$\left[A_{i1,n-1} \ldots A_{i(i-1),n-1} A_{i(i+1),n-1} \ldots A_{iK,n-1}\right]^T. \tag{33}$$

A simple (but tedious) verification using Equations (9) and (10) with these choices for $\mathbf{F}_n$, $\mathbf{G}_n$, $\mathbf{y}_n$, and $\boldsymbol{\theta}_n$ ensures that the model in Equations (5), (6), (7), and (8) is recovered. It is also established that when $K = 1$ the model in Equations (1) and (2) is recovered, with the improvement of dynamic α and β.

When compared to the bGLS algorithm [14, 15], the state-of-the-art to estimate parameters in the scenario of sensorimotor synchronization, the KF model presents a great advantage, that is the possibility of performing online estimation as more data become available: this feature can be important if one desires to implement real-time synchronization schemes. When the complete time-series of onset times/IOIs is available, one can apply the smoothing equation (21), in order to dynamically estimate the parameters of interest throughout the performance, as well as estimate them by applying the filtering equations (11), (12), and (13), for example, to simulate an online scenario.

Notice that the dimensions of $\mathbf{F}_n$, $\mathbf{G}_n$, and $\boldsymbol{\theta}_n$ scale quadratically with the number of performers, which may render the model overly complicated or cause computational issues when computing the KF update/filtering equations. [5] However, due to the sparsness of matrices $\mathbf{F}_n$ and $\mathbf{G}_n$, block-multiplication will highly reduce the number of operations when computing Equations (14) to (20), mitigating the latter issue. Regarding the complexity of the model, notice that in real large-scale scenarios (eg. a symphony orchestra) it is not realistic to assume that each musician synchronizes with every other, thus allowing for potential simplifications, like considering a group of instruments as a single unity and synchronizing with every other group. This procedure would diminish the value of K from approximately 100 to less than 20. A useful topic for future research would be to investigate the possibility of modeling the synchronization scheme between performers (or group of performers) in a graph, in order to decrease even more the number of relevant connections.

Another issue that is important to point out is the design of the covariance matrices for the observation and process noises, $\mathbf{V}_n$ and $\mathbf{W}_n$ respectively. On a first view, it makes sense to consider $\mathbf{V}_n$ as diagonal matrices, for simplicity, since the interaction between the performers is already "captured" by the correction gains in the hidden variables; however, it is not clear if $\mathbf{W}_n$ should be a sequence of diagonal matrices, since it makes sense to consider at least correlations between both correction gains of the same performer. This work employs a particular choice for these covariance matrices, as will be further discussed in Section 5. Further investigation on this question could involve coupling the Expectation-Maximization algorithm with the KF in order to estimate not only matrices $\mathbf{V}_n$ and $\mathbf{W}_n$ but also $\mathbf{F}_n$ and $\mathbf{G}_n$ [20]. However a disadvantage would be that these estimates would need to be static through time, requiring a large amount of data, and being highly dependent of the piece of music being analyzed.

[4] The identity matrix of dimensions $L \times L$ is denoted by $\mathbf{I}_L$; the matrix of dimensions $L \times M$ filled with zeros is denoted by $\mathbf{0}_{L\times M}$; a square null matrix of dimensions $L \times L$ is abbreviated by $\mathbf{0}_L$.

[5] Other computational issues on the Kalman Filter are largely discussed in [18].

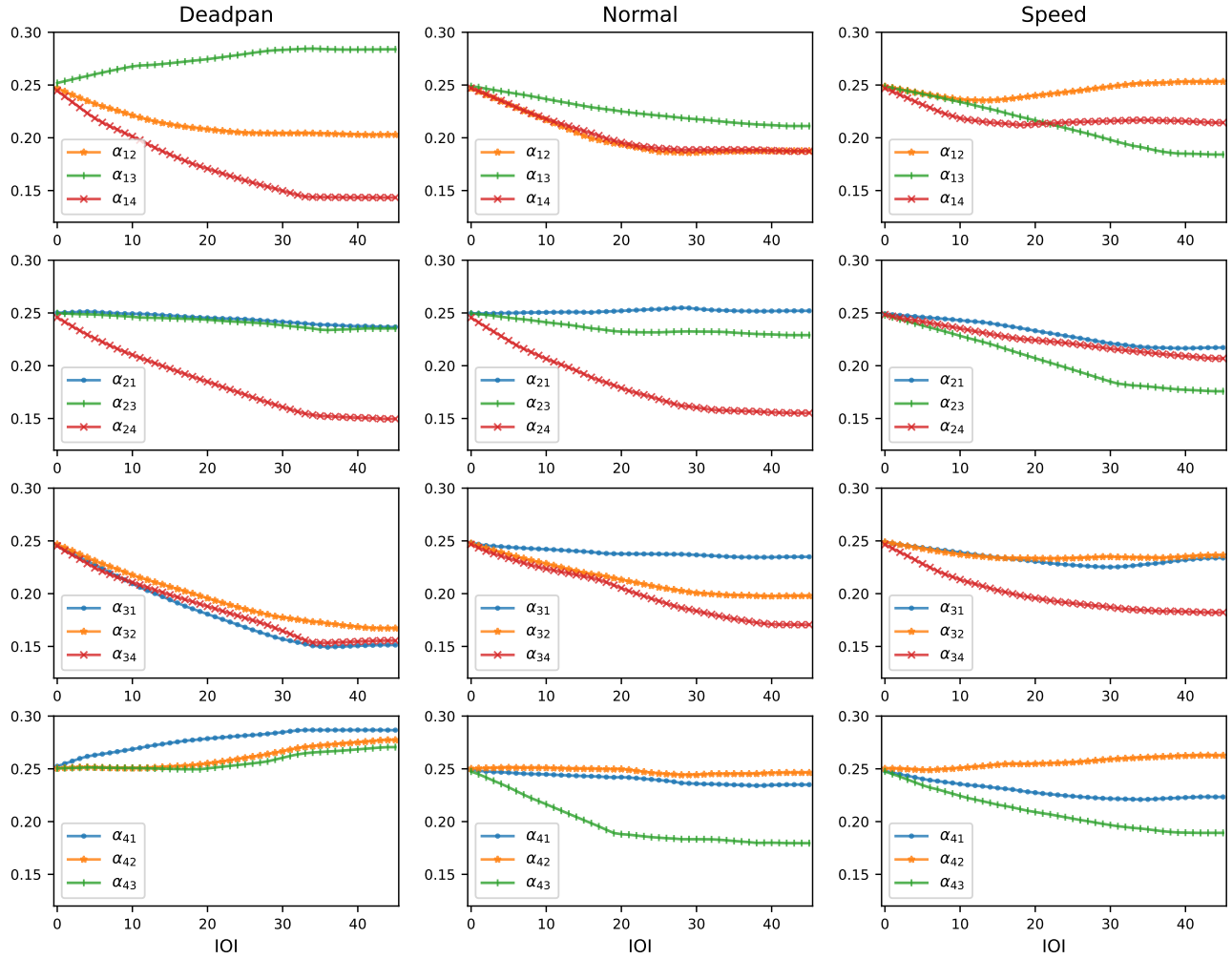

Figure 1. Smoothed time-series for the phase correction gain on three performance styles of an excerpt of the fourth movement of the string quartet Op. 74 no. 1, by Joseph Haydn. See Section 5 for discussion.

5. RESULTS

To illustrate the effectiveness of the proposed model, a set of simulations was performed, using an excerpt of the dataset presented in [21], similar to the one used in [8]: the homophonic section from the fourth movement of the string quartet Op. 74 no. 1 by Joseph Haydn, from bars 13 to 24. In this excerpt the instruments play a sequence of 47 quarter notes in rhytmic unison, with the first violin breaking the pattern near the end with an adornment of four sixteenth notes, which are disregarded in this study.

Three performance styles are considered: *Normal* condition (concert-style performance); *Speed* condition (including a spontaneous *accelerando* and *ritardando* initiated by a single musician – the designated leader, that can be the first or second violin); and *Deadpan* condition (performances with minimal expression in tempo and articulation). All the simulation were performed on a computer equipped with a 12th Generation Intel Core™ i7 processor and 16GB of RAM, running Windows 11 Pro™; the implementations were conducted in Python version 3.11.7. [6]

Regarding the parameters of the KF, the covariance matrix for the process noise, $\mathbf{W}_n$, plays an important role, since it indicates how the variables in $\boldsymbol{\theta}$ interact. Based on the interpretation of the hidden variables, a reasonable choice for all the $\mathbf{W}_n$ is the block-diagonal matrix in Equation (34): [7]

$$\begin{bmatrix} \mathbf{W}^{(T)} & & & \\ & \mathbf{W}^{r} & & \\ & & \mathbf{W}^{\alpha} & \\ & & & \mathbf{W}^{\beta} \end{bmatrix}, \tag{34}$$

where $\mathbf{W}^{(T)}$ and $\mathbf{W}^{r}$ are given respectively by $\sigma_T^2\mathbf{I}_K$ and $\sigma_r^2\mathbf{I}_K$, being σ_T^2 the *timekeeper variance* and σ_r^2 the *motor variance*. Since it is known that the motor variance is way smaller than the timekeeper variance [8, 14, 15], the conservatively high values $\sigma_T^2 = 500$ and $\sigma_r^2 = 25$ were considered. Both $\mathbf{W}^{\alpha}$ and $\mathbf{W}^{\beta}$ are also block-diagonal matrices, consisting of K blocks, each of dimen-

[6] Codes available at https://github.com/arme-project/ismir-2024.

[7] Off-diagonal blocks are null matrices, that were omitted exceptionally here, to avoid a line-break in the number of the equation. Moreover, notation $\mathbf{W}^{(T)}$ means to avoid confusion with the transpose matrix.

sions $(K-1)\times(K-1)$ and as in Equation (35):

$$\begin{bmatrix} v & c & \cdots & c \\ c & v & \cdots & c \\ \vdots & \vdots & \ddots & \vdots \\ c & c & \cdots & v \end{bmatrix}, \tag{35}$$

where v represents the variance of each α_{ij} (or β_{ij}) and c is the covariance between two distinct α_{ij} (or β_{ij}).

The rationale behind this construction for $\mathbf{W}^{\alpha}$ and $\mathbf{W}^{\beta}$ is that it makes sense to assume that for performer i there is a correlation only between the α_{ij} (or β_{ij}) for $j \neq i$. This means that all the correction gains for performer i interact among themselves, but not directly with the correction gains of other performers. Also, it is expected that the correlation between two distinct α_{ij} (or β_{ij}) is negative, since increasing correction towards a specific performer may cause a decrease of the synchronization towards the others. With this in mind, for matrix $\mathbf{W}^{\alpha}$ the value of v was chosen as 10^{-4}; the value of c was chosen such that the correlation between any two distinct α_{ij} is equal to -0.1. [8]

In this preliminary set of experiments with the KF, the effect of the β_{ij} was disregarded, by considering $\mathbf{W}^{\beta}$ a null matrix. It is known that the effect of the phase correction is way more relevant than the effect of the period correction [8, 14, 15], with the β_{ij} coefficients being usually much smaller than the α_{ij}. Also, preliminary experiments with artificial data also indicate that the dynamic values of the phase correction may render the period correction unnecessary. Since this is a point to be further investigated, it seemed safe to first experiment only with phase correction.

Matrices $\mathbf{V}_n$ were chosen to be constant and equal to $10^{-5}\mathbf{I}_K$: since the block $\mathbf{W}^r$ in matrix $\mathbf{W}_n$ already captures the motor variance, $\mathbf{V}_n$ should be a sequence of null matrices, but a negligible diagonal term was added to avoid numerical errors. Finally, the initialization of vector $\boldsymbol{\theta}$ was done by choosing its first K components and the components from $K+1$ to $2K$ to be equal to the first IOI of each of the K instruments, all the α_{ij} were initially set to 0.25, and all the β_{ij} to zero. This initialization of α_{ij} is supported by [8], where optimal correction values for ensembles of size K were derived.

Figure 1 summarizes one experiment performed in the aforementioned scenario. Three repetitions of the Haydn quartet excerpt were analyzed, being one for each of the three performance conditions, having the second violin as the leader in the "Speed" case. Each performance consists of a sequence of 46 four-dimensional vectors containing the IOIs for each instrument. Since it is not the goal of this set of experiments to evaluate online performance of the proposed model, these three sequences were smoothed by the KF, [9] according to Equation 21. Each panel of Figure 1 displays the evolution of the α_{ij}, for $j \neq i$, organized as follows: each column contains a performance condition (made explicit at its top), and each row displays the evolution of α_{ij} for $j \neq i$ and a fixed value of i. The conditioning of each α_{ij} on $\mathbf{y}_{1:N}$ is omitted, and the instruments are abbreviated by numbers, where 1, 2, 3, and 4 refers to the first violin, second violin, viola, and cello, respectively. On each panel of Figure 1 the values of α_{ij} promptly deviates from the optimal initialization of 0.25 (but still varies around it), and their respective behavior are now discussed.

In "Speed" condition (third column in Figure 1), on each panel the phase correction parameter toward the second violin (α_{i2}, for $i = 1, 3, 4$) shows a small increase by the end of the performance, when the change in speed occurs, since the second violin is assigned as the leader to initiate this change in speed. Notice also that his/her phase correction parameters towards the other performers (α_{2j}, for $j = 1, 3, 4$) decrease through time, specially near the last notes, reinforcing its leadership in this tempo change.

In the "Normal" condition (second column in Figure 1) it is noticeable that the second violin, viola, and cello are systematically synchronizing mainly to the first violin, which plays the melody in this excerpt: notice the almost constant value for α_{i1}, for $i = 2, 3, 4$. While the cello is synchronizing mainly with the first and second violin, it presents the weaker "synchronization attractor", as seen by the significant decrease in α_{i4} through time, for $i = 1, 2, 3$.

Finally, in the "Deadpan" condition (first column in Figure 1) the first and second violin and the cello are synchronizing mainly to the viola (steady increase of α_{i3} for $i = 1, 2, 4$, and decrease in α_{3j} for $j = 1, 2, 4$), which may be the cause of the cello synchronizing systematically with all the three other instruments.

This experiment indicates that the proposed model is capable of capturing local fluctuations in tempo, reinforces the role of the phase correction gain in interpreting synchronization mechanisms in musical ensembles, and assess qualitatively the validity of a time-varying model to the problem of ensemble synchronization. As a next step in this new direction for the field, the proposed model will be broadly tested and systematically compared with other models. Some issues to be addressed in future work are: perform experiments with other data contained on [21]; compare filtering and smoothing procedures, as well as investigate if the filtered estimates make sense from a music cognition perspective; implement tools from the theory of dynamic linear models to automatically estimate the covariance matrices $\mathbf{V}_n$ and $\mathbf{W}_n$ [18]; perform a systematic comparison with the bGLS and ADAM algorithms.

6. CONCLUSION

This paper presented a novel model, based on the Kalman Filter, for analysing asynchrony correction in music ensemble performances. The proposed model is founded on well-established models in the literature, and has the advantage of considering dynamic phase and period correction gains. A set of experiments (using only phase correction) on a homophonic section of a string quartet by J. Haydn was conducted, illustrating the capabilities of the model in explaining synchronization schemes within musical ensembles.

[8] This procedure will not always lead to a positive-definite matrix, for sufficiently high value of K and depending on c – not the case here.

[9] The computational time of each smoothing is less than 100ms.

7. ACKNOWLEDGMENTS

The ARME Project (Augmented Reality Music Ensemble – https://arme-project.co.uk/) is funded by the EPSRC grant with reference EP/V034987/1. We would like to thank the University of Birmingham for the scholarship awarded to the first author through the Brazil Visiting Fellows Scheme. We also would like to thank the reviewers and meta-reviewers for the useful insights and suggestions to improve this paper.

8. REFERENCES

[1] A. D. Patel, J. R. Iversen, M. R. Bregman, and I. Schulz, "Experimental evidence for synchronization to a musical beat in a nonhuman animal," *Current Biology*, vol. 19, no. 10, pp. 827–830, 2009.

[2] A. M. Wing and C. Woodburn, "The coordination and consistency of rowers in a racing eight," *Journal of Sports Sciences*, vol. 13, no. 3, pp. 187–197, 1995.

[3] E. Goodman, "Ensemble performance," in *Musical performance: a guide to understanding*, J. Rink, Ed. Cambridge, UK: Cambridge University Press, 2002, pp. 153–167.

[4] C. Palmer, "Music performance," *Annual Review of Psychology*, vol. 48, pp. 115–138, 1997.

[5] J. W. Davidson and J. M. M. Good, "Social and musical co-ordination between members of a string quartet: An exploratory study," *Psychology of Music*, vol. 30, no. 2, pp. 186–201, 2002.

[6] J. K. Murnighan and D. E. Conlon, "The dynamics of intense work groups: A study of british string quartets," *Administrative Science Quarterly*, vol. 36, no. 2, pp. 165–186, 1991.

[7] J. Gibbon, C. Malapani, C. L. Dale, and C. R. Gallistel, "Toward a neurobiology of temporal cognition: advances and challenges," *Current Opinion in Neurobiology*, vol. 7, no. 2, pp. 170–184, 1997.

[8] A. M. Wing, S. Endo, A. Bradbury, and D. Vorberg, "Optimal feedback correction in string quartet synchronization," *Journal of the Royal Society Interface*, vol. 11, no. 20131125, 2014.

[9] D. Vorberg and H. H. Schulze, "Linear phase correction in synchronization: predictions, parameter estimation, and simulations," *Journal of Mathematical Psychology*, vol. 46, no. 1, pp. 56–87, 2002.

[10] D. Vorberg and A. M. Wing, "Modeling variability and dependence in timing," in *Handbook of perception and action, vol. 2*, H. Heuer and S. Keele, Eds. New York, USA: Academid Press, 1996, pp. 181–262.

[11] J. Mates, "A model of synchronization of motor acts to a stimulus sequence," *Biological Cybernetics*, vol. 71, p. 186, 1994.

[12] B. H. Repp, "Processes underlying adaptation to tempo changes in sensorimotor synchronization," *Human Movement Science*, vol. 20, no. 3, pp. 277–312, 2001.

[13] B. H. Repp and P. E. Keller, "Adaptation to tempo changes in sensorimotor synchronization: effects of intention, attention, and awareness," *Quarterly Journal of Experimental Psychology*, vol. 57, no. 3, pp. 499–521, 2004.

[14] N. Jacoby, N. Tishby, B. H. Repp, M. Ahissar, and P. E. Keller, "Parameter estimation of linear sensorimotor synchronization models: Phase correction, period correction, and ensemble synchronization," *Timing & Time Perception*, vol. 3, no. 1–2, pp. 52–87, 2015.

[15] N. Jacoby, P. E. Keller, B. H. Repp, M. Ahissar, and N. Tishby, "Lower bound on the accuracy of parameter estimation methods for linear sensorimotor synchronization models," *Timing & Time Perception*, vol. 3, no. 1–2, pp. 32–51, 2015.

[16] M. C. van der Steen and P. E. Keller, "The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization," *Frontiers in Human Neuroscience*, vol. 7, 2013.

[17] B. Harry and P. E. Keller, "Tutorial and simulations with ADAM: an adaptation and anticipation model of sensorimotor synchronization," *Biological Cybernetics*, vol. 113, pp. 397–421, 2019.

[18] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic Linear Models with R*. Berlin/Heidelberg, Germany: Springer, 2009.

[19] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice with MATLAB*. New Jersey, USA: Wiley-IEEE Press, 2014.

[20] W. Mader, Y. Linke, M. Mader, L. Sommerlade, J. Timmer, and B. Schelter, "A numerically efficient implementation of the expectation maximization algorithm for state space models," *Applied Mathematics and Computation*, vol. 241, pp. 222–232, 2014.

[21] M. Tomczak, M. S. Li, and M. D. Luca, "Virtuoso strings: A dataset of string ensemble recordings and onset annotations for timing analysis," in *Extended Abstracts for the Late-Breaking Demo Session of the 24th Int. Society for Music Information Retrieval Conf.*, Milan, Italy, 2023.