Technical Appendix for *The ACCompanion: Combining Reactivity, Robustness, and Musical Expressivity in an Automatic Piano Accompanist*

Carlos Cancino-Chacón[1] Silvan Peter[1] Patricia Hu[1]
Emmanouil Karystinaios[1] Florian Henkel[2] Francesco Foscarin[1]
Nimrod Varga[1] Gerhard Widmer[1]

[1]Johannes Kepler University Linz, Austria
[2]SiriusXM + Pandora, USA

Version: May 20, 2023

1 Sensorimotor Synchronization Models

This section follows the notation introduced in Section 3 (in particular Figure 3 in the paper), but we drop the superscript $\texttt{p}$ of performed onsets to unclutter notation. In order to make this document more readable on its own, we include Figure 3 in the paper here (Figure 1). In the following description of the tempo models, we denote the observed performed onset time of the n-th onset in the score as o_n and the onset time predicted by the synchronization models as $\hat{o}_n$. The asynchrony between these onsets is denoted as $A_n = \hat{o}_n - o_n$ and the observed beat period is given as $\tau_n = \frac{\delta_n^{\texttt{perf}}}{\delta_n^{\texttt{score}}}$, with τ_0 being the initial tempo set by the performer as a hyper parameter.

1.1 Reactive Sync Model (R)

The next accompaniment onset is "estimated" at the same time as the last observed performed onset, making this model purely reactive.

$$\hat{o}_{n+1} = o_n + b_n \delta_n^{\texttt{score}} \tag{1}$$

$$b_{n+1} = \tau_n \tag{2}$$

1.2 Moving Average Sync Model (MA)

This model is very similar to the first baseline, except for its estimation of the global

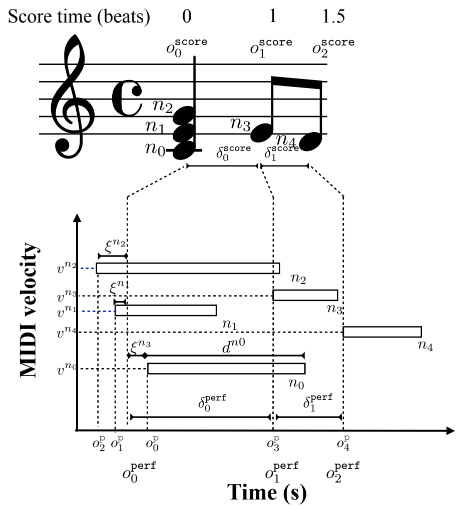

Figure 1: Excerpt of a MIDI performance (as a piano roll where the x-axis describes performance time in seconds and the y-axis is the MIDI velocity of the note) and its corresponding score, showcasing the elements for encoding/decoding an expressive performance. **This figure is identical to Figure 3 in the paper.**

tempo, which uses a weighted average of the last and current tempo estimate.

$$\hat{o}_{n+1} = o_n + b_n \delta_n^{\texttt{score}} \tag{3}$$

$$b_{n+1} = \eta_{\mathrm{MA}} b_n + (1 - \eta_{\mathrm{MA}}) \tau_n \tag{4}$$

where η_{MA} is a constant parameter.

1.3 Linear SMS Model (L)

Onset and beat period estimates o_{n+1} and b_{n+1} are computed as follows:

$$\hat{o}_{n+1} = \hat{o}_n + b_n \delta_n^{\texttt{score}} - \eta^o A_n \tag{5}$$

$$b_{n+1} = \begin{cases} b_n - \eta^b A_n, & \text{if } A_n < 0, \\ b_n - 2\eta^b A_n, & \text{else} \end{cases} \tag{6}$$

where η^o and η^b are the learning rates for the onset and beat period, respectively.

1.4 Linear Tempo Expectation Model (LTE)

This model is similar to the previous one, but includes an anticipation (expectation) term in the computation of the beat period:

$$\hat{o}_{n+1} = \hat{o}_n + b_n \delta_n^{\texttt{score}} - \eta^o A_n \tag{7}$$

$$b_{n+1} = \phi(o_{n+1}^{\texttt{score}}) - \eta^b A_n, \tag{8}$$

where $\phi(o_n^{\texttt{score}})$ is a function that computes an estimate of the beat period at score onset time $o_n^{\texttt{score}}$ based on the tempo of the reference performance(s) and η^o and η^b are constant parameters that represent the learning rates for the onset and beat period, respectively.

1.5 Joint Adaptation Anticipation Model (JADAM)

This model includes both an error-correction term (the adaptation part) and a moving average estimate of the observed beat period (the anticipation part).

1. Adaptation Module

$$\hat{o}_{n+1}^{\mathrm{ad}} = \hat{o}_n + b_n \delta_n^{\texttt{score}} - \eta_{\mathrm{J}}^o A_n \tag{9}$$

$$b_{n+1} = b_n - \eta_{\mathrm{J}}^b A_n \tag{10}$$

2. Anticipation Module

$$\hat{\tau}_n = \eta_{\mathrm{J}}^b (2\tau_n - \tau_{n-1}) + (1 - \eta_{\mathrm{J}}^b)\tau_n \tag{11}$$

$$\hat{o}_{n+1}^{\mathrm{an}} = o_n + \hat{\tau}_n \delta_n^{\texttt{score}} \tag{12}$$

3. Joint Module

$$\hat{A}_n = \hat{o}_{n+1}^{\mathrm{ad}} - \hat{o}_{n+1}^{\mathrm{an}} \tag{13}$$

$$\hat{o}_{n+1} = \hat{o}_{n+1}^{\mathrm{an}} - \eta_{\mathrm{J}}^a \hat{A}_n \tag{14}$$

Parameters η_{J}^o and η_{J}^a and η_{J}^b are learning rates for the prediction of the onset of the adaptation and joint modules and the learning rate for the beat period, respectively.

1.6 Kalman Tempo Model (KT)

The observed variable of the Kalman filter is the performed onset time, and the beat period is the latent variable. The updates are computed as follows:

$$\hat{b}_n = \alpha_{\mathrm{K}} b_n \tag{15}$$

$$v_n = \gamma_{\mathrm{K}}^2 \hat{v}_n + \beta_{\mathrm{K}} \tag{16}$$

$$\hat{A}_n = \delta_n^{\mathtt{perf}} - \hat{b}_n \delta_n^{\mathtt{score}} \tag{17}$$

$$\kappa_n = \frac{v_n \delta_n^{\mathtt{score}}}{v_n \delta_n^{\mathtt{score}^2} + \lambda_{\mathrm{K}}} \tag{18}$$

$$b_{n+1} = \hat{b}_n + \kappa_n \hat{A}_n \tag{19}$$

$$\hat{v}_{n+1} = (1 - \kappa_n \delta_n^{\mathtt{score}}) v_n \tag{20}$$

$$\hat{o}_{n+1} = \hat{o}_n + b_{n+1} \delta_n^{\mathtt{score}} \tag{21}$$

where α_{K}, β_{K}, γ_{K} and λ_{K} are the parameters of the model.