

A general linear framework for the comparison and evaluation of models of sensorimotor synchronization

Nori Jacoby · Bruno H. Repp

Received: 3 September 2010 / Accepted: 27 February 2012 / Published online: 24 April 2012
© Springer-Verlag 2012

Abstract Sensorimotor synchronization (SMS), the temporal coordination of a rhythmic movement with an external rhythm, has been studied most often in tasks that require tapping along with a metronome. Models of SMS use information about the timing of preceding stimuli and responses to predict when the next response will be made. This article compares the theoretical structure and empirical predictions of four two-parameter models proposed in the literature: Michon (Timing in temporal tracking, Van Gorcum, Assen, 1967), Hary and Moore (Br J Math Stat Psychol 40:109–124, 1987b), Mates (Biol Cybern 70:463–473, 1994a; Biol Cybern 70:475–484, 1994b), and Schulze et al. (Mus Percept 22:461–467, 2005). By embedding these models within a general linear framework, the mathematical equivalence of the Michon, Hary and Moore, and Schulze et al. models is demonstrated. The Mates model, which differs from the other three, is then tested empirically with new data from a tapping experiment in which the metronome alternated between two tempi. The Mates model predictions are found to be invalid for about one-third of the trials, suggesting that at least one of the model's underlying assumptions is incorrect. The other models cannot be refuted as easily, but they do not predict some features of the data very accurately. Comparison of the models' predictions in a training/test pro-

cedure did not yield any significant differences. The general linear framework introduced here may help in the formulation of new models that make better predictions.

Keywords Synchronization · Tapping · Linear Models · Model Comparison · Action-Perception Loop · Timing

1 Introduction

Sensorimotor synchronization (SMS), usually in the form of tapping along with a metronome, has been studied extensively in recent decades (for a review, see Repp 2005), and various models have been proposed to explain the experimental results. These models can be grouped into two categories: (1) linear models (Hary and Moore 1985, 1987a,b; Mates 1994a,b; Michon 1967; Pressing 1998a,b, 1999; Pressing and Jolley-Rogers 1997; Schulze et al. 2005; Schulze and Vorberg 2002; Thaut et al. 1998a,b; Vorberg and Schulze 2002; Vorberg and Wing 1996) and (2) nonlinear dynamic systems models (e.g., Large and Jones 1999; McAuley and Jones 2003). This article is concerned only with linear models, which have dominated SMS research and lend themselves to analytical treatment. Linear models usually contain two parts: (a) a “deterministic” prediction of the time of the next tap based on previous stimulus and response times, and (b) added stochastic terms, usually Gaussian, that reflect noisy internal processes such as clock (timekeeper) and motor variance (Wing and Kristofferson 1973a,b). We focus here on the deterministic parts of several models. The role of the stochastic parts is considered briefly in Appendix 2.

Michon (1967) proposed a model that is completely deterministic in its original version. Later models of Hary and Moore (1985, 1987a,b) and Mates (1994a,b) have sometimes been used with the noise terms set to zero, as may be done when considering averaged data (e.g., Repp and Keller 2004).

N. Jacoby (✉)

The Edmond and Lily Safra Center for Brain Sciences and Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel
e-mail: nori.viola@gmail.com

N. Jacoby

Music Department, Bar Ilan University, Ramat Gan, Israel

B. H. Repp

Haskins Laboratories, 300 George Street, New Haven, CT 06511-6624, USA

More recently, Schulze et al. (2005) proposed a new variant of the Mates model. In this article, we will compare these four different linear models by identifying and evaluating their properties. We will introduce a formal framework that generalizes the assumptions underlying these models, with the individual models presented as specific cases. Then, using a newly collected set of data, we will use the accuracy of each model's predictions as an objective measure of its success.

We will begin by considering the model proposed by Mates (1994a,b), which postulates two internal processes (period correction and phase correction) that are based on specific perceptual information. Next, we will analyze Michon's (1967) pioneering linear prediction model, which is based on recent asynchronies between taps and metronome ticks. We will then examine one of the models proposed by Hary and Moore (1985, 1987a,b), which makes a prediction based on two reference points, either the last metronome beat or the last response ("mixed phase resetting"). Finally, we will consider the variant of the Mates model employed by Schulze et al. (2005), in which period correction is based on different perceptual information than in the Mates model.

As we will demonstrate, we can embed these models within a general linear framework similar to the ones envisioned by Thaut et al. (1998a,b), Hasan and Thaut (1999), and Pressing (1999). More specifically, we will see that each model has two adjustable parameters that can be expressed within what we will call the Canonical XY model. Furthermore, we will see that the Michon, Hary and Moore, and Schulze et al. models are equivalent in the sense that we can map their parameters onto those of the Canonical XY model. Conversely, we can map any parameters of the Canonical XY model to parameters of the Michon, Hary and Moore, and Schulze et al. models. The models differ in the "naming" and psychological interpretation of the parameters, but they are equivalent mathematically and therefore make the same predictions. However, the Mates model is not equivalent to the other three, and makes further assumptions. Therefore, we can try to validate those further assumptions by experimental data, and test the accuracy of the model.

This article is divided into four parts: An introduction presenting the general linear framework is followed by a review of specific models from the literature and their representation within the Canonical XY model. In the third section, we compare the success of these different models in explaining newly collected data, using the prediction error as an objective measure. The fourth section is a general discussion. Two appendices provide additional details about parameter estimation methods and introduce a new method that we apply to our data.

2 Introducing the linear framework

Figure 1 shows the overall structure of our framework. The model takes as input the times of occurrence of past stimuli

$S(i)$ and responses $R(i)$, or the stimulus inter-onset intervals $s(i)$ and response inter-onset intervals $r(i)$, where i is an integer index running from 1 to k , the present. Let us further denote by $e(i)$ the asynchrony or synchronization error $R(i) - S(i)$.

The stimuli and responses are input information available to our brain. The brain processes the input and makes a prediction regarding the timing of the next response to an expected stimulus. This prediction is executed through our motor system, which produces the next response $R(k + 1)$. The aim of any model is to mimic this behavior and accurately predict the next response by deriving $R'(k + 1)$, the predicted time of the response, from the past $S(i)$ and $R(i)$ for $i = k, k - 1, k - 2, \dots$. We can empirically estimate the prediction error of a model by calculating $R'(k + 1) - R(k + 1)$. This gives us an objective measure by which to evaluate models: A better model has a smaller prediction error for a given set of experimental data.

The deterministic part of a linear model consists of input variables and parameters. The parameter values are the weights that are given to the input variables, and are usually determined by fitting the model to data. The best values are those that achieve minimal prediction error. Given those weights, the prediction is calculated by multiplying each of the input variables by its appropriate weight (a, b, c, \dots) and taking the sum. For example, if the model structure consists of the variables $S(k), S(k - 1), R(k)$, and $R(k - 1)$, the prediction of the model would be

$$R'(k + 1) = a * S(k) + b * S(k - 1) + c * R(k) + d * R(k - 1). \quad (1)$$

The root mean squared prediction error of the model in (1) is

$$E = \text{sqrt} \left\{ \sum_{i=3,4,\dots,N} [R'(i) - R(i)]^2 / (N - 2) \right\}, \quad (2)$$

where N is the number of taps in the experimental data and sqrt is the square root.

Note the following combinatorial identity between $s(k)$, $r(k)$, $e(k)$, and $e(k - 1)$:

$$s(k) = e(k - 1) + r(k) - e(k). \quad (3)$$

Proof

$$e(k - 1) + r(k) - e(k) = [R(k - 1) - S(k - 1)] + [R(k) - R(k - 1)] - [R(k) - S(k)] = s(k).$$

Figure 2 depicts a graphical representation of the preceding proof: The sum of any closed path between nodes in the graph is zero. For example, we start at $S(k)$ and go counter-clockwise to $S(k - 1)$, followed by $R(k - 1)$, then continuing to $R(k)$, and ending at $S(k)$, closing the cycle where we began.

Fig. 1 The general linear framework. S stimulus; R response; s stimulus interval; r response interval; e synchronization error (asynchrony); f, g, h parameter weights in a general linear model

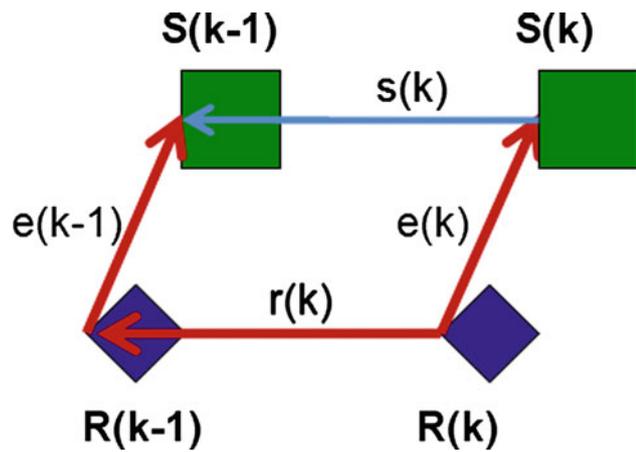
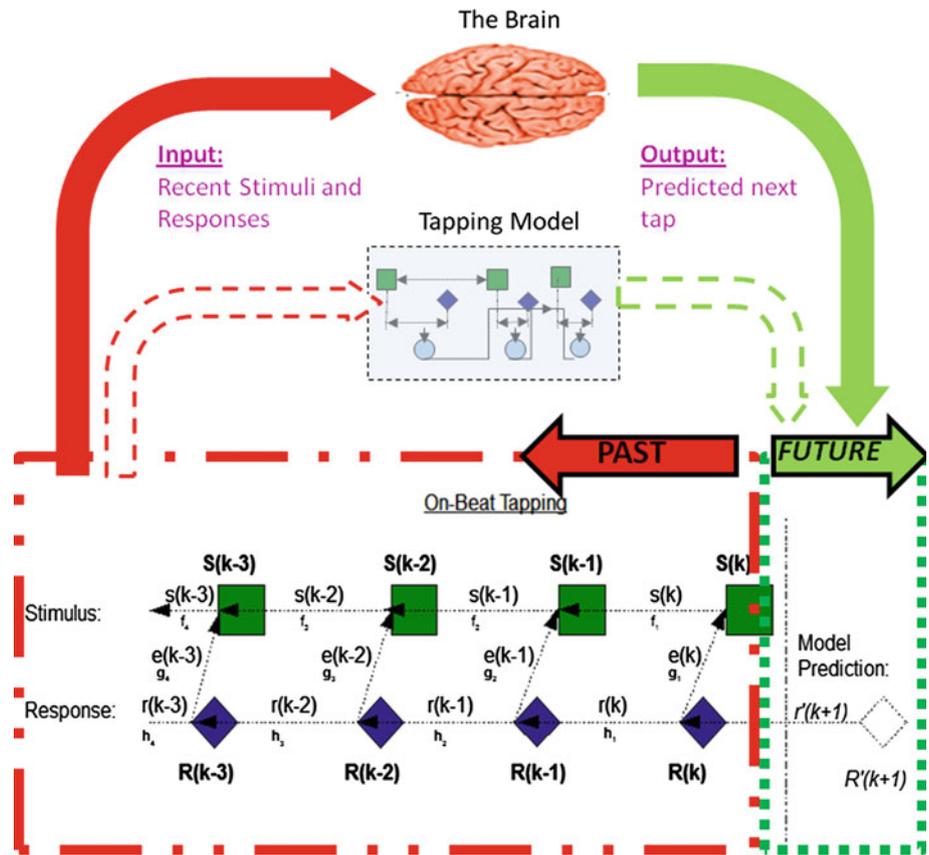


Fig. 2 Detail of a closed path in the general linear model

Therefore, $s(k) - e(k - 1) - r(k) + e(k) = 0$, leading to the combinatorial identity of Eq. 3. Note that the signs in Eq. 3 are determined by the direction of the arrows in Fig. 2: We have a minus sign when our path takes us against the direction of the arrow, and a plus sign when our path accords with the direction of the arrow.

This principle holds also for any other closed path in the larger structure that results from extending the graph in Fig. 2

backward in time (cf. Fig. 1), leading to further combinatorial identities, for example

$$s(k) + s(k - 1) - e(k - 2) - r(k - 1) - r(k) + e(k) = 0. \tag{4}$$

The combinatorial identity in Eq. 3 relates the asynchronies $e(k)$ and $e(k - 1)$ and the response interval $r(k)$ to the stimulus interval $s(k)$. Therefore, if we know $e(k)$, $r(k)$, and $e(k - 1)$, then $s(k)$ is determined by Eq. 3. If we know all past asynchronies and response intervals, we can reconstruct the stimulus interval sequence, and the most general model we can construct will take into account as variables the entire asynchrony and response sequence. The formula for a general linear model is

$$r(k + 1) = \sum_{i=0 \dots k-1} [g_i * e(k - i) + h_i * r(k - i)] + Z(k + 1), \tag{5}$$

where $Z(k + 1)$ is a stochastic error term (usually assumed to be Gaussian), and $r'(k + 1) = r(k + 1) - Z(k + 1)$.

Note that any other linear model can be nested within this model. As we will see, the Mates, Michon, Hary and Moore, and Schulze et al. models can be presented as specific choices of g_i, h_i , and $Z(k + 1)$.

The combinatorial identity (Eq. 3) also enables us to choose other variables for our general model. For example, we could have chosen $s(k)$ and $e(k)$ instead of $e(k)$ and $r(k)$, obtaining equivalent results. In this case, the formula for the general linear model would have been

$$r(k+1) = \sum_{i=0 \dots k-1} [f_i * s(k-i) + g_i * e(k-i)] + Z'(k+1), \quad (6)$$

where again $Z'(k+1)$ is a stochastic variable. Note that also for this model, any other linear model can be written as a specific choice of f_i , g_i , and $Z'(k+1)$. In the context of this article, we are going to work with Eq. 5, and we will select the asynchronies and response intervals as the variables for our models.

3 Linear models from the literature and their representation

In this section, we will present the Mates, Michon, Hary and Moore, and Schulze et al. models, and see how they fit into the general framework introduced in Part 1. Specifically, we are going to represent those models' predictions $r'(k+1)$ as a function of $e(k)$, $e(k-1)$, and $r(k)$.

Note that the Mates, Hary and Moore, and Schulze et al. models have a stochastic and possibly autocorrelated noise term $Z(k+1)$. In the entire following discussion, we will consider only the deterministic part of those models. This is equivalent to $Z(k)$ being uncorrelated or $\text{cov}[Z(k), Z(k+j)] = 0$ for $j > 0$ because then the best estimate of the next step is the deterministic part (see Rao and Toutenburg 1999). See Appendix 2 for a more general discussion that takes the error term $Z(k+1)$ into account.

3.1 The Mates model

The Mates (1994a,b) model considers two different error correction processes: period correction and phase correction. This model creates a control system similar to a phase-locked loop (Best 2003; Dorf 1993). It includes both external (observable) and internal (hypothetical psychological) variables.

Period correction relies on an internal variable $t_1(k)$ called a timekeeper. (The subscript indicates that it is internal.) The timekeeper is similar to a metronome, and its interval is continually updated by the current stimulus interval. The next timekeeper interval $t_1(k)$ is a function of the difference between the previous timekeeper interval $t_1(k-1)$ and the stimulus interval $s(k) = S(k) - S(k-1)$, multiplied by β , the period correction parameter. As β approaches 1, the internal timekeeper is increasingly affected by the last stimulus

interval, and less by the previous timekeeper interval. The formula for internal timekeeper updating thus is

$$t_1(k) = t_1(k-1) - \beta * [t_1(k-1) - s(k)]. \quad (7)$$

Phase correction is an instantaneous correction mechanism dependent on the last asynchrony $e(k)$. The phase correction parameter α determines the degree by which the model prediction will be affected by the asynchrony. The formula for the prediction of the next response interval $r'(k+1) = R'(k+1) - R(k)$ is therefore

$$r'(k+1) = t_1(k) - \alpha * e(k). \quad (8)$$

We propose writing the Mates model as a function of $e(k)$, $e(k-1)$, and $r(k)$, so that the model can be rewritten as

$$r'(k+1) = (-\alpha - \beta) * e(k) + (\alpha + \beta - \alpha * \beta) * e(k-1) + r(k). \quad (9)$$

Proof From Eq. 8, we can infer that $r'(k) = t_1(k-1) - \alpha * e(k-1)$, and from this we get

$$t_1(k-1) = r'(k) + \alpha * e(k-1).$$

By substituting $t_1(k-1)$ in Eq. 7 we get

$$t_1(k) = (1 - \beta) * [r'(k) + \alpha * e(k-1)] + \beta * s(k), \quad (10)$$

or

$$t_1(k) = (1 - \beta) * r'(k) + \alpha * (1 - \beta) * e(k-1) + \beta * s(k). \quad (11)$$

We substitute $t_1(k)$ in Eq. 8 with Eq. 11 and get

$$r'(k+1) = (1 - \beta) * r'(k) + \alpha * (1 - \beta) * e(k-1) + \beta * s(k) - \alpha * e(k). \quad (12)$$

We can also use the combinatorial identity of Eq. 3 and substitute $s(k)$ in Eq. 12:

$$r'(k+1) = (1 - \beta) * r'(k) + \alpha(1 - \beta) * e(k-1) + \beta * [e(k-1) + r(k) - e(k)] - \alpha * e(k), \quad (13)$$

or

$$r'(k+1) - (1 - \beta) * r'(k) = (\alpha + \beta - \alpha * \beta) * e(k-1) + \beta * r(k) - (\alpha + \beta) * e(k). \quad (14)$$

At $k+1$, $r'(k)$ is already known; therefore, we can substitute it with the real value $r(k)$. This step significantly simplifies the calculations.¹ After substituting $r'(k)$ with $r(k)$, we get

$$r'(k+1) = (-\alpha - \beta) * e(k) + (\alpha + \beta - \alpha * \beta) * e(k-1) + r(k), \quad (15)$$

which completes the proof (cf. Eq. 9). \square

¹ We have carried out experimental analyses with and without this extra assumption, and it seems that the difference is slight. An alternative calculation without this further assumption is described in Appendix 2.

Note that this formula is symmetric in α and β . This means that if we swap the values of α and β , we will get the same predictions, a fact pointed out previously by Repp (2001a, 2005). This also means that if we fit the model parameters, we cannot determine which value belongs to which parameter, except perhaps by external criteria. We can also see that even though the model is linear in the input variables $e(k)$, $e(k - 1)$, and $r(k)$, the coefficients multiplying them are not linear in α and β .

To be more concise, if we write the Mates model as

$$r'(k + 1) = x * e(k) + y * e(k - 1) + r(k), \tag{16}$$

then

$$x = -\alpha - \beta \tag{17}$$

and

$$y = \alpha + \beta - \alpha * \beta. \tag{18}$$

The parameter y has the term $\alpha * \beta$ and is therefore a nonlinear function of α and β .

By writing $r'(k + 1)$ as a function of $e(k)$, $e(k - 1)$, and $r(k)$, we have expressed $r'(k + 1)$ without explicitly using the internal timekeeper. Instead, its correction parameter β together with the phase correction parameter α determines the model coefficients x and y .

3.2 The Michon model

Michon (1967) (see also Michon and van der Valk 1967) introduced a simple model that bases its prediction upon the last few asynchronies. Michon specifically mentioned one- and two-parameter versions of this model, while his general model states that the predicted response interval is a function of the last stimulus interval and all past asynchronies. The formula for the general model is $r'(k + 1) = s(k) + \sum_{i=0 \dots k-1} [f_k * e(k - i)]$. In this article, we will consider the two-parameter version of the Michon model.

In the notation we previously introduced, this model can be written as

$$r'(k + 1) = s(k) + a * e(k) + b * e(k - 1). \tag{19}$$

Using Eq. 3, we can substitute $s(k)$ with $e(k - 1) + r(k) - e(k)$ and get

$$r'(k + 1) = e(k - 1) + r(k) - e(k) + a * e(k) + b * e(k - 1) \tag{20}$$

or

$$r'(k + 1) = (a - 1) * e(k) + (1 + b) * e(k - 1) + r(k). \tag{21}$$

We can rewrite Eq. 21 as

$$r'(k + 1) = x * e(k) + y * e(k - 1) + r(k), \tag{22}$$

where

$$x = a - 1 \tag{23}$$

and

$$y = 1 + b. \tag{24}$$

Note the striking similarity between Eqs. 22–24 and the Mates model Eqs. 16–18. However, there is an important difference: In the Michon model, x and y are linear functions of parameters a and b , whereas in the Mates model, y is multiplicative in parameters α and β .²

3.3 The Hary and Moore model

The Hary and Moore (1985, 1987a,b) model consists of an internal delay time, $t(k)$, similar to Mates' internal timekeeper in that it is updated by an "error" signal, in this case the most recent asynchrony. Using the equation from Hary and Moore (1987b, p. 121, Formula 40):

$$t(k + 1) = t(k) - f * [e(k) + n_e(k)], \tag{25}$$

where $t(k + 1)$ is the updated internal delay, and f is a constant called the "feedback fraction." Also, $n_e(k)$ is a "noise" factor that we omit in the following equations (see Appendix 2).

This model prediction is dependent on the choice of reference for each phase reset (i.e., the point from which the internal delay $t(k + 1)$ is measured). The phase could be reset to the metronome, so that the next prediction will be calculated from the last metronome time, $S(k)$, or to the last response, in which case the delay is calculated from the response time, $R(k)$. Therefore, the Hary and Moore model prediction is

$$R'(k + 1) = R(k) + t(k) + n_r(k) \text{ with probability } g, \text{ and} \\ R'(k + 1) = S(k) + t(k) + n_r(k) \text{ with probability } (1 - g), \tag{26}$$

where $n_r(k)$ is an additional independent noise factor (such as motor peripheral noise) that we again omit in the following equations.³

Later, the authors transformed this model into a deterministic model. Instead of a probabilistic selection between a metronome and a response resetting, they made the reset a linear function of $R(k)$ and $S(k)$, rewriting their Formulae 38–40 (Hary and Moore 1987b, p. 121) to produce a "combined-reset" strategy:

$$R'(k + 1) = g * R(k) + (1 - g) * S(k) + t(k) \tag{27}$$

$$t(k + 1) = t(k) - f * e(k) \tag{28}$$

² This means that all parameters of the Mates model can be mapped to parameters of the Michon model (while the reverse may not be true). For a detailed explanation and discussion of this claim, see Sect. 3.6.

³ Note that Hary and Moore denote our $S(k)$ by M_i and our g by a .

Note that this is similar but not identical to the Mates model. As we did in Mates' case, we will now rewrite the model. From Eq. 27, we can deduce that

$$t(k) = R'(k+1) - g * R(k) - (1-g) * S(k). \quad (29)$$

Substituting $t(k)$ in Eq. 28 with Eq. 29 we get

$$t(k+1) = R'(k+1) - g * R(k) - (1-g) * S(k) - f * e(k), \quad (30)$$

or

$$t(k) = R'(k) - g * R(k-1) - (1-g) * S(k-1) - f * e(k-1). \quad (31)$$

Substituting $t(k)$ in Eq. 27 with Eq. 31 we get

$$R'(k+1) = g * R(k) + (1-g) * S(k) + R'(k) - g * R(k-1) - (1-g) * S(k-1) - f * e(k-1). \quad (32)$$

Again substituting $R'(k)$ with $R(k)$ (see our discussion of Mates' model), and rearranging the terms

$$R'(k+1) = R(k) + g * r(k) + (1-g) * s(k) - f * e(k-1). \quad (33)$$

Using Eq. 3 to substitute for $s(k)$

$$R'(k+1) = R(k) + g * r(k) + (1-g) * [e(k-1) + r(k) - e(k)] - f * e(k-1), \quad (34)$$

or

$$R'(k+1) = R(k) + r(k) + (1-g-f) * e(k-1) + (g-1) * e(k), \quad (35)$$

or

$$r'(k+1) = (g-1) * e(k) + (1-g-f) * e(k-1) + r(k). \quad (36)$$

This is almost identical to Michon's model (Eq. 21). Note that we can again write the model in the form:

$$r'(k+1) = x * e(k) + y * e(k-1) + r(k), \quad (37)$$

where

$$x = g - 1 \quad (38)$$

$$y = 1 - g - f \quad (39)$$

These parameters are linear functions of f and g , and therefore we can "translate" every model of the form of Eq. 36 (Hary and Moore) to a model of the form of Eq. 21 (Michon) and vice versa. We can say that those models are *equivalent*,

since after the translation the models make the same predictions and therefore also have the same empirical prediction error.

3.4 The Schulze et al. model

Formulae 2 and 4 in Schulze et al. (2005, p. 462) present an alternative to the Mates model. For Schulze et al., phase correction depends on dual temporal references (as in the Hary and Moore model), which is formally equivalent to depending on the preceding asynchrony, but period correction depends on the preceding asynchrony⁴, too, not on a comparison of $s(k)$ and the internal timekeeper interval $t_1(k)$. A deterministic version of their model can therefore be expressed as

$$t_1(k) = t_1(k-1) - \beta * e(k-1), \quad (40)$$

$$r'(k+1) = t_1(k+1) - \alpha * e(k). \quad (41)$$

As in the other three cases presented above, we can compute

$$r'(k) = t_1(k) - \alpha * e(k-1) \quad (\text{from Eq. 41}) \quad (42)$$

$$t_1(k) = r'(k) + \alpha * e(k-1) \quad (\text{reordered}) \quad (43)$$

$$t_1(k+1) = t_1(k) - \beta * e(k) \quad (\text{from Eq. 40}) \quad (44)$$

$$t_1(k+1) = r'(k) + \alpha * e(k-1) - \beta * e(k) \quad (\text{substitution of } t_1(k)) \quad (45)$$

$$r'(k+1) = r'(k) + \alpha * e(k-1) - \beta * e(k) - \alpha * e(k). \quad (\text{substitution of } t_1(k+1) \text{ in Eq. 41}) \quad (46)$$

And again, replacing $r'(k)$ by $r(k)$, we get

$$r'(k+1) = (-\alpha - \beta) * e(k) + \alpha * e(k-1) + r(k). \quad (47)$$

As is readily evident, this model is also equivalent to the Michon and Hary and Moore models. Here,

$$x = -\alpha - \beta \quad (48)$$

$$y = \alpha \quad (49)$$

3.5 The Canonical XY model

As we have seen, all models discussed so far can be represented by the same formula:

$$r'(k+1) = x * e(k) + y * e(k-1) + r(k), \quad (50)$$

where x and y are functions of the model parameters (a, b) in Michon, of (α, β) in Mates and Schulze et al., and of (f, g) in Hary and Moore. Let us consider a Canonical XY model, where (x, y) are arbitrary constants. As we saw, the models of Michon, Hary and Moore, and Schulze et al. all make

⁴ Please, note that we believe Eq. 2 in Schulze et al. (2005) to be erroneous. In their indexing scheme, which is different from ours, their term A_{i-1} refers to our $e(k-2)$. We believe they actually meant $A_i = e(k-1)$.

the same prediction and thus are equivalent to the Canonical XY model, after their parameters have been properly mapped. However, while we can map any parameter (α, β) of the Mates model onto a Canonical XY model with parameters (x, y) if $x = -\alpha - \beta$ and $y = \alpha + \beta - \alpha * \beta$, the reverse is *not* true; there are parameters (x, y) of the Canonical XY model such that there is no (α, β) where the Mates model will have the same predictions. This is significant, as it means that the Mates model restricts the possible (x, y) values, meaning that we can use empirical data to check whether the Mates model is indeed correct.

3.6 The Mates model versus other models

We can determine the range of possible (x, y) values in the Mates model by the following calculation of the inverse function $F_{\text{Mates}}^{-1}(x, y)$. The inverse of the function $F_{\text{Mates}}(\alpha, \beta) = (-\alpha - \beta, \alpha + \beta - \alpha * \beta) = (x, y)$ has two possible values:

$$(\alpha, \beta) = \left\{ \left[\begin{array}{l} -x + \sqrt{x^2 + 4x + 4y} \\ -x - \sqrt{x^2 + 4x + 4y} \end{array} \right] / 2 \right\} \tag{51}$$

and

$$(\alpha, \beta) = \left\{ \left[\begin{array}{l} -x - \sqrt{x^2 + 4x + 4y} \\ -x + \sqrt{x^2 + 4x + 4y} \end{array} \right] / 2 \right\}. \tag{52}$$

Proof From Eq. 17, we can deduce that $\beta = -x - \alpha$. Substituting this in Eq. 18, we get

$$y = -x + \alpha * x + \alpha^2$$

or

$$-x - y + \alpha * x + \alpha^2 = 0.$$

This is a quadratic equation in α , and its two solutions are

$$\alpha = \left[-x + / - \sqrt{x^2 + 4x + 4y} \right] / 2. \tag{53}$$

Since $\beta = -\alpha - x$,

$$\beta = \left[-x - / + \sqrt{x^2 + 4x + 4y} \right] / 2, \tag{54}$$

which is the parameter symmetry we have noted earlier. \square

A very important consequence of this calculation is that when $y < -x - x^2/4$ or $(x^2 + 4x + 4y < 0)$ there is no inverse $F_{\text{Mates}}^{-1}(x, y)$. It means that not all pairs (x, y) can be obtained from any α and β . When (x, y) are in the range $y > -x - x^2/4$, there are two solutions (see Eqs. 42, 43). For example, for $x = -0.7$ and $y = 0.58$, the two solutions are $\alpha = 0.4, \beta = 0.3$ and $\alpha = 0.3, \beta = 0.4$. When $y = -x - x^2/4$, there is one solution in which $\alpha = \beta$. Figure 3 shows the two-dimensional space of all the pairs (x, y) . This space is divided by the function $y = -x - x^2/4$ into two distinct regions: A “possible Mates region” ($y \geq -x - x^2/4$)

where (x, y) of the Mates model can be obtained by some α and β , and an “impossible Mates region” ($y < -x - x^2/4$) where (x, y) cannot be obtained by any α and β .

As the Mates model predicts that (x, y) will be in the possible Mates region while the Canonical XY model does not make this prediction, we can empirically distinguish between them. If the Mates model is correct, then the (x, y) that we would get by fitting the Canonical XY model to data would be in the possible Mates region. The importance of this result lies in that we can test empirically if the Mates model is correct. We do this by simply finding the best Canonical XY model for an experimental data set, and seeing if its (x, y) lie in the possible Mates range.

4 Experimental comparison of models

Let us now investigate whether the Mates model is correct by comparing its predictions with empirical data. If the model holds, then we expect that the true response intervals $r(k + 1)$ will be the predicted $r'(k + 1)$, plus some independent noise $n(k)$.⁵ Therefore, we expect:⁶

$$r(k + 1) = (-\alpha - \beta) * e(k) + (\alpha + \beta - \alpha * \beta) * e(k - 1) + r(k) + n(k).$$

If we then fit the Canonical XY model to data and find that $(x, y) = (-\alpha - \beta, \alpha + \beta - \alpha * \beta)$ for some α and β , then these (x, y) are in the possible Mates region.

In practice, we can fit the Canonical XY model to many data sets and plot the resulting (x, y) on a diagram similar to Fig. 3 to see whether the (x, y) represented as points fall inside the possible Mates region. If we get many (x, y) outside the possible Mates region, this will question the validity of the Mates model. We will also compare the Mates and Canonical XY models with regard to their prediction error.⁷

To compare these model predictions, we used data from a synchronization task that employed metronome sequences containing abrupt changes in tempo, so-called step changes (Michon 1967; Repp 2001b; Repp and Keller 2004). These changes were introduced in order to induce updating of the internal timekeeper (i.e., period correction), which is the distinctive component of the Mates model that accounts for its

⁵ Here, we assume that all $n(k)$ are i.i.d. (independent and identically distributed) Gaussian random variables with zero mean. Appendix 2 discusses the more general case where this assumption may not hold.

⁶ Assuming that the noise $n(k)$ is Gaussian, and in order to evaluate empirically the parameters that minimize the prediction error $||r' - r||$, we need to use the formulae given in Appendix 1.

⁷ Since the Canonical XY model is equivalent to the Michon, Hary and Moore, and Schulze et al. models, we are effectively comparing all these models to the Mates model.

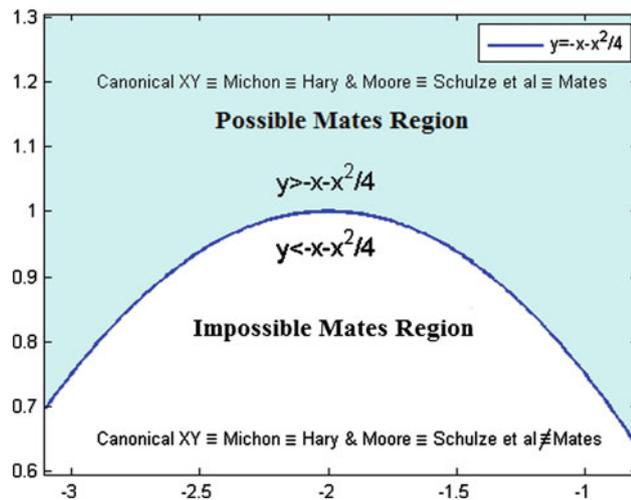


Fig. 3 The two-dimensional Space of (x, y) values, illustrating the boundary on possible parameter values imposed by the Mates model

parameters' nonlinearity⁸. Without such tempo changes, the β parameter would most likely be close to zero, and the Mates model would be indistinguishable from the other models.⁹

4.1 Methods

The participants were all musically trained: they included 8 graduate students of the Yale School of Music (4 men and 4 women, ages 22–26) and author BHR (age 65). The musicians' primary instruments were piano (2), violin (3), viola, cello, and oboe, which they had studied for 13–21 years. Author BHR is a lifelong amateur pianist with 10 years of training in childhood. All were regular participants in synchronization experiments, and the students were paid for their participation.

Auditory sequences (trials) consisted of 99 digital piano tones with a fundamental frequency of 261 Hz (C4) and a nominal duration of 40 ms. The implied baseline inter-onset interval (IOI) was 500 or 800 ms. Each trial alternated between two tempi: fast and slow. Trials varied in the magnitude of the deviations from the baseline IOI, ranging from ± 2 , ± 6 , ± 10 , ± 14 to ± 18 %. For example, a trial with baseline IOI = 800 ms and ± 10 % deviations alternated between actual IOIs of 720 and 880 ms. Each trial started with the faster tempo, followed by about ten step changes (abrupt changes in IOI from one value to the other). Each new step change occurred randomly within 8–12 metronome beats after the previous change. Trials were separated by short pauses determined by the participant. Each participant completed two

⁸ Note that the term “nonlinearity” here refers to the parameters and not to the Mates model.

⁹ Schulze et al. (2005) used gradual tempo changes to compare their model with the Mates model and found smaller prediction errors with their model.

blocks of ten trials each (2 baseline IOIs \times 5 deviation magnitudes). The task was to tap in synchrony with each tone, starting with the third tone in each sequence, and to adjust as quickly as possible to any tempo change.

The tone sequences were generated on-line by a program written in MAX 4.0.9, running on an Intel iMac computer. The tones were produced by a Roland RD-250s digital piano according to musical instrument digital interface (MIDI) instructions from the MAX program and were presented over Sennheiser HD280 pro headphones. Participants tapped on a Roland SPD-6 electronic percussion pad, held on the lap. Finger impacts were audible as thuds, the loudness of which depended on individual tapping force but was attenuated considerably by the circumaural headphones.

4.2 Analysis of the experimental data

Figure 4 shows the first 30 s of a typical trial. Stimulus intervals $s(k)$ (=IOIs) and response intervals $r(k)$ are displayed in the top graph, and asynchronies are displayed in the bottom graph. Also shown here are the predictions of the Mates model and of the Canonical XY model. The parameters for the models were obtained by finding the minimal root mean squared prediction error $\sqrt{(|R' - R|^2)}$. An analytic formula for the fitted parameters of both the Mates and the Canonical XY model is given in Appendix 1. The figure also illustrates the initial “overshoot” in inter-response intervals that is typically observed in response to perceptible step changes (Michon 1967; Repp 2001b).

As another example, Fig. 5 displays the average response (across all participants and trials but only for a step change from 545 to 455 ms) aligned to beat 0, the beginning of the step change. We can see that neither model predicts the observed mean values exactly, but the Canonical XY model does so somewhat better than the Mates model.

Figure 6 show the average response (linearly normalized by setting the step change equal to ± 1 and expressing the response relative to it) across all participants for different step changes around baseline IOIs of 500 and 800 ms, respectively. This graph may be compared with one shown in Repp and Keller (2004, Fig. 3A), although there the data were not normalized. We can see that the amount of initial “overshoot” (Michon 1967) was larger for large than for small step changes, with no overshoot for the smallest changes. This difference has been observed previously (Repp 2001b; Repp and Keller 2004; Thaut et al. 1998a,b) and in the context of the Mates model has been attributed to a dependence of β (period correction) on participants' awareness of the tempo change.

The Canonical XY model states that

$$r'(k+1) = x * e(k) + y * e(k-1) + r(k). \quad (55)$$

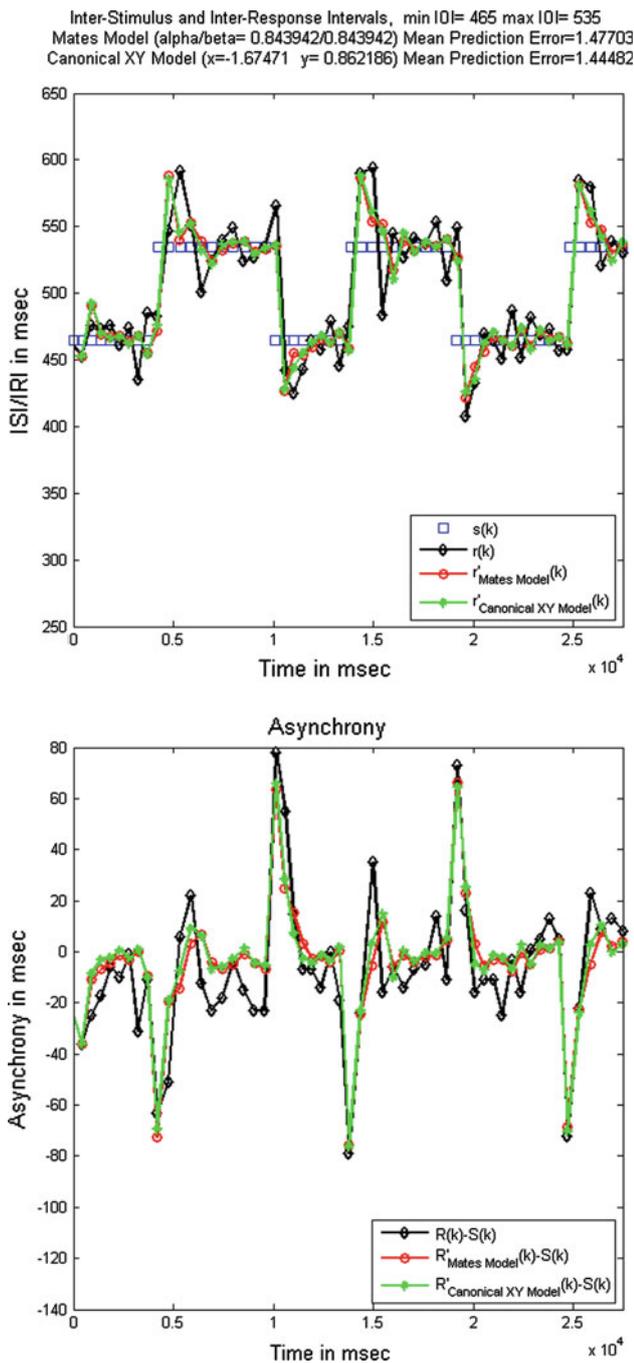


Fig. 4 The first 30s of a typical trial, with best fits of the Mates and Canonical XY models. *Upper panel* Intervals [ISI = inter-stimulus interval = $s(k)$ = IOI; IRI inter-response interval = $r(k)$]. *Lower panel* Asynchronies

Note that if the perturbation (a step change from $s = s_1$ to $s = s_2$) happened at beat k , we expect $r(k) = s_1$, because the change could not be anticipated (Michon 1967). Therefore, the expected $e(k) = s_1 - s_2$. Before the perturbation, subjects are synchronized with the stimulus so that $e(k - 1) \sim 0$. Therefore, at the time of the perturbation

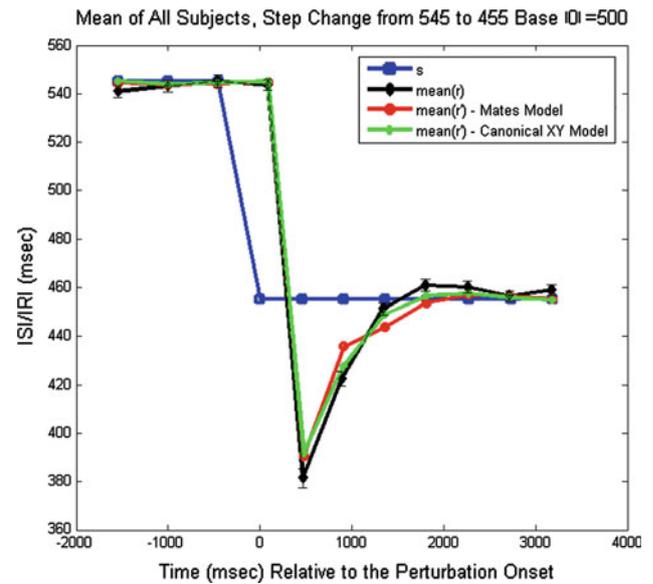


Fig. 5 Mean of all participants' responses to a step change from 545 to 455 ms, with best fits of Mates and Canonical XY models

$$r'(k + 1) = x * (s_1 - s_2) + s_1$$

or

$$\text{overshoot factor} = [r'(k + 1) - s_1] / (s_2 - s_1) = -x. \quad (56)$$

Therefore, $-x$ determines the “overshoot factor”.

Note that the Mates model maps $-\alpha - \beta$ onto x of the Canonical XY model, so we can also say that

$$\text{overshoot factor} = \alpha + \beta, \quad (57)$$

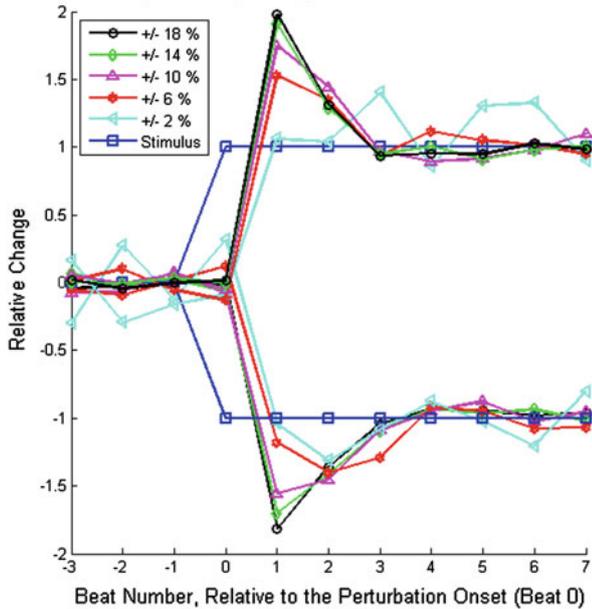
where α and β are the Mates model's parameters. This corresponds to previous assumptions that the overshoot represents the additive effects of phase and period correction (Repp 2001b; Repp and Keller 2004). When $(\alpha + \beta) < 1$, there is no overshoot.

4.3 Experimental evaluation of the Mates model

Were the Mates model “correct,” we would expect to see all pairs of (x, y) obtained by fitting¹⁰ the Canonical XY model to be within the possible Mates region (see Sect. 3.6). Figures 7 and 8 show the best-fitting (x, y) pairs for the Canonical XY model. Each point represents the parameters obtained from a single trial. Out of the 180 trials (9 participants \times 2 base IOIs \times 5 step change magnitudes \times 2 blocks), 59 (32.7%) were located in the impossible Mates region. This challenges the Mates model, which predicts that all (x, y) pairs of the Canonical XY model should lie within the possible Mates region. Of course, the parameter estimation may be subject to error, but in Appendix 2 we demonstrate that the

¹⁰ See Appendix 1 regarding the fitting procedure.

Mean of All Subjects, Step Change +/- 18,14,10,6,2 % BaseIOI = 500



Mean of All Subjects, Step Change +/- 18,14,10,6,2 % BaseIOI = 800

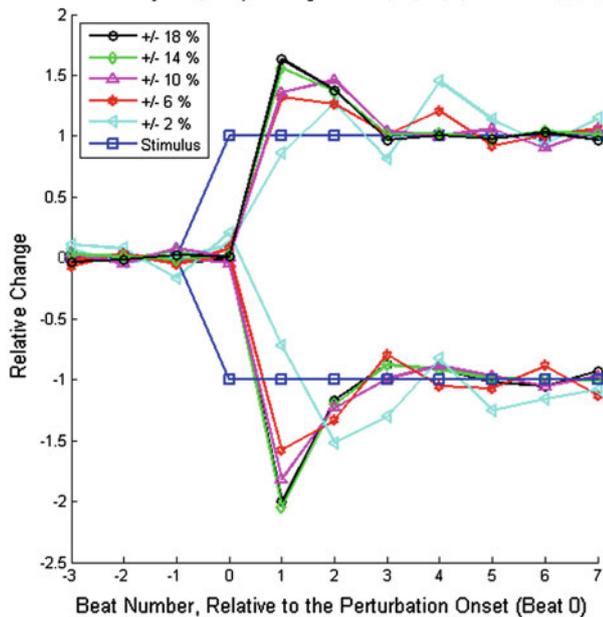


Fig. 6 Mean linearly normalized responses to step changes of different sizes, with base inter-onset intervals (IOIs) of 500 ms (upper panel) and 800 ms (lower panel)

chances that the Mates model parameters are in the impossible Mates region just as a consequence of estimation error is extremely small ($p < 1e-5$).

We can see also different distributions of (x, y) values for the two different base IOIs. For base IOIs of 500 and 800ms, the mean (x, y) values were $(-1.58, 1.00)$ and $(-1.66, 1.11)$, respectively. Both means fall within the possible Mates region, but the first one just barely. The

All Subjects' x and y Values

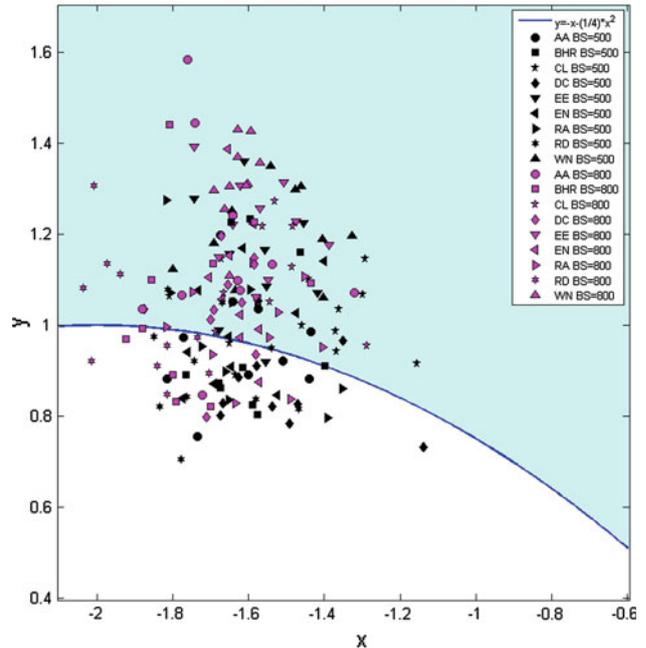


Fig. 7 Mean (x, y) parameter estimates of the Canonical XY model for individual participants (BHR is the second author), separately for each of the two base IOIs (500 and 800 ms)

All Subjects' x and y Values

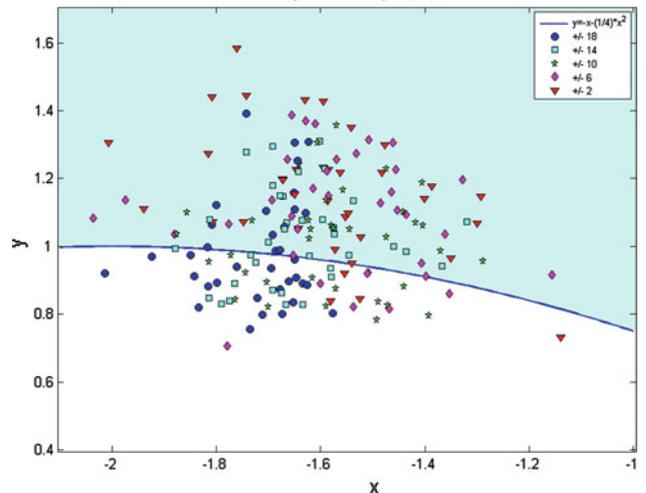


Fig. 8 Mean (x, y) parameter estimates of the Canonical XY model for individual participants, separately for each step size

x and y parameters also depended on the step size of the tempo changes we introduced, with both parameters decreasing as step size increased. We assessed the statistical reliability of these effects in separate 2 (base IOIs) $\times 5$ (step sizes) repeated-measures ANOVAs (with Greenhouse–Geisser correction) on the two parameters. The main effects of base IOI and of step size, were significant for x ($F(1, 8) = 8.62, p = 0.019; F(4, 32) = 6.22, p = 0.004$) and for y ($F(1, 8) = 28.57, p = 0.001; F(4, 32) = 6.88,$

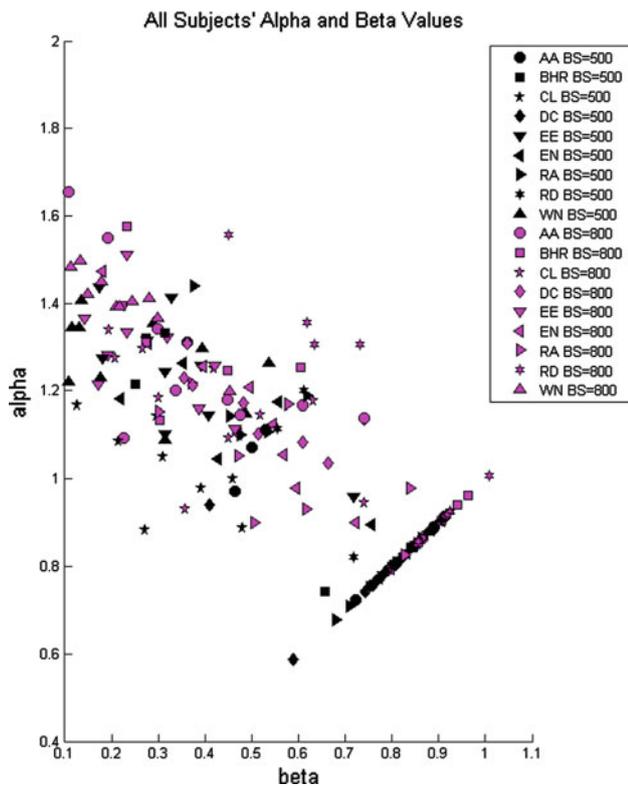


Fig. 9 Mean (α , β) parameter estimates of the Mates model for individual participants, separately for each of the two base IOIs (500 and 800 ms)

$p = 0.013$), whereas the interaction was not significant for either.

In Fig. 9, we can see the parameters obtained for the Mates model. Note that since α and β are interchangeable, we cannot determine which is which, except by external criteria. The points clustered on the $\alpha = \beta$ line are the ones that are in the impossible Mates region in Fig. 7. These points are on this line because the best fit without constraining parameters to the possible Mates region is outside the possible Mates region. Therefore, adding the Mates constraint gives a best fit that is on the boundary of the impossible Mates and possible Mates regions.¹¹ If we optimized the Mates model instead of the Canonical XY model, yet still displayed the results in an (x, y) diagram as in Fig. 7, all the points would have been either in the possible Mates region or on the boundary $y = -x - x^2/4$. This boundary is mapped to the $\alpha = \beta$ line when applying the transformation from the (x, y) plane to the (α, β) plane.

¹¹ A formal proof of this is given in Appendix 1.

4.4 Mates model versus Canonical XY model: evaluation of the prediction error

We return now to our evaluation of the Mates model. Let us consider the prediction error of a model m , $E_m = ||R - R'_m||$. An analytic formula for the fitted parameters of both the Mates and the Canonical XY model is given in Appendix 1. The prediction error of the Canonical XY model must be smaller than or equal to that of the Mates model, since in the parameter estimation process in both cases we minimize E_m , but in the Mates case the optimization process has more constraints, which implies $E_{\text{Mates}} \geq E_{\text{Canonical } XY}$. (See Appendix 1 for proof.) This does not necessarily mean that the Canonical XY model prediction is actually better, as it could be a consequence of a phenomenon called “overfitting” (Hastie et al. 2001; Vapnik 1998). Overfitting occurs if the model actually captures noise in the data, in addition to the systematic variation of interest. If the data set used for estimating the prediction error (the test data set) is different from the one used to estimate the parameters (the training data set), we obtain a better estimate of the “true” prediction error (or the “generalization error” of Vapnik 1998). As this is a common practice in the statistical learning literature (Hastie et al. 2001), we adopted it here also. We trained the model on one half of the data and then applied the estimated parameter values to the other half. This procedure could be done in our case, since we had two identical blocks of trials for every participant, and therefore we used the first block as training set and the second block as a test set.

Mathematically speaking, the Canonical XY model will always yield a smaller prediction error than the Mates model on the training data set. However, this may not apply to the test set. In theory, the prediction error could be smaller for the Mates model than for the Canonical XY model in the test set. In fact, the percentage of the test set trials where the Mates model had a smaller prediction error (18%) was almost the same as the percentage of trials where the Canonical XY model had the smaller error (16%)¹².

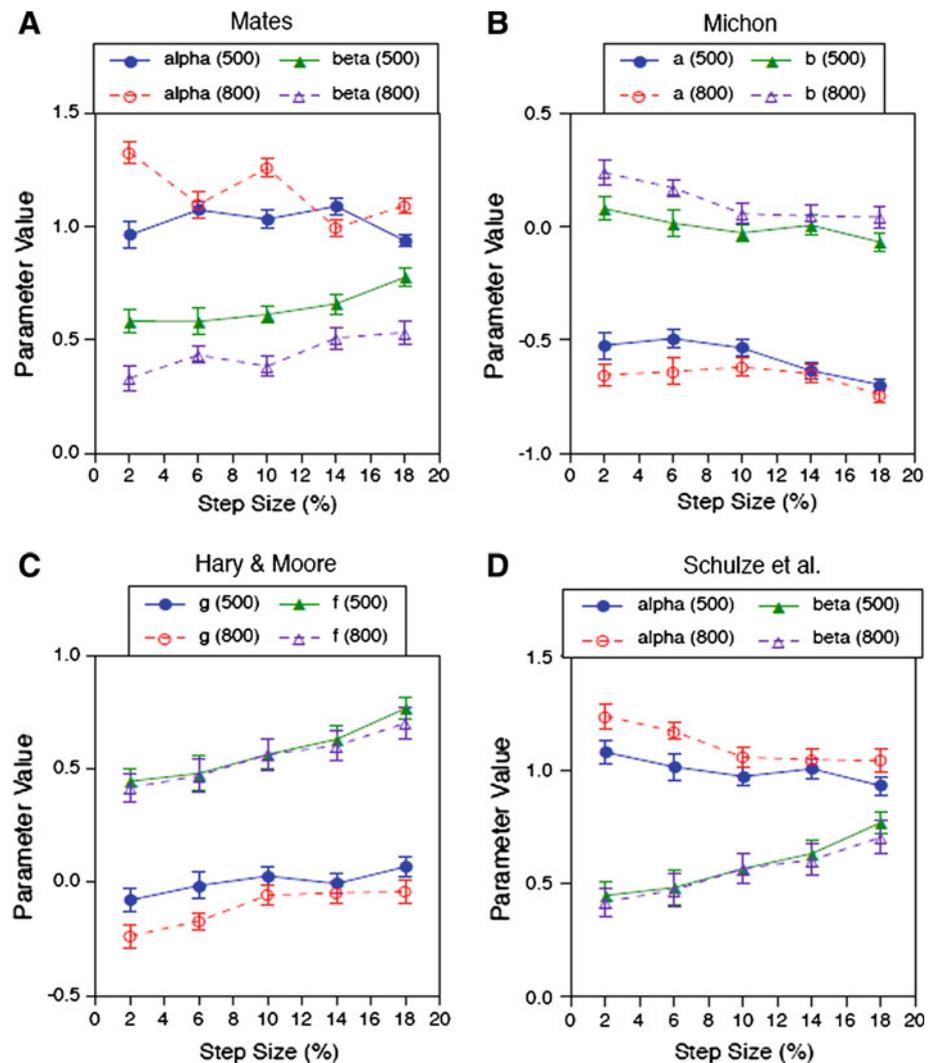
For the remaining trials, the two models made the same predictions because the (x, y) parameters were in the possible Mates region. We also trained the models on the trials from the second block and tested them on the trials from the first block. This led to very similar results.

4.5 Parameter estimates for all models

In Fig. 10, we display the mean parameter values of all four models from the literature as a function of base IOI and step

¹² Note that the percentage of trials where the two models differ is determined by the training set (since we estimate the model parameters from the training set, and if they provide an equivalent solution on the training set they will be equivalent also on the test set).

Fig. 10 Mean parameter values of four models as a function of IOI and step size



size. They were subjected to two-way ANOVAs like the one on the parameters of the Canonical XY model. The parameter values were derived from the best-fitting parameters of the Canonical XY model by simple linear transformations, as described in Sect. 3, except in the case of the Mates model, for which the detailed parameter values are those shown in Figs. 7 and 8.

Although α and β are interchangeable in the Mates model, reference to earlier empirical findings (e.g., Repp and Keller 2004) allows us to interpret the larger parameter as α (phase correction) and the smaller one as β (period correction). Both parameters had positive values, as expected. The results (Fig. 10a) indicate that α was surprisingly large and decreased as step size increased, $F(4, 32) = 5.73$, $p = 0.022$, an unexpected finding. It was also larger at the slower base tempo, $F(1, 8) = 26.93$, $p = 0.001$, which is expected (cf. Repp 2008). By contrast, β increased with step size, $F(4, 32) = 10.85$, $p < 0.001$, as expected (cf. Repp 2001b), but was smaller at the slower base tempo, $F(1, 8) = 8.29$, $p = 0.021$. The large percentage of estimates in the impos-

sible Mates region, which forced $\alpha = \beta$, may have distorted these trends. We examined the data for three participants who had very few values in the impossible Mates region. Their Mates parameters were almost identical to their Schulze et al. parameters (to be discussed shortly), as they should be.

The Michon parameters (Fig. 10b) were similar to those of the Canonical XY model because $a = 1 + x$ and $b = y - 1$. Therefore, the trends and statistical results were the same. However, the parameters can be interpreted more easily within the Michon model. Their values indicate that the timing of the next tap depended primarily (and negatively) on the preceding asynchrony (weight a) and only little (but positively) on the asynchrony one position back (weight b). This asymmetry increased with step size. Surprisingly, however, it decreased as the base IOI increased.

The Hary and Moore parameters (Fig. 10c) show a somewhat different pattern. The g parameter is straightforwardly related to y because $g = 1 - y$, but $f = -x - y$. The former increased with step size and was smaller at the slower base tempo (statistics as for y), whereas the latter clearly

increased with step size, $F(4, 32) = 11.58$, $p < 0.001$, but did not depend on base IOI. The results seem puzzling initially because g is a probability and thus should have only positive values. Instead, its values tended to be negative, suggesting at the very least that phase resetting was entirely stimulus based. Indeed, the negative values of g may indicate overshoot in stimulus-based phase resetting, equivalent to a phase correction parameter greater than 1. The f parameter represents period correction, and both its mean values and its increase with step size are as expected.

Finally, the parameters of the Schulze et al. model (Fig. 10d) are readily interpretable as phase correction ($\alpha = y$) and period correction ($\beta = -x - y$), respectively. The β parameter is equal to f in the Hary and Moore model, whereas $\alpha = 1 - g$ and thus reveals phase overcorrection at small step sizes and at the slower base tempo. The statistics are the same as for the Hary and Moore model. The Schulze et al. parameters make perfect sense, and comparison with the Mates parameters (Fig. 10a) reveals the distortions inflicted on the latter by the (x, y) estimates in the impossible Mates region, which forced $\alpha = \beta$.

4.6 PCA analysis of the coordinate systems best describing the empirical data

If we only go by the model predictions, the Michon, Hary and Moore, and Schulze et al. models are all equivalent because the two parameters that describe them can be linearly changed one to another. Yet, it is reasonable to assume that if the two parameters reflect independent psychological processes (phase and period correction in the Schulze et al. case and phase resetting and internal timekeeper update in the Hary and Moore case) then the two parameters would be statistically uncorrelated.

A well-established method to find the best uncorrelated mapping or best orthogonal coordinate system is Principal Components Analysis or PCA (see Hastie et al. 2001). PCA finds two orthogonal projections P_1 , P_2 such that the variance of the data mapped by P_1 is maximal and the variance of the data mapped by P_2 is maximal (subject to the constraint that P_2 is orthogonal to P_1). The resulting coordinate system $P_1(x, y)$, $P_2(x, y)$ is often considered as a good (canonical) representation of the data.

We performed a PCA analysis on the (x, y) pairs for all participants and trials and obtained the projection $P_1(x, y) = -0.12x + 0.99y$, $P_2(x, y) = 0.99x + 0.12y$. As we can clearly see, the coordinate system that maps (x, y) pairs to independent components is similar to the XY coordinate system (the angular difference is only about 7°), and far from the one of the Schulze et al. or Hary and Moore models.

We also calculated the correlation coefficient of the parameter estimates of the different models. As expected from the PCA analysis, the correlation coefficient of the two variables

was small in the case of the Canonical XY model and the Michon model (-0.02), but much larger and negative in the cases of Hary and Moore (-0.66) and Schulze et al. (-0.73). The largest correlation coefficient was obtained for the Mates model (-0.80).

To conclude, the XY model is an elegant description of the experimental data since its parameters are (almost) uncorrelated. However, the model suffers from the fact that it has no direct psychological interpretation. The models of Hary and Moore and Schulze et al. have parameters estimates that are more correlated, but they lend themselves more readily to a psychological interpretation.

5 Discussion

We began by presenting a linear framework in which the prediction of the next inter-response interval is a linear function of previous inter-stimulus intervals, inter-response intervals, and recent asynchronies, and embedded four models from the literature (Mates, Michon, Hary and Moore, Schulze et al.) in it. We further suggested a Canonical XY model that is also embedded in this framework. We then defined a notion of equivalence, and showed that except for the Mates model, all the models are equivalent. The Mates model made a stronger prediction, one that could be tested experimentally.

Taking this as our departure point, we analyzed experimental data, and saw that the Mates model prediction was not validated for about a third of the trials. We also saw that the Mates model prediction was not significantly better than that of the Canonical XY model when the models were compared using a “fair” training and testing scheme.

Appendix 2 further generalizes these results for models including the original stochastic (and possibly correlated) noise terms that we had omitted for simplicity. It presents a method for computing an unbiased estimate of the model parameters in the presence of correlated noise and also evaluates the estimation error. The results of the Appendix further support the conclusion that the Mates model assumptions are too restrictive by showing that neither correlated noise nor estimation error can explain the violation of the Mates model by the empirical data.

The Mates model differs from the other models only because of its assumption that period correction is based on a comparison of stimulus and response intervals. Therefore, our findings suggest that this specific assumption is incorrect, and that the alternative assumption of Schulze et al. that period correction is based on perception of asynchronies is to be preferred, even though it results in a model that is equivalent to the Michon and Hary & Moore models.

The agreement of these three models may suggest that all these researchers, perhaps without knowing it, were on

the right path. However, these equivalent models also have weaknesses:

- (1) They are limited in their ability to predict results when experimental conditions are changed. The model parameters can be adjusted empirically to obtain a best fit to the experimental data, but the models cannot predict effects of step size or of base tempo. Accounting for such effects will require additional variables and parameters.
- (2) Although the models originate from very different intuitions, they produce equivalent empirical predictions. This suggests that the intuitive or theory-based approach to model building may not be the most efficient one, and that a more systematic approach within a general linear framework, such as taken here, might lead to faster progress. The Canonical XY model can easily be augmented by adding variables from farther back in the time series (see Eq. 5), and such expanded models may be able to account for additional systematic variance in the data.

However, the general linear framework is not a panacea. Adding variables from farther back in the time series can increase predictive power, but if the number of parameters gets too large, the model will begin to explain noise in the data. Moreover, such models still cannot account for the just-mentioned effects of step size or base tempo. Prediction of these effects needs to rely on findings from independent empirical research on the functional relationships between these variables and model parameters. Furthermore, while a non-intuitive, purely statistical approach to modeling offers great freedom, it has the drawback that the resulting models cannot easily be related back to theories about how the human mind operates. Thus, they may contain parameters that have no psychological interpretation or that are implausible considering what is known about human perception and memory. To prevent the modeling effort from being merely an algebraic exercise, we need to remain in contact with psychological theories and strive to interpret the parameters in a meaningful way.

Finally, we need to acknowledge that linear models are limited by their very linearity, and that nonlinear dynamic models may offer an even more general and powerful approach than even the most general linear model. Regimes of linear control are often nested locally within a broader nonlinear landscape (cf. Pressing 1999), and the approach we have pursued here may be limited to experimental situations in which regularity is high, changes are small, and sensorimotor coordination is simple. Future model comparisons should therefore include not only expanded linear models but also nonlinear models, and should assess their ability to

account for performance in coordination tasks that go beyond simple one-to-one synchronization. (For a recent step in that direction, see Loehr et al. 2011.)

Acknowledgements This research was supported in part by National Science Foundation grant BCS-0924206 to BHR. NJ is grateful to Carmel Raz for her help with the English text.

Appendix 1

Finding the best parameters for the Canonical XY model

In this section, we explain how to find the “best” parameters—the parameters that minimize the squared prediction error.

Let n be the total number of metronome beats in the experiment.

Let \mathbf{A} be the $n - 2$ by 2 matrix:

$$\begin{aligned} \mathbf{A}(k, 1) &= e(k + 1) (k = 1 \dots n - 2) \\ \mathbf{A}(k, 2) &= e(k) (k = 1 \dots n - 2) \end{aligned} \quad (58)$$

Let \mathbf{b} , \mathbf{c} , and \mathbf{d} be the column vectors of length $n - 2$:

$$\begin{aligned} \mathbf{b}(k) &= r(k + 2) (k = 1 \dots n - 2) \\ \mathbf{c}(k) &= r(k + 1) (k = 1 \dots n - 2) \\ \mathbf{d}(k) &= r'(k + 2) (k = 1 \dots n - 2) \end{aligned} \quad (59)$$

Let \mathbf{w} be the column vector of length 2:

$$\begin{aligned} \mathbf{w}(1) &= x \\ \mathbf{w}(2) &= y \end{aligned} \quad (60)$$

Lemma

$$\mathbf{A} * \mathbf{w} + \mathbf{c} = \mathbf{d} \quad (61)$$

(“*” here stands for matrix multiplication).

Proof

$$\begin{aligned} \mathbf{A} * \mathbf{w} + \mathbf{c} &= \mathbf{A}(k, 1) * x + \mathbf{A}(k, 2) * y + \mathbf{c}(k) \\ &= e(k + 1) * x + e(k) * y + r(k + 1) \\ &= r'(k + 2) = \mathbf{d}(k) \end{aligned}$$

(See Eq. 50)

Lemma The mean squared prediction error is $\|\mathbf{A} * \mathbf{w} + \mathbf{c} - \mathbf{b}\|^2 / (n - 2)$, where $\|x\|$ is the Euclidean norm of the vector x .

Proof From the last lemma

$$\begin{aligned} &\|\mathbf{A} * \mathbf{w} + \mathbf{c} - \mathbf{b}\|^2 \\ &= \|\mathbf{d} - \mathbf{b}\|^2 = \sum_{k=1 \dots n-2} [r'(k + 2) - r(k + 2)]^2 \\ &= \sum_{k=3 \dots n} [R'(k) - R(k)]^2 \end{aligned}$$

We assume that the following initial conditions hold: $R'(1) = R(1)$ and $R'(2) = R(2)$.

Therefore, given that $\mathbf{w} = (x, y)$, minimizing the mean squared error means minimizing $\|\mathbf{A} * (x, y) - (\mathbf{b} - \mathbf{c})\|^2$ over all the possible (x, y) . This has the well-known least-mean-square solution (see Rao and Toutenburg 1999):

$$(x, y)^T = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T (\mathbf{b} - \mathbf{c}), \tag{62}$$

where \mathbf{A}^T is the transpose of \mathbf{A} , and $(\mathbf{A}^T \mathbf{A})^{-1}$ is the inverse of $(\mathbf{A}^T \mathbf{A})$. Equation (62) is a ‘‘closed’’ formula, one that enables us to calculate the ‘‘best’’ (x, y) of the Canonical XY model directly, without needing an iterative algorithm.

Closed formula for the best Mates model parameters

In this section we will obtain a ‘‘closed’’ formula for the best Mates model coefficients, similar to Eq. 62 for the case of the Canonical XY model. In order to find the best Mates model coefficients, we start by finding the best (x, y) for the Canonical XY model. If we are in the possible Mates region, where $y \geq -x - (x^2)/4$, then the pair (x, y) can be translated into a pair (α, β) of Mates parameters (see Eqs. 53, 54):

$$(\alpha, \beta) = \left\{ \left[\begin{array}{l} -x + \sqrt{x^2 + 4x + 4y} \\ -x - \sqrt{x^2 + 4x + 4y} \end{array} \right] / 2 \right\} \tag{63}$$

If $y < -x - (x^2)/4$, then we are in the impossible Mates region. Now we need to find (x', y') in the *possible* Mates region that minimize the mean squared error. So the best Mates model parameters are the (x', y') which minimize $\|\mathbf{A} * (x', y')^T - (\mathbf{b} - \mathbf{c})\|^2$ subject to the constraint $y' \geq -x' - (x'^2)/4$. From the convexity of minimized norm (see Rockafellar 1970), we know that in this case the optimal solution is obtained at the boundary of the possible Mates and impossible Mates regions, where $y = -x - (x^2)/4$ or $x + y + (x^2)/4 = 0$. Therefore, (x', y') is the solution to the following optimization problem:

Minimize $f(x, y)$ such that $g(x, y) = 0$;

$$f(x, y) = \left[\mathbf{A} * (x, y)^T - (\mathbf{b} - \mathbf{c}) \right]^T \times \left[\mathbf{A} * (x, y)^T - (\mathbf{b} - \mathbf{c}) \right]$$

$$g(x, y) = x + y + (x^2)/4$$

This problem can be solved using the Lagrange multipliers technique (see Rudin 1987). The solution is obtained when $\nabla f = -\lambda \nabla g$ (and $g = 0$).

λ is the Lagrange multiplier, and ∇ is the gradient operator $\nabla = (\partial x_1, \partial x_2)$.

Now

$$\nabla f = 2(\mathbf{A})^T (\mathbf{A}(x, y)^T - (\mathbf{b} - \mathbf{c})) \tag{64}$$

$$\nabla g = (1 + x/2, 1)^T \tag{65}$$

Therefore

$$\begin{aligned} \nabla f &= -\lambda \nabla g \rightarrow 2(\mathbf{A}^T \mathbf{A})(x, y)^T - 2(\mathbf{A}^T (\mathbf{b} - \mathbf{c})) \\ &= (-\lambda(1 + x/2), -\lambda)^T \end{aligned}$$

Now since \mathbf{A} is an $(n - 2)$ by 2 matrix, $\mathbf{A}^T \mathbf{A}$ is a 2 by 2 matrix. Let \mathbf{B} be this matrix ($\mathbf{B} = \mathbf{A}^T \mathbf{A}$).

We denote by \mathbf{C} the 2 by 1 vector $\mathbf{C} = \mathbf{A}^T (\mathbf{b} - \mathbf{c})$.

$$\mathbf{B}_{ij} = \sum_{k=1..n-2} e(2 + k - i)e(2 + k - j) \tag{66}$$

$$\mathbf{C}_i = \sum_{k=1..n-2} e(2 + k - i)[r(k + 2) - r(k + 1)] \tag{67}$$

Now we can write $\nabla f = -\lambda \nabla g \rightarrow$

$$2(\mathbf{B}_{11}x + \mathbf{B}_{12}y - \mathbf{C}_1) = -\lambda(1 + x/2) \tag{68}$$

$$2(\mathbf{B}_{21}x + \mathbf{B}_{22}y - \mathbf{C}_2) = -\lambda$$

We can compute λ from the second equation, and by substituting it in the first equation we get

$$\mathbf{B}_{11}x + \mathbf{B}_{12}y - \mathbf{C}_1 = (\mathbf{B}_{21}x + \mathbf{B}_{22}y - \mathbf{C}_2)(1 + x/2) \tag{69}$$

or

$$\mathbf{B}_{11}x + \mathbf{B}_{12}y - \mathbf{C}_1 + (\mathbf{C}_2 - \mathbf{B}_{21}x - \mathbf{B}_{22}y)(1 + x/2) = 0 \tag{70}$$

Let us recall that $y = -x - x^2/4$.

Substituting in the last equation leads us to a cubic equation that can be solved analytically or numerically

$$\begin{aligned} &(\mathbf{B}_{22}/8)x^3 + (-\mathbf{B}_{12}/4 - \mathbf{B}_{21}/2 + 3\mathbf{B}_{22}/4)x^2 \\ &+ (-\mathbf{B}_{12} - \mathbf{B}_{21} + \mathbf{B}_{11} + \mathbf{C}_2/2 + \mathbf{B}_{22})x + \mathbf{C}_2 - \mathbf{C}_1 \end{aligned} \tag{71}$$

This equation has at least one real root. Thus we know that a solution to the original problem exists. Let x be a real root of this cubic equation; then $(x, y) = (x, -x - x^2/4)$ is a solution to the original problem.

Note that in this way we can find the best (x, y) for the Mates model directly without using an iterative algorithm. Comparing the results by computer simulation, we obtained the same (x, y) (up to numeric precision differences), but the speed of the iterative algorithm was slower by a factor of over 100.

Appendix 2

The purpose of this appendix is to deal with two issues that were not considered so far:

- (a) The stochastic part of the linear models discussed, and especially correlations in the noise terms appearing in the non-deterministic versions of the models.
- (b) The estimation error of the model parameters.

One of this article’s main findings is that the Mates model is less successful than the Canonical XY model, since a substantial percentage of the estimated parameters are empirically found to be in the impossible Mates region, contrary to the Mates model’s assumptions. Here we examine whether either of two sources of error might have caused these parameter estimates (but not the “true” parameters) to be in the impossible Mates region.

Another issue that remained unspecified until now is the exact link between intensive and extensive variables, and the stochastic structure of the residual noise $Z(k)$ of Eq. 5.

The residual noise term $Z(k)$ reflects the difference between the intensive variable $r'(k)$ and the extensive variable $r(k)$ [in Eq. 5 $r'(k + 1) = r(k + 1) - Z(k + 1)$].

In the section “Calculation of $Z(k)$ in the case of the Schulze et al. model” of this appendix, we will show the specific correlation structure implied by the Schulze et al. model. Similar correlation structures can be found in all other models, and a general algorithm to find model parameters with Gaussian and possibly correlated $Z(k)$ is presented in the section “The bGLS method” of this appendix.¹³

In order to deal with the concern discussed above, we therefore present a variant of a well-known method (generalized least squares, GLS) that provides an unbiased estimate of the model parameters, even in the presence of correlated noise. This method is much simpler than the estimation methods used by previous authors (Pressing 1998a,b; Repp and Keller 2008; Vorberg and Schulze 2002), yet we prove that it indeed provides an unbiased estimate of the model parameters. As we will see, the method will also provide us with an estimate of the covariance of the estimation error. Based on this, we can calculate the probability that the estimation error caused the parameters to fall inside the impossible Mates region.

The bGLS method¹⁴

Until now we ignored the stochastic part of the four models discussed.

¹³ Note that the process of presenting an ARMA model with intensive parameters (“state-space model”) as an equivalent model with only extensive parameters is well established in the literature (see Ljung 1987).

¹⁴ The method is also described briefly in another recent paper (Repp et al. 2012).

All four models can be written as

$$r(k + 1) = x * e(k) + y * e(k - 1) + r(k) + Z(k + 1) \tag{72}$$

where $Z(k + 1)$ is a Gaussian random variable with short autocorrelation structure:

$$\gamma_z(j) = 0 \text{ for } j \geq K \tag{73}$$

where $\gamma_z(j)$ is the autocorrelation function: $\gamma_z(j) = \text{cov}[Z(k), Z(k + j)]$, and K is some small integer (1, 2, or 3).

The key difference between Eqs. 72 and 50 is that we have replaced $r'(k)$ with $r(k)$. We now demonstrate how to find the parameters (x, y) without making this substitution.

We recall that $r'(k) = r(k) - Z(k)$, and following Appendix 1 Eq. 61 we can write Eq. 72 as

$$\mathbf{A} * \mathbf{w} + \mathbf{Z} = \mathbf{B} \tag{74}$$

where \mathbf{A} is the matrix \mathbf{A} of Appendix 1, \mathbf{B} is a column vector: $\mathbf{B}(k) = r(k + 2) - r(k + 1)$, and $\mathbf{Z}(k)$ is the prediction error vector: $\mathbf{Z} = [Z(3), Z(4), \dots Z(n)]^T$. Note that in computing \mathbf{A} (eq. 58) we replace $e(k)$ by $e(k) - \text{mean}(e)$ and $e(k + 1)$ by $e(k + 1) - \text{mean}(e)$, where $\text{mean}(e)$ is the empirical mean of the asynchrony. In this way we can assume without loss of generality that \mathbf{Z} has zero mean.

Assuming that \mathbf{Z} is Gaussian with covariance matrix Σ , the likelihood of \mathbf{Z} given \mathbf{A} , \mathbf{B} , and \mathbf{w} is

$$\text{Prob}(\mathbf{Z} = \mathbf{B} - \mathbf{A}\mathbf{w} | \mathbf{B}, \mathbf{A}, \Sigma, \mathbf{w}) = \frac{1}{(2\pi)^{N/2}} (|\Sigma|)^{-1/2} \exp(-1/2 (\mathbf{B} - \mathbf{A}\mathbf{w})^T (\Sigma)^{-1} (\mathbf{B} - \mathbf{A}\mathbf{w})) \tag{75}$$

If $\mathbf{Z}(3), \mathbf{Z}(4), \dots \mathbf{Z}(n)$ are i.i.d. and Gaussian with zero mean (Σ is diagonal), then the linear regression of the parameters, as described in Appendix 1 (Eq. 62), that minimizes the mean squared error also maximizes the likelihood of Eq. 75, and therefore provides an unbiased estimate of the model parameters. This is a well-known result (see Rao and Toutenburg 1999). However, we need to consider that $Z(k)$ might not be i.i.d.

In case \mathbf{Z} is correlated and has a known covariance matrix Σ , then an unbiased estimate of the model parameters is also given by the maximal likelihood estimator and is called GLS (see Aitken 1935):

$$(x, y)^T = \mathbf{w}^T = (\mathbf{A}^T \Sigma^{-1} \mathbf{A})^{-1} (\mathbf{A}^T \Sigma^{-1}) \mathbf{B} \tag{76}$$

In our case, the covariance matrix is unknown but has a special form:

$$\Sigma = \gamma_z(0)\mathbf{I} + \gamma_z(1) \begin{bmatrix} \Delta_1 + \Delta_{-1} \\ \Delta_2 + \Delta_{-2} \\ \dots \\ \Delta_{K-1} + \Delta_{-(K-1)} \end{bmatrix} + \gamma_z(2) \begin{bmatrix} \Delta_2 + \Delta_{-2} \\ \dots \\ \Delta_{K-1} + \Delta_{-(K-1)} \end{bmatrix} + \dots + \gamma_z(K - 1) \begin{bmatrix} \Delta_{K-1} + \Delta_{-(K-1)} \end{bmatrix} \tag{77}$$

where \mathbf{I} is the identity matrix, Δ_k is a matrix with 1 on the k th diagonal, and $\gamma_z(k)$ is the autocorrelation function of Z : $\gamma_z(k) = \text{cov}(Z(n), Z(n+k))$, where $\gamma_z(j) = 0$ for $j \geq K$.

Since in our case $\gamma_z(k)$ are not known, an iterative algorithm can be applied, called sometimes Feasible Generalized Least Squares (see Ljung 1987). In each iteration, Σ is estimated by calculating the empirical autocovariance function of the residual noise. Later, using the final estimate of Σ , an estimate of alpha is computed based on Eq. 76.

We applied this algorithm and compared the results with the ones obtained by simple linear regression. Numerical stability is improved if we “bound” the parameter covariance. This leads to a variant of the GLS method that we call “bounded GLS method.” Vorberg and Schulze (2002) mentioned that the parameter estimates may suffer from “interdependence,” a reduced numerical stability of the estimate, and this actually relates to the lower bound on an unbiased estimator (Cramér–Rao bound see Kay 1993). To resolve this issue, Vorberg and Schulze suggested placing some restriction on the parameter range. Following this suggestion, in each iteration we apply certain constraints on the covariance matrix by adjusting the moments. Note that we also assume that all large autocovariances vanish ($\gamma_z(j) = 0$ for $j \geq K$).

Algorithm (bounded generalized least squares method)

The input of this algorithm is the matrices \mathbf{A} and \mathbf{B} , and a certain constraint on the autocovariance matrix¹⁵. The output of the algorithm is the estimate for the parameters: x^{bGLS} , y^{bGLS} .

- a. Start by setting $\Sigma_1 = \mathbf{I}$ (the identity matrix).
- b. Iterate the following steps until a stop criterion is obtained¹⁶.
 - (i) Compute an estimate: $(x_n, y_n)^T = (\mathbf{A}^T \Sigma_n^{-1} \mathbf{A})^{-1} (\mathbf{A}^T \Sigma_n^{-1} \mathbf{B})$.
 - (ii) Compute the residual noise $D_n = \mathbf{B} - \mathbf{A}(x_n, y_n)^T$.
 - (iii) Estimate $\gamma^{n+1}(0), \gamma^{n+1}(1), \dots, \gamma^{n+1}(K-1)$ by the autocovariance of the residual noise:

$$\gamma^{n+1}(j) = \gamma_{D_n}(j).$$
 - (iv) For all $j < K$ adjust $\gamma^{n+1}(j)$ by increasing or decreasing it so that it meets the constraints.

¹⁵ In this paper we used $-5/8 < \gamma_z(1)/\gamma_z(0) < -1/2$ and $0 < \gamma_z(2)/\gamma_z(0) < 1/8$. We will soon explain why these specific constraints were selected.

¹⁶ In this paper, we used a fixed number of 20 iterations and among all solutions we took the one that maximized the likelihood $(1/(2\pi)^{N/2}) (|\Sigma|)^{-1/2} \exp(-(1/2)(\mathbf{B} - \mathbf{A}\mathbf{w})^T (\Sigma)^{-1} (\mathbf{B} - \mathbf{A}\mathbf{w}))$ and yielded (x, y) estimates in the “reasonable range” $-0.8 < x < -2.3$ and $0.5 < y < 2.8$.

- (v) Compute $\Sigma_{n+1} = \gamma^{n+1}(0)\mathbf{I} + \gamma^{n+1}(1)[\Delta_1 + \Delta_{-1}] + \gamma^{n+1}(2)[\Delta_2 + \Delta_{-2}] + \dots + \gamma^{n+1}(K-1)[\Delta_{K-1} + \Delta_{-(K-1)}]$.

Calculation of $Z(k)$ in the case of the Schulze et al. model

We will show an implicit formula for $Z(k)$ in the case of the Schulze et al. model, though similar formulae could be obtained also for the other models. A full discussion of correlation structure in different models is outside the scope of this paper, but the following calculation supports our claim that the main results of this paper hold in the presence of correlated noise.

Note that from the original paper of Schulze et al. (2005), Eq. (1) (p. 463) can be written as

$$e(k) = (1 - \alpha)e(k-1) + t_1(k) - s(k) + T(k) + M(k) - M(k-1) \tag{78}$$

or

$$r(k) = t_1(k) - \alpha * e(k-1) + T(k) + M(k) - M(k-1) \tag{79}$$

where $T(k)$, $M(k)$, and $M(k-1)$ are Gaussian noises reflecting the motor and internal time keeper noises, and $t_1(k)$ is τ_k of Schultze et al. (the mean of the noise term). The standard deviations of T and M are (α, β) , respectively.

So instead of our Eq. 41

$$r'(k+1) = t_1(k+1) - \alpha * e(k), \tag{80}$$

we can write

$$r(k+1) = t_1(k+1) - \alpha * e(k) + h(k) \tag{81}$$

where

$$h(k) = T(k) + M(k) - M(k-1). \tag{82}$$

Note that repeating the calculations in Eqs. 42–47 without “replacing” $r'(k)$ and $r(k)$ yields:

$$r(k+1) = t_1(k+1) - \alpha * e(k) + h(k) \tag{83}$$

$$t_1(k) = t_1(k-1) - \beta * e(k-1), \tag{84}$$

and then

$$r(k) = t_1(k) - \alpha * e(k-1) + h(k-1) \tag{85}$$

$$t_1(k) = r(k) + \alpha * e(k-1) - h(k-1) \tag{86}$$

$$t_1(k+1) = t_1(k) - \beta * e(k). \tag{87}$$

Therefore

$$r(k+1) = r(k) + \alpha * e(k-1) - h(k-1) - \beta * e(k) - \alpha * e(k) + h(k) \rightarrow \tag{88}$$

$$r(k+1) = (-\alpha - \beta) * e(k) + \alpha * e(k-1) + r(k) + h(k) - h(k-1). \tag{89}$$

Compare that with Eq. 47: $r(k+1) = (-\alpha - \beta) * e(k) + \alpha * e(k-1)$.

The difference is in the new noise term:

$$z(k) = h(k) - h(k-1) = T(k) - T(k-1) + M(k) - 2M(k-1) + M(k-2). \quad (90)$$

Interestingly, this term is in general a Gaussian noise with three non-trivial autocovariances (and not two like in the [Vorberg and Schulze 2002](#) model) since:

$$\gamma_z(0) = \text{Var}(z(k)) = \text{Cov}(z(k), z(k)) = 2\sigma_T^2 + 6\sigma_M^2 \quad (91)$$

$$\gamma_z(1) = \text{Cov}(z(k), z(k+1)) = -\sigma_T^2 - 4\sigma_M^2 \quad (92)$$

$$\gamma_z(2) = \text{Cov}(z(k), z(k+2)) = \sigma_M^2 \quad (93)$$

$$\gamma_z(k) = 0 \quad \text{for } k > 2. \quad (94)$$

Now we need to see that we can estimate the model parameters

$$r(k+1) = (-\alpha - \beta) * e(k) + \alpha * e(k-1) + r(k) + z(k) \quad (95)$$

or

$$r(k+1) = (-\alpha - \beta) * e(k) + \alpha * e(k-1) + r(k) + T(k) - T(k-1) + M(k) - 2M(k-1) + M(k-2). \quad (96)$$

Algorithm 1 described here provides the solution known as Feasible Least Squares or GLS, dating back to [Aitken \(1935\)](#). To get improved numerical stability, we need to specify restrictions on the covariance matrix. A reasonable assumption that has often been used in the literature (see for example [Repp and Keller 2008](#)) is to assume

$$\sigma_T^2 > \sigma_M^2 > 0. \quad (97)$$

This results in $-5/8 < \gamma_h(1)/\gamma_h(0) < -1/2$ and $0 < \gamma_h(2)/\gamma_h(0) < 1/8$ (since σ_M^2 is positive), which limits the amount of correlation possible within this model and decreases the estimation variance. We created a simulation with $x = -1.8$, $y = 1.6$, $\sigma_T^2 = 49$, $\sigma_M^2 = 25$ using Eqs. 95–96 and a typical stimulus of our experiment. (The two alternating intervals were 465 and 535 ms.) For each simulated trial, we ran algorithm 1 to find the bGLS estimates. The mean estimates of 1,000 of those simulations were -1.79 and 1.58 for the x and y parameters, respectively (compare with the true results $x = -1.80$, $y = 1.60$) with standard deviations of 0.18 and 0.21 . This shows that the algorithm does provide an unbiased estimate of the model parameters, and the estimation error for a single trial is relatively small.

With this method we can estimate the estimation error magnitude of x and y for every trial i , namely σ_x^i, σ_y^i .

Validation of the main result of this article

One of the main results of this paper was that a significant number of trials fell outside the possible Mates region. Since until Appendix 2 we did not take into account correlated noise, it is theoretically possible that in the presence of correlated estimation error all the trials will be inside the possible region. However, computations based on Algorithm 1 show that this is not the case.

Indeed, 62% of the trials (112 of 180) were *outside* the possible Mates region when parameters were estimated using the bGLS method. Compare that with 32.7% (59 of 180) without the bGLS method.

Another concern was that the model parameters are actually inside the possible Mates region, but only seem to fall outside the possible region because of estimation error. In order to show that this claim is not likely to be true, we used the estimates of the standard deviation of the estimation error obtained earlier in this appendix (σ_x^i, σ_y^i).

Specifically, for each estimated parameter we ran 100 simulations using the parameters and Eqs. 95–96, and $\sigma_T^2 = 49$, $\sigma_M^2 = 25$. Then we calculated for each of the simulated trials the bGLS estimate and computed the estimated standard deviation of (σ_x^i, σ_y^i). Based on this, we estimated the probability that all points were actually in the possible region but were found in the impossible region because of estimation error. This probability turned out to be extremely small ($p < 1e-5$), supporting one of the main claims of this paper.

Computing the probability that estimates are outside the possible region only because of estimation error

We now estimate the probability that estimation error caused our data to fall outside the possible Mates region. We assume the following:

An observed data point (x_i, y_i) represents a “true” point (x_i^0, y_i^0) plus noise: $(x_i, y_i) = (x_i^0, y_i^0) + (n_x^i, n_y^i)$. We assume that (n_x^i, n_y^i) is Gaussian with mean 0 (an unbiased estimate!) and standard deviation (σ_x^i, σ_y^i) (the estimation error estimated by simulations, as explained before). Now let H_1 be the hypothesis that all (x_i^0, y_i^0) are in the possible Mates region, and let H_0 be the hypothesis that (x_i^0, y_i^0) can be anywhere. In order to compute this probability, we need a further Bayesian assumption on the prior of the distribution of (x_i^0, y_i^0) . We will assume that (x_i^0, y_i^0) are Gaussian with mean μ_G and covariance Σ_G identical to the grand average across all participants and trials.

Applying the Bayesian argumentation:

$$P(H_1 | (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)) = \frac{P((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) | H_1) P(H_1)}{[P((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) | H_1) P(H_1)]}$$

$$+P((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)|H_0)P(H_0)]. \quad (98)$$

From independence of n_x^i :

$$P((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)|H_1) = \prod_i \int p((n_x^i, n_y^i)|H_1)P((x_i^0, y_i^0)|H_1)dx_i^0 dy_i^0 \quad (99)$$

$$P(H_1|(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)) = 1/(1 + L), \quad (100)$$

where L is the likelihood ratio

$$L = \prod_i \int p((n_x^i, n_y^i)|H_0)P((x_i^0, y_i^0)|H_0)dx_i^0 dy_i^0 / \prod_i \int p((n_x^i, n_y^i)|H_1)P((x_i^0, y_i^0)|H_1)dx_i^0 dy_i^0 \quad (101)$$

Since (n_x^i, n_y^i) are Gaussian with mean 0 and covariances σ_x^i, σ_y^i and $(x_i^0, y_i^0) \sim N(\mu_G, \Sigma_G)$ the log likelihood can be easily computed.

The Matlab script written to estimate the probability computed a probability smaller than $1e-5$, allowing us to reject the hypothesis that all true parameters were in the possible region of the Mates model. In other words, it is highly improbable that our parameter estimates in the impossible Mates region were due to estimation error alone. To conclude, our main findings concerning the Mates model hold also when correlated noise terms and estimation errors are considered.

References

Aitken AC (1935) On least squares and linear combinations of observations. *Proc Roy Soc Edinburgh* 55:42–48

Best RE (2003) Phase-locked loops: design, simulation, and applications. McGraw-Hill, New York

Dorf RC (1993) The electrical engineering handbook. CRC Press, Boca Raton

Hary D, Moore G (1985) Temporal tracking and synchronization strategies. *Hum Neurobiol* 4:73–77

Hary D, Moore G (1987a) Synchronizing human movement with an external clock source. *Biol Cybern* 56:305–311. doi:10.1007/bf00319511

Hary D, Moore G (1987b) On the performance and stability of human metronome-synchronization strategies. *Br J Math Stat Psychol* 40:109–124

Hasan MA, Thaut MH (1999) Autoregressive moving average modeling for finger tapping with an external stimulus. *Percept Motor Skills* 88:1331–1346

Hastie T, Tibshirani R, Friedman JH (2001) The elements of statistical learning: data mining, inference, and prediction. Springer, New York

Kay SM (1993) Fundamentals of statistical signal processing: estimation theory. Prentice Hall, Englewood Cliffs

Large EW, Jones MR (1999) The dynamics of attending: How we track time-varying events. *Psychol Rev* 106:119–159

Ljung L (1987) System identification: theory for the user. Prentice Hall, Englewood Cliffs

Loehr JD, Large EW, Palmer C (2011) Temporal coordination and adaptation to rate change in music performance. *J Exp Psychol Hum Percept Perform*. doi:10.1037/a0023102

Mates J (1994a) A model of synchronization of motor acts to a stimulus sequence. I. Timing and error corrections. *Biol Cybern* 70:463–473. doi:10.1007/bf00203239

Mates J (1994b) A model of synchronization of motor acts to a stimulus sequence. II. Stability analysis, error estimation and simulations. *Biol Cybern* 70:475–484. doi:10.1007/bf00203240

McAuley JD, Jones MR (2003) Modeling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short-interval timing. *J Exp Psychol Hum Percept Perform* 29:1102–1125

Michon JA (1967) Timing in temporal tracking. Van Gorcum, Assen

Michon JA, van der Valk NJL (1967) A dynamic model of timing behavior. *Acta Psychol* 27:204–212

Pressing J (1998a) Error correction processes in temporal pattern production. *J Math Psychol* 42:63–101

Pressing J (1998b) Referential behavior theory: A framework for multiple perspectives on motor control. In: Piek JP (ed) *Motor behavior and human skill: a multidisciplinary approach*. Hum Kinetics, Champaign pp 357–384

Pressing J (1999) The referential dynamics of cognition and action. *Psychol Rev* 106:714–747

Pressing J, Jolley-Rogers G (1997) Spectral properties of human cognition and skill. *Biol Cybern* 76:339–347. doi:10.1007/s004220050347

Rao CR, Toutenburg H (1999) Linear models: least squares and alternatives. Springer, New York

Repp BH (2001a) Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *J Exp Psychol Hum Percept Perform* 27:600–621

Repp BH (2001b) Processes underlying adaptation to tempo changes in sensorimotor synchronization. *Hum Mov Sci* 20:277–312

Repp BH (2005) Sensorimotor synchronization: A review of the tapping literature. *Psychonom Bull Rev* 12:969–992. doi:10.3758/bf03206433

Repp BH (2008) Perfect phase correction in synchronization with slow auditory sequences. *J Motor Behav* 40:363–367

Repp BH (2010) Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Hum Mov Sci* 29:200–213

Repp BH, Keller PE (2004) Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *Quart J Exp Psychol* 57:499–521

Repp BH, Keller PE (2008) Sensorimotor synchronization with adaptively timed sequences. *Hum Mov Sci* 27:423–456

Repp BH, Keller PE, Jacoby N (2012) Quantifying phase correction in sensorimotor synchronization: empirical comparison of three paradigms. *Acta Psychol* 139:281–290

Rockafellar RT (1970) Convex analysis. Princeton University Press, Princeton

Rudin W (1987) Real and complex analysis. 3rd edn. McGraw-Hill, New York

Schulze H-H, Cordes A, Vorberg D (2005) Keeping synchrony while tempo changes: accelerando and ritardando. *Mus Percept* 22:461–477

Schulze H-H, Vorberg D (2002) Linear phase correction models for synchronization: parameter identification and estimation of parameters. *Brain Cog* 48:80–97

Semjen A, Schulze H-H, Vorberg D (2000) Timing precision in continuation and synchronization tapping. *Psychol Res* 63:137–147 doi:10.1007/pl00008172

Thaut MH, Miller RA, Schauer LM (1998a) Multiple synchronization strategies in rhythmic sensorimotor tasks: phase vs period correction. *Biol Cybern* 79:241–250. doi:10.1007/s004220050474

Thaut MH, Tian B, Azimi-Sadjadi MR (1998b) Rhythmic finger tapping to cosine-wave modulated metronome sequences: evidence of subliminal entrainment. *Hum Mov Sci* 17:839–863

- Torre K, Delignières D (2008) Unraveling the finding of $1/f^\beta$ noise in self-paced and synchronized tapping: a unifying mechanistic model. *Biol Cybern* 99:159–170
- Vapnik VN (1998) *Statistical learning theory: adaptive and learning systems for signal processing, communications, and control*. Wiley, New York
- Vorberg D, Schulze H (2002) A two-level timing model for synchronization. *J Math Psychol* 46:57–87
- Vorberg D, Wing A (1996) Modeling variability and dependence in timing. In: Keele S, Heuer H (eds) *Handbook of perception and action*, vol 2. Academic Press, New York pp 181–262
- Wing AM, Kristofferson AB (1973) Response delays and the timing of discrete motor responses. *Percept Psychophys* 14:5–12
- Wing AM, Kristofferson AB (1973) The timing of interresponse intervals. *Percept Psychophys* 13:455–460