BRILL *Timing & Time Perception* (*2015*) DOI:10.1163/22134468-00002047 brill.com/time

TIME

Lower Bound on the Accuracy of Parameter Estimation Methods for Linear Sensorimotor Synchronization Models

Nori Jacoby [1, 2, *], **Peter E. Keller** [3], **Bruno H. Repp** [4], **Merav Ahissar** [1, 5] **and Naftali Tishby** [1, 6]

[1] The Edmond & Lily Safra Center for Brain Science, Hebrew University of Jerusalem, Israel

[2] Music Department, Bar Ilan University, Israel

[3] MARCS Institute, University of Western Sydney, Sydney, Australia

[4] Haskins Laboratories, New Haven, CT, USA

[5] Department of Psychology Hebrew University of Jerusalem, Israel

[6] School of Engineering and Computer Science, Hebrew University of Jerusalem, Israel

Received 05 May 2014; accepted 18 February 2015

Abstract

The mechanisms that support sensorimotor synchronization — that is, the temporal coordination of movement with an external rhythm — are often investigated using linear computational models. The main method used for estimating the parameters of this type of model was established in the seminal work of Vorberg and Schulze (2002), and is based on fitting the model to the observed auto-covariance function of asynchronies between movements and pacing events. Vorberg and Schulze also identified the problem of *parameter interdependence*, namely, that different sets of parameters might yield almost identical fits, and therefore the estimation method cannot determine the parameters uniquely. This problem results in a large estimation error and bias, thereby limiting the explanatory power of existing linear models of sensorimotor synchronization. We present a mathematical analysis of the *parameter interdependence* problem. By applying the Cramér–Rao lower bound, a general lower bound limiting the accuracy of any parameter estimation procedure, we prove that the mathematical structure of the linear models used in the literature determines that this problem cannot be resolved by any unbiased estimation method without adopting further assumptions. We then show that adding a simple and empirically justified constraint on the parameter space — assuming a relationship between the variances of the noise terms in the model — resolves the problem. In a follow-up paper in this volume, we present a novel estimation technique that uses this constraint in conjunction with matrix algebra to reliably estimate the parameters of almost all linear models used in the literature.

Keywords

Sensorimotor synchronization, linear models, phase correction, period correction, Cramér–Rao lower bound

* To whom correspondence should be addressed. E-mail: jacoby@mit.edu.

 DOI: 10.1163/22134468-00002047

1. Introduction

Precision in the timing of movements with respect to internally specified goals and external events is essential for many forms of sequential behavior, including music performance, dance, and sports such as rowing. Over the years, several experimental paradigms, typically involving simple movements such as finger taps, have been repeatedly used to study movement timing and sensorimotor synchronization. The use of computational modeling in conjunction with these paradigms provides a powerful approach for understanding the mechanisms underlying timing and synchronization.

The current article is the first of two papers in this volume that address the problem of estimating the parameters for linear sensorimotor synchronization models. A concise description of the evolution of these models in the sensorimotor synchronization literature is provided in the second paper (for a general review of the sensorimotor synchronization literature, see Repp, 2005; Repp & Su, 2013). Here we focus on the methodological limitations of previous estimation techniques, while in the second paper (Jacoby et al., 2015) we suggest a unified, efficient, and reliable method for parameter estimation of almost all linear sensorimotor synchronization models extant in the literature.

Early efforts to model human movement timing targeted the simplest paradigm, which entails a scenario in which the subject freely taps a finger in the absence of an external stimulus. Often this is studied using a task where the base tempo is given to the subject by an external metronome, which eventually stops, while the subject is required to continue tapping. Early research using this synchronization-continuation paradigm (e.g., Stevens, 1886) showed that subjects are able to maintain the tempo with remarkable precision, leading to the idea that humans have an internal 'timekeeper' akin to an adjustable metronome-like cognitive process. Based on such data, Wing and Kristofferson (1973) proposed that two different noise components contribute to the overall variability of tapping: the variability of the internal timekeeper (σ_T^2) and the variability of motor delays (σ_M^2), where the latter reflects the inherent inaccuracies of the motor system. On this basis, they suggested that the next inter-stimulus interval at tap number $k+1$, R_{k+1} takes the form of the following equation:

$$R_{k+1} = T_k + M_{k+1} - M_k \tag{1}$$

where T_k and M_k are the timekeeper and motor noise at tap k. Wing and Kristofferson (1973) further assumed that T_k, and M_k are independent, and that the mean of T_k satisfies $E(T_k) = \tau$, where τ is the experiment base tempo. Note that this equation implies a specific correlation structure of the inter-response intervals.

The auto-covariance function (*acvf*) of R_k, $\gamma_R(j) \equiv \mathrm{Cov}(R_k, R_{k+j})$, has the following form:

$$\gamma_R(0) = \mathrm{Var}(R_k) = \sigma_T^2 + 2\sigma_M^2 \tag{2}$$

$$\gamma_R(1) = \mathrm{Cov}(R_k, R_{k+1}) = -\sigma_M^2 \tag{3}$$

$$\gamma_R(j) = \mathrm{Cov}\left(R_k, R_{k+j}\right) = 0 \quad \text{for every } j \geq 2 \tag{4}$$

This correlation structure can be used to extract the parameters of the model by simply inputting the empirical estimates of the *acvf*, and then using eqns (2) and (3) to extract the model's parameters.

Two experimental observations from Wing and Kristofferson's research are noteworthy. First, motor variance does not scale with base tempo (for a review, see Wing, 2002). Second, the timekeeper variance has been consistently found to be larger (usually much larger) than the motor variance. As we will see shortly, the second observation will be crucial for obtaining a reliable estimation method.

Later work extended the Wing and Kristofferson model to sensorimotor synchronization, namely, to a scenario where the subject synchronizes finger taps with an isochronous metronome (Vorberg & Wing, 1996). To handle this case, Vorberg and Wing (1996) and Vorberg and Schulze (2002) suggested the following linear phase correction model:

$$A_{k+1} = (1 - \alpha)A_k + T_k + M_{k+1} - M_k - S_{k+1} \tag{5}$$

where S_k is the inter-stimulus interval between the onset at beat number k and the onset at beat number $(k-1)$[1]; A_k is asynchrony, namely the difference between the response and the stimulus onsets at time k; and α is a phase correction parameter. Note that by convention the asynchrony is *negative* when the subject taps *before* the stimulus beat.

In fact, this model of sensorimotor synchronization is an amalgamation of Wing and Kristofferson's noise term with a hypothesized phase correction process controlled by parameter α (phase correction constant). The same exact model can be used for sequences of stimuli with occasional timing perturbations (slight delays or advances of events), as long as the general tempo of the stimulus remains constant (Repp et al., 2012). Tempo changing stimuli are discussed in our second, companion paper (Jacoby et al., 2015).

[1] Note the small inconsistency between this notation and the metronome intervals C_n as defined in Vorberg and Schulze (2002): $C_n = S_{n+1}$.

For tapping experiments involving an isochronous metronome, Vorberg and Schulze (2002) derived explicit analytical formulae for the auto-covariance function (*acvf*) of the asynchronies of model. They then extracted the model parameters by computing the estimates from the empirical *acvf*. This method also works for metronomes in which inter-onset intervals are stochastic (sampled from independent and identically distributed random variables; see Feller, 2008). However, if the metronome sequence has any other type of structure (e.g. abrupt phase changes, as in the paradigm used in Repp et al., 2012), there is no simple analytical solution. In this case the estimation process is relatively slow, requiring one to compute the empirical *acvf* of the asynchronies from the data and compare it with simulations.

Furthermore, the problem of *parameter interdependence* (Li et al., 1996) limits the accuracy of this method (this has been analyzed by Vorberg & Schulze, 2002), sometimes even rendering it impossible to use (Vorberg & Schulze, 2013).

Vorberg and Schulze used standard moment-estimation (Pearson, 1896) to find the first terms of the empirical asynchrony *acvf*, which can then be matched with the model predictions. This is effectively equivalent to searching the parameter space until a good match is found, thereby estimating the unknown parameters. The problem of parameter independence occurs when there is more than one solution to this process. In this case, the method cannot determine which of the possible solutions is correct, resulting in an increased estimation error.

Thus far it was unclear whether there might be another method based on a different fitting procedure that allowed this ambiguity to be resolved. As we will later see, the *acvf* method runs into difficulty exactly when the likelihood function of the observed data is 'flat'. Namely, that there is a relatively large zone in the parameter space around the global maximum where the likelihood function 'does not change too much' (we will shortly formally define these concepts).

The likelihood function is generally defined as the probability (assuming a model M) of the observed data (X) given the parameters (θ). Formally:

$$P_M(X|\theta) \tag{6}$$

A natural candidate for the model's fit is the point in the parameter space that maximizes the likelihood function. Formally:

$$\hat{\theta} = \text{argmax}_\theta \, P_M(X|\theta) \tag{7}$$

Figure 1a schematically depicts a contour plot of the likelihood function in a case where there is no large flat zone. In this case there is a small area in the parameter space whose $P_M(X|\theta)$ values approach the maximum.

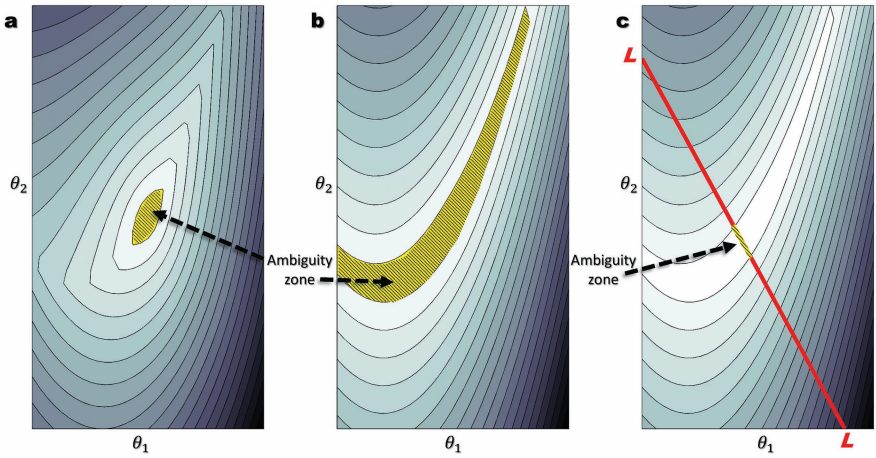

Figure 1. Schematic illustration of the likelihood function of the model $P_M(X|\theta)$. In this example, the parameter space is two-dimensional $\theta = (\theta_1, \theta_2)$. Colors in the contour plot correspond to the values of $P_M(X|\theta)$ where brighter colors represent higher values. In Fig. 1a there is relatively small zone in the space where $P_M(X|\theta)$ is close to the maximal value. Here the likelihood function yields a single maximum and the ambiguity zone of parameter estimation is small (depicted with tilted lines). Figure 1b represents a more difficult scenario. In this case there is a large zone where the likelihood is close to the maximum. The ambiguity zone is large and in this case the Cramér–Rao lower bound predicts a large estimation error. Figure 1c shows the effect of constraining the parameter space. If we assume that parameters are necessarily restricted to a line (the diagonal line L), then the ambiguity zone of values of the likelihood function close to the maximum is significantly reduced. This figure is published in color in the online version.

In the case of linear models, this maximum likelihood estimation procedure results in an unbiased and reliable estimator, at least when the number of samples is large (Ljung, 1998). However, computing this estimate in practice might be non-trivial. Indeed, our companion paper (Jacoby et al., 2015) is dedicated to developing a practical algorithm that approximates the maximum likelihood estimator for this problem.

Figure 1b schematically depicts a contour plot of the likelihood function for a more difficult scenario. In this case, there is a zone in the parameter space in which $P_M(X|\theta)$ obtains nearly maximal value, and therefore the maximum likelihood estimator would fail in retrieving the correct parameter from the data (it can pick any value in the ambiguous zone, and not necessarily the correct parameters). This is similar to the difficulty encountered by the method that uses moments estimates due to parameter interdependence.

A well-known result from the theory of statistics called the Cramér–Rao lower bound (Rao, 1992; Cramér, 1999 — henceforth CRLB) asserts that in this case there is *no other* unbiased estimation method that can succeed. Namely, the CRLB points to a direct link between the property of the likelihood functions

$P_M(X|\theta)$ and *any* unbiased estimator. More precisely, the CRLB asserts that the variance of the estimation error of any unbiased estimator (var $g(X)$) satisfies:

$$\text{var } g(X) \geq \frac{1}{I(\theta)} \tag{8}$$

where $I(\theta)$ is the Fisher information, defined as:

$$I(\theta) = -E_X\left[\frac{\partial^2 \log P_M(X|\theta)}{\partial \theta^2}\right] \tag{9}$$

Geometrically this means that when the likelihood function (and therefore also the logarithm of the likelihood function) is 'flat' as in Fig. 1b, there is a small increase in the rate of change of the log-likelihood function near the maximal value. This results in a small second derivative of the log-likelihood function. In this case, the CRLB asserts a large estimation error [eqn. (8)].

Figure 1c illustrates how constraining the parameter space could remediate the effect of this problem. If we further assume some limitation on the geometric structure of the problem (in the figure we assume that θ is constrained to the line L), then the likelihood function constrained to this subspace is no longer 'flat' and has a smaller ambiguity zone. In this case, the maximal likelihood estimator will result in one uniquely defined solution. Note however that if the assumption used is 'wrong' in the sense that it does not really hold for the data, then even larger estimation errors could be generated, due to the fact that the search space is reduced to a sub-space that does not contain the real solution.

As we will see shortly, it is possible to supplement linear sensorimotor synchronization models with an empirically justified assumption that acts exactly like the line L in Fig. 1c. This single assumption removes the redundancy from the likelihood function and therefore resolves the problem in the estimation procedure.

In the current study, we applied the CRLB to show that any method of estimating the linear phase correction model parameters is prone to failure unless further assumptions are made. We next demonstrate that the CRLB predicts the magnitude of the estimation error of the method based on estimation of moments, and provides a measure of the magnitude of the problem created by parameter interdependence. We propose a solution to this problem by showing that a reliable estimate can be obtained if the simple assumption that the motor variance is smaller than the timekeeper variance is adopted:

$$\sigma_M^2 < \sigma_T^2 \tag{10}$$

We show that a modified version of the original estimation method that relies on this assumption no longer suffers from parameter interdependence, and provides an unbiased estimator of phase correction and timekeeper and motor noise variance. In the case of the moments estimation method, this means that interdependence problem is generated from points in parameter space in which $\sigma_M^2 < \sigma_T^2$. If we ignore these points, there is no parameter interdependence, and the original estimation procedure works.

We end by comparing these results with results obtained using a novel estimation method termed *bounded Generalized Least Squares* (bGLS), developed in the companion paper (Jacoby et al., 2015). Using the constraint of eqn. (10) in conjunction with the tools of matrix algebra, this method can reliably estimate the parameters of most linear models described in the literature (e.g., Hary & Moore, 1987a, b; Mates, 1994a, b; Michon, 1967; Schulze et al., 2005; Vorberg & Wing, 1996).

2. The *acvf* Method for Estimating the Parameters of the Phase Correction Model

The parameters of the phase correction model can be estimated using the empirical *acvf*. This leads to the following simple method, which is depicted schematically in Fig. 2 (see Vorberg & Schulze, 2002; Vorberg & Wing, 1996).

Algorithm 1 – the *acvf* method
Input: *nseq* sequences of N asynchronies $\{A_{k=1\ \ldots\ N}\}_{i\ =\ 1\ \ldots\ nseq.}$
Output: the estimated parameters $(\overline{\alpha}, \overline{\sigma}_T, \overline{\sigma}_M)$.

1. Compute the mean over the *nseq* sequences of the biased *acvf* estimator: $\hat{\gamma}_A(k) = \frac{1}{N-k}\Sigma_t^{N-k}[A_t - E(A)][A_{t-k} - E(A)]$
2. Perform a numerical search over parameter $\theta = (\alpha,\ \sigma_T,\ \sigma_M)$. For each $\theta = (\alpha, \sigma_T, \sigma_M)$:
   i. Compute the asymptotic *acvf*: $\gamma_A(0) = \frac{\sigma_T^2 + 2\alpha\sigma_M^2}{1-(1-\alpha)^2}$; $\gamma_A(k) = \left[\frac{(1-\alpha)\sigma_T^2 + 2(1-\alpha)\alpha\sigma_M^2}{1-(1-\alpha)^2} - \sigma_M^2\right](1-\alpha)^{k-1}$ for $k > 0$.
   ii. Use $E[\hat{\gamma}_A(k)] = \gamma_A(k) - \frac{2}{N(N-k)}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N-k}\gamma_A(i-j) + \frac{1}{N^2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\gamma_A(i-j)$, and compute the mean of the biased estimator.
   iii. Find $\overline{\theta}$ such that $S(\theta) = \Sigma_{k=0}^{m}\{\hat{\gamma}_A(k) - E[\hat{\gamma}_A(k)]\}^2$ is minimized ($m = 2$).

3. Performance of the *acvf* Method

Figure 3 shows the mean and standard deviation of the estimation for parameters $(\alpha, \sigma_T^2, \sigma_M^2)$ based on 1000 iterations of the process obtained by the *acvf* method, for multiple values of α (on the x-axis). The dashed line represents

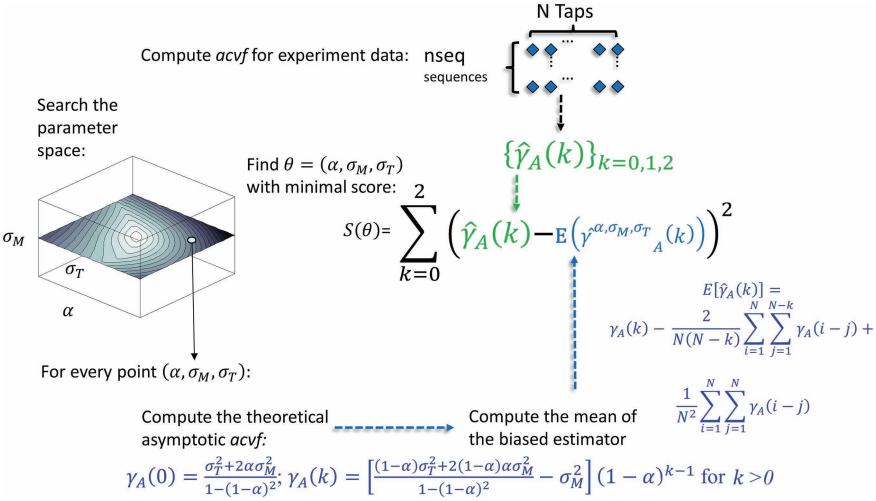

Figure 2. Schematic illustration of *acvf* method for parameter estimation of the phase correction model. The input of this procedure is *nseq* sequences of *N* taps. From this data we can compute the empirical estimate of the auto-covariance function (*acvf*): $\hat{\gamma}_A(k)$. To find the fitted parameters we scan the parameter space (all possible α, σ_M, σ_T values) and for each parameter compute the theoretical asymptotic *acvf*: $\gamma_A(k)$. We further compute the mean of the biased estimator: $E[\hat{\gamma}_A(k)]$. We then select the parameters that provide the best fit to the score function $S(\theta)$. This figure is published in color in the online version.

the correct value of the parameter, and the diamond represents the data from Table 1 in Vorberg and Schulze (2002) ($\alpha = 0.2$, $\sigma_T^2 = 100$, $\sigma_M^2 = 25$). Note that $nseq = 15$, as in the second line of Vorberg and Schulze's Table 1.

The first observation is that the results obtained with the *acvf* method with $\alpha = 0.2$ are identical to the results reported in Table 1 in Vorberg and Schulze (2002). This was to be expected, since this algorithm was used to create this table. The estimate at this point is very accurate and the mean of α, σ_T^2, and σ_M^2 in our simulation (0.2084, 103.14, and 24.25, respectively) is similar to the mean reported in Vorberg and Schulze's Table 1 (0.193, 100.80, and 25.13, respectively). Our standard deviation of the estimate for σ_T^2 and σ_M^2 (0.0705, 21.94, and 13.07, respectively) is also similar to that of Vorberg and Schulze (0.067, 21.3, and 12.0, respectively).

However, for some alpha values, the estimator clearly become biased and the estimation error increases. Such inaccuracies have been reported previously in the literature; see for example Diedrichsen et al. (2003), and Vorberg and Wing (1996). Vorberg and Schulze (2002) argued that the estimates for the different parameters become interdependent, and that the *acvf* for different parameters is similar, thus resulting in numerical instability. Vorberg and Wing (1996) noted that this phenomenon occurs around $\alpha \approx \alpha_{OPT}$ (the alpha value that minimizes

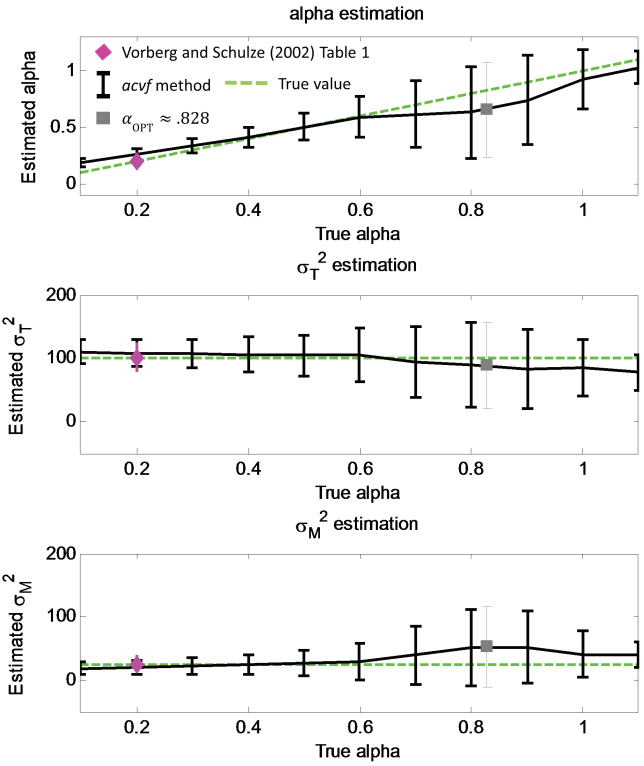

Figure 3. Simulation results of parameter estimation using the *acvf* method. In these simulations, we scanned multiple values of α, while setting $\sigma_T^2 = 100$ and $\sigma_M^2 = 25$. For each α we computed 1000 iterations of the simulation and estimated α (top figure, thick line), the timekeeper variance (middle figure, thick line), and the motor variance (bottom figure, thick line). In each graph the correct underlying parameters were plotted as the dashed line. For example, in the top diagram the dashed line is defined by the equation 'estimated α = true α'. In addition, we plotted the results of Vorberg and Schulze's (2002) Table 1 (diamonds). Error bars represent standard deviation of the estimates. We also computed the standard deviation and bias for the special value $\alpha_{\mathrm{OPT}} \approx 0.828$ (the α value that minimized the asynchrony variance for given parameters σ_T^2 and σ_M^2; Vorberg & Wing, 1996). This figure is published in color in the online version.

the asynchrony variance for given parameters σ_T^2 and σ_M^2). Note that empirical research shows that subjects tend to adjust their alpha towards the optimal value (Repp et al., 2012; Wing et al., 2014). Thus, we expected that the parameter estimated in real experiments would be near the optimal value, exactly where the estimation procedure may be considerably inaccurate.

However, it remains unclear whether this is a particular problem of the methods applied, or whether there is some limitation on the accuracy of the estimation that applies to all estimation methods.

We now show that all unbiased estimators suffer from a quantifiable accuracy limitation, and therefore the numerical problem in the estimation process cannot be bypassed by unbiased estimators unless further assumptions are made.

4. The Cramér–Rao Lower Bound Limits any Unbiased Estimation Method

It is useful to rewrite the phase correction model as an ARMA model (see Diedrichsen et al., 2003):

The model:

$$A_{k+1} = (1 - \alpha)A_k + T_k + M_{k+1} - M_k - S_{k+1} \tag{11}$$

can be written as an ARMA(1,1) model:

$$A_{k+1} = -aA_k + \lambda \cdot W_{k+1} + \lambda c \cdot W_k + \tau \tag{12}$$

where $c = \sqrt{r(2+r)} - r - 1$, $r = \sigma_T^2/2\sigma_M^2$, $a = \alpha - 1$, $\lambda^2 = -\sigma_M^2/c$, $\tau > 0$ and W_k is white Gaussian noise with unit variance (Diedrichsen et al., 2003).

Note that the new parameters satisfy the following conditions:

$$\begin{aligned} a &= \alpha - 1, \\ c &= \frac{-2\sigma_M^2 - \sigma_T^2 + \sigma_T\sqrt{4\sigma_M^2 + \sigma_T^2}}{2\sigma_M^2}, \\ \lambda &= \sqrt{\frac{2\sigma_M^4}{2\sigma_M^2 + \sigma_T^2 - \sigma_T\sqrt{4\sigma_M^2 + \sigma_T^2}}} \end{aligned}$$

The CRBL is a well-known mathematical limit to the variance of the estimation error, as described in the introduction (see also Fig. 1). For the specific case of an ARMA model, Friedlander (1984) derived an explicit formula for the bound.

Let us consider the following ARMA(n,m) model discussed in Friedlander (1984):

$$A_k = -\sum_{i=1}^{n} a_i A_{k-i} + \lambda \sum_{i=0}^{m} c_i W_{k-i};\ c_0 = 1 \tag{13}$$

where W_k is zero mean, unit-variance Gaussian white noise.

Note that when $n=1$ and $m=1$ this model is equivalent to the model described by eqns (10) and (11) and has the following form:

$$A_k = -aA_{k-1} + \lambda(W_k + cW_{k-1}) \tag{14}$$

In the general case, given N' data points A_k where $k = 1, \ldots, N'$, we want to estimate the vector of the unknown parameters:

$[\overline{A}, \overline{C}, \lambda] = [a_1, a_2, \ldots, a_n, c_1, c_2, \ldots, c_m, \lambda]$, where $\overline{A} = [a_1, a_2, \ldots, a_n]^T$, $\overline{C} = [c_1, c_2, \ldots, c_m]^T$.

M^{T} is henceforth used to denote the transpose of a matrix M, and M^{-1} represents the inverse of a regular square matrix M (Strang, 2006).

Let $\hat{A}$, $\hat{C}$, and $\hat{\lambda}$ be an unbiased estimators (based on N' points) of $\overline{A}$, $\overline{C}$, and λ, and let: $\tilde{A}$, $\tilde{C}$, and $\tilde{\lambda}$ be the estimation error: $\left[\tilde{A}, \tilde{C}, \tilde{\lambda}\right] = [\overline{A}, \overline{C}, \lambda] - [\hat{A}, \hat{C}, \hat{\lambda}]$.

Then the following theorem limits the minimal estimation error:

Theorem (Friedlander, 1984)

$$\mathrm{var}\left(\begin{bmatrix}\tilde{A}\\ \tilde{C}\end{bmatrix}\left[\tilde{A}^T\ \tilde{C}^T\right]\right) \geq \frac{1}{N'}\begin{bmatrix} R_{xx} & -R_{xz} \\ -R_{xz} & R_{zz}\end{bmatrix}^{-1} = \frac{1}{N'}G \tag{15}$$

$$\mathrm{var}\left(\tilde{\lambda}\right) \geq \frac{\lambda^2}{2N'} \tag{16}$$

$$\mathrm{var}(\tilde{A}) \geq \frac{1}{N'}\left[R_{xx} - R_{xz}R_{zz}^{-1}R_{zx}\right]^{-1} = \frac{1}{N'}F_A \tag{17}$$

$$\mathrm{var}(\tilde{C}) \geq \frac{1}{N'}\left[R_{zz} - R_{zx}R_{xx}^{-1}R_{xz}\right]^{-1} = \frac{1}{N'}F_C \tag{18}$$

where the matrices R_{xz}, R_{xx}, and R_{zz} can be computed from $\overline{A}$ and $\overline{C}$, and for any two matrices A and B, $A \geq B$ if and only if $A - B$ is a positive semi-definite matrix (Strang, 2006).

In the specific case where $n = m$:[2]

$$R_{xx} = \left[\overline{A}_1\overline{A}_1^T - \overline{A}_2\overline{A}_2^T\right]^{-1} \tag{19}$$

$$R_{zz} = \left[\overline{C}_1\overline{C}_1^T - \overline{C}_2\overline{C}_2^T\right]^{-1} \tag{20}$$

$$R_{zx} = \left[\overline{A}_1\overline{C}_1^T - \overline{C}_2\overline{A}_2^T\right]^{-1} \tag{21}$$

[2] The general formula for $n \neq m$ can be found in Friedlander (1984) and is not necessary for the current discussion.

where:

$$\overline{A}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_1 & & \ddots & \vdots \\ \vdots & & & 0 \\ a_{q-1} & \cdots & a_1 & 1 \end{bmatrix}, \overline{A}_2 = \begin{bmatrix} a_q & 0 & & \cdots & 0 \\ a_{q-1} & & \ddots & & \vdots \\ \vdots & & & & 0 \\ a_1 & \cdots & & a_{q-1} & a_q \end{bmatrix} \tag{22}$$

$$\overline{C}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ c_1 & & \ddots & \vdots \\ \vdots & & & 0 \\ c_{q-1} & \cdots & c_1 & 1 \end{bmatrix}, \overline{C}_2 = \begin{bmatrix} c_q & 0 & & \cdots & 0 \\ c_{q-1} & & \ddots & & \vdots \\ \vdots & & & & 0 \\ c_1 & \cdots & & c_{q-1} & c_q \end{bmatrix} \tag{23}$$

and $q = \max(n, m) = n = m$.

Applying the theorem for the case $n = 1$ and $m = 1$ results in:

$$\overline{A} = a,\ \overline{C} = c,\ \overline{A}_1 = 1,\ \overline{A}_2 = a,\ \overline{C}_1 = 1,\ \overline{C}_2 = c.$$

It follows that: $R_{xx} = (1 - a^2)^{-1}$, $R_{zz} = (1 - c^2)^{-1}$, $R_{zx} = (1 - ac)^{-1}$.

Therefore:

$$F_A = \left[R_{xx} - R_{xz}R_{zz}^{-1}R_{zx}\right]^{-1} = \left[\frac{1}{1-a^2} - \frac{(1-c^2)}{(1-ac)^2}\right]^{-1} = \frac{(1-a^2)(1-ac)^2}{(1-ac)^2 - (1-c^2)(1-a^2)} = \frac{(1-a^2)(1-ac)^2}{(a-c)^2} \tag{24}$$

$$F_C = \left[R_{zz} - R_{zx}R_{xx}^{-1}R_{xz}\right]^{-1} = \left[\frac{1}{(1-c^2)} - \frac{(1-a^2)}{(1-ac)^2}\right]^{-1} = \frac{(1-c^2)(1-ac)^2}{(a-c)^2} \tag{25}$$

According to Friedlander (1984), from eqn. (17): $\mathrm{var}\left(\tilde{\lambda}\right) \geq \frac{\lambda^2}{2N'}$, the estimation of λ is independent of the other parameters. This corresponds to the well-known square root convergence of the estimate of the standard deviation.

However, the CRLB bound generates more complicated results when applied to other parameters of the model.

Figure 4 shows the CRLB estimation for the prediction of the standard deviation of the estimation error for parameters a and c $\left(\sqrt{\frac{1}{N'}F_A}\text{ and }\sqrt{\frac{1}{N'}F_C}\text{, respectively}\right)$ where $N' = 450$. Here it can be seen that it predicts large deviations when $a \approx c$.

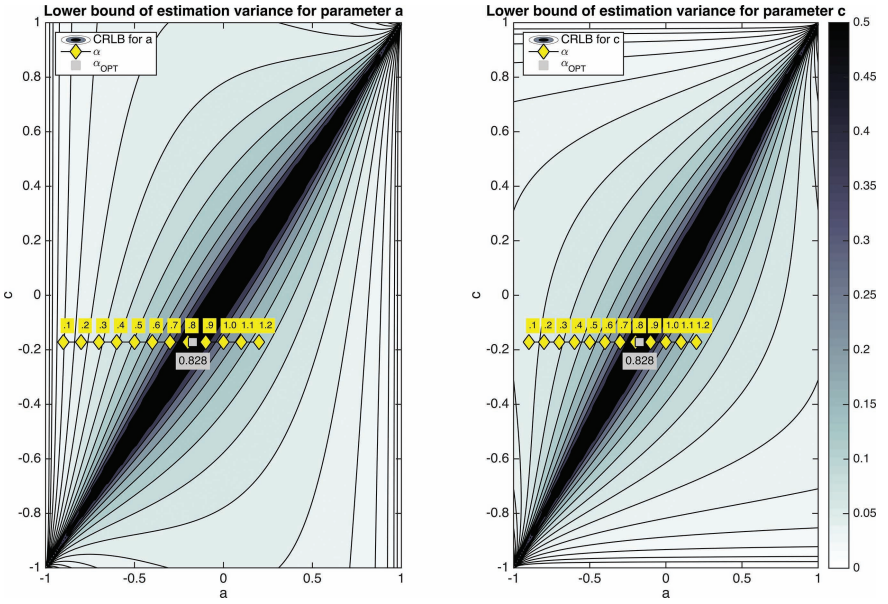

Figure 4. Contour plot of the predicted CRLB standard deviation of the estimation error for $\sqrt{\frac{1}{N'}F_A}$ and $\sqrt{\frac{1}{N'}F_C}$ when $N' = 450$. The x-axis is parameter a (determined by parameter α of the original model) and the y-axis is parameter c (determined by σ_M^2, and σ_T^2). The left and right figures show the predicted estimation error for parameters a and c, respectively. The diamonds show the CRLB computed for α value in the range of 0.1 to 1.2 and $\sigma_T^2 = 100$, $\sigma_M^2 = 25$ (or $a = \alpha - 1$ and $c \cong -0.1716$). Note that the prediction gets very large when $a \approx c$, therefore this graph only displays CRLBs of less than 0.5. Note also that $\alpha_{\mathrm{OPT}} \approx 0.828$ (the α value that minimized the asynchrony variance for given parameters σ_T^2 and σ_M^2; Vorberg and Wing 1996) is in the center of the problematic zone. This figure is published in color in the online version.

Figure 4 also shows the predicted deviations for the case $\sigma_T^2 = 100$, $\sigma_M^2 = 25$, and α in the range from 0.1 to 1.2 (marked with diamonds). In this case, $c = \frac{-2\sigma_M^2 - \sigma_T^2 + \sigma_T\sqrt{4\sigma_M^2 + \sigma_T^2}}{2\sigma_M^2} \cong -0.1716$, and the parameter $a = \alpha - 1$ ranges between -0.9 and 0.2. Note that if $c \cong -0.1716$ then the singularity occurs when $a \approx c$ which corresponds to $\alpha = a + 1 \cong 0.828 = \alpha_{\mathrm{OPT}}$.

Figure 5 shows the CRLB predictions, and compares the results to those obtained for 100 simulations applying the *acvf* method (again we used the parameters: $\sigma_T^2 = 100$, $\sigma_M^2 = 25$, $N = 30$, $nseq = 15$). Note that for the CRLB we used $N' = nseq * N = 30 * 15 = 450$.

For alpha in the range of 0.1–0.6, there is agreement between the CRLB and the variance obtained by the *acvf* method. The method is only slightly less efficient than the predicted bound in this parameter range, which is to be expected because here the estimation method is unbiased, and therefore must have equal or slightly larger standard error than the CRLB. Around

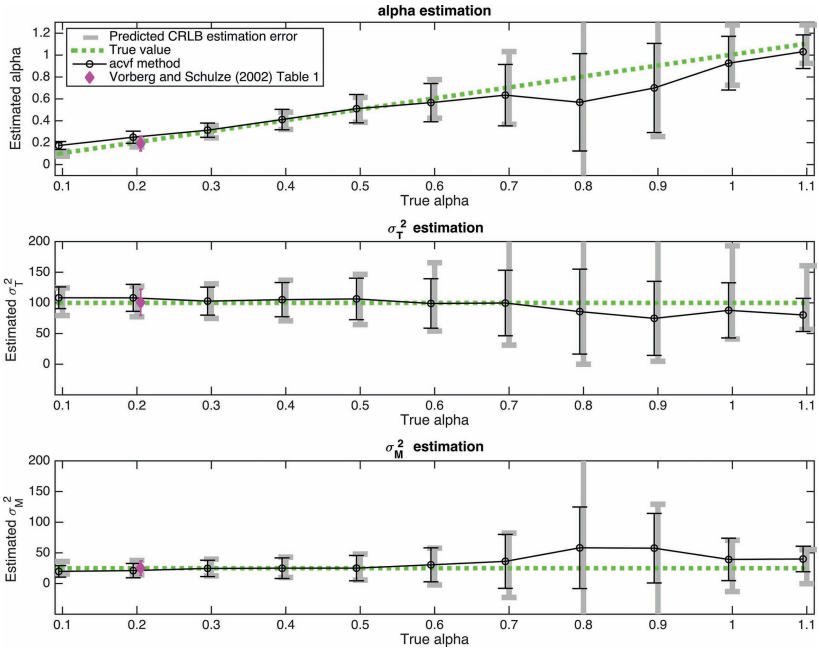

Figure 5. CRLB with estimated alphas based on simulation results (acvf method). The circle shapes represents alpha estimates for *acvf* method. The thick lines represent the CRLB error estimates. The diamonds show results from Vorberg and Schulze (2002) table 1. The simulations use the same parameters as Fig. 4 ($\sigma_T^2 = 100$, $\sigma_M^2 = 25$, $nseq = 15$, $N = 30$). Error bars represent the standard error of estimates. This figure is published in color in the online version.

the singularity point at F_A ($\alpha_{\mathrm{OPT}} = 0.828$), the CRLB predicts the increase in the variance of the estimation, as we can clearly see in Fig. 6. Note that for these points, the *acvf* method provides *biased* results, and therefore the CRLB does not directly predict *the magnitude* of the algorithm deviation. Note that the CRLB applies solely to unbiased estimators, so these results do not contradict the Friedlander (1984) formula.

Figure 5 shows also similar results for parameters σ_M^2 and σ_T^2. Since σ_M^2 and σ_T^2 are nonlinear functions of the model parameters c and λ, we cannot use the CRLB formula directly.[3] However, a relatively good estimate of the confidence interval predicted by the CRLB can be obtained if we compute a confidence interval for each parameter, based on perturbing each parameter (c, λ) by one standard deviation, according to $c \pm \sqrt{\frac{1}{N'} F_C}$ and $\lambda \pm \sqrt{\frac{1}{2N'} \lambda^2}$ and then apply the transformation equations below to the perturbed values:

$$\sigma_T^2 = \lambda^2 (1 + c)^2 \tag{26}$$

[3] It is possible to derive direct formulae of the bound for this case; however, this is beyond the scope of this article.

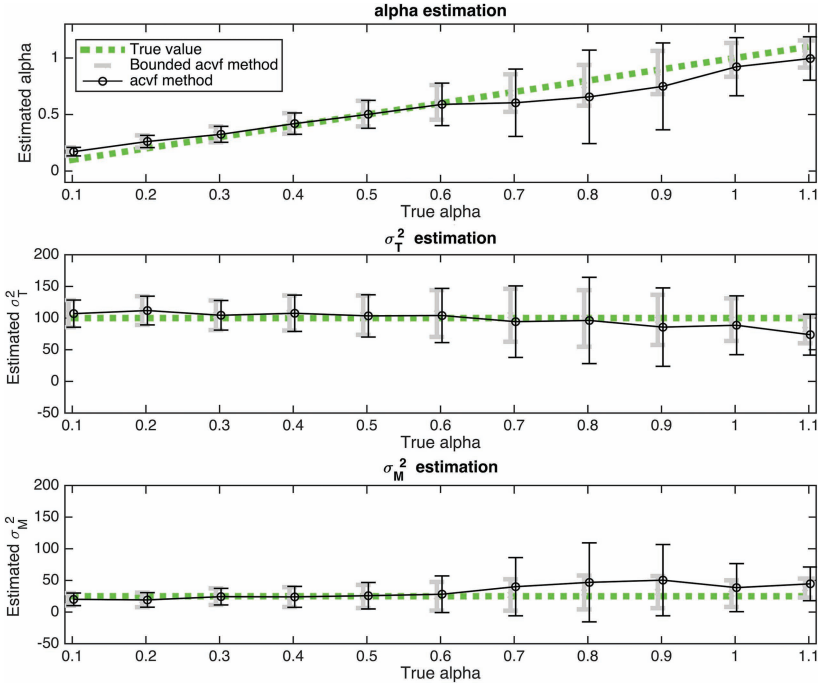

Figure 6. A comparison of the *acvf* and bounded *acvf* method. We compared the simulation results for the estimation of phase correction (top), timekeeper variance (middle), and motor variance (bottom). The simulations use the same parameters as Fig. 4 ($\sigma_T^2 = 100$, $\sigma_M^2 = 25$, $nseq = 15$, $N = 30$). The bounded version (thick lines) outperforms the unbounded version (thin circles), and is almost an unbiased estimator of the correct underlying parameters (dashed line). Error bars represent the standard error of estimates. This figure is published in color in the online version.

$$\sigma_M^2 = -\lambda^2 c \tag{27}$$

The result of this procedure is displayed in the two bottom panels of Fig. 5, where the CRLB confidence interval predicts the simulation results quite well when the estimators are unbiased.

These simulation results show that the CRLB provides a good prediction of the estimation error when the estimator is unbiased, and provides an approximation of the estimation error when the estimator is biased. Furthermore, it suggests that the *acvf* method is nearly optimal in performance for some values of alpha. However, the CRLB also predicts the large estimation error of this method for alpha near α_{OPT}. These results fully characterize the expected estimation error of the linear phase correction model.

5. Solving the Estimation Problem with Additional Assumptions: Circumventing the Problem Predicted by the CRLB

In the previous sections we identified and quantified the source of the inaccuracy for the previous parameter estimation methods. We showed that the relatively large inaccuracy of the parameter estimation is not unique to these methods and is in fact expected for any unbiased method. However, the analysis in the previous section suggests a possible strategy to circumvent the problem. The idea developed in this section is to use additional information about the parameter range when computing the estimate, thereby avoiding the CRLB inaccuracy.

Recall that the linear phase correction model originated from the Wing and Kristofferson (1973) model. In this context it is natural to assume that $\sigma_T^2 > \sigma_M^2$ (see for example Vorberg & Wing, 1996; Wing, 2002; Wing & Kristofferson, 1973). This single assumption greatly impacts the accuracy of parameter estimation. We can modify the *acvf* method to search only the parameter space that satisfies the constraint. This algorithm that we call the 'bounded acvf method' is completely identical to the standard acvf method but we constrained to θ values that satisfy $\sigma_M < \sigma_T$. As explained in Fig. 1, constraining the parameter space can reduce the ambiguity.

The bounded *acvf* method has almost the same running time as the standard *acvf* method; however the estimation errors are much smaller. Note, however, that if the assumption that $\sigma_T^2 > \sigma_M^2$ does not hold, there is a risk of an extremely large estimation error; i.e., obtaining a severely biased estimator.

Figure 6 shows simulation results that compare the standard (un-bounded) and bounded *acvf* method for the same set of parameters $\sigma_T^2 = 100$, $\sigma_M^2 = 25$, $nseq = 15$, $N = 30$.

For parameter $\alpha = 0.2$ from Table 1 in Vorberg and Schulze (2002) there is not much difference between the methods, but when α approaches $\alpha_{\mathrm{OPT}} = 0.828$ the bounded method is significantly better and unbiased.

This simple assumption $(\sigma_T^2 > \sigma_M^2)$[4] and modification of the original algorithm fully resolves the estimation problem, and the resulting bounded algorithm is efficient and robust. Note that when the stimulus sequence has large changes (transients) we cannot use the asymptotic formula and therefore

[4] Note that according to Wing and Kristofferson (1973), it is to be predicted that $\sigma_T^2 > 2\sigma_M^2$, which is even a stronger. In this work we mostly use the weaker assumption $\sigma_T^2 > \sigma_M^2$. We have also tested other constraints, for example $\sigma_T^2 > k\sigma_M^2$ where k is a multiplier constant. When k is large (e.g., $k = 9$), then $\sigma_T^2 \gg \sigma_M^2$. In a companion paper (Jacoby et al., 2015) we fully analyse this case and show that the model then can be estimated with standard least squares (simple regression). We also show that while this does produce bias, it is of relatively small magnitude. A multiplier smaller than 1 (e.g., $\sigma_T^2 > \frac{1}{2}\sigma_M^2$) is not enough to guarantee reliable estimates. We conclude that a multiplier of $k = 1$, as we suggest, is a good compromise between generality and estimation error magnitude. Note also that our simulations show that applying the stronger assumption $\sigma_T^2 > 2\sigma_M^2$ has a negligible effect on performance.

cannot use the method. Instead, we will present an efficient method for this case in the companion paper (Jacoby et al., 2015).

6. Performance of the bounded General Least Squares method

In the companion paper (Jacoby et al., 2015), we develop a computational procedure that is an efficient and unbiased estimator of the linear phase correction model. This method utilizes the same assumption about the relation between motor and timekeeper variances, namely that $\sigma_T^2 > \sigma_M^2$. The details of this method are explained in the subsequent paper. Here we only report a select few encouraging results obtained with the method. The advantage of this procedure is that it applies to most general scenarios, including cases in which the *acvf* method could be applied, while being even faster than bounded *acvf* method. Figure 7 shows that this method performs similarly to the bounded *acvf* method.

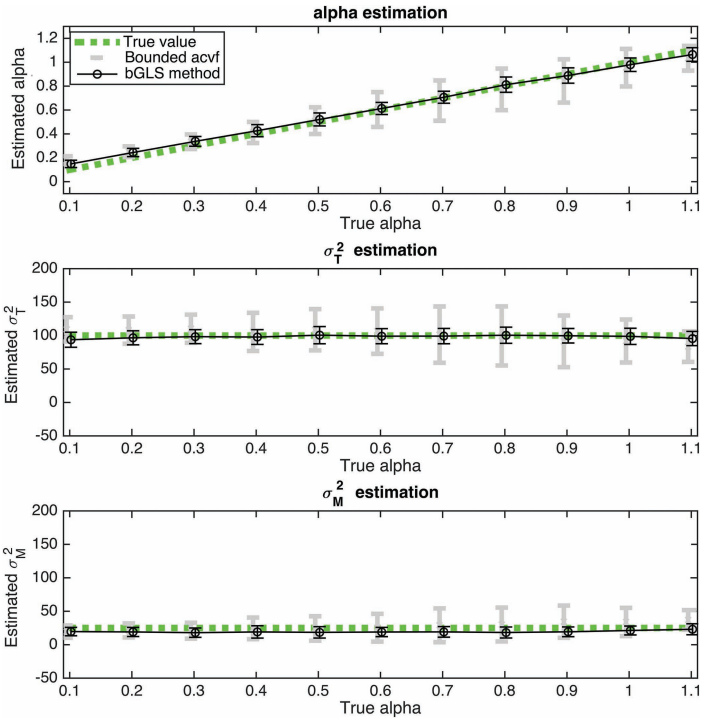

Figure 7. Bounded *acvf* method (thick line) compared to the bGLS method (circles). The correct underlying parameters are plotted as a dashed line. 1000 iterations were simulated for different α values (x-axis) with parameters ($\sigma_T^2 = 100$ and $\sigma_M^2 = 25$, $nseq = 15$, $N = 30$). The top, middle, and bottom graphs show the estimates for phase correction, timekeeper variance, and motor variance, respectively. This figure is published in color in the online version.

7. Conclusion

We used the CRLB to quantify the problem of large estimation errors of linear phase correction models. Our approach proved that this problem cannot be resolved by any estimation procedure without further assumptions. We showed that adding a constraint on the parameter space, a simple assumption about the relationship between the two noise components in the model, resolves the problem.

Even though this paper was focused on analyzing the phase correction model, other models that extend the phase correction model will also suffer from similar problems. For example, the period correction model suggested by Schulze et al. (2005) is a generalization of this model, and therefore has the same covariance structure of noise parameters as in eqns (2)–(4). Estimation procedures for this model are therefore also constrained by the same problem to an even larger extent, since when more parameters are used the likelihood of redundancies in the model increases. Even more importantly, the method based on moments estimates (the *acvf* method) completely fails in the case of multi-person ensemble synchronization, as in the dataset of Wing et al. (2014). This is due to parameter interdependence and the inherited noise in the estimation of a multi-dimensional *acvf* that is in this case necessary for the computation (Vorberg & Schulze, 2013).

This issue is the starting point for the sequel to this paper in this volume. In the next paper we show that the assumption that resolves the problem of parameter interdependence for the phase correction model, also helps in generating a reliable estimation procedure for period correction and ensemble synchronization.

Acknowledgments

The first author would like to thank Carmel Raz, Shira Balter, and Esther Singer for reviewing the manuscript. We also would like to thank the editors Timo Fischinger and Mark Eliott and the reviewers Dirk Vorberg and Hans-Henning Schulze for their valuable suggestions to the content and form of this paper.

References

Cramér, H. (1999). *Mathematical methods of statistics*. Princeton, NJ, USA: Princeton University Press.

Diedrichsen, J., Ivry, R. B., & Pressing, J. (2003). Cerebellar and basal ganglia contributions to interval timing. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 457–481). Boca Raton, FL, USA: CRC Press.

Feller, W. (2008). *An introduction to probability theory and its applications*. New York, NY, USA: John Wiley & Sons.

Friedlander, B. (1984). On the computation of the Cramer–Rao bound for ARMA parameter estimation. *IEEE Trans. Acoust.*, *32*, 721–727.

Hary, D., & Moore, G. (1987a). On the performance and stability of human metronome — Synchronization strategies. *Br. J. Math. Stat. Psychol.*, *40*, 109–124.

Hary, D., & Moore, G. (1987b). Synchronizing human movement with an external clock source. *Biol. Cybern.*, *56*, 305–311.

Jacoby, N., Tishby, N., Repp, B. H., Ahissar, M., & Keller, P. E. (2015). Parameter estimation of linear sensorimotor synchronization models: Phase correction, period correction and ensemble synchronization. *Timing Time Percept.*, *3*, 52–87.

Li, S., Lewandowsky, S., & DeBrunner, V. E. (1996). Using parameter sensitivity and interdependence to predict model scope and falsifiability. *J. Exp. Psychol. Gen.*, *125*, 360–369.

Ljung, L. (1998). *System identification*. Berlin, Germany: Springer.

Mates, J. (1994a). A model of synchronization of motor acts to a stimulus sequence. *Biol. Cybern.*, *70*, 463–473.

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.

Michon, J. (1967). *Timing in temporal tracking*. Soesterberg, The Netherlands: Institute for Perception RVO-TNO.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. R. Soc. Lond.*, *187*, 253–318.

Rao, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In S. Kotz & N. L. Johnson (Eds), *Breakthroughs in statistics* (pp. 235–247). New York, NY, USA: Springer.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonom. Bull. Rev.*, *12*, 969–992.

Repp, B. H., & Su, Y. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonom. Bull. Rev.*, *20*, 403–452.

Repp, B., Keller, P., & Jacoby, N. (2012). Quantifying phase correction in sensorimotor synchronization: Empirical comparison of three paradigms. *Acta Psychol.*, *139*, 281–290.

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

Stevens, L. T. (1886). On the time-sense. *Mind*, *11*, 393–404.

Strang, G. (2006). *Linear algebra and its applications*. Belmont, CA, USA: Thomson Brooks Cole/Boston, MA, USA: Cengage Learning.

Vorberg, D., & Schulze, H. (2002). Linear phase-correction in synchronization: Predictions, parameter estimation, and simulations. *J. Math. Psychol.*, *46*, 56–87.

Vorberg, D., & Schulze, H. H. (2013). *Modeling synchronization in musical ensemble playing: Parameter estimation and sensitivity to assumptions*. Fourteenth Rhythm Production and Perception Workshop, University of Birmingham, Birmingham, UK.

Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer & S.W. Keele (Eds), *Handbook of perception and action*, Vol. *2* (pp. 181–262). London, UK: Academic Press.

Wing, A. M. (2002). Voluntary timing and brain function: An information processing approach. *Brain Cogn.*, *48*, 7–30.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Percept. Psychophys.*, *14*, 5–12.

Wing, A. M., Endo, S., Bradbury, A., & Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *J. R. Soc. Interface*, *11*, 1125. DOI: 10.1098/rsif.2013.1125.