

Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach

Rasmus Bååth¹

© Psychonomic Society, Inc. 2015

Abstract The sensorimotor synchronization paradigm is used when studying the coordination of rhythmic motor responses with a pacing stimulus and is an important paradigm in the study of human timing and time perception. Two measures of performance frequently calculated using sensorimotor synchronization data are the average offset and variability of the stimulus-to-response asynchronies—the offsets between the stimuli and the motor responses. Here it is shown that assuming that asynchronies are normally distributed when estimating these measures can result in considerable underestimation of both the average offset and variability. This is due to a tendency for the distribution of the asynchronies to be bimodal and left skewed when the interstimulus interval is longer than 2 s. It is argued that (1) this asymmetry is the result of the distribution of the asynchronies being a mixture of two types of responses—predictive and reactive—and (2) the main interest in a sensorimotor synchronization study is the predictive responses. A Bayesian hierarchical modeling approach is proposed in which sensorimotor synchronization data are modeled as coming from a right-censored normal distribution that effectively separates the predictive responses from the reactive responses. Evaluation using both simulated data and experimental data from a study by Repp and Doggett (2007) showed that the proposed approach produces more precise

estimates of the average offset and variability, with considerably less underestimation.

Keywords Bayesian statistics · Sensorimotor synchronization · Hierarchical models · Finger tapping

The experimental study of human timing and time perception has a long history in psychology, with the sensorimotor synchronization (SMS) task being one of the most important experimental paradigms (Roetzheim, 2008). Following Stevens (1886), this task requires a participant to produce periodic movements synchronized to a regular pacing stimulus such as a metronome (Schulze, 1992). Sensorimotor synchronization is often studied in a musical context, because the ability to engage in SMS is central to musical activities, especially in ensemble music, in which many musicians are required to follow the same rhythm and coordinate their movements together (Repp, 2006). Yet, SMS performance is also a relevant measure in many other fields. For example, SMS performance has been shown to correlate with personality traits (Forsman, Madison, & Ullén, 2009) and measures of intelligence for both children (Corriveau & Goswami, 2009) and adults (Madison et al., 2009). It is also correlated with performance in other experimental paradigms related to timing, such as simple reaction time (Holm, Ullén, & Madison, 2011) and eye blink conditioning (Green, Ivry, & Woodruff-Pak, 1999).

What is most often measured in an SMS task is the stimulus-to-response asynchronies—that is, the offset of a participant's responses from the stimulus onsets, where a negative asynchrony indicates that a participant's response preceded the stimulus (Repp, 2005). The two basic parameters estimated in an SMS task are the constant error—the average deviation from the target stimuli—and the timing variability of the asynchronies. The most straightforward and common

Electronic supplementary material The online version of this article (doi:10.3758/s13428-015-0591-2) contains supplementary material, which is available to authorized users.

✉ Rasmus Bååth
rasmus.baath@lucs.lu.se

¹ Department of Philosophy, Lund University, Box 192, 22100 Lund, Sweden

way to estimate these parameters is to calculate the sample mean and the sample standard deviation (*SD*) of the asynchronies. However, it is shown here that this approach results in negatively biased estimates of constant error and timing variability; that is, these parameters are generally underestimated by the sample mean and sample *SD*. The concern is with the nonnormal distribution of timing asynchronies at slow tempi and the failure of moment estimators to incorporate task-relevant information.

The present article describes a Bayesian model with which to estimate the constant error and timing variability in SMS tasks that do not suffer from the problems associated with using the sample mean and *SD*. The text is organized as follows. First the distribution of SMS data is reviewed, and it is shown that the distribution is approximately normal when the tempo of the pacing stimuli is moderate, but that the distribution becomes increasingly nonnormal when the interstimulus interval (ISI) exceeds 2 s. It is argued that this is due to a tendency for participants to react to the stimulus onset and that, because of this, SMS data are modeled better using a right-censored normal distribution. A Bayesian model is then developed that uses a censored normal distribution to model the distribution of timing asynchronies. The model is developed in two stages, first as a basic nonhierarchical model, and then as a fully hierarchical model. Finally, the model is compared with the traditional methods of using the sample mean and *SD* to estimate constant error and timing variability. This is done using both simulated data and the experimental data from a study by Repp and Doggett (2007). It is shown that using a censored normal distribution to model timing asynchronies results in considerably less bias than using traditional methods. Furthermore, using a hierarchical Bayesian approach outperforms both traditional methods and a nonhierarchical Bayesian model with regard to accuracy. This article advances the study of human timing and time perception by giving researchers a better tool to measure SMS performance. A further advance is that the Bayesian methods proposed facilitate analyzing SMS data when the ISI exceeds 2 s.

The distribution of sensorimotor synchronization data

In a typical SMS task, a participant is asked to produce responses in time with a recurring, isochronous (equally spaced in time) stimulus sequence. The commonly employed stimuli are sequences of equally spaced auditory tones that the participant synchronizes to by tapping a button using the index finger, although there are many variations of this basic experimental procedure (Repp, 2005). Stimulus-to-response asynchronies, that is, the time offsets between the stimulus onset and the participant's timed response, are of primary interest in SMS tasks (Repp & Su, 2013). Such asynchronies can be positive or negative, where a negative asynchrony indicates that the corresponding response preceded the stimulus, and a positive asynchrony indicates that the corresponding response followed the stimulus onset.

Under many circumstances the distribution of the asynchronies is approximately normal (Chen, Ding, & Kelso, 1997; Mates, Müller, Radil, & Pöppel, 1994; Moore & Chen, 2010), but it is rarely normal when the ISI of the pacing stimuli is longer than 2 s (cf. Mates et al., 1994, and Miyake, Onishi, & Pöppel, 2004). An example of timing asynchrony distributions at different ISIs is shown in Fig. 1 using data from a study by Bååth and Madison (2012), in which 30 participants were asked to tap with their index finger in synchrony with a pacing tone sequence. In Fig. 1A, the asynchrony distributions for a representative participant are shown; for ISIs of 600 ms and 1,200 ms, the distributions can be seen to be heap shaped and symmetric. The central tendencies of the distributions are not centered around zero (i.e., at the onset of the pacing tone) but are slightly negative, a well-known phenomenon termed the *mean negative asynchrony* (Aschersleben, 2002). At ISIs of 1,800 ms and above, a visible peak from 100 to 200 ms makes the distribution left skewed, or even bimodal. This peak coincides with where auditory reaction time responses to the stimulus onset would be likely to occur (Gottsdanker, 1982). These deviations from normality not only can be seen by visual inspection, but a Shapiro–Wilk normality test is also rejected, with $p < .01$ as the

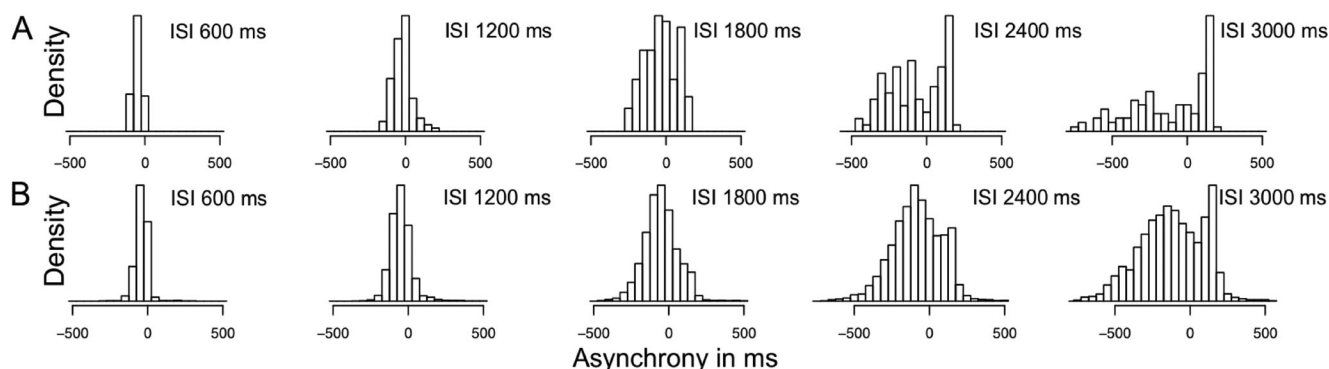


Fig. 1 Tone-to-response asynchrony distributions (A) for a single participant and (B) for all participants in Bååth and Madison (2012)

rejection criterion, for ISIs of 1,800 ms ($p = .004$), 2,400 ms ($p < .001$), and 3,000 ms ($p < .001$), but not for ISIs of 600 ms ($p = .62$) and 1,200 ms ($p = .90$). This was even after using Tukey's (1977) boxplot method to label and remove outliers. A similar pattern can be seen when looking at the distributions of the asynchronies from all participants. The percentages of the 30 participants who produced asynchronies at different ISI levels that resulted in rejection by a Shapiro–Wilk test were 7% (600 ms), 3% (1,200 ms), 7% (1,800 ms), 27% (2,400 ms), and 47% (3,000 ms). This resonates with the finding of Mates et al. (1994) that asynchronies produced at ISIs above 1,800 ms tend to be increasingly nonnormal.

According to Repp and Doggett (2007), the distribution of timing asynchronies departs from normality at ISIs longer than 2 s because participants occasionally overshoot the target stimuli and instead react to them. At moderately fast ISIs, around 600 ms (see Fig. 1), this rarely happens, because the asynchronies tend to be within 100 ms of the stimulus onset. At longer ISIs the variability of the asynchronies increases, and at an ISI of 2 s many asynchronies are smaller than –300 ms but rarely exceed 300 ms. This asymmetry results in a skewed, bimodal asynchrony distribution with a long left tail and a short right tail at long ISIs.

If Repp and Doggett's (2007) interpretation is correct, the observed distribution of timing asynchronies at ISIs greater than 2 s is then a mixture of predictive asynchronies, which generally are the responses of interest, and contaminating reactive asynchronies. The mean and *SD* of the whole asynchrony distribution is then uninformative, and methods should be used to separate out these distributions so as to estimate the mean and *SD* of the predictive distribution adequately. If this is not done, different phenomena are measured at long and short ISIs: At short ISIs, a participant's ability to synchronize to a pacing sequence is being measured, whereas at longer ISIs, a mixture of reaction time and timing ability may be measured. The data from an SMS task can then be thought of as being sampled from two different distributions. At short ISIs, the predictive timing asynchronies can be seen as samples from a normally distributed random variable X_P . At longer ISIs, at which the timing variability is large enough that participants sometimes make reactive responses, the timing asynchronies can be seen as samples from a random variable $X_A = \min(X_P, X_R)$, where the random variable X_R is distributed as a reaction time distribution. There are many proposed models for the distribution of reaction time responses, all of which are right skewed and, for practical purposes, left bounded (Ulrich & Miller, 1994; van Zandt, 2000). For the present purpose of modeling the predictive responses, the distribution of the reactive responses could be assumed to be any of those—for example, the ex-Gaussian distribution.

An illustration of the distribution of X_A is shown in Fig. 2. The rationale for taking the *minimum* of X_P and X_R is that these two random variables can be thought of as representing two

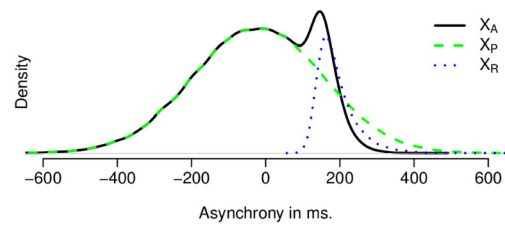


Fig. 2 Theoretical distribution of timing asynchronies (X_A , solid line), modeled as a combination of a distribution of predictive responses (X_P , dashed lines) and reactive responses (X_R , dotted line)

independent processes that can either trigger a response (e.g., a buttonpress or drum stroke), with the response being initiated by whichever process triggers first. For example, if 300 ms is an outcome of X_P and 200 ms is an outcome of X_R , then the outcome of X_A , which represents a participant's response, would be $\min(300 \text{ ms}, 200 \text{ ms}) = 200 \text{ ms}$. Because X_A is the minimum of X_P and another random variable, X_A will always have a shorter right tail than X_P . Note that there are other combinations of X_P and X_R that do not result in this behavior—for example, the average of X_P and X_R , or a mixture distribution constructed from X_P and X_R . The average of X_P and X_R would instead result in a distribution with a positively shifted mean, as compared to X_P . This would not be in agreement with the well-established finding that the mean asynchrony from timing tasks tends to be slightly negative (Aschersleben, 2002). The mixture distribution defined by $X_{\text{mix}} = wX_P + (1 - w)X_R$ would not resemble X_A in that, depending on the location of X_R and the mixture weight w , the distribution of X_{mix} could have a longer, rather than a shorter, right tail than X_P .

In an SMS task, the interest is in the distribution of X_P , and whereas X_P is possible to measure at short ISIs, at long ISIs X_P can be considered a latent variable. Using the sample mean and *SD* to estimate the distribution of X_P is then problematic because reactive responses may confound estimates of constant error and timing variability, resulting in considerable negative bias. This happens because the distribution of X_A has a shorter right tail than the distribution of X_P . Due to the asymmetry of X_A , a sample mean estimate will be biased toward a negative mean asynchrony, and a sample *SD* estimate will be smaller than the actual *SD* of X_P , due to the right tail of the distribution of X_A being less spread out than the distribution of X_P . In other words, using moment estimators will make it appear that participants are responding earlier and more accurately than they actually are.

At first, it might seem that the distributions of X_P and X_R could be separated using a standard mixture-model approach, by modeling an outcome as being generated by first selecting one of the underlying distributions and then using that distribution to generate the outcome. This approach does not consider that the distribution of X_A was the result of taking the *minimum* of two random variables, and so will not result in a consistent estimator of the parameters underlying X_A . Another

approach would be to implement the fully generative model shown in Fig. 2. This would, however, require specifying not only the distribution for X_P , but also that for X_R . Although many possible distributions could be assumed for reactive responses (van Zandt, 2000), it is not known whether these are applicable for the reactive responses in an SMS task, and misspecification of this distribution would impact the parameter estimates of both X_P and X_R .

One way of estimating the distribution of X_P is by noticing that the distribution of X_R does not depend on the ISI and is left-bounded by participants' ability to react to the pacing tones. Therefore, the distribution of X_P can be retrieved by modeling the asynchronies as coming from a right-censored normal distribution (see Fig. 3) with the censoring threshold being selected to exclude reactive taps—for example, set to 100 ms. That is, all asynchronies below 100 ms would be assumed to be direct observations of X_P , whereas asynchronies above 100 ms would also be assumed to be observations of X_P , but the actual asynchrony would be disregarded and the only information retained would be that the observation fell in the range $(100, \infty)$ ms. The rest of this article will focus on a model implementing these assumptions and Bayesian estimation of the parameters of this model.

Bayesian modeling of sensorimotor synchronization data

Bayesian methods of data analysis are becoming increasingly popular in psychology and other fields (Andrews & Baguley, 2013; Kruschke, 2011b). The rationale for using Bayesian modeling is the ease with which nonnormal data can be modeled, hierarchical dependencies in the data can be specified, and prior knowledge, such as task-specific constraints, can be included (Kruschke, 2011a). The model described here is Bayesian and is implemented using Markov chain Monte Carlo (MCMC) methods. Both the terminology and the philosophy of Bayesian statistics are different from those of classical frequentist statistics. The following sections assume some acquaintance with Bayesian statistics and MCMC methods, and many good text cover these topics—for example, the books by Kruschke (2011a), Lunn, Jackson, Best, Thomas, and Spiegelhalter (2012), and Lee and

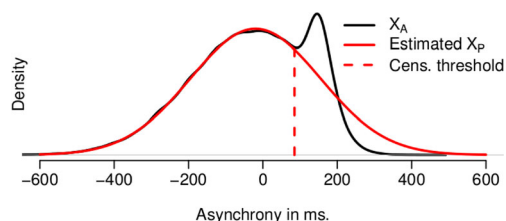


Fig. 3 Schematic diagram of the fit of a right-censored normal distribution to the theoretical distribution of timing asynchronies shown in Fig. 2

Wagenmakers (2014). Because the data from an SMS experiment often are hierarchically organized—that is, they consist of many participants performing multiple trials—the model is presented in a hierarchical as well as in a nonhierarchical version. The advantages of hierarchical modeling are described well by Gelman and Hill (2006).

A number of Bayesian models are designed to analyze reaction time data (Craigmile, Peruggia, & van Zandt, 2010; Farrell & Ludwig, 2008; Rouder, Sun, Speckman, Lu, & Zhou, 2003), and these models and the model presented in this article have much in common, in that they are hierarchical, deal with timing data, and model data as coming from nonnormal distributions. Differences include the types of distributions used and what type of task-related information is incorporated in the model. The present model makes two important assumptions. First, the asynchronies are treated as independent observations. This is strictly not true, because asynchronies tend to be autocorrelated when they are considered as forming a time series (Chen et al., 1997). Estimates of constant error and timing variability do not, then, constitute an exhaustive description of the underlying data. Still, these measures are two of the most common in the literature (Mates et al., 1994), and together they form a useful summary of timing performance. Second, the asynchronies are assumed to be from trials with the same ISI. In an experimental setup in which many ISI levels are used, the constant error and timing variability have to be estimated separately for each ISI level.

Below, the nonhierarchical version of the model is first described, which is then extended into a fully hierarchical model. The model uses priors that are vague and noninformative, except with regard to certain task-specific constraints that are used to inform the priors. How to extend the model in order to use more informative priors and how to model interresponse intervals instead of asynchronies is described in the [supplementary text](#), and software implementing the model is freely available at https://github.com/rasmusab/bayes_timing. The software is implemented using the R statistical environment (R Development Core Team, 2012) and the JAGS framework (Plummer, 2003). The JAGS framework takes a model definition and automatically generates a sampling scheme using Gibbs sampling. Because the technical details regarding sampling schemes are handled by JAGS, it is relatively straightforward to extend and modify the models given below—for example, by changing the priors or the distributional assumptions.

The nonhierarchical model

The timing asynchronies (Y) are modeled as coming from a normal distribution in which asynchronies exceeding a threshold c are assumed to be right censored. Strictly, it is the values that are censored, and not the actual distribution, but the procedure of first censoring values above a threshold and then

modeling them as being normally distributed could be seen as defining a right-censored normal distribution with three parameters: μ , σ , and c . In the present model, the threshold c is fixed at a value that is assumed to censor asynchronies that could be the results of reactive responses. A conservative value for c would be, for example, 100 ms, which is a reasonable tradeoff between censoring reactive responses and maintaining predictive responses.

The prior distributions for both μ and σ use noninformative Jeffreys (1946) priors, but with additional constraints imposed by the SMS task. In an SMS task, a single asynchrony cannot be farther away from a stimulus onset time than half the ISI. For example, if the ISI is 600 ms and a response is registered 250 ms before a stimulus onset, this will be encoded as a –250-ms asynchrony, but if that response was registered 375 ms before the same stimulus onset, it would instead be encoded as a 225-ms asynchrony with respect to the preceding stimulus onset. If asynchronies are bounded by the interval $[-\text{ISI}/2, \text{ISI}/2]$ ms, then μ is also bounded by this interval. The Jeffreys prior on μ is an unbounded uniform distribution (Lunn et al., 2012), but when using the task-specific information, the prior becomes a uniform distribution bounded on the interval $[-\text{ISI}/2, \text{ISI}/2]$ ms. Given that the difference between any two asynchronies can be at most the length of the ISI, σ can be at most $\text{ISI}/2$ ms.¹ The Jeffreys prior on σ is an unbounded uniform distribution on $\log(\sigma)$ (Lunn et al., 2012), but using the present task-specific constraints bounds the distribution on the interval $[\log(1), \log(\text{ISI}/2)]$ ms. The lower bound on $\log(\sigma)$ is set to $\log(1)$ because the probability that the *SD* of a participant's asynchronies would be close to 1 ms is negligible. The full specification of the nonhierarchical model is then:

$$\begin{aligned} Y_i &\sim \text{Right-Censored-Normal}(\mu, \sigma, c), \\ \mu &\sim \text{Uniform}(-\text{ISI}/2, \text{ISI}/2), \\ \log(\sigma) &\sim \text{Uniform}[\log(1), \log(\text{ISI}/2)], \end{aligned}$$

where Y_i is the i th asynchrony.

Point estimates of the parameters μ and σ can be calculated by taking the mean or the median of their respective posterior distributions (Robert, 2007). These estimates can then be used as a “drop-in” replacement for the sample mean and *SD* estimates of constant error and timing variability. This approach is useful in the case in which a researcher wants to avoid the bias associated with using the sample mean and *SD* but prefers a classical analysis of the point estimates rather than a fully Bayesian analysis.

It should be noted that a point estimate generated using the Bayesian model above and a maximum likelihood approach would be very similar. This is because the maximum a posteriori estimate from a Bayesian model with flat priors

and a maximum likelihood estimate are identical (Hastie, Tibshirani, & Friedman, 2009). When only a point estimate is required, it can therefore be convenient to use a maximum likelihood method, which is computationally more efficient and, perhaps, a better-known method than Bayesian estimation. Ulrich and Miller (1994) have described a maximum likelihood approach for fitting a right-truncated normal distribution that is applicable in this case. This method has also been implemented and is available at https://github.com/rasmusab/bayes_timing.

The hierarchical model

Hierarchical modeling is an elegant solution to the problem of analyzing a data set with repeated measurements (Kruschke, 2011a). It is an increasingly used technique in psychological research and is variously known as *hierarchical modeling*, *multilevel modeling*, or *mixed modeling* (Baayen, Davidson, & Bates, 2008). One of the reasons to use a hierarchical model is to better describe data that have a multilevel structure (Gelman & Hill, 2006), and the typical SMS experiment is inherently multileveled. Timing responses can be assumed to be related within a trial, between trials, within a participant, and between participants. Furthermore, it is reasonable to assume that relations exist between the timing responses at different tempi; for example, a participant who produces highly variable asynchronies at an ISI of 500 ms will probably produce highly variable asynchronies at longer ISIs. The hierarchical model described below does not take all of the possible multilevel relations into account. Asynchronies from different trials produced by the same participant are assumed to have the same distribution; that is, no training or exhaustion effect is assumed to exist. Although it is well known that timing variability increases as the tempo gets slower, this relation does not seem to be linear (Grondin, 2012; Repp & Su, 2013), and it is not known whether it follows any simple function. Therefore, the relationship between the asynchronies produced at different tempi is not part of the model.

What this model adds over the nonhierarchical version is that the relation between participants' timing performance is modeled, allowing measurements made on all participants to inform the parameter estimates of single participants. The hierarchical formulation also facilitates investigating individual differences as constant error and timing variability are estimated at both the individual and group levels. The mean of the censored normal distribution of the j th participant, μ_j , is modeled as coming from a normal distribution with mean μ_μ and *SD* σ_μ . The subscript μ is used to indicate that μ_μ and σ_μ are hyperparameters of the prior distribution on μ_j . The *SD* of the censored normal distribution, σ_j , is modeled as coming from a log-normal distribution with the parameters μ_σ and σ_σ (as was proposed by Lunn et al., 2012). Note that these parameters are not the mean and *SD* of the log-normal

¹ This maximum *SD* would occur in the unlikely event that half of the asynchronies were $-\text{ISI}/2$ ms and half were $\text{ISI}/2$ ms.

distribution. Through reparameterization, this distribution can be described by its mean, m_σ , and SD , s_σ (Limpert, Stahel, & Abbt, 2001). The rationale for using the log-normal distribution is that it allows for modeling participants' timing variability on the same scale as σ_j , which makes the hyperparameters m_σ and s_σ easy to interpret; the posterior distributions of m_σ and s_σ can be directly used to estimate the population's mean timing variability and the SD of the population's mean timing variability. A popular choice for the prior distribution on a variability parameter is otherwise $1/\sigma^2 \sim \text{Gamma}(\cdot, \cdot)$ (Gelman, 2006), but the problem with using this prior is that it is not defined on the scale of the SD of the asynchronies, which makes it hard to interpret the posterior distributions of its parameters. This is fine in the case in which the constant error is the main interest and the SD of the asynchronies is considered a nuisance parameter. When analyzing SMS data, however, it is often the case that timing variability is the main interest (Repp, 2005).

The prior distributions on μ_μ and m_σ are the same as those used for the parameters μ and σ in the nonhierarchical model. The prior distributions for the SD parameters σ_μ and s_σ are uniform distributions (as was proposed by Gelman, 2006) with their left boundaries at 0, and with task-related constraints defining the right boundaries as for μ and σ in the nonhierarchical model. The left boundary for σ_μ is set to $ISI/2$ because this is the largest possible between-participants mean asynchrony SD , since all μ_j s are bounded within the interval $[-ISI/2, ISI/2]$. The largest possible within-participants asynchrony SD is $ISI/2$, which implies that the largest possible value of σ_σ , the between-participants SD of the asynchrony SD , is $ISI/4$. The full specification of the hierarchical model is then

$$\begin{aligned} Y_{ij} &\sim \text{Right-Censored-Normal}(\mu_j, \sigma_j, c_j), \\ \mu_j &\sim \text{Normal}(\mu_\mu, \sigma_\mu), \\ \sigma_j &\sim \text{Log-Normal}(\mu_\sigma, \sigma_\sigma), \\ \mu_\mu &\sim \text{Uniform}(-ISI/2, ISI/2), \\ \sigma_\mu &\sim \text{Uniform}(0, ISI/2), \\ \mu_\sigma &\sim \log(m_\sigma) - \sigma_\sigma^2/2, \\ \sigma_\sigma &\sim \sqrt{\log(s_\sigma^2/m_\sigma^2 + 1)}, \\ \log(m_\sigma) &\sim \text{Uniform}[\log(1), \log(ISI/2)], \\ s_\sigma &\sim \text{Uniform}(0, ISI/4), \end{aligned}$$

where Y_{ij} is the i th asynchrony of the j th participant. After Kruschke (2011a), a graphical model diagram of this model is shown in Fig. 4.

An example using the hierarchical model

The data from Bååth and Madison (2012) were analyzed using the hierarchical model. The study included responses from 30 participants who synchronized finger taps to isochronous tone sequences with five different ISIs: 600, 1,200, 1,800, 2,400,

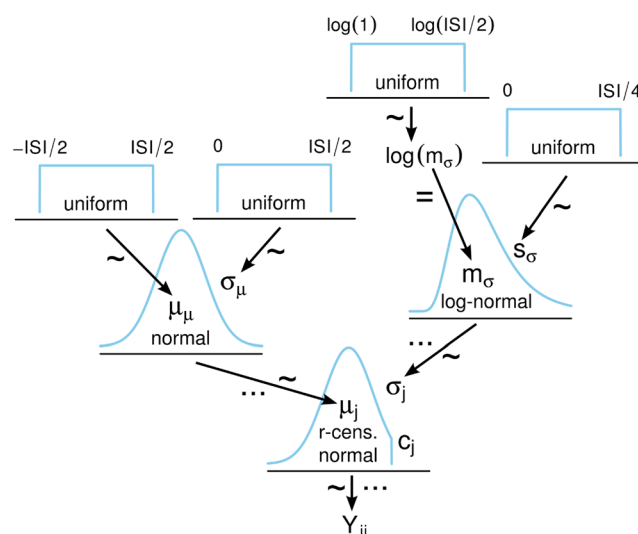


Fig. 4 Model diagram showing the specification of the fully hierarchical model. Here, Y_{ij} is the i th asynchrony of the j th participant

and 3,000 ms. Because the data include repeated measurements, with each participant producing many asynchronies at each ISI level, the hierarchical model was fitted to the data separately for each ISI level.

After a Bayesian model has been fitted, the estimated parameters can be investigated in many ways. Depending on where the interest lies, the parameters can be examined at the participant level (σ_j and μ_j) or the group level (m_σ and μ_μ). If individual differences are of interest, the variance components s_σ and σ_μ can be inspected, because they index the degree to which the participant-level parameters differ. Because the model is fully Bayesian, estimates for all of the parameters are readily available after the model has been fitted, including reliability measures in the form of credible intervals.

Because tapping variability is often of interest in SMS studies (Repp, 2005), the group-level timing variability and the variability on the participant level were investigated further. Figure 5 shows the point estimates and 95% credible intervals, extracted from the model, of the group mean asynchrony SD (the m_σ parameter in the model) and the participant-level asynchrony SD s for three participants (σ_1 , σ_2 , and σ_3) at the five ISI levels. The group mean SD increases almost linearly as a function of ISI, but we also see that there are large differences between the three participants. Using the posterior distribution of the fitted model, it was now possible to investigate any relation between the parameters. For example, there was an 89% probability that Participant 1 had a lower timing variability than Participant 2 at an ISI of 1,800 ms, and there was a 99% probability that the increase in timing variability between the ISIs of 1,800 and 2,400 ms was larger for Participant 1 than for Participant 2. A script that fully replicates these calculations and the fitting of the model can be found at https://github.com/rasmusab/bayes_timing, together with the full data set from Bååth and Madison (2012).

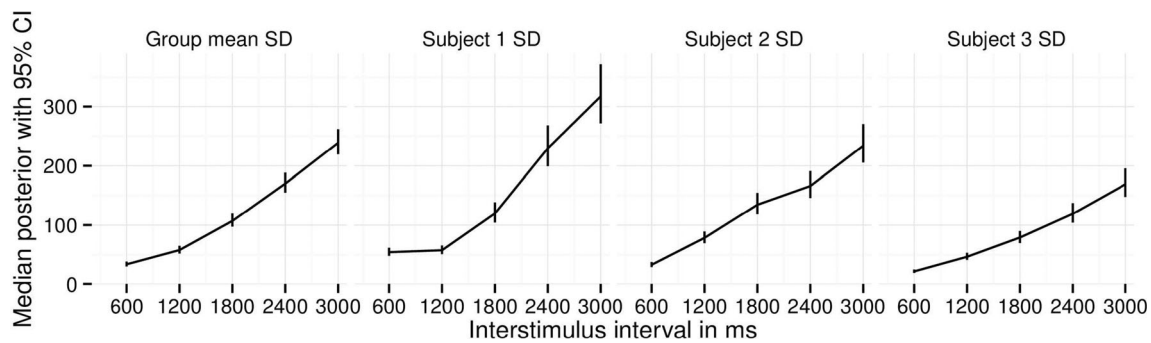


Fig. 5 Point estimates with 95% credible intervals of the group timing variability (measured as the asynchrony *SDs*) and the participant timing variability for three participants from the hierarchical model, fitted to the data from Bååth and Madison (2012)

Evaluation of the model

In order to evaluate the model, it was applied to both simulated data and the experimental data from a study by Repp and Doggett (2007). The estimates generated by the model were compared with the common estimates of constant error and timing variability, the sample mean, and the sample *SD*. These estimators will be referred to as the *moment estimators*. The Bayesian model can be used to generate point estimates of the constant error and timing variability for further analysis, or the hyperparameters of the model can be used directly to draw inferences. In order to be able to compare the Bayesian model with the moment estimators, evaluations were made using the first approach of analyzing estimated point values. There are different methods for calculating point values from the marginal posterior distributions of a Bayesian model (see Robert, 2007). Here the median of the marginal posterior distribution of the parameter of interest was taken as the point estimator. When fitting the Bayesian models, the censoring limit c was fixed to 100 ms.

Evaluation using a simulated data set

The point with a simulation study is to simulate data using a distribution in which the parameters are known, since comparisons can then be made as to how well different estimators retrieve the true parameters using the simulated data. This makes it possible to compare different estimators and to gauge the magnitude of estimation error when the models are applied to real data.

Simulation of timing asynchronies In order to simulate timing asynchronies, a number of assumptions have to be made. As was argued earlier, the timing responses will be assumed to come from two sources: Either a response is predictive, resulting from a prediction of the timing of the target stimulus, or a response is reactive, resulting from a reaction to the target stimulus. Furthermore, as is shown in Fig. 2, the distribution of predictive responses is assumed to be a normal distribution with a mean and *SD* that are dependent on the ISI. The reactive responses are assumed to be distributed as an

exponentially modified Gaussian (ex-Gaussian) distribution—a right-skewed distribution that has been used to describe the distribution of reaction time responses (Hohle, 1965; Palmer, Horowitz, Torralba, & Wolfe, 2011). The distribution of the reactive responses is assumed to be independent of the ISI. If the predictive and reactive responses are represented by the random variables X_P and X_R , the actual timing responses are distributed as $\min(X_P, X_R)$. This assumes that a timing response is initiated by whichever of the reactive and predictive responses is triggered first, and also that participants tend not to respond twice to target stimuli.

The distribution of the timing responses has five parameters; μ_P and σ_P of the normal distribution for the predictive responses, and μ_R , σ_R , and λ_R of the ex-Gaussian distribution for the reactive responses. In order to simulate the timing responses at different ISIs, these parameters needed to be assigned reasonable values. For μ_P and σ_P , such values were generated by taking the sample mean and *SD* of the asynchronies at different ISIs using the finger-tapping data from the group of musicians in the study by Repp and Doggett (2007) (see Table 1). Although it has been argued in the present article that there are better ways to estimate these parameters than using the sample mean and *SD*, the performance of

Table 1 Values of μ_P and σ_P that were used when simulating the timing asynchronies at different interstimulus intervals (ISIs)

ISI	μ_P	σ_P
1,000 ms	−19 ms	28 ms
1,250 ms	−23 ms	37 ms
1,500 ms	−30 ms	49 ms
1,750 ms	−30 ms	61 ms
2,000 ms	−17 ms	81 ms
2,250 ms	−22 ms	91 ms
2,500 ms	−12 ms	106 ms
2,750 ms	−20 ms	123 ms
3,000 ms	−9 ms	155 ms
3,250 ms	−41 ms	157 ms
3,500 ms	−24 ms	209 ms

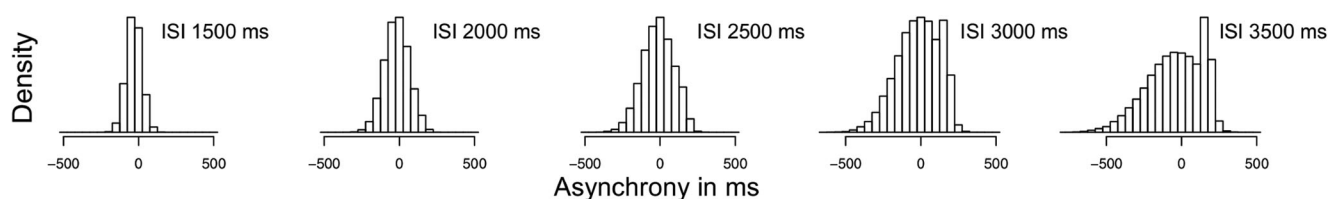


Fig. 6 Distribution of the simulated timing asynchronies for a subset of the ISI levels, generated according to the procedure described in the Simulation of Timing Asynchronies section

the estimators used in this specific case were not of great importance; what mattered was that the values of μ_P and σ_P should be likely values of the constant error and timing variability. Reasonable values for μ_R , σ_R , and λ_R were generated by fitting an ex-Gaussian distribution to the simple auditory reaction time data from a study by Löwgren, Bååth, Lindgren, Sahlén, and Hesslow (2014). The parameter values used for the ex-Gaussian distribution were $\mu_R = 157$ ms, $\sigma_R = 12.5$, and $\lambda_R = 0.031$ ms, which are in agreement with response distributions found in the literature (e.g., Ulrich & Stapf, 1984). Distributions of the simulated data using these parameter values are shown in Fig. 6 (cf. the actual asynchrony distributions from Repp & Doggett, 2007, in Fig. 8 below). For each of the 11 ISI levels, 500 batches of 90 timing responses each were simulated.

Comparison with the moment estimators Because the simulated data were not hierarchical, all data points from the same ISI level shared the same true parameter values, and only the nonhierarchical model and the moment estimators were compared. For each of the in total $500 \times 11 = 5,500$ batches, the Bayesian model was fit with the JAGS framework (Plummer, 2003), using 1,000 burn-in steps followed by 5,000 MCMC samples, and the resulting fits were used to calculate point estimates for μ_P and σ_P . Similarly, the moment estimators were also used to calculate point estimates for μ_P and σ_P . Figure 7 shows the mean differences between the estimated parameters and the true parameters, with a relative parameter estimate of 0

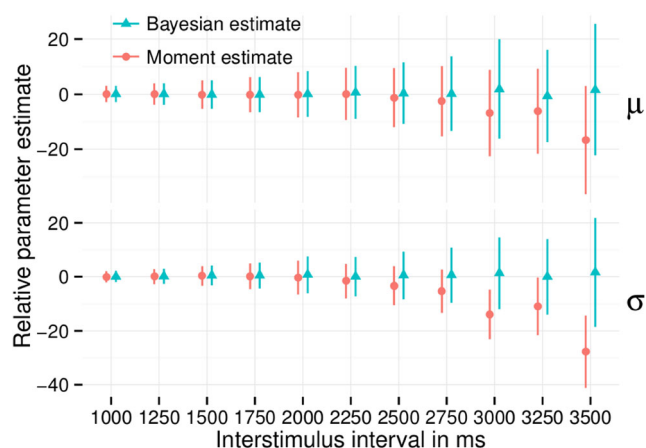


Fig. 7 Mean errors of the moment estimates and the Bayesian estimates compared to the actual parameter values. The error bars show the SDs of the estimates

indicating no difference between the true parameter and the mean of the estimated parameters. Up to an ISI of 2,000 ms, both the Bayesian model and the moment estimators performed similarly, but from an ISI of 2,000 ms the moment estimators increasingly underestimated the true values of μ and σ . At an ISI of 3,500 ms, the mean differences between the Bayesian estimates and the true parameter values were 1.6 ms for both μ and σ , but for the moment estimators the mean differences were 16.6 ms for μ and 27.8 ms for σ .

Reanalyzing the data of Repp and Doggett (2007)

In a study by Repp and Doggett (2007), finger-tapping data were collected from eight musicians and 12 nonmusicians synchronizing to isochronous sound sequences, using ISI levels ranging from 1,000 to 3,500 ms.² The distributions of timing asynchronies for a subset of those ISI levels are shown in Fig. 8. Notice that the distributions of asynchronies for the long ISI levels exhibit the same pattern shown in Fig. 1; that is, large numbers of responses occur around 200 ms after the stimulus onset.

The data of Repp and Doggett (2007) were reanalyzed by calculating point estimates of the constant error and timing variability for all participants and ISI levels, using both the moment estimators, as in the original article, and the hierarchical Bayesian model. This model was fit separately to the data from the musicians and the data from the nonmusicians with the JAGS framework, using 1,000 burn-in steps followed by 15,000 MCMC samples. The differences between the moment and Bayesian estimates are shown in Figs. 9 and 10. On the basis of the simulation study, the moment estimates should have a tendency to underestimate the constant error and timing variability, and the results of the present analysis support this notion, because the estimates of the Bayesian model are higher than the moment estimates at the long ISI levels. At short ISI levels, this underestimation will be negligible, but at longer ISI levels it will have more of an impact. Whether an analysis would benefit from avoiding underestimation, then, would depend on the ISI range of the study and whether underestimation would impact the conclusions of the study. The Bayesian model estimates and the moment estimates start to diverge when the

² Due to a technical error, the timing asynchronies are shifted +15 ms in the original data (Repp, 2008). In the subsequent analysis, this shift has been corrected.

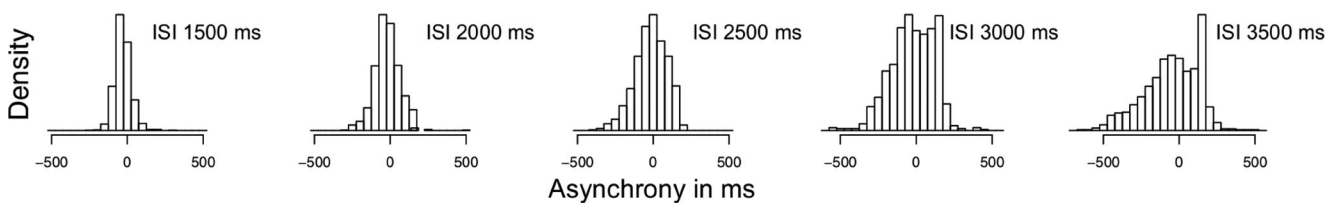


Fig. 8 Distribution of the timing asynchronies from Repp and Doggett (2007) for a subset of the ISI levels

ISI is longer than 2,500 ms, both for the simulated data and for the data from Repp and Doggett. Thus, one might want to consider using a method that avoids underestimation when analyzing a data set that includes ISIs slower than 2,500 ms.

Except for resulting in better parameter estimates, how does the interpretation of Repp and Doggett's (2007) data change when Bayesian estimates are used rather than the moment estimates? One example of what the Bayesian estimates change, relative to the moment estimates, is the interpretation of the source of the reactive responses. Mates et al. (1994) argued that the reason that reactive responses start occurring at long ISIs is a qualitative change in participants' response strategy, due to participants trying to minimize synchronization error. When the ISIs are long enough that predictive responses result in a large expected synchronization error, a better strategy might be to react to the stimulus tone, since this would result in an expected synchronization error smaller than around 200 ms, the average auditory reaction time. Alternatively, Repp and Doggett (2007) argued that reactive responses are not due to a change in response strategy, but rather to participants tending to produce reactive responses when failing to produce a predictive response long enough after the stimulus tone that a reactive response is possible. To evaluate this possible explanation, they used each participant's estimated constant error and timing variability to predict the percentage of reactive responses, under the

assumption that the predictive responses would be normally distributed. The percentages of predicted reactive responses were then compared with the actual percentages of reactive responses at the different ISI levels, labeling all responses later than 100 ms as reactive. The predicted percentage of reactive responses was found to be similar to the actual percentage, and Repp and Doggett concluded that "no special strategy of reacting to the tones needs to be assumed" (p. 371). The predicted percentage was, however, slightly lower than the actual percentage, and this difference could still be explained by, for example, a change in response strategy.

A reason for the slightly lower estimates of the percentage of reactive responses might be that the constant error and timing variability were underestimated due to the use of the moment estimators. Using the Bayesian estimates to predict the number of reactive responses yielded a much closer correspondence with the actual percentage of reactive responses, especially at slow ISIs, as is shown in Fig. 11. Consequently, the Bayesian estimates change the interpretation of the data to more strongly support Repp and Doggett's (2007) explanation than their original analysis based on moment estimates.

The power of a hierarchical model

In order to investigate the utility of a hierarchical model, the data from the musicians in Repp and Doggett (2007) were

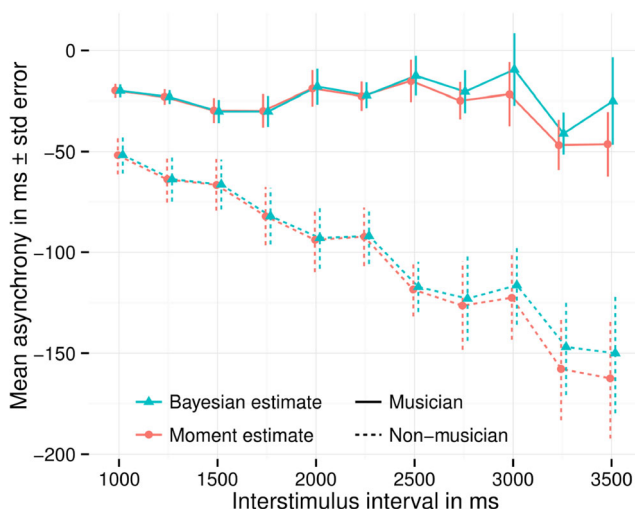


Fig. 9 Grand means of the sample mean estimates and the Bayesian estimates of the mean asynchronies for the musicians and nonmusicians in Repp and Doggett (2007)

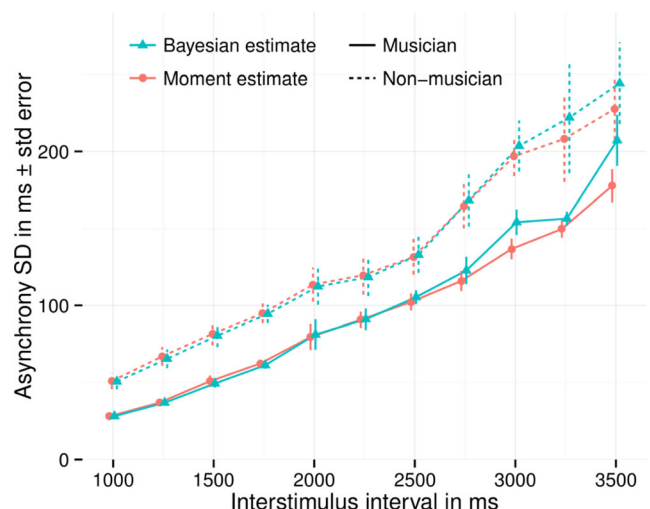


Fig. 10 Grand means of the sample SD estimates and the Bayesian estimates of asynchrony SDs for the musicians and nonmusicians in Repp and Doggett (2007)

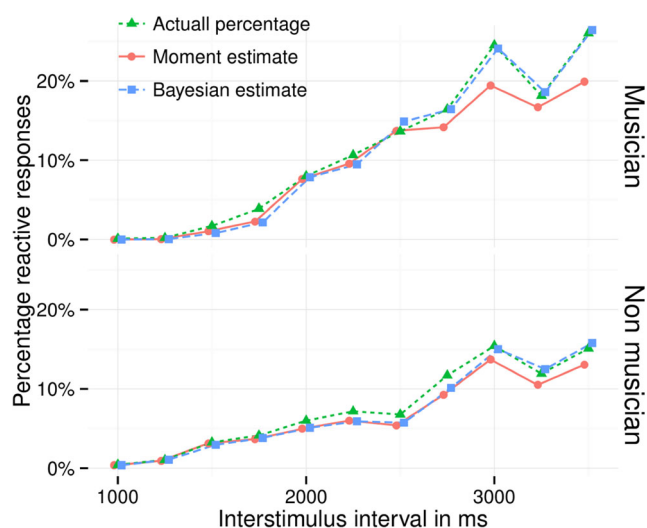


Fig. 11 Actual numbers of reactive responses for the musicians and nonmusicians in Repp and Doggett (2007), compared to the predicted numbers of reactive responses using the moment estimates and the Bayesian estimates

used. Four partitions of the data set were formed, such that each of the four tapping trials at each ISI level for each participant was randomly assigned to a partition. For each partition, the constant error and tapping variability were estimated using the moment estimators, the nonhierarchical model, and the hierarchical model. The Bayesian models were fit with the JAGS framework, using 1,000 burn-in steps followed by 15,000 MCMC samples. For each estimation method, this yielded 4 (number of partitions) \times 8 (number of participants) \times 11 (number of ISI levels) = 352 estimates of constant error and tapping variability. In the simulation study, it was possible to compare the estimated parameter values with the actual parameter values, but here there were no “true” parameter values. An alternative would be to compare the parameter estimates from the partitions, which are estimated using only one fourth of the available data, with estimates that used the whole data set. The estimates calculated using the whole data set are assumed to be closer to the true parameters, and by comparing these with the estimates calculated using the partitioned data, it was possible to evaluate how well the three estimation methods retrieve the whole-data estimates. The whole-data estimates were estimated using the nonhierarchical Bayesian model.

Figure 12 shows the mean absolute errors of the three estimation methods applied to the partitioned data, as compared to the estimates based on the whole data. The mean errors of the hierarchical estimates are consistently lower than the errors of the two other estimation methods. Averaged over the ISI levels, the medians³ of the absolute errors of the hierarchical Bayesian estimates were 12% less than for both the

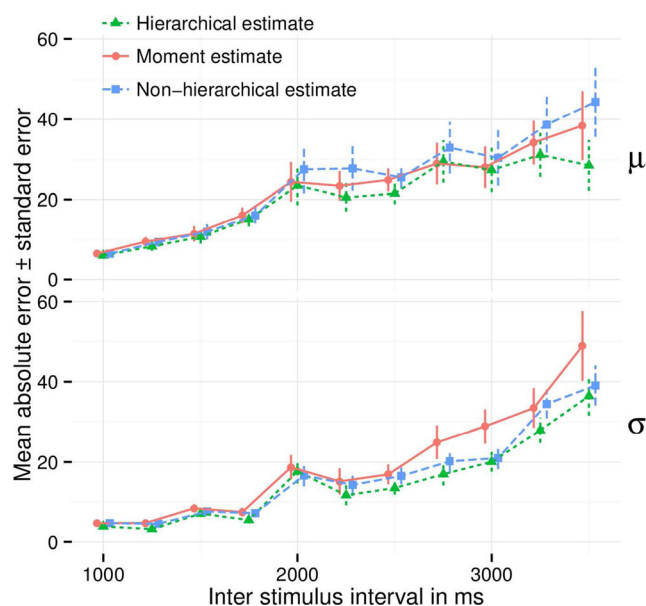


Fig. 12 Mean absolute errors of the three estimators applied to the partitioned data from Repp and Doggett (2007), as compared to the estimates using the full data set. The errors of the hierarchical Bayesian estimates are consistently lower than the errors of the moment estimates and of the nonhierarchical Bayesian estimates

nonhierarchical Bayesian and the moment estimates. Because a hierarchical model benefits from there being many participants in the data set, this better performance of the hierarchical model occurred in spite of only eight participants being included in the analysis. For a data set with even fewer participants, or for participants that perform very differently from each other, a hierarchical model is not likely to improve the estimates much over a nonhierarchical model. However, for the common case in which many participants have comparable performance, using a hierarchical model will likely result in better estimates.

Discussion

In studies dealing with sensorimotor synchronization (SMS) and rhythm production, two of the main parameters of interest are constant error and timing variability. It is common to estimate these parameters by calculating the sample mean and *SD*, but using these moment estimators is problematic in two respects. First, the moment estimators tend to underestimate the constant error and timing variability at long ISIs. This is due to a tendency of participants to overshoot the target interval and instead to react to the target stimulus, resulting in a left-skewed and bimodal distribution for both the stimulus-to-response asynchronies and interresponse intervals. Second, when a data set includes many participants, moment estimators fail to model the hierarchical structure of the data, and as a result all available data are not used when estimating the parameters.

³ The median was used here as the measure of central tendency because the distributions of relative errors were heavily right skewed.

The Bayesian model presented in this article addressed the first problem by treating the predictive timing responses as a partially latent variable and by modeling the timing asynchronies using a right-censored normal distribution. The second problem was addressed by modeling the hierarchical structure of the data. The model was compared with the moment estimators and was shown to be less biased toward low estimates of constant error and timing variability and to yield more accurate estimates when applied to a hierarchical data set with multiple participants. It was also shown that the Bayesian estimates changed the interpretation of the data from a study by Repp and Doggett (2007).

The focus of the evaluation of the Bayesian model was on contrasting it with the moment estimators. This choice was made because moment estimators are arguably the most common estimators of constant error and timing variability in the literature. The case made against using these estimators when dealing with SMS data could also be made against other methods that do not consider the skewness and hierarchical structure of such data. Some methods other than Bayesian modeling could also address these two issues. A right-truncated normal distribution could, for example, be fit to the data from an SMS task using the approach described by Ulrich and Miller (1994), or the hierarchical structure could be modeled using mixed-effects modeling (Baayen et al., 2008). The combination of these two approaches is, however, more straightforward in a Bayesian framework.

All comparisons between the Bayesian model and the moment estimators were made using point estimates generated by the Bayesian model. This was done to facilitate the comparison with the moment estimators. Although it is certainly possible to use the Bayesian model in this way, it disregards the much more useful approach of using the posterior probability distributions of the parameters for inference. In many cases, the latter approach would make more sense. Why go through the trouble of estimating point values and analyzing them when it is possible to directly analyze the distributions of the hyperparameters already specified in the model? Using the hierarchical version of the model, it is also possible to make inferences regarding the population timing variability by using the posterior probability distribution for the m_σ parameter, and regarding the population constant error by using the posterior probability distribution of the μ_μ parameter. In order to compare two groups of participants, the data of each group could be fit separately using the hierarchical Bayesian model, and then the credible differences between the group parameters could be investigated. One of the main advantages of doing a full Bayesian analysis is that all model parameters are estimated, including measures of uncertainty, so that comparisons and inferences can readily be made regarding any parameter or generated quantities.

Because the model is implemented in the flexible modeling language JAGS, it will be straightforward to extend it. A

possible extension would be to include additional predictor variables in the model, allowing the timing variability and constant error to vary not only by ISI, but also by, for example, participant group or task condition. Another extension would be to introduce a functional dependency on the timing variability or constant error between ISI levels. For example, one could assume that the scalar property (Gibbon, Church, & Meck, 1984) holds for timing variability by adding the assumption that the asynchrony *SD* increases linearly as a function ISI. The [supplementary text](#) describes how one could add such an assumption to the hierarchical model. That text also describes how one could model a correlation in a participant's timing performance between ISI levels. The purpose of the model presented here was to model the distributional properties of SMS data. In doing that, it did not consider the time series properties of the data, such as the serial dependency of the responses. A further extension of the model would be to combine it with a time series model of SMS, such as the one developed by Vorberg and Wing (1996). Because that model does not separate predictive responses from reactive responses, it should, like the moment estimators, be biased toward low estimates of timing variability when the timing responses include reactive responses.

Author Note This research was supported by the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, Swedish Research Council Grant Number 349-2007-8695. I thank Bruno H. Repp and Rebecca Doggett for providing their data for reanalysis. In addition, I thank Geoffrey R. Patching and Christian Balkenius for useful comments on various aspects of this work.

References

- Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, *66*, 1–7. doi:10.1111/bmsp.12004
- Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, *48*, 66–79. doi:10.1006/brcg.2001.1304
- Bååth, R., & Madison, G. (2012). The subjective difficulty of tapping to a slow beat. In E. Cambouropoulos, C. Tsourgas, P. Mavromatis, & C. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition* (pp. 82–85). Jyväskylä, Finland: European Society for the Cognitive Sciences of Music. Retrieved from icmpc-escom2012.web.auth.gr/?q=node/67.
- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005
- Chen, Y., Ding, M., & Kelso, J. (1997). Long memory processes ($1/f^\alpha$ type) in human coordination. *Physical Review Letters*, *79*, 4501–4504. doi:10.1103/PhysRevLett.79.4501
- Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, *45*, 119–130. doi:10.1016/j.cortex.2007.09.008

- Craigmile, P. F., Peruggia, M., & van Zandt, T. (2010). Hierarchical Bayes models for response time data. *Psychometrika*, *75*, 613–632. doi:10.1007/s11336-010-9172-6
- Development Core Team, R. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*, 1209–1217. doi:10.3758/PBR.15.6.1209
- Forsman, L. J., Madison, G., & Ullén, F. (2009). Neuroticism is correlated with drift in serial time interval production. *Personality and Individual Differences*, *47*, 229–232. doi:10.1016/j.paid.2009.02.020
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 515–534. doi:10.1214/06-BA117A
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, *423*, 52–77. doi:10.1111/j.1749-6632.1984.tb23417.x
- Gottsdanker, R. (1982). Age and simple reaction time. *Journal of Gerontology*, *37*, 342–348. doi:10.1093/geronj/37.3.342
- Green, J. T., Ivry, R. B., & Woodruff-Pak, D. S. (1999). Timing in eye-blink classical conditioning and timed-interval tapping. *Psychological Science*, *10*, 19–23. doi:10.1111/1467-9280.00100
- Grondin, S. (2012). Violation of the scalar property for time perception between 1 and 2 seconds: Evidence from interval discrimination, reproduction, and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 880–890. doi:10.1037/a0027188
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. New York, NY: Springer.
- Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*, 382–386. doi:10.1037/h0021740
- Holm, L., Ullén, F., & Madison, G. (2011). Intelligence and temporal accuracy of behaviour: Unique and shared associations with reaction time and motor timing. *Experimental Brain Research*, *214*, 175–183. doi:10.1007/s00221-011-2817-6
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society A*, *186*, 453–461. doi:10.1098/rspa.1946.0056
- Kruschke, J. K. (2011a). *Doing Bayesian data analysis*. Amsterdam, The Netherlands: Elsevier.
- Kruschke, J. K. (2011b). Introduction to special section on Bayesian data analysis. *Perspectives on Psychological Science*, *6*, 272–273. doi:10.1177/1745691611406926
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, *51*, 341. doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2
- Löwgren, K., Bååth, R., Lindgren, M., Sahlén, B., & Hesslow, G. (2014). *The cerebellum, timing and age: Eyeblink conditioning, isochronous interval tapping and prism adaptation in children and adults*. (Manuscript in preparation).
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: Chapman & Hall/CRC.
- Madison, G., Forsman, L., Blom, B., Karabanov, A., Ullén, F., & Blom, O. (2009). Correlations between intelligence and components of serial timing variability. *Intelligence*, *37*, 68–75. doi:10.1016/j.intell.2008.07.006
- Mates, J., Müller, U., Radil, T., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, *6*, 332–340. doi:10.1162/jocn.1994.6.4.332
- Miyake, Y., Onishi, Y., & Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta Neurobiologiae Experimentalis*, *64*, 415–426.
- Moore, G. P., & Chen, J. (2010). Timings and interactions of skilled musicians. *Biological Cybernetics*, *103*, 401–414. doi:10.1007/s00422-010-0407-5
- Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 58–71. doi:10.1037/a0020747
- Plummer, M. (2003, March). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. Article presented at the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria. Retrieved from www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/Plummer.pdf
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*, 969–992. doi:10.3758/BF03206433
- Repp, B. H. (2006). Musical synchronization. In E. Altenmüller, M. Wiesendanger, & J. Kesselring (Eds.), *Music, motor control and the brain* (pp. 55–76). Oxford, UK: Oxford University Press.
- Repp, B. H. (2008). Metrical subdivision results in subjective slowing of the beat. *Music Perception*, *26*, 19–39. doi:10.1525/mp.2008.26.1.19
- Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: A comparison of musicians and nonmusicians. *Music Perception*, *24*, 367–376. doi:10.1525/mp.2007.24.4.367
- Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*, 403–452. doi:10.3758/s13423-012-0371-2
- Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Heidelberg, Germany: Springer.
- Roeckelein, J. E. (2008). History of conceptions and accounts of time and early time perception research. In S. Grondin (Ed.), *Psychology of time* (pp. 1–50). Bingley, UK: Emerald.
- Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, *68*, 589–606. doi:10.1007/BF02295614
- Schulze, H. (1992). The error correction model for the tracking of a random metronome: Statistical properties and an empirical test. In M. Françoise, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 275–286). Dordrecht, The Netherlands: Kluwer.
- Stevens, L. T. (1886). On the time-sense. *Mind*, *11*, 393–404.
- Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Pearson.
- Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*, 34–80. doi:10.1037/0096-3445.123.1.34
- Ulrich, R., & Stapf, K. H. (1984). A double-response paradigm to study stimulus intensity effects upon the motor system in simple reaction time experiments. *Perception & Psychophysics*, *36*, 545–558. doi:10.3758/BF03207515
- van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*, 424–465. doi:10.3758/BF03214357
- 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). San Diego, CA: Academic Press.