Detection of irregularities in auditory sequences: A neural-network approach to temporal processing

Detection of irregularities in auditory sequences: A neural-network approach to temporal processing *

Joachim Haß[1,2†], Stefan Blaschke[1,2], Thomas Rammsayer[1,3] and J. Michael Herrmann[1,4]

[1] *Bernstein Center for Computational Neuroscience Göttingen, Germany;*
[2] *University of Göttingen, Göttingen, Germany;*
[3] *University of Bern, Bern, Switzerland;*
[4] *University of Edinburgh, Edinburgh, UK*
[†] *E-mail: joachim@nld.ds.mpg.de*

Combining experiments and modeling, we study how the discrimination of time intervals depends both on the interval duration and on contextual stimuli. Participants had to judge the temporal regularity of a sequence of standard intervals that contained a deviant interval. We find that the performance to detect the deviant increases with the number of standards preceeding the deviant and decreases with the duration of the standard. While the effect of the standard duration can be explained by an neural network model that realizes the concept of multiple synfire chains, the position effect is incorporated into the model by an in-situ averaging process. Furthermore, experiments are discussed that are critical for the predictions of the model.

Keywords: time perception; sequence experiment; synfire chains; adaptation; serial memory system.

1. Introduction

Whenever we listen to somebody talking, or to a piece of music, we are presented with a sequence of stimuli that contain information in their duration and timing. For instance, the phonemes /ba/ and /pa/ differ by only 25 to 50 ms in their onset time but can still be reliably discriminated. This discrimination is even better when the phonemes are embedded in a sequence that forms natural speech.

While speech is a quite complex example of a sequence including semantic information, the neural mechanisms that enable discrimination of

*This study was supported by a grant from the BMBF in the framework of the Bernstein Center for Computational Neuroscience Göttingen, grant number 01GQ0432.

interval durations are not well understood even for much simpler sequences with purely temporal context, or even single intervals. Despite of numerous experimental and theoretical studies on the topic,[1–3] many ambiguities even about the psychophysical regularities remain. For instance, it is established that variability of time estimates σ_T increases as the intervals T to be estimated get longer,[1] but it is debated whether this increase is linear in T (Weber's law),[4] even steeper[5] or less steep than linear.[6] Similarly, for the question of whether context information enhance discrimination performance, there is both supporting[7,8] and contradicting evidence.[9,10]

We approach these two questions with an experimental paradigm (Sec. 2) where participants discriminate the duration of a variable interval from the constant standard durations of a number of previously presented intervals. The more standards are presented before the variable interval, the more context information is available. Varying the standard duration between blocks, we can simultaneously assess the decrease of discrimination performance with the interval duration. This experiment can be seen as a critical test between two classes of models: Static models like the classical pacemaker-accumulator system[11] predict no context effects at all, while dynamic models such as the multiple look model[7] predict an improved performance with increasing context information. Our results support the latter class, as performance increases at later positions of the variable interval.

In Sec. 3, we formalize the concept of the multiple look model[7] that improved performance results from averaging previous temporal information to reduce discrimination errors. The model provides a statistical framework for perception of both single intervals and sequences of intervals, as judgements about sequences are based on comparison of the individual intervals it is composed of. The model can be readily extended to account for more complex aspects of time perception and has partly been implemented as a biological neural network.[12] This implementation is based on general connection principles in the neocortex and does not depend on any modality-specific properties. Finally, Sec. 4 discusses the results and gives an outlook on further experiments.

2. Sequence experiment

2.1. *Method*

23 psychology undergraduates (mean age 23 years, 17 female) participated in the experiment for partial fullfillment of course requirements. They were naive to the purpose of the experiment, but were debriefed afterwards, in-

cluding feedback about their performance. In each trial, a sequence of seven intervals filled with white noise was presented via headphones. Six of these intervals were standard intervals (STI) with a constant duration, while the seventh was a variable interval (VTI). All intervals were separated by an inter-stimulus interval (ISI) with a duration identical to the STI. Participants were instructed that a deviating interval could be presented at any of the seven positions in the sequence and that if there was a deviating interval in the sequence, it would be the only one. The task was to decide whether the presented sequence was regular or irregular. As independent variables, we used the position of the VTI within the sequence (position 1 to position 7), which was randomized from trial to trial, and three different STI durations (50 ms, 150 ms and 250 ms), which were tested in separate blocks. The duration of the VTI was adjusted by a weighted up-down method.[13] Starting from an initial value, the duration was increased (step-up) if the participant had judged the sequence as "regular" and decreased (step-down) after an "irregular" judgment. The adjustments were done independently for each position of the variable interval. We chose the step sizes such that the VTI converged to the .75 percentile of the answer "irregular". As the dependent variable, we used the 75% detection threshold V_{75}, which can be computed from the percentile by subtracting the respective standard duration. The smaller the threshold, the better is the performance in detecting a deviant interval.

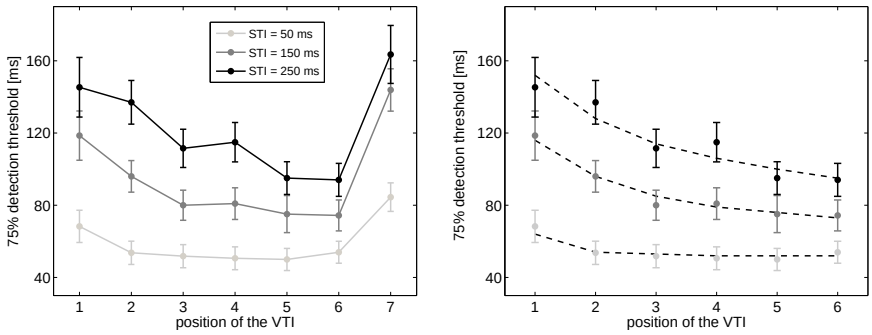

Fig. 1. 75% detection thresholds as a function of the position of the variable interval for three standard durations. The dots are means over participants with standard error bars. (Left) Data for all seven positions. (Right) Data for position one to six. The lines are fits of this data to Eqn. 6 (see Sec. 3.2.4). The color map is the same in both figures.

2.2. *Results*

Fig. 1 shows the mean values of V_{75} as a function of both the position of deviant and the duration of the STI. Three effects are apparent: The threshold increases with the standard duration, decreases from position one to six, and finally, increases again at the last position. To confirm these effects statistically, we performed a two-way ANOVA with the factors position and standard duration (levels as indicated above). The ANOVA showed highly significant effects for both factors, $F(6, 132) = 35.61$, $p < .001$, $\eta = 0.62$ and $F(2, 44) = 68.97$, $p < .001$, $\eta = 0.76$, respectively, and also an interaction, $F(12, 264) = 8.02$, $p < .001$, $\eta = .27$ (η is short for partial eta-squared). These results did not qualitatively change when the seventh position was excluded from the analysis (data not shown).

To further analyse the increase of V_{75} with the STI duration, we take the mean over all seven positions within an STI duration and calculate the Weber fraction $\bar{V}_{75}/STI$ for each STI duration. The values were 1.18, 0.64 and 0.49 for $S = 50$ ms, 150 ms and 250 ms, respectively. Decreasing Weber fractions are in accordance with standard theories of temporal perception[1] within this range of relatively short durations.

The significant decrease of the detection threshold from position one to six established that the number of STIs presented before a VTI indeed improves the performance to detect the deviant. This rules out static models of time perception[11] that would predict no such effect. However, also adaptive models that predict improved performance with increasing number of standards do not predict the decrease in performance at the final position.

3. Serial memory model

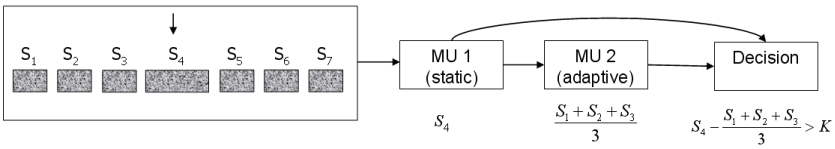

Fig. 2. Illustration of the model structure.

We now develop a model that aims to explain the findings in the present experiment. The model composes the representation of a temporal sequence from the representations of the individual intervals. The basic structure

of this model (Fig. 2) is similar to the classical pacemaker-accumulator system,[11] although its elements include mechanisms of adaptivity. Each interval is first encoded in a single-interval representation. We proposed a neural model for this encoding,[12] which we briefly present in the next section.

The second stage in the model is a memory system with two units (MU). These units also exist in the original model, but we make two modifications. First, the units are arranged in serial, e.g. the representation of interval one is first stored in MU1, but as the second interval is encoded, interval one is shifted to MU2 and interval two is stored in MU1 and so on (cf. Fig. 2). And second, while MU1 always contains a representation of the individual intervals, in MU2 the representations of all presented intervals are *averaged* to decrease variability.

Finally, in the third stage, the intervals represented in the two units are compared, and whenever the difference between the two exceeds a certain criterion, a deviant interval is detected. In this respect, the framework is similar to classical signal detection theory.

3.1. *Single interval representations by synfire chains*

A neural correlate of an interval representation should consist of a neural network that is able to store a wide range of time intervals with high precision. A neural structure that fullfills these requirements is the synfire chain,[14] a layered network of spiking neurons with feed-forward connectivity. This type of network has been shown to enable stable propagation of neuronal activity: If a sufficient number of neurons in the first layer is activated, neurons in the second layer also start spiking after some time, and this activation in turn is transmitted to the third pool, and so on. It has been shown that under broad conditions on the strength and timing of the initial activation[15,16] and the model parameters,[12] this propagation is stable, and activity travels along the layers like a wave. The propagation is linear in time and the temporal spread σ_L of the wave at each layer converges to a constant fixed point value in the range of milliseconds,[15,16] even in the presence of synaptic background noise. Therefore, the system is able to translate temporal information into a precise quasi-spatial code: The time elapsed since the initiation of the wave is represented in the position of the layer that is currently most active.[12]

Variability in the representation arises from the remaining temporal spread σ_L in the spikes. This constant error in each layer accumulates to smear the arrival time T of the wave at layer i to a standard deviation σ_T

proportional to the square root of i. Therefore, the Weber fraction σ_T/T decreases with the interval length like $1/\sqrt{T}$, consistent with the results of our experiment.

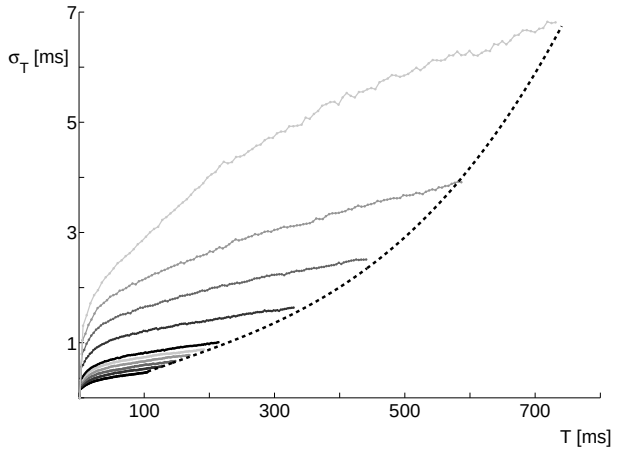

Fig. 3. Timing error, e.g. standard deviation σ_T of the total runtime of an activity wave as a function of time T for various transmission speeds of the chains. The solid curves depict simulation data and the dotted line represents the optimal timing error $\sigma_T^*(T)$ from Eqn. 1. It is close to the lower envelope of the simulation data.

For short intervals up to a few hundred milliseconds, this result of a decreasing Weber fraction has also been found previous experiments.[1,6] However, the steeper increase found at longer intervals (linear or even superlinear with duration[1,3–5]) is not easily reconciled with the accumulation of neuronal noise. For the steeper increase at longer intervals, there must be an additional constraint. In a synfire chain, the most obvious of such constraints is a finite chain length L. With a given mean transmission delay Δt from one pool to the next, the maximal interval to be represented is $T = \Delta t \cdot L$. For longer intervals, a chain with a higher value of Δt must be used. We could show that the speed of the activity wave can be manipulated by various model parameters,[12] but that any change in the synfire model that increases Δt also increases the spread of the spike times σ_L and thus, results in a larger timing error σ_T (Fig. 3).[12] From Fig. 3, it is also apparent that there exists an optimal chain for each interval of time to be encoded, meaning that the timing error σ_T is minimal. As the increase of this error with Δt is much larger (order 3) than the increase along the

layers (order 1/2), it is always optimal to use the entire length of the chain with the lowest ΔT that is able to encode the current interval. The form of the optimal timing error is[12]

$$\sigma_T^*(T) = \begin{cases} \sigma_{\min(\Delta t)} \cdot \sqrt{T} + D & \text{for } T \leq \min(\Delta t) \cdot L \\ A\,T^3 + B\,T^2 + C\,T + D & \text{otherwise,} \end{cases} \quad (1)$$

where $\sigma^2_{\min(\Delta t)}$ is the variance of the minimal transmission delay Δt. The dotted line in Fig. 3 shows a fit of the simulated data to Eqn. 1, which is close to the lower envelope of all chains.

The data in our experiment shows a decreasing Weber fraction, so all intervals can be assumed to be encoded by the fastest synfire chain available. We thus use the first column of Eqn. 1 to fit the data, resulting in values of $\sigma_{\min(\Delta t)} = 7.13$ ms and $D = 6.87$ ms. The fit gives a very good description of the data averaged over participants (97.5% of variance explained).

3.2. *Memory and decision stage*

3.2.1. *Stochastic framework*

We now formalize the adaptation in the serial memory system as an information processing model. A neural implementation of this system is in progress. The central stochastic variable is the difference $X_i(I)$ between the contents of the first and the second unit, where I is the time index of the intervals and i is the position of the deviant interval within the sequence. We use a general number of N intervals (set to seven to fit the present data). The intervals are presented during the first N time steps, while the computation of the difference $X_i(I)$ starts with the arrival of the second interval ($I = 2$) and is finished after $I = N + 1$ to complete a total of N comparisons.

Each interval represented by the spike patterns of the synfire chains is denoted by S_i and can be considered as a Gaussian random variable with the actual interval duration as the mean and the variance determined by the timing error σ_T (cf. Sec. 3.1). Assuming the same σ_T for both standard and deviant interval, the VTI is given by

$$S_i = S_v = \mathcal{N}(\bar{S}_v, \sigma_T^2) \quad (2)$$

and the STI is

$$S_j = S_s = \mathcal{N}(\bar{S}_s, \sigma_T^2), \qquad j \neq i. \quad (3)$$

With these definitions, we can write $X_i(I)$ in the general form

$$X_i(I) = S_I - \sum_{j=1}^{I} \frac{S_j}{I}. \tag{4}$$

The first term is the content of MU1 (the interval presented at position I), and the second term is the average in MU2 over all intervals presented before position I.

The difference $X_i(I)$ between the two units can now be used to evaluate the current interval in MU1 based on the information accumulated in MU2: If the difference exceeds a decision criterion K, the interval is judged to be irregular, otherwise it is judged to be regular. The probability for an "irregular" judgment is thus given by

$$P(X_i(I) > K) = \Phi\left(\frac{\bar{X}_i(I) - K}{\sqrt{\mathrm{Var}(X_i(I))}}\right), \tag{5}$$

where Φ is the standard normal distribution function and $\bar{X}_i(I)$ and $\mathrm{Var}(X_i(I))$ are the mean and the variance of $X_i(I)$, respectively.

In the framework of signal detection theory, the probability of a "irregular" response given the VTI in MU1 would correspond to the hit rate, while the probability of the same response, given an STI in MU1 would be the false positive rate. However, we are more interested in the joint probability to judge the entire sequence of N intervals as "irregular", since this response determines the 75% detection thresholds $\bar{S}_v$ in the experiment. This probability is given by

$$P(\text{“irreg}'') = 1 - P\left(\bigcap_{I=2}^{N+1} (X_i(I) < K)\right) = 1 - \prod_{I=2}^{N+1} P\left(X_i(I) < K\right). \tag{6}$$

The second equality holds under the assumption that all events are statistically independent. Note that Eqn. 6 gives an implicit equation for the 75% detection thresholds $\bar{S}_v$ at each of the positions of the VTI, given the probability $P(\text{“irreg}'')$, and the set of parameters$\{\sigma_T, K\}$. $P(\text{“irreg}'')$ is set to 0.75 in the current experiment, and the parameter set $\{\sigma_T, K\}$ can be used to fit the model to the experimental data.

3.2.2. *Results*

To use Eqn. 6 for determining the $\bar{S}_v$ values, we must calculate the probabilities $P(X_i(I) > K)$ for each value of i and I, and thus, the means and

variances of the respective variables $X_i(I)$. However, we can divide all possible combinations of i and I in three groups, each of which have the same mean and variance for all its respective members:

1) The VTI has not yet been presented at position I ($i > I$). In this case, both MU1 and MU2 contain only STIs. Thus,

$$X_i^{(1)}(I) = S_s - \sum_{j=1}^{I} \frac{S_s}{I}; \qquad \bar{X}_i^{(1)}(I) = 0. \tag{7}$$

2) The VTI is presented at time I ($i = I$). Now MU1 contains the VTI, while MU2 is the same as in 1):

$$X_i^{(2)}(I) = S_v - \sum_{j=1}^{I} \frac{S_s}{I}; \qquad \bar{X}_i^{(2)}(I) = \bar{S}_v - \bar{S}_s. \tag{8}$$

3) The VTI has already been presented at an earlier position than I ($i < I$). MU1 contains an STI, again, but one of the intervals in MU2 is the VTI. Thus,

$$X_i^{(3)}(I) = S_s - \sum_{j=1}^{I} \frac{S_s}{I} - \frac{S_v}{I}; \qquad \bar{X}_i^{(2)}(I) = \frac{\bar{S}_s - \bar{S}_v}{I}. \tag{9}$$

The variance of the $X_i(I)$ does is the same in all three cases, because the variance σ_T does not differ for the STIs and the VTI:

$$\mathrm{Var}(X_i^{(1)}(I)) = \mathrm{Var}(X_i^{(2)}(I)) = \mathrm{Var}(X_i^{(3)}(I)) = \sigma_T \frac{I^2 + I}{I^2}. \tag{10}$$

Additionally, it must be noted that the criterion K can not be entirely freely chosen. Specifically, it must be ensured that the probability of an "irregular" judgment is below the defined $P(\text{“irreg}'')$ if the sequence does not contain a deviant interval, or $\bar{S}_s - \bar{S}_v = 0$. Otherwise, the adaptive method would make the detection thresholds converge to zero, as a sequence of regular intervals would be sufficient to elicit "irregular" responses with the defined probability. Together with Eqn. 5 and 7, this requirement results in the following condition on K:

$$K > \sigma_T \left(1 - P(\text{“irreg}'')\right)^{1/N} \approx 0.915\, \sigma_T, \tag{11}$$

where the second equality holds for $P(\text{“irreg}'') = 0.75$ and $N = 7$.

3.2.3. *Approximation*

Plugging the results Eqn. 7, 8, 9 and 10 into Eqn. 6 yields an equation that only depends on σ_T, K and the detection thresholds $\bar{S}_v - \bar{S}_s$. This equation can be used to fit σ_T and K to the experimentally obtained thresholds. However, the relative contributions of the two parameters to the data will not be apparent in these equations. Here, we derive an approximation where these contributions can be more clearly seen.

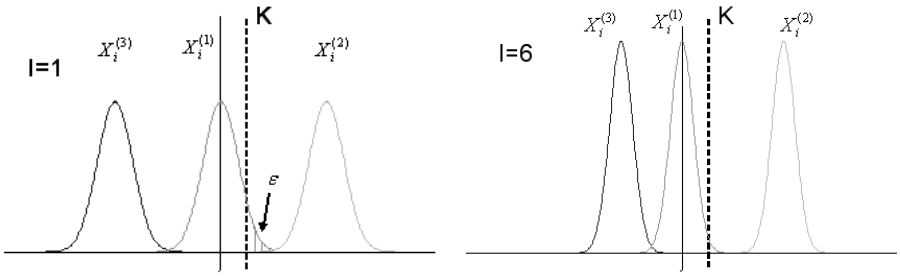

Fig. 4. Distributions of $X_i(I)$ for the three cases (see text) and two values of I.

Fig. 4 illustrates the three distributions of differences $X_i(I)$ for two values of I. All distributions become more peaked for later Is as a result of the averaging process. Furthermore, the mean value of $X_i^{(2)}$ always reflects the actual difference between the VTI and the STI, while the mean of $X_i^{(3)}$ is the negative of this difference for $I = 1$ and decreases in its absolute value for later I. Therefore, it is apparent that the false positive rate $\epsilon = P(X_i(I) < K \,|\, S_i = S_s)$ (shaded area in Fig. 4) is maximal for $X_i^{(1)}$.

Now assume that we have chosen the criterion K such that ϵ never exceeds a certain value ϵ^* for $X_i^{(1)}$. Then, from the above observations, we see that ϵ^* is also the upper bound for the false positive rate for $X_i^{(3)}$, so we can consider $\epsilon \leq \epsilon^*$ for all $N-1$ false positive cases and approximate Eqn. 6 by

$$P(\text{“irreg''}) \geq 1 - (1-\epsilon^*)^{N-1} \left(1 - \Phi\left(\frac{\bar{S}_v - \bar{S}_s - K}{\sigma_T(I)}\right)\right), \qquad (12)$$

where $\sigma_T^2(I)$ is the position-dependent variance common to all three cases, as given in Eqn. 10. Thus, the detection threshold is given for each i by

$$\bar{S}_v - \bar{S}_s \geq \Phi^{-1}\left(1 - \frac{1 - P(\text{“irreg''})}{(1-\epsilon^*)^{N-1}}\right) \sigma_T \sqrt{\frac{I^2 + I}{I^2}} + K. \qquad (13)$$

From this equation, one can see that the threshold decreases with I like $\sqrt{1+1/T}$, while the steepness of the decrease is governed by σ_T and a factor depending on ϵ^*, $P(\text{“irreg}'')$ and the number of intervals N. Additionally, there is an offset that is equal to the criterion K.

3.2.4. *Fit to data*

We use Eqn. 6 together with the results on mean and variance Eqn. 7, 8, 9 and 10 and the constraint on K, Eqn. 11 to fit the parameters σ_T and K to the data set of the three different standard durations. The fits are depicted as solid lines in Fig. 1. The model gives a good description of the data averaged over participants. (see Tab. 1).

Table 1. Fit parameters for Eqn. 6 (and others, see text).

STI duration [ms]	σ_T [ms]	K/σ_T	variance explained [%]
50	45	0.923	87
150	90	0.965	97
250	110	1.0	89

4. Discussion

We presented a model that can explain context effects on interval discrimination performance that we observed in a sequence experiment, and also the decrease of performance with increasing standard interval durations. Apart from the indivdiual effects, the model also explains the interaction of the two: σ_T increases with the STI durations and enters as a factor in Eqn. 13. Thus, longer STI durations increase the steepness of the adaptation curve, and thus enhances the position effect.

Fitting Eqn. 1 to the data suggested a very high temporal spread, $\sigma_{\min(\Delta t)} = 7.13$ ms. This is about one order of magnitude higher than the values that we found to be realistic.[12] However, this may be a specificity of sequence experiments, as the Weber fractions (0.49 to 1.18) are also very high compared to interval discrimination, where fractions between 0.05 and 0.2 are typical. A possible explanation lies in the rapid presentation of the stimuli: The ISI of maximally 250 ms might not be long enough to allow the intervals to be completely processed, causing an additional error.

The detection thresholds decrease with the position I of the variable interval like $\sqrt{1+1/I}$. Therefore, even for very long sequences, the vari-

ablity will not be eliminated, but only decreased to a value close to σ_T (cf. Eqn. 13), the variability of a single interval. Therefore, the model could be falsified by data showing a drastically different form of decrease, e.g. linear or superlinear. Moreover, the model predicts that i) the saturation of the detection threshold should be apparent in longer sequences, and ii) that there should only be a limited effect in single-interval task such as interval production. We already confirmed the first prediction in an experiment with nine intervals.[17]

On the other hand, the model is not directly falsified by the fact that it does not explain the end effect. Like other more complex effects,[17] this could be included by introducing a decay of the representations in the MUs. At the final time step, no new interval is represented in MU1, so the comparison has to rely on the partly decayed memory trace of the second-to-last interval. Because of the decay, the variability of this representation will be increased, which explains the poor discrimination performance at the final position.

References

1. S. Grondin, *Psych Bull* **127**, 22 (2001).
2. J. Gibbon, C. Malapani, C. L. Dale and C. R. Gallistel, *Curr Opin Neuro* **7**, 170 (1997).
3. T. H. Rammsayer and S. Grondin, *Psychophysics of human timing*, in *Time and the brain*, ed. R. Miller (Harwood Academic, 2000), pp. 157–167.
4. J. Gibbon, *Psychol Rev* **84**, 279 (1977).
5. L. A. Bizo, J. Y. Chua, F. Sanabria and P. R. Killeen, *Behav Process* **71**, 201 (2006).
6. D. J. Getty, *Perception and Psychophysics* **20**, 191 (1976).
7. C. Drake and M.-C. Botte, *Perception and Psychophysics* **54**, 277 (1993).
8. R. B. Ivry and E. Hazeltine, *J Exp Psychol [Hum Percept]* **21**, 3 (1995).
9. G. ten Hoopen, R. Hartsuiker, T. Sasaki, Y. Nakajima, M. Tanaka and T. Tsumura, *Perception* **24**, 577 (1995).
10. H. Pashler, *J Exp Psychol [Hum Percept]* **27**, 485 (2001).
11. C. D. Creelman, *J Acoust Soc Am* **34**, 582 (1962).
12. J. Haß, S. Blaschke, T. Rammsayer and J. M. Herrmann, *J Comput Neurosci* **25**, 449 (2008).
13. C. Kaernbach, *Perception and Psychophysics* **49**, 227 (1991).
14. M. Abeles, *Corticonics: Neural circuits of the cerebral cortex* (Cambridge University Press, 1991).
15. J. M. Herrmann, J. A. Hertz and A. Prügel-Bennet, *Network-Comp Neural* **6**, 403 (1995).
16. M. Diesmann, M.-O. Gewaltig and A. Aertsen, *Nature* **402**, 529 (1999).
17. S. Blaschke, J. Haß, J. M. Herrmann and T. H. Rammsayer, *submitted* (2008).