GAMIT-Net: Retrospective and prospective interval timing in a single neural network.

Caspar Addyman (c.addyman@bbk.ac.uk),
Denis Mareschal (d.mareschal@bbk.ac.uk)
CBCD, Department of Psychological Sciences, Birkbeck,
University of London, Malet Street, London, WC1E 7HX, UK

Abstract

The neural network version of the Gaussian Activation Model of Interval Timing (GAMIT-Net) is a simple recurrent network that unifies retrospective and prospective timing in a single framework. It has two parts. Firstly, a time-dependent signal is generated by a spreading Gaussian activation. Next, a simple recurrent network (SRN) combines information from the Gaussian and its own internal state during a timing task to generate time estimates. This model captures the scalar property of interval timing (Gibbon, 1977). Furthermore, under high cognitive load the Gaussian fades faster while the internal state is updated less often. These factors interact to account for the surprising finding that retrospective estimates increase under cognitive load while prospective estimates decrease (Block, Hancock & Zakay, 2010).

Keywords: interval-timing, activation-based model, time-perception, retrospective and prospective timing.

Introduction

Our sense of time is ubiquitous and yet enigmatic. Interval timing is central to cognition in humans (e.g., Grondin, 2008; Zakay & Block, 1997) and animals (Gibbon and Allan, 1984). It may even underlie conditioned learning in animals (Gallistel & Gibbon, 2000). Over intervals in the range from half a second to several minutes humans and other animals show very similar abilities. Interval timing judgments by humans, rats and pigeons obey a version of Weber's Law known as the scalar property (Gibbon, 1977*).* Yet three mysteries exist. What explains the scalar property? Why do retrospective and prospective timing show opposite effects of cognitive load (Block, Hancock & Zakay, 2010)? What explains the long developmental trajectory for interval timing abilities (e.g., Szelag et al, 2002; Droit-Volet, Tourret & Wearden, 2004)? The current paper builds on our previous research (Addyman et al. 2011; French et al. 2014) to provide a unified answer to all these questions.

Our Gaussian Activation Model of Interval Timing (GAMIT-Net) is a stochastic learning model. GAMIT-Net has two parts a columnar memory trace that produces decay and a connectionist simple recurrent network (Elman, 1990) that samples the decaying trace to provide time estimates. The traces decay in statistically predictable (Gaussian) manner permitting timing estimates. Mathematical constraints on the accuracy of these estimates leads to the linear growth in errors characteristic of the scalar property. The differences between retrospective and prospective timing with cognitive load are explained through the interaction of two factors. (1) High cognitive load causes memory traces to decay faster leading to longer estimates. This factor alone explains increases in retrospective timing paradigms where a time estimate must be made without prior warning (and hence without intermediate sampling). (2) In prospective timing where it is known in advance that a time estimate will be required, participants will be estimating time as the task progresses. Increased cognitive load will lead to reduced number of these intermediate estimates. This gives the sense that time is passing more quickly and results in shorter estimates. Finally, the connectionist nature of the model provides a framework for explaining developmental effects.

The paper is organized as follows. We begin by describing existing models of interval timing and discuss two recent key findings that a good model must address. We then describe of our connectionist model of timing based on memory-trace decay and demonstrate mathematically why our model is constrained to show linear growth in errors. Finally, we present the simulations of the model. In particular, we show how prospective and retrospective estimates can be made within a single framework and yet have opposite effects of cognitive load. We also show how this model could provide developmental predictions.

Existing models of interval timing

There are three major classes for interval-timing model. (1) Pacemaker-accumulator models rely on an internal pacemaker that emits regular, short pulses that are counted by an accumulator. The number of pulses stored in the accumulator gives the measure of the time that has passed (Church, 1984; Gibbon et al. 1984; Taatgen et al, 2007). (2) Multiple oscillator-coincidence detector models (also sometimes called timestamp models) rely on multiple neuronal oscillators started simultaneously with coincidence detectors associating particular patterns of firing with given time intervals, effectively time-stamping when an event occurs (Church & Broadbent, 1990; Matell & Meck, 2000) (3) In memory or neural process models the passage of time is derived from the activation of a neural process that is decaying (Staddon & Higa, 1999) or increasing (Reutimann et al., 2004).

The Scalar Property

The scalar property or time scale invariance (Gibbon, 1977) is a very widely replicated effect with humans, rats and pigeons (Gibbon & Allan, 1984; Gibbon et al, 1997;

Matell & Meck, 2000). It states that participant responses in an interval timing task will have an approximately normal (right skewed) distribution peaked at the target time with the width of the distribution directly proportional to the length of the interval. In other words the growth of error is constant (scalar) such that if estimates for interval T have error $\pm E$, an interval of 2T will have errors ±2E. This is an instance of Weber's Law, which states that the confusability of two stimuli is proportional to their magnitude. It places several important restrictions on the nature of any interval timing mechanism (Hass & Hermann, 2012).

In particular, it implies that the neural process underlying time perception must measure growing variance in the system. Only variance-based processes will lead to the scalar growth of error. Accumulator models base their estimates on the mean number of accumulated ticks. According to the Central Limit Theorem, such estimates have errors that grow with the square root of the total. Pacemaker-accumulator models must introduce assumptions as to why the cognitive system cannot use these more precise quantities (e.g. Gibbon, 1992). Other models introduce arbitrary Gaussian thresholds on otherwise linear (Reutimann et al., 2004) or logarithmic processes (Staddon & Higa, 1999). Early multiple oscillator models required perfectly correlated sets of oscillators (Matell & Meck, 2004). Recent work addresses this using more realistic, noisy neural oscillators and neural network architecture (Buhusi & Oprisan 2013). However, no model that we are aware of accounts for the scalar property as an unavoidable consequence of the way the timing mechanism works (Hass & Hermann, 2012; Hass et al. 2008).

Retrospective and prospective time estimation

One of the biggest distinctions within interval timing is between retrospective and prospective paradigms. *Retrospective* time keeping concerns our *estimates* of time in the recent past, while *prospective* time keeping concerns our *predictions* about the near future. In the former awareness that time must be judged comes without warning at the end of the interval, while in the latter it is known from the beginning of the interval that a time judgment will be required. Zakay and Block (2004) refer to this as the difference between remembered and experienced duration. The majority of models of time perception are models of prospective timekeeping only. They are concerned with predictions about future events, such as rats and pigeons learning to respond maximally at the target time in a fixed interval paradigm. One reason is they are built around a counter (e.g. an accumulator, a set of oscillators or a climbing neural process) that must be explicitly started for each trial, something that is (by definition) not possible in retrospective paradigms.

The other reason is that retrospective and prospective estimates show a striking interaction with cognitive load. Block et al. (2010) analyzed the results from over one hundred studies with human subjects. They found that high cognitive load increases your estimates in the case of retrospective timing, whereas it decreases your estimates in the case of prospective timing. This interaction is a challenge to clock and timestamp models. They provide no a priori reason to expect a difference between these two conditions. Furthermore, this interaction suggests that cognitive load is not just an additive factor (e.g., damping responses across the board). This is a challenge for all existing models of interval timing. The main aim of the GAMIT-Net model is to explain the scalar property and seemingly disparate retrospective and prospective timing in a single framework.

GAMIT-NET

In this section we describe GAMIT-Net, our model of interval timing. A MATLAB implementation of the model is available at http://github.com/YourBrain/GAMIT.

GAMIT-Net is built on the intuition that our sense of time arises from our fading memory for events. The longer ago an event happened the fuzzier the memory associated with it will be. We claim that this relationship is statistically predictable and that our interval timing abilities are acquired by process of learning from our experience of changes in the world around us. We use estimates of the variance of the spreading Gaussian activation trace as a measure of how much time has passed. Furthermore, inescapable errors in the estimation process lead to the scalar property.

Two factors account for differences between retrospective and prospective timing. First, we assume that the memory decay is affected by cognitive load. Decay occurs faster under high cognitive load, perhaps due to global inhibition from competing processes. This factor alone accounts for longer estimates in retrospective timing. Secondly, in prospective timing, we must additionally take into account the fact that participants will be making intermediate estimates as the task progresses. We assume that the cognitive system makes a number of 'attentional saccades' to the activation trace during a given interval. The final estimate will take account of both the final pattern of activation and these intermediate estimates. Under high cognitive load the trace is decay faster as before but in addition there will be fewer attentional saccades that are more spread out in time, giving the sense of time that less time has passed, leading to shorter estimates.

Network architecture

GAMIT-Net has two distinct components; a fading Gaussian memory and a connectionist learning network. The network is schematically represented in Figure 1. In this section we first describe the Gaussian activation curve and explain why it constrains our model to show *at best* linear growth in errors. Next we explain how the model uses a simple recurrent network (SRN) architecture to capture both retrospective and prospective timing within a single framework.

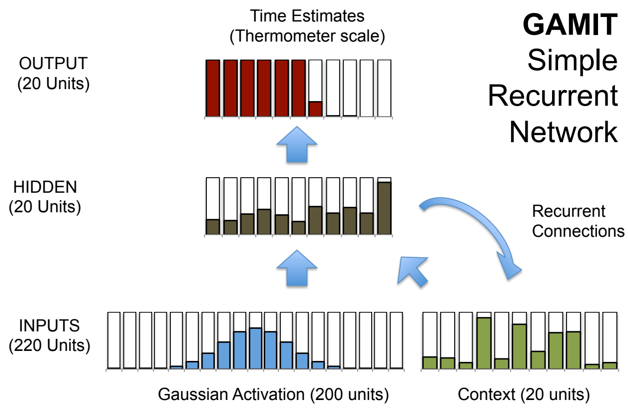

Figure 1: GAMIT consists of an Elman-style simple recurrent network that learns to convert time-decaying Gaussian activation curves into linear time estimates.

Time as Gaussian Activation Decay

To implement the GAMIT-Net model, we begin with a cluster of cortical columns. The activation in the central column corresponds to an event in the world that is registered in memory. Activation then spreads across the cortical columns as follows. If we designate the activation of the i^{th} column at time step *t* by $A_i(t)$, its activation at time *t+1* is determined by the following equation:

$$A_i(t+1) = \alpha A_i(t) + \beta\big(A_{i-1}(t) + A_{i+1}(t)\big) + \xi \qquad (1)$$

where α is the fraction of activation that remains in column *i* on each time step; β is the fraction of activation spread from each immediate neighbor of *i* on each time step; ξ is a noise parameter. The values of α and β must be chosen so that the total activity over time of the system neither rapidly decreases to zero nor increases exponentially. Unless otherwise stated, we used values of α = 0.7, β = 0.14952 and *ξ* = 0.00025. The evolution of activation in this cluster of columns is illustrated by the series of graphs in Figure 2. Note that the difference equation presented here is an approximation to an underlying stochastic process. There is ample neurobiological evidence for this type of spreading-activation mechanism (e.g., Amari, 1980; Grossberg, 1980; Herman et al., 1993; Koch & Segev, 1998; Capaday et al., 2011).

Mathematical contraints on temporal estimates.

Time estimates in GAMIT-Net are based on the growing variance in the system as an initially localised activation spreads through the columns. Here we show, following Hass & Hermann (2012), that if the underlying process has a linear growth in variance (i.e. X(t) ~ N(μ, σ), X(2t) ~ N(μ, 2σ)), X(3t) ~ N(μ, 3σ), then the scalar property arises because these estimates can only be made with limited accuracy. Namely that if the uncertainty at time t is Δ, then at time 2t it will be 2Δ.

The Kullback–Leibler divergence between two normal distributions $X_1 \sim N(\mu_1, \sigma^2{}_1)$ and $X_2 \sim N(\mu_2, \sigma^2{}_2)$ provides a measure of their confusability (Kullback & Leibler, 1951) and is given by:

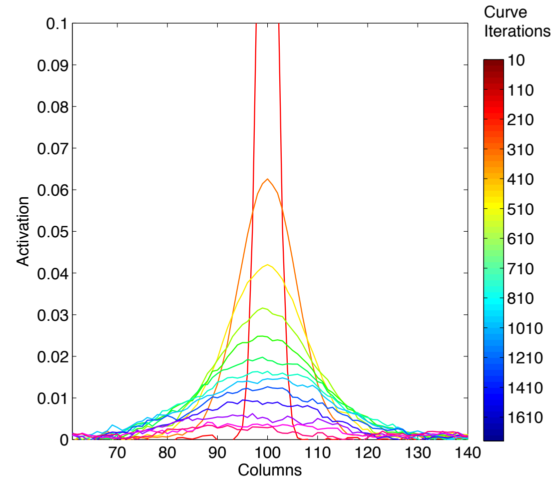

Figure 2: An initially localized activation fades and spreads over time as equation 1 iterates through time. Curves are color coded according to the number of iterations indicated by the scale on the right hand side.

$$D_{KL}(X_1 \| X_2) = \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2} + \frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_2^2} - 1 - ln\frac{\sigma_1^2}{\sigma_2^2}\right).$$

We wish to show that in general

$$D_{KL}(X(t) \,\|\, X(t+\Delta)) = D_{KL}(X(2t) \,\|\, X(2t+2\Delta)).$$

This is easy to see since, in all cases, μ is constant so the first term cancels to 0 and other term is in ratio of σ_1^2 to σ_2^2 which is the same in both cases; on the left hand side the ratio $\sigma : (1+\Delta)\sigma$ while on the right it is $2\sigma : (2+2\Delta)\sigma$.

In other words, the confusability of Gaussian-like curves grows linearly with variance (when the means are the same). Hence any process based on the discriminating two such curves cannot do better than a scalar error. Unlike in other models, in GAMIT-Net scalar errors are a lower bound on accuracy. Furthermore, as we will see below, the implemented version of GAMIT-Net shows a broadly linear growth in error.

GAMIT-Net Simple Recurrent Network

The current paper uses an SRN (Elman, 1990) to combine two of our previous modelling efforts (Addyman et al., 2011; French et al., 2014) in a single framework. Addyman et al. (2011) showed that a feedforward neural network built on top of the Gaussian spreading activation function could model how timing abilities are acquired in infancy. We used motor signals to calibrate an embodied timing mechanism across multiple sensory modalities. This model showed developmental effects (Szelag et al., 2003) and the scalar property but could not capture retrospective and prospective effects (Block et al 2010). In a separate cognitive model (GAMIT - French et al., 2014) we demonstrated how to capture those effects. That paper introduced the idea that prospective timing involves 'attentional saccades' during the timing task which affect estimates by using compensatory parameter which reduced prospective estimates when saccades were less frequent (i.e. when time appeared to passing more quickly). However, that model could not capture learning.

In GAMIT-Net each time related event corresponds to an updating of the network. In retrospective case, the SRN receives just two updates; one for the initial event and one at the test time. In the prospective case, the network is also updated at intermediate points when the cognitive system monitors the passage of time (attentional saccade to the timing task). Higher cognitive load increases the decay of the Gaussian but decreases the number of saccades. In all cases, we coded the initial event by a localised activation and empty context units and at each subsequent update the inputs are the current Gaussian activations and a copy of the previous hidden representation in the context units.

RESULTS

We report three simulations results. First, we show that a single network can learn to perform both retrospective and prospective timing and learning could be a rich source of developmental predictions. Second, we show that a trained network has scalar errors in both paradigms. Third, we show that increased cognitive load has a differential effect on retrospective and prospective time estimates in line with empirical findings (Block et al. 2010).

Simulation 1 – Retrospective and prospective estimates in a single network.

Ten SRNs with 220 inputs (200 curve + 20 context), 20 hidden and 20 output units were initialized with small random weights and were trained for 20 epochs each. For each epoch of training we generated 50 fading Gaussian curves by iterating an initially localized input using Equation 1 with the parameters given above. For a given curve we randomly picked 50 target times and presented the network a random mixture of retrospective (p=0.5) and prospective (p=0.5) timing events. Hence each epoch consisted of 2500 training events. In a retrospective timing event, the network received an input at the start of the trial and second input and hidden layer context at the end. In a prospective timing event the network also received inputs of the curve shape and context at several random attentional saccades. Saccades were generated by a Poisson process with parameter $\lambda = 100$. Target times were coded on a thermometer scale and learning was via backpropagation of errors. The learning rate was 0.05 and the momentum was 0.005. At each epoch of training we tested the networks across the full range of possible time intervals on each of the two timing tasks. Figure 3 shows the average output of the 20 networks. As can be seen the network learns both tasks well, approaching the idealised performance (dotted line) as they mature. Immature networks overestimate short intervals and underestimate long ones in line with experimental results with children (Szelag et al. 2003).

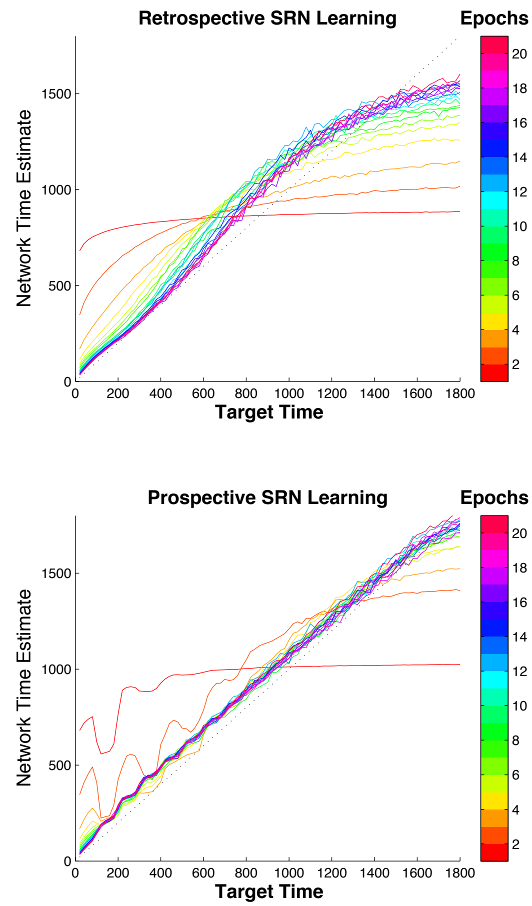

Figure 3: 10 naïve networks were trained for 20 epochs. Each epoch contained 2500 randomly determined retrospective and prospective timing events

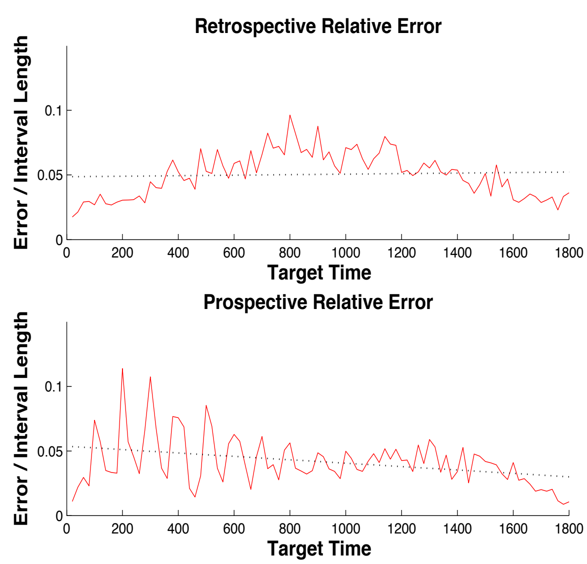

Figure 4: The relative error on time estimate tasks. Average of 20 fully trained networks. Dotted line show best linear fit.

Simulation 2 – The Scalar Property

Ten SRNs were fully trained as in Simulation 1 with 20 epochs of 2500 training events. Figure 4 shows the average relative error for 20 networks across the full range of possible time intervals on each of the two timing tasks, calculated by dividing the absolute error by the target time. These plots show that relative error was broadly constant proportion at all time intervals in line with the scalar property.

Simulation 3 – The effects of cognitive load

A single SRN was fully trained as in Simulation 1 with 20 epochs of 2500 training events with normal cognitive load parameters. . To simulate high cognitive load conditions we decreased the value of decay parameter β to 0.14946 and increased the sampling parameter λ to 110. For lower-than-typical cognitive load, β was increased to 0.14955 and λ decreased to 90. We then tested the networks performance on 20 estimates of with target time t = 600. The results are shown in Figure 6. The pattern of performance matched that found in Block et al. (2010), as shown in Figure 5. It is important to note that prospective underestimates are an emergent property of the network. They are a result of the weighting that the network learns to give these two competing pieces of information under normal cognitive load.

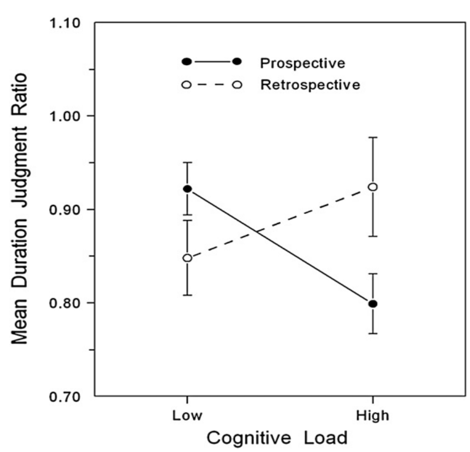

Figure 5: The effects of cognitive load on interval timing based on a meta-analysis of 82 prospective and 31 retrospective tasks (Block et al., 2010).

Conclusion

We have developed GAMIT-Net to address three goals. First, we sought to build a model of interval timing based on measured variance that would give rise to the scalar property as direct consequence of the way the timing mechanism works without ad-hoc assumptions or modifications (Hass & Herrmann, 2012). Second, we wished to unify prospective and retrospective interval timing within a single model while still being able to account for the differential effects of cognitive load (Block et al, 2010).

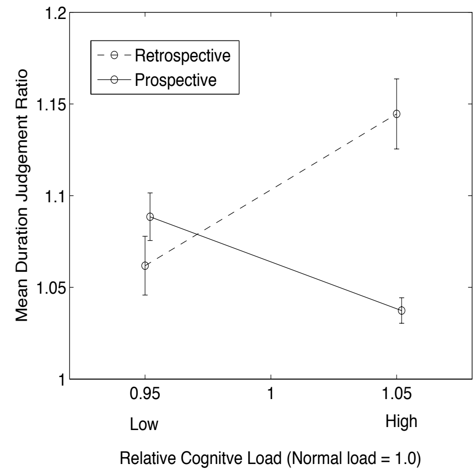

Figure 6: Performance of GAMIT under cognitive load, Results averaged over 20 runs of the program.

Finally, the neural network architecture of makes GAMIT-Net unique as a developmental model of time perception that allows for research and predictions about infant timing abilities.

GAMIT-Net addresses all these issues and, we believe, it has other features to recommend it. The model has good explanatory power. It is important to note that the network was not specifically designed to show the cognitive load interaction (Block et al. 2010). Philosophically, we believe our model is an advance over clock-based models in that gets more directly at the experience of time passing. Longer intervals correspond to greater memory decay. While greater cognitive load has two complementary effects of making memories fade faster and making time seem to pass more quickly.

Moreover, activation decay and growth processes are ubiquitous and well understood and can account for evidence that timing and memory use the same cognitive resources (Fortin and Rousseau, 1997; Fortin, 1999) and both recruit the dorso-lateral prefrontal cortex (Wager and Smith, 2003, Genovesio et al., 2006). Related to this, the neural network architecture GAMIT-Net provides an approach to timing in which is based on a lifetime of past experiences of observing change in the world. As a result information about the passage of time is embedded in the world rather than constructed in an abstract cognitive module.

There remain many issues to address. In Simulation 2, the fit to linear growth in error is not perfect and future work should investigate it. Some studies report a greater than linear increase of the timing errors (reviewed in Gibbon et al., 1997; Grondin, 2001; Hass et al., 2008). At present model fits data from a meta-analysis. Future work must simulate results from individual experiments. Likewise, the current work is restricted to recognition tasks where an estimate is given at the end of an interval. The model should also be used to production tasks where participants generate time. We believe that the attentional saccade mechanism built into GAMIT-Net provides a natural means of simulating production tasks. Much further work is needed but we believe GAMIT-Net represents a powerful new paradigm in interval timing research.

Acknowledgments

This work was supported by UK ESRC grant (RES-062-23-0819).

References

Addyman, C., French, R.M., Mareschal, D. & Thomas, E. (2011) Learning to perceive time: A connectionist, memory-decay model of the development of interval timing in infants. In, *Proc. of the 33rd Annual Conference of the Cognitive Science Society,* 354-359.

Amari, S. I. (1980) Topographic organization of nerve fields. *Bulletin of Mathematical Biology, 42*, 339–364.

Block, R. A., Hancock, P. A., & Zakay, D. (2010) How cognitive load affects duration judgments: A meta-analytic review**.** *Acta Psychologica, 134*, 330–343.

Buhusi, C. V., & Oprisan, S. A. (2013). Time-scale invariance as an emergent property in a perceptron with realistic, noisy neurons. *Behavioural Processes*, *95,* 1–11. doi:10.1016/j.beproc.2013.02.015

Capaday C, van Vreeswijk C, Ethier C, Ferkinghoff-Borg J and Weber D (2011) Neural mechanism of activity spread in the cat motor cortex and its relation to the intrinsic connectivity. *Journal of Physiology, 589,* 2515-2528.

Church, R. (1984). Properties of the internal clock. *Annual Proceedings of the New York Academy of Science*, 423, 566–582.

Church, R., & Broadbent, H.(1990) Alternative representations of time, number and rate. *Cognition*, *37*, 55–81.

Droit-Volet, S. (2003). Alerting attention and time perception in children. *Journal of Experimental Child Psychology*, *85*(4), 372–384.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–221.

Fortin, C. (1999) Short-term Memory in Time Interval Production. *International Journal of Psychology 34*, 308-316.

Fortin, C., & Rousseau, R. (1998). Interference from short-term memory processing on encoding and reproducing brief durations. *Psychological Research*, 61, 269−276.

French, R. M., Addyman, C., Mareschal, D., & Thomas, E. (2014). Unifying prospective and retrospective interval-time estimation: A new fading-Gaussian activation-based model of interval-timing, *Procedia - Social and Behavioral Sciences*, *126*, 141–150.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, *107*(2), 289–344.

Genovesio, A., Tsujimoto, S., & Wise, S. P. (2006). Neuronal activity related to elapsed time in prefrontal cortex. *Journal of Neurophysiology*, *95*(5), 3281–3285.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84(3), 279-325.

Gibbon, J. (1992). Ubiquity of Scalar Timing with a Poisson Clock. *Journal of Mathematical Psychology*, *36*, 283–293.

Gibbon, J. & Allan, L. (Eds.) (1984) *Timing and time perception. (Vol. 423)* New York, NY: New York Academy of Sciences.

Gibbon, J., Church, R., and Meck, W. (1984). Scalar timing in memory. *New York Academy of Sciences.*, 423, 52–77.

Gibbon, J., Malapani, C., Dale, C.L., & Gallistel, C. (1997). Toward a neurobiology of temporal cognition: advances and challenges. *Current Opinion in Neurobiology*, *7*(2), 170–184.

Grossberg, S. (1980) How does the brain build a cognitive code? *Psychological Review*, *87*, 1–51.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psycholigcal Bulletin*, *127*(1), 22–44.

Hass, J., & Herrmann, J. M. (2012). The neural representation of time: an information-theoretic perspective. *Neural Comp.*, *24*, 1519–1552.

Hass, J., Blaschke, S., Rammsayer, T., & Herrmann, J. M. (2008). A neurocomputational model for optimal temporal processing. *Journal of Computational Neuroscience*, *25*, 449–464.

Herman, M., Ruppin, E. & Usher, M. (1993) A neural model of the dynamic activation of memory. *Bio. Cyber., 68*, 455-463.

Koch, E. & Segev, I. (1998). *Methods in Neuronal Modeling : From Ions to Networks.* Cambridge, MA: The MIT Press.

Matell, M. S. & Meck, W. H. (2000) Neuropsychological mechanisms of interval time behavior *Bioessays, 22*, 94-103.

Reutimann, J., Yakovlev, V., Fusi, S., & Senn, W. (2004). Climbing neuronal activity as an event-based cortical representation of time. *Journal of Neuroscience*, *24*, 3295–3303.

Staddon, J.E.R. & Higa, J.J. (1999) Multiple time scales in simple habituation. *Psychological Review, 103*, 720-733.

Szelag, E., Kowalska, J., Rymarczyk, K., & Pöppel, E. (2002). Duration processing in children as determined by time reproduction: implications for a few seconds temporal window. *Acta Psychologica*, *110*(1), 1–19.

Taatgen, N. A., van Rijn, H., & Anderson, J. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review*, *114*, 577–598.

Wager, T.D. and Smith, E.E. (2003) Neuroimaging studies of working memory: a meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3,* 255–274.

Zakay , D. & Block, R. A. (1997) Temporal cognition. *Current Directions in Psycholgical Science*, 6, 12-16.

Zakay, D., & Block, R. A. (2004). Prospective and retrospective duration judgments: an executive-control perspective. *Acta neurobiologiae experimentalis*, *64*, 319–328.