



# A Readout Mechanism for Latency Codes

Oran Zohar<sup>1,2\*</sup> and Maoz Shamir<sup>2,3,4</sup>

<sup>1</sup> Department of Brain and Cognitive Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel, <sup>2</sup> Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel, <sup>3</sup> Department of Physiology and Cell Biology, Ben-Gurion University of the Negev, Beer-Sheva, Israel, <sup>4</sup> Department of Physics, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Response latency has been suggested as a possible source of information in the central nervous system when fast decisions are required. The accuracy of latency codes was studied in the past using a simplified readout algorithm termed the temporal-winner-take-all (tWTA). The tWTA is a competitive readout algorithm in which populations of neurons with a similar decision preference compete, and the algorithm selects according to the preference of the population that reaches the decision threshold first. It has been shown that this algorithm can account for accurate decisions among a small number of alternatives during short biologically relevant time periods. However, one of the major points of criticism of latency codes has been that it is unclear how can such a readout be implemented by the central nervous system. Here we show that the solution to this long standing puzzle may be rather simple. We suggest a mechanism that is based on reciprocal inhibition architecture, similar to that of the conventional winner-take-all, and show that under a wide range of parameters this mechanism is sufficient to implement the tWTA algorithm. This is done by first analyzing a rate toy model, and demonstrating its ability to discriminate short latency differences between its inputs. We then study the sensitivity of this mechanism to fine-tuning of its initial conditions, and show that it is robust to wide range of noise levels in the initial conditions. These results are then generalized to a Hodgkin-Huxley type of neuron model, using numerical simulations. Latency codes have been criticized for requiring a reliable stimulus-onset detection mechanism as a reference for measuring latency. Here we show that this frequent assumption does not hold, and that, an additional onset estimator is not needed to trigger this simple tWTA mechanism.

## OPEN ACCESS

### Edited by:

David Hansel,  
Paris Descartes University, France

### Reviewed by:

Carl Van Vreeswijk,  
Centre National de la Recherche  
Scientifique (CNRS), France  
Mattia Rigotti,  
IBM Thomas J. Watson Research  
Center, USA

### \*Correspondence:

Oran Zohar  
oranzohar@gmail.com

**Received:** 17 February 2016

**Accepted:** 28 September 2016

**Published:** 20 October 2016

### Citation:

Zohar O and Shamir M (2016) A  
Readout Mechanism for Latency  
Codes.  
*Front. Comput. Neurosci.* 10:107.  
doi: 10.3389/fncom.2016.00107

**Keywords:** spike latency, winner takes all, fast readout, temporal code, conductance based model, rate model

## INTRODUCTION

Neuronal response latency has been shown to be tuned to various features of external stimuli across many sensory modalities including the somatosensory (Panzeri et al., 2001; Johansson and Birznieks, 2004; Panzeri and Diamond, 2010), auditory (Brugge et al., 1996, 2001; Klug et al., 2000; Leibold and van Hemmen, 2005; McAlpine, 2005; Zhou et al., 2005; Chase and Young, 2007; Joris and Yin, 2007; Goodman and Brette, 2010; Grothe et al., 2010; Ashida and Carr, 2011; Grothe and Koch, 2011; Lüling et al., 2011; Zohar et al., 2011; Shamir, 2014), and visual systems (Hubel and Wiesel, 1968; Gawne et al., 1996; Van Rullen and Thorpe, 2001; Gollisch and Meister, 2008; Shriki et al., 2012; Shamir, 2014).

In addition, latency tuning has been observed in various stations along the information processing pathway from the receptor level (Hall et al., 1995; Van Rullen and Thorpe, 2001; Gollisch and Meister, 2008; McGinley et al., 2012), the brainstem (Klug et al., 2000; Joris and Yin, 2007; Grothe and Koch, 2011; Zohar et al., 2011) and up to the cerebral cortex (Hubel and Wiesel, 1968; Brugge et al., 1996, 2001; Shriki et al., 2012). For this reason, response latency has been suggested as a possible source of information in cases where fast decisions are required (Gawne et al., 1996; Gautrais and Thorpe, 1998; Van Rullen and Thorpe, 2001; Johansson and Birznieks, 2004; Chase and Young, 2007; Gollisch and Meister, 2008; Shamir, 2009, 2014; Goodman and Brette, 2010; Panzeri and Diamond, 2010; Zohar et al., 2011; Shriki et al., 2012).

The accuracy of latency codes has been investigated using the framework of a simple latency based competitive readout, the “temporal-winner-take-all” (tWTA), in a modeling study (Shamir, 2009) and more recently in the auditory (Zohar et al., 2011) and the visual systems (Shriki et al., 2012). The tWTA estimates the stimulus on the basis of the preferred stimulus of the neuron (or population) that fired first (or reached a certain decision threshold). The utility of the tWTA is that it is sufficiently simple and well-defined to enable analytical investigation of readout speed and accuracy. It was shown that the tWTA can account for fast and accurate discrimination between a small number of alternatives. However, it remains unclear how the central nervous system implements such a readout, if at all. This raises crucial questions as to the utility of latency coding in the brain.

The conventional rate winner-take-all (WTA) readout that estimates the stimulus based on the preferred stimulus of the neuron that fired the most spikes (rather than the first as in tWTA) has been widely used and studied in neuroscience (Fukai and Tanaka, 1997; White et al., 1998; Jin and Seung, 2002; Laing and Chow, 2002). There is a general consensus that the computation of the WTA decision is based on reciprocal inhibition between groups of neurons with similar preferences.

Here we show that the basic WTA architecture is also sensitive to the temporal order of its inputs such that it is able to implement the tWTA readout. This paper is organized as follows. First we define the basic reciprocal inhibition architecture of the tWTA mechanism and analyze the system using the framework of a rate toy model to facilitate the analytical investigation. Then, we study the generalization of our analysis to a model of spiking neurons in a numerical study of conductance-based neurons. Finally, we discuss the issue of estimating the stimulus onset time and the implications of the findings.

## METHODS

### Spiking Neurons Model

In the numerical simulations of spiking neurons we used the following conductance-based model with typical parameters taken from Shriki et al. (2003) and Shamir et al. (2009).

$$C \frac{dV_i}{dt} = I_i^{leak} + I_i^{net} + I_i^{ext} + I_i^{active}$$

where  $I_i^{leak}$ ,  $I_i^{net}$ ,  $I_i^{ext}$ ,  $I_i^{active}$  denote the leak, reciprocal inhibition, upstream input and active currents, respectively, and  $C = 1 \mu\text{F}/\text{cm}^2$  is the membrane capacitance. The currents obey  $I_i^{leak} = g_L(E_L - V_i(t))$ ,  $I_{i \in \{popA\}}^{net} = \frac{2g_s}{N} \sum_{j \in \{popB\}} s_{ij}(t)(E_s - V_j)$ ,  $I_i^{active} = I_{Na} + I_K + I_A$  with  $I_{Na} = g_{Na}m_\infty^3 h(V - E_{Na})$ ,  $I_K = g_K n^4(V - E_K)$ ,  $I_A = g_A a_\infty^3 b(V - E_K)$ . The parameters  $E_L$ ,  $E_s$ ,  $E_{Na}$ ,  $E_K$  denote the reversal potentials of the ionic currents in mV and are  $-65$ ,  $-80$ ,  $55$ ,  $-80$  respectively. Conductance is given in units of  $\text{ms}/\text{cm}^2$ , and  $g_L$ ,  $g_{Na}$ ,  $g_K$ ,  $g_A$  are  $0.05$ ,  $100$ ,  $40$ ,  $20$  respectively. The value of  $g_s$  is discussed below. The variables  $s_{ij}$ ,  $m$ ,  $h$ ,  $n$ ,  $a$ ,  $b$  obey the following dynamics;

$$\begin{aligned} ds_{ij}/dt &= 1 + (\tanh(V_{pre}/10)/2)(1-s)/(\tau_R) - s/\tau_D, \quad db/dt \\ &= (b_\infty - b)/\tau_b \end{aligned}$$

$dx/dt = (x_\infty - x)\tau_x$  for  $x = n, h, b$  with  $b_\infty = 1/(\exp((V+80)/6) + 1)$ ,  $a_\infty = 1/(\exp(-(V+50)/20) + 1)$  and  $x_\infty = \alpha_x/\alpha_x + \beta_x$  for  $x = m, h, n$ .

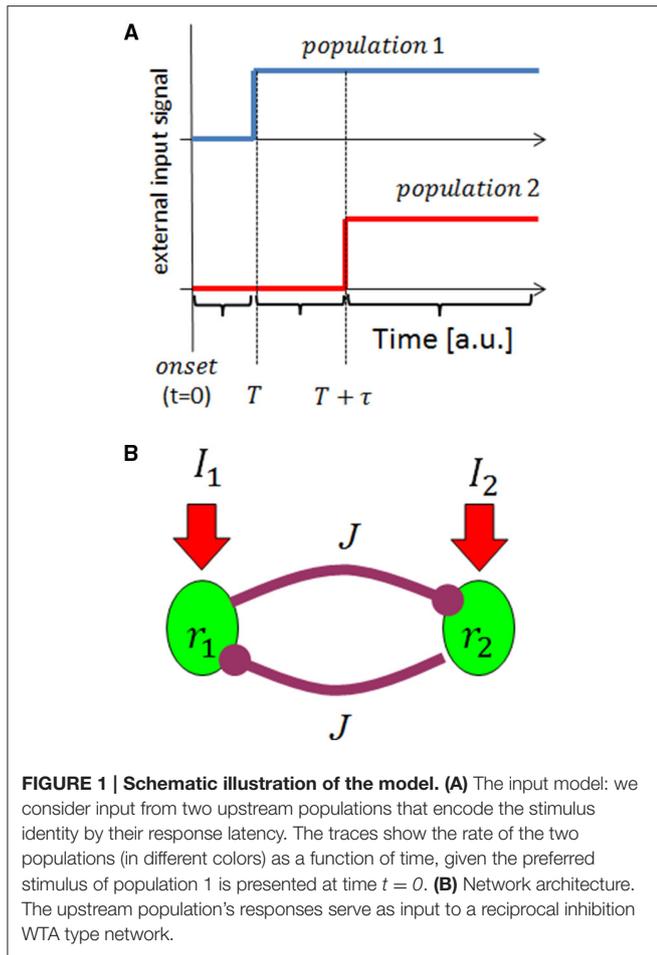
Time constants are measured in ms, and  $\tau_b$ ,  $\tau_D$ ,  $\tau_R$  are  $20$ ,  $10$ ,  $0.5$  respectively, and  $\tau_x = \phi/\alpha_x + \beta_x$  for  $x = m, h, n$ , where  $\phi = 0.1$  and

$$\begin{aligned} \beta_m &= 4 \exp(-(V+55)/18), \\ \alpha_m &= -(0.1(V+30)/(\exp(-0.1(V+30)-1)), \\ \beta_h &= 1/(\exp(-0.1(V+14))+1), \\ \alpha_h &= 0.07 \exp(-(V+44)), \\ \alpha_n &= -(0.01(V+34)/(\exp(-0.1(V+34)-1)), \\ \beta_n &= 0.125 \exp(-(V+44)/80). \end{aligned}$$

$I_{x,i}^{ext}(t) = r\Theta(t - [T + \tau_x + \delta_{x,i}]) + \xi_{x,i}(t)$  is the input current to neuron  $i$  in population  $x \in \{1, 2\}$  from the upstream population, where  $\Theta$  is the Heaviside step function,  $T$  is an absolute delay in the response and  $\tau_x$  is stimulus selective delay (see A with  $\tau_i = 0$  at the preferred stimulus, and  $\tau_i = \tau$ ). The input current contained two sources of noise. One was a random jitter in the input latency to each neuron, denoted by  $\delta_{x,i}$ , that were taken to be i.i.d. Gaussian random variables with zero mean and standard deviation  $\Delta$ ,  $\delta_{x,i} \sim N(0, \Delta^2)$ . The second source was an additive Gaussian white noise term,  $\xi_{x,i}(t)$ , that was drawn i.i.d with  $\xi(t) \sim N(0, \Xi^2)$ . In the numerical simulations we took  $r = 5 \text{ mA}/\text{cm}^2$ , and the white noise,  $\xi_{x,i}(t)$ , was binned at 1 ms time intervals with  $\Xi = 5 \text{ mA}/\text{cm}^2$ .

## RESULTS

We first studied the tWTA readout mechanism using the framework two-alternative forced choice discrimination in a rate model. The readout mechanism receives two inputs,  $I_1$  and  $I_2$ . These inputs represent the activities of two upstream populations of neurons responding to an external stimulus, each of which is characterized by its own preferred stimulus. For example,  $I_1$  can be thought of as the response of a population of inferior colliculus auditory neurons with a preferred sound source azimuth to the right of the animal, whereas  $I_2$  is the response to the left. The



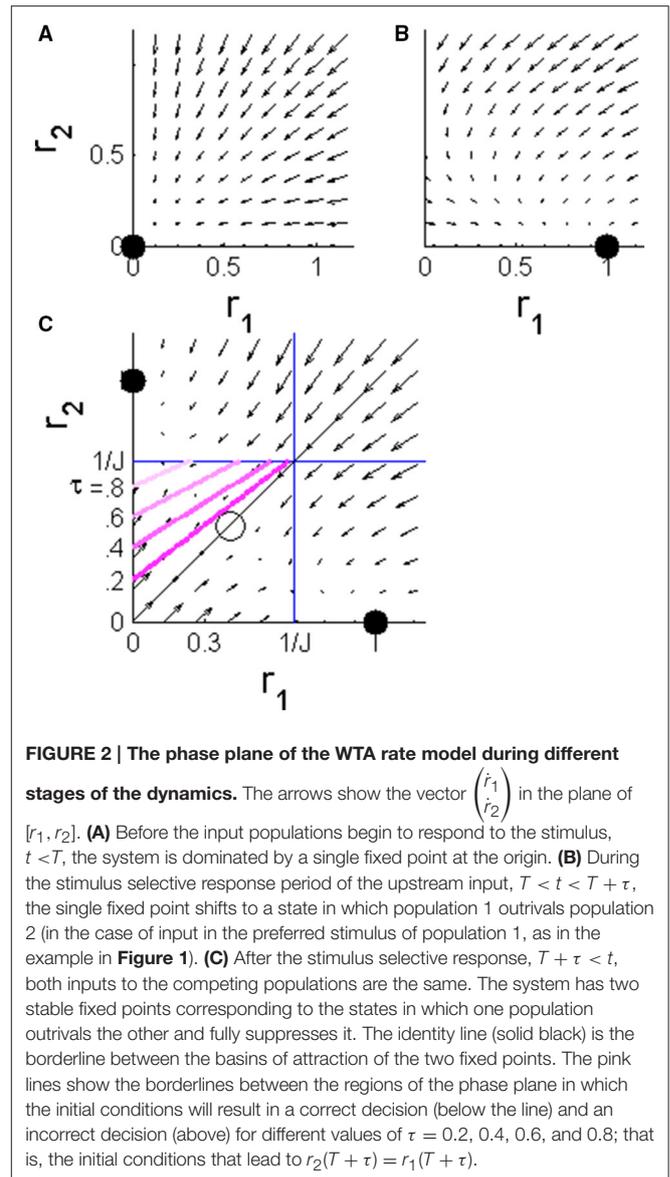
response latency of the two inputs is assumed to be tuned to the external stimulus. The task of the readout mechanism is to infer the stimulus based on the inputs it receives. For purposes of studying the ability of the readout to select according to the latency of its inputs, we further assume that both inputs have the same strength; however, the input corresponding to the “correct choice” responds faster in some sense. Specifically, we modeled the inputs as a step function,  $I_i(t) = \Theta(t - T - \tau_i)$ , where  $\Theta$  is the Heaviside step function,  $T$  is an absolute delay in the response and  $\tau_i$  is the stimulus selective delay (latency tuning:  $\tau_i = 0$  at the preferred stimulus, and  $\tau_i = \tau > 0$  otherwise); **Figure 1A**.

Our tWTA mechanism is based on reciprocal inhibition between the two populations, **Figure 1B**. Denoting by  $r_i$  the mean firing rate of neurons in the  $i$ th population ( $i = 1, 2$ ), the neuronal dynamics in our model obey

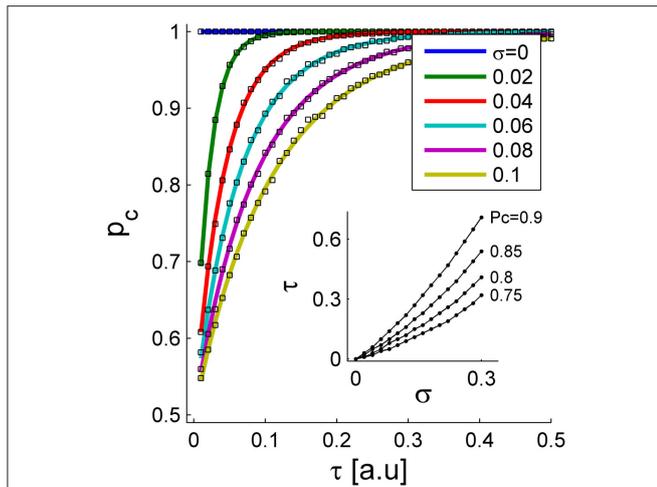
$$\dot{r}_1 = -r_1 + g(I_1 - Jr_2) \tag{1}$$

$$\dot{r}_2 = -r_2 + g(I_2 - Jr_1) \tag{2}$$

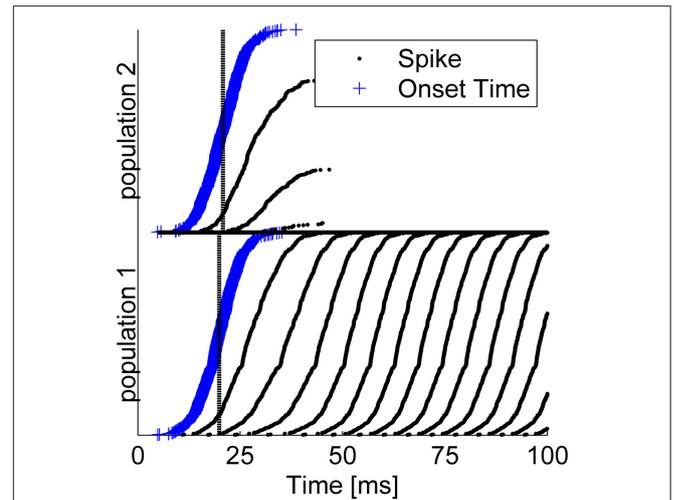
where  $J$  is the strength of the reciprocal inhibition, and  $g(x)$  is typically modeled by a sigmoidal function. Here, for simplicity of the analysis, we used a threshold-linear function,  $g(x) = x$  for  $x > 0$  and  $g(x) = 0$  otherwise. We now consider the dynamics in the case where the stimulus is the preferred stimulus of population 1,



consequently population 1 receives the “faster” input. During the absolute delay period (i.e., from stimulus onset at time  $t = 0$  to time  $T$ , see **Figure 1A**) the downstream populations have yet to respond to the stimulus,  $I_1 = I_2 = 0$ . In this case, the system is governed by a single stable fixed point at the origin,  $r_1 = r_2 = 0$ , to which the system converges exponentially; **Figure 2A**. For the duration of the stimulus selective response,  $t \in [T, T + \tau]$  the single fixed point changes to  $(1, 0)$  in which population 1 “outrivals” population 2, in this case; **Figure 2B**. This fixed point will serve as an attractor and the system will be drawn toward it. At longer times,  $t > (T + \tau)$ , both populations receive the same level of input from the upstream population. If the reciprocal inhibition is weak,  $J < 1$ , no population can outrival the other and the system will converge to a symmetric fixed point in which  $r_1 = r_2$ . On the other hand, for sufficiently strong inhibition,  $J > 1$ , the system has two stable fixed points in which one



**FIGURE 3 |** The probability of correct discrimination is shown as a function of the stimulus selective delay,  $\tau$ , for different initial conditions of noise levels,  $\sigma$  (different colors). The solid lines depict the analytical approximation of Equation (3). The squares show the numerical estimation obtained by averaging the results of the simulation decision over  $10^4$  repetitions. The inset shows the “signal”,  $\tau$ , required to obtain a specific level of performance,  $P_C$ , as a function of noise level,  $\sigma$ . Here we used  $J = 1.1$ .



**FIGURE 4 |** Typical example of tWTA competition between two populations of 1000 neurons during a single trial. The Figure depicts a “population raster” where each row shows the spike times of a single neuron (black dots). Bottom half shows neurons from population 1, and upper half population 2. The blue “+” signs depict the specific input latency time for each neuron ( $T + \tau_a + \delta_{a,i}$ ), and the neurons within each group are arranged according to their input latency. The dashed vertical line shows the mean latency for each group,  $T + \tau_a$ . Here we used  $T = 20$  ms,  $\tau_1 = 0$  ms,  $\tau_2 = 1$  ms,  $\Delta = 5$  ms, and  $g_L = 0.71$  mS/cm<sup>2</sup>.

population outrivals the other, **Figure 2C**. Due to symmetry, the basin of attraction of the two fixed points is separated by the identity line,  $r_1 = r_2$ .

In the absence of noise, the initial conditions are expected to be at the origin. Consequently, any amount of stimulus selective delay,  $\tau > 0$ , will tip the system into the correct basin of attraction. Thus, this trivial architecture produces infinite sensitivity to response latency, in the sense that it achieves correct discrimination for every positive  $\tau$ .

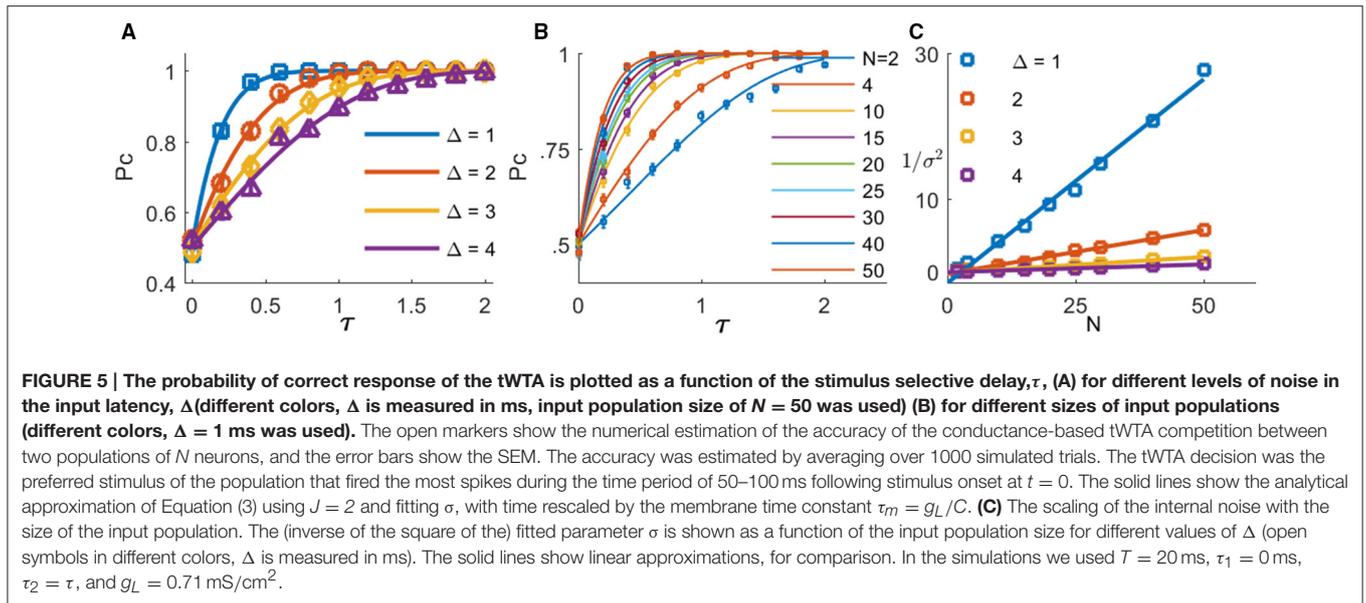
This hypersensitivity results from bi-stable dynamics where the initial conditions are exactly on the borderline between the basins of attraction of both fixed points. However, neural activity is inherently stochastic and it is unreasonable to assume the WTA type competition will start from such fine-tuned initial conditions. To study the robustness of this readout mechanism we added noise to the initial conditions and investigated the dependence of the tWTA accuracy on the noise level. Thus, instead of starting the WTA competition exactly at the origin,  $r_1 = r_2 = 0$ , we now assume that prior to the stimulus selective response period the activity of both populations fluctuates such that the WTA dynamics starts from a “cloud” of distribution close to the origin. Specifically, for the sake of analytical simplicity, we assume the initial conditions  $r_1(T)$  and  $r_2(T)$  to be independent and identically distributed exponential random variables with a standard deviation  $\sigma$ , at the onset of the tWTA competition at the beginning of the stimulus selective response, time  $t = T$ . Assuming further that the noise level is small relative to  $1/J$ , one obtains that for  $t + T \in [T, T + \tau]$

$$\begin{aligned} r_1(T + t) &= 1 - r_1(T)Je^{-t} + (r_1(T) - 1)e^{-t} \\ r_2(T + t) &= r_2(T)e^{-t} \end{aligned}$$

where we have assumed without loss of generality that stimulus 1 was presented first. During the stimulus selective response the dynamics attracts the system toward the “correct” fixed point, “outrival 1” in this case. An incorrect decision will occur when the system is at the basin of attraction of “outrival 2” at the end of the stimulus selective period, namely  $r_2(T + \tau) > r_1(T + \tau)$ , which is translated to  $r_2(T) > ar_1(T) + b$ , where  $a = 1/(1 + J\tau)$ , and  $b = (e^\tau - 1)/(1 + J\tau)$ ; see pink lines in **Figure 2C**. Averaging over the distribution of the initial conditions yields the probability of a correct discrimination,  $P_C$ , in this approximation

$$P_C = 1 - \frac{e^{-\frac{b}{\sigma}}}{a + 1} \tag{3}$$

**Figure 3** shows the probability of correct discrimination as a function of  $\tau$  for different noise levels in the initial conditions,  $\sigma$ , in different colors. The solid lines show the analytical approximation of Equation (3), and the squares depict the numerical estimation of  $P_C$ . Note that according to Equations (1) and (2),  $\tau$  is measured in units of the neuronal rate dynamics, which are typically in the range of about 10 ms. In the absence of noise (blue line), the readout mechanism discriminates correctly,  $P_C = 1$ , for any positive  $\tau$ . As the noise level increases, the probability of a correct discrimination deteriorates for any given  $\tau$ , and higher values of  $\tau$  are required to obtain the same level of accuracy. Thus, the neural noise sets the scale of the sensitivity that this readout mechanism can achieve. In other words, for larger noise levels the signal,  $\tau$ , has to be scaled up to achieve the same level of performance; **Figure 3** inset. Note that the accuracy of the tWTA readout has already been addressed in



the past, (Shamir, 2009; Zohar et al., 2011; Shriki et al., 2012). Here we focus on the dynamical system that can implement this computation. However, the parameter  $\sigma$  does not reflect noise in the input that limits the information in it, but rather an inherent variability of the readout mechanism itself that limits its sensitivity.

Rate models describe neural activity by a continuous parameter, and as a result can react immediately to changes in their input. Real spiking neurons have an inherent delay before emitting a spike, and even then their response is sparse in time. This makes it crucial to test the generalization of our results beyond the threshold-linear rate model to a model of spiking neurons. For this purpose, we simulated a tWTA readout mechanism based on competition by reciprocal inhibition between two populations of  $N$  Hodgkin-Huxley neurons. This was facilitated by the mapping between rate and conductance-based models (Shriki et al., 2003). The dynamic equation for the membrane potential of neuron  $i$  in population  $a \in \{1, 2\}$  is as follows

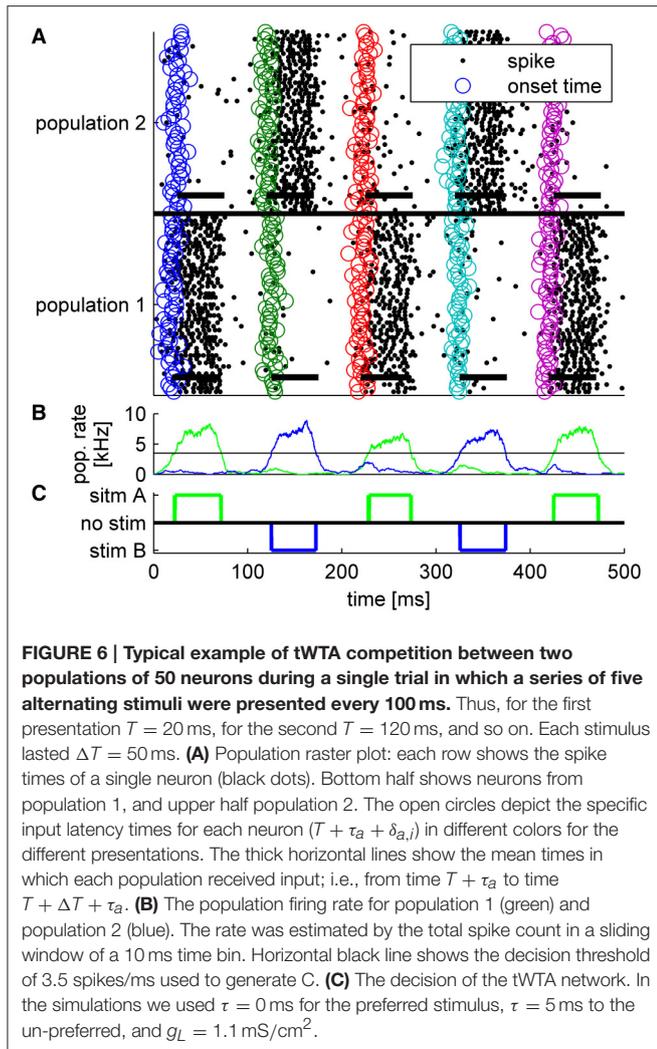
$$C \frac{dV_{a,i}}{dt} = g_L (E_L - V_{a,i}(t)) - I_{a,i}^{\text{active}} + I_{a,i}^{\text{ext}} + I_{a,i}^{\text{net}}$$

where  $g_L$  and  $E_L$  are the leak conductance and reversal potential,  $C$  is the membrane capacitance, and  $I_{a,i}^{\text{active}}$  is the voltage dependent current  $I_{a,i}^{\text{active}} = I_{Na} + I_K + I_A$ . The term  $I_{a,i}^{\text{net}}$  denotes the reciprocal inhibition from the competing population. Unless otherwise stated, the parameters of the model follow (Shriki et al., 2003); see Methods. The term  $I_{a,i}^{\text{ext}}$  denotes the input to cell  $i$  in population  $a \in \{1, 2\}$  from the input layer that is latency-tuned to the stimulus. In our simulations,  $I_{a,i}^{\text{ext}}$  is the input current from the upstream population to neuron  $i$  in population  $a$ , and is modeled by  $I_{a,i}^{\text{ext}}(t) = r \Theta(t - [T + \tau_a + \delta_{a,i}]) + \xi_{a,i}(t)$ , where  $r$  is the strength of the response of the upstream population to the stimulus,  $\xi_{a,i}(t)$  is a stimulus-independent white noise of the upstream population,  $T$  and  $\tau_a$  are the absolute and stimulus

selective delays, respectively, and  $\delta_{a,i}$  represents the trial-to-trial variability in the delay of the input to the neuron. We modeled  $\{\delta_{a,i}\}$  by i.i.d. Gaussian random variables with zero mean and a standard deviation  $\Delta$ . The term  $I^{\text{net}}$  denotes currents resulting from the lateral connection. Specifically, in this case  $I^{\text{net}}$  is the reciprocal inhibition, and its strength is governed by the synaptic strength  $g_s$  (see Methods for more details).

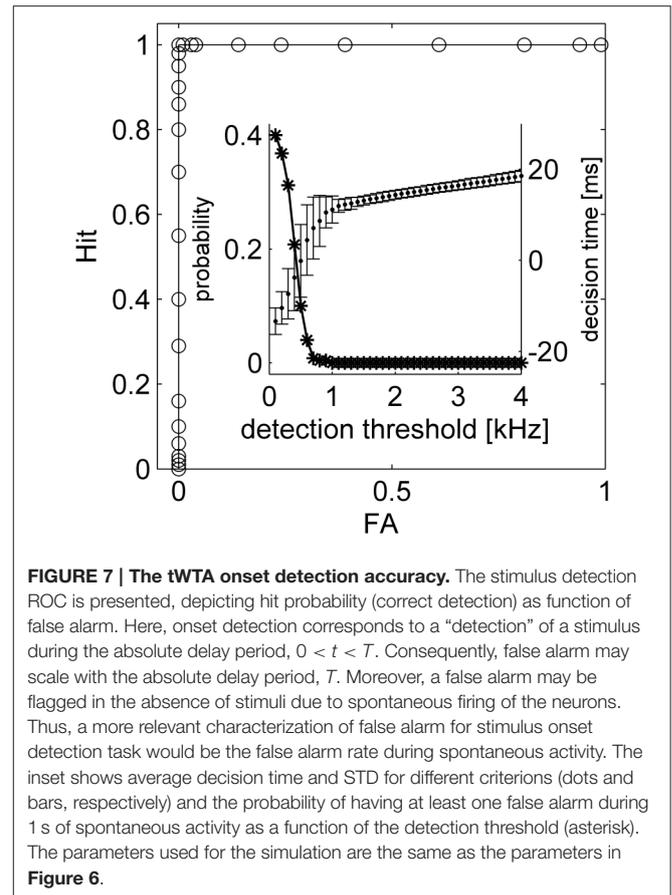
As we are interested in a WTA-type competitive readout mechanism, the reciprocal inhibition must be sufficiently strong to enable a bi-stable regime in which one population can outlive the other and fully suppress its activity. For this reason, we need to set the total maximum conductance of population internetwork (inhibitory) connection to a level that causes the network to be in a winner-take-all state (just over  $0.7 \mu\text{F}/\text{cm}^2$ ).

Figure 4 shows the result of a competition between the two populations of the readout mechanism during a single trial with:  $T = 20$  ms,  $\tau_1 = 0$  ms,  $\tau_2 = 1$  ms ( $T + \tau$  shown by the dashed lines), and  $\Delta = 5$  ms. The blue plus signs (“+”) denote the specific onset time of the input to each cell within this trial,  $T + \tau_a + \delta_{a,i}$ . The neurons in each group are arranged in the Figure according the latency of their input (early to late from bottom to top). Note that the mean latency difference between the inputs to the two populations was  $\tau \equiv \tau_2 - \tau_1 = 1$  ms, whereas within each group input latency varied with a much greater standard deviation of 5 ms. Given the considerable variability in the input times to single cells, some neurons in population 2 fired up to three spikes in response to the stimulus. Nevertheless, within several tens of milliseconds the reciprocal inhibition dynamics converged to a state in which population 1 fired with a high rate and the activity of population 2 was fully suppressed (Figure 4). Thus, even though in their steady state both populations received the same input level, this relatively small mean latency difference was sufficient to enable the reciprocal inhibition mechanism to “choose” the correct group in this specific example.



To evaluate the extent to which the result in **Figure 4** is typical, we estimated numerically the probability of correct discrimination. To do so, we first needed to define an objective criterion for a “correct” response by the system. In our simulations we chose the winner in the WTA competition to be the population that fired more spikes during the time interval of  $50 \text{ ms} \leq t \leq 100 \text{ ms}$ . Other choices yielded qualitatively similar results.

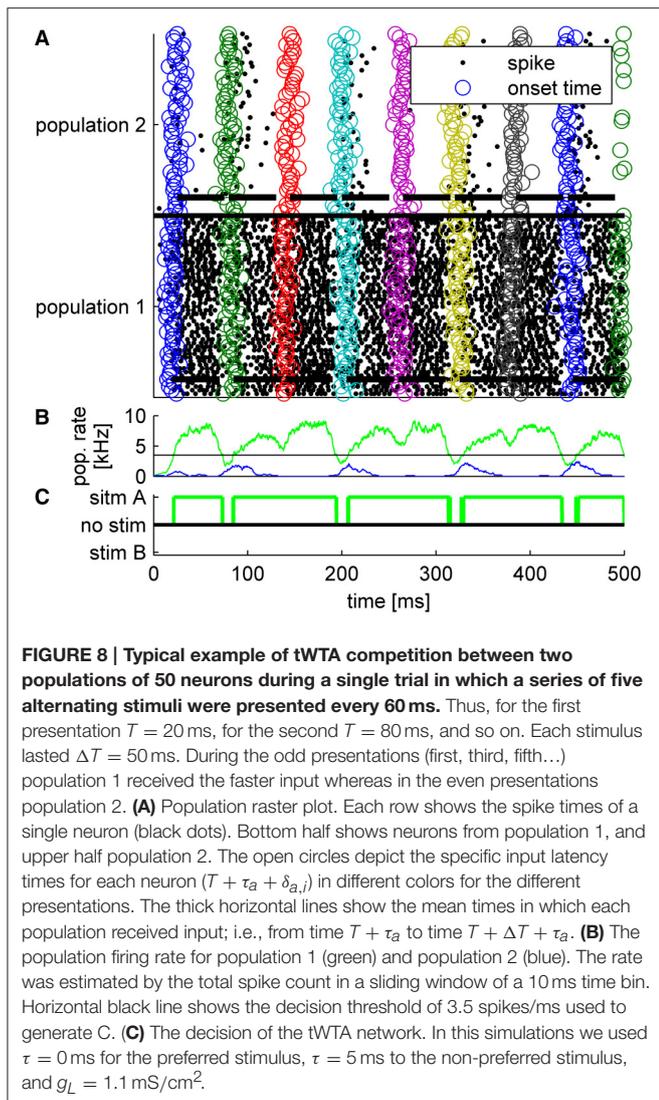
**Figure 5A** shows the probability of the Hodgkin-Huxley type model to discriminate correctly between the two alternatives as function of the mean input latency difference between the two populations,  $\tau \equiv \tau_2 - \tau_1$ , for different levels of within-group onset variation  $\Delta$  (shown by the different colors). The solid lines are the analytical approximation of Equation (3) with  $J = 2$  and fitted values of  $\sigma$  (note that time in Equation (3) is measured in units of  $\tau_m = g_L/C$ ). As can be seen from the figure, even small mean latency differences of less than  $\tau \approx 1$  ms can be detected with a high reliability by this simple mechanism. Note, that it is expected that in the limit of large input population sizes the results of the rate model will hold, and in particular Equation (3).



**Figure 5B** shows the probability of the spiking neurons model to discriminate correctly between the two alternatives as function of the mean input latency difference,  $\tau \equiv \tau_2 - \tau_1$ , for different sizes of input populations (shown by different colors). As can be seen from the figure, even though discrimination accuracy deteriorates with the decrease in population size, this simple readout mechanism is still able to extract information from the neuronal response latencies. Furthermore, fitting the analytical approximation of Equation (3) (solid lines), we find that the parameter that reflects the noise in the input population in the rate model,  $\sigma$ , decays to zero as one over the square root of the input population size, **Figure 5C**. Hence, one may attribute the deterioration in discrimination accuracy to a decrease in signal-to-noise ratio in the responses of the input populations for smaller population sizes.

## DISCUSSION

Previous studies of the WTA algorithm have focused on the conventional rate-WTA decision mechanism. However, a number of studies reported that the WTA is also sensitive to the temporal structure of its inputs (Coultrip et al., 1992; Lee et al., 1999; Jin and Seung, 2002; Sheliga et al., 2006; Standage et al., 2005; Kurt et al., 2008). Here we focused on the ability of a dynamical system to implement a specific latency-based decision



mechanism; namely, the tWTA readout. We found that the simple reciprocal inhibition architecture of the WTA mechanism is sufficient to implement the tWTA. To implement the tWTA we assumed that the system was in the “strong inhibition” regime ( $J > 1$ ), such that it was bi-stable when the inputs to the two competing population were similar. This condition was needed, as we assumed that the input strength to both populations was similar and that response latency alone depended on the stimulus. However, in many cases neurons tend to respond with a higher firing rate to the same stimuli to which they respond with a shorter latency (Zohar et al., 2011; Shriki et al., 2012). In these cases the assumption of “strong inhibition” can be relaxed.

The robustness of the reciprocal inhibition mechanism for latency coding was tested against noise in the initial conditions. Additional parameters that govern the tWTA dynamics may also vary. In particular, we assumed that the reciprocal inhibition between the two populations is identical. This assumption contributed to the symmetry that is underlying the infinite sensitivity of the tWTA mechanism in the absence of noise. How

will asymmetry in the reciprocal inhibition affect the tWTA? Assume, for example that the effective inhibition from population 2 to 1,  $J_{1 \leftarrow 2}$ , is stronger than from population 1 to 2,  $J_{2 \leftarrow 1}$ . We find that as long the system is in the strong inhibition regime, i.e.,  $J_{2 \leftarrow 1}, J_{1 \leftarrow 2} > 1$ , then for sufficiently large  $\tau$  the system will converge to the correct fixed point (in the absence of noise). However, for any positive difference,  $(J_{2 \leftarrow 1} - J_{1 \leftarrow 2}) > 0$ , there exists a critical value  $\tau_c$ , such that for any  $\tau < \tau_c$  (here positive  $\tau$  denotes input to population 1 preceding input to population 2) population 2 will win the tWTA competition. The critical value  $\tau_c$  decays to zero as the system approaches symmetry,  $(J_{2 \leftarrow 1} - J_{1 \leftarrow 2}) \rightarrow 0$ . Consequently, an additional “learning mechanism” that can fine-tune the reciprocal inhibition is required if high sensitivity for very short latencies is necessary. Note that the parameter  $J$  reflects the effective inhibition strength that is a product of the strength of a single synapse by the number of synapses.

It has been argued that latency code readouts require an additional stimulus onset detector which serves as a reference from which latency can be measured. A number of studies have reported that onset can be detected by pooling information from neurons that show poor latency tuning (Zohar et al., 2011; Brasselet et al., 2012; Shriki et al., 2012). Here we suggest that such an additional mechanism is not essential. **Figure 6A** shows the spiking response of the tWTA network to a series of stimuli. In the first, third and fifth presentations the mean latency of the inputs to population 1 was shorter, whereas in the second and fourth presentations population 2 received the fastest input (on average). By examining the population firing rate in **Figure 6B**, stimulus 1, stimulus 2, and no stimulus can easily be discriminated, by adding a decision threshold (black horizontal line in **Figure 6B**). Setting a reasonable value for the decision threshold results in correct detection and discrimination for all 5 presentations without the use of an additional neural population or detectors; **Figure 6C**. **Figure 7** quantifies the tradeoff between generating false alarms at a high rate by setting the decision threshold too low, and failing to detect the stimulus by setting the threshold too high. The considerable difference between the spontaneous firing rate and the firing rate in response to the stimulus allows for a high detection probability and a negligible false alarm rate, as shown by the sharp ROC curve. However, decision threshold also affects the speed of the response and the probability of false alarm during spontaneous activity; **Figure 7** inset.

The tWTA mechanism does not operate like a working memory but rather more like a sensory system. Thus, in the example in **Figure 6** the neural activity decays to a spontaneous firing rate after each stimulus offset. However, this decay is not instantaneous. Therefore, if the inter-stimulus-interval is very brief, the tWTA decision may be affected by its past. This scenario is illustrated in **Figure 8**, in which due to the short inter-stimulus-interval the tWTA decision is dictated by the preceding decision rather than by the stimulus itself. This is a hallmark of tWTA competition which may serve as an empirical prediction both on the psychophysical as well as the neural level. In this respect the well-studied psycho-acoustical phenomenon termed the *precedence effect* (Litovsky et al., 1999) may have some

bearing. In the precedence effect, when estimating sound source location, the perception of the delayed stimulus is suppressed. This is believed to assist in overcoming the corrupting effects of echoes in the computation of the sound source location.

Conventional rate-WTA mechanisms have been suggested to play an important role in various computations in the brain and in numerous systems (Lee et al., 1999; Standage et al., 2005; Sheliga et al., 2006, 2007; Kurt et al., 2008). Their implementation only requires a very rudimentary reciprocal inhibition design. Nevertheless, even this simple architecture yields a mechanism that is intrinsically sensitive to the response latency of its inputs. Consequently, by construction, almost every conventional WTA mechanism is highly sensitive to latency cues as well. This

highlights the possible role of latency as a source of information in the central nervous system.

## AUTHOR CONTRIBUTIONS

OZ, MS: Methodological, analysis, and interpretation of results.

## FUNDING

This work was supported in part by Israel Science Foundation ISF grants No 722/10 and 300/16, and by the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative of Ben-Gurion University of the Negev.

## REFERENCES

- Ashida, G., and Carr, C. E. (2011). Sound localization: jeffress and beyond. *Curr. Opin. Neurobiol.* 21, 745–751. doi: 10.1016/j.conb.2011.05.008
- Brasaset, R., Panzeri, S., Logothetis, N. K., and Kayser, C. (2012). Neurons with stereotyped and rapid responses provide a reference frame for relative temporal coding in primate auditory cortex. *J. Neurosci.* 32, 2998–3008. doi: 10.1523/JNEUROSCI.5435-11.2012
- Brugge, J. F., Reale, R. A., and Hind, J. E. (1996). The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *J. Neurosci.* 16, 4420–4437.
- Brugge, J. F., Reale, R. A., Jenison, R. L., and Schnupp, J. (2001). Auditory cortical spatial receptive fields. *Audiol. Neurootol.* 6, 173–177. doi: 10.1159/000046827
- Chase, S. M., and Young, E. D. (2007). First-spike latency information in single neurons increases when referenced to population onset. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5175–5180. doi: 10.1073/pnas.0610368104
- Coultrip, R., Granger, R., and Lynch, G. (1992). A cortical model of winner-take-all competition via lateral inhibition. *Neural Netw.* 5, 47–54. doi: 10.1016/S0893-6080(05)80006-1
- Fukui, T., and Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: from winner-take-all to winners-share-all. *Neural Comput.* 9, 77–97. doi: 10.1162/neco.1997.9.1.77
- Gautrais, J., and Thorpe, S. (1998). Rate coding versus temporal order coding: a theoretical approach. *Biosystems* 48, 57–65.
- Gawne, T. J., Kjaer, T. W., and Richmond, B. J. (1996). Latency: another potential code for feature binding in striate cortex. *J. Neurophysiol.* 76, 1356–1360.
- Gollisch, T., and Meister, M. (2008). Rapid neural coding in the retina with relative spike latencies. *Science* 319, 1108–1111. doi: 10.1126/science.1149639
- Goodman, D. F., and Brette, R. (2010). Spike-timing-based computation in sound localization. *PLoS Comput. Biol.* 6:e1000993. doi: 10.1371/journal.pcbi.1000993
- Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiol. Rev.* 90, 983–1012. doi: 10.1152/physrev.00026.2009
- Grothe, B., and Koch, U. (2011). Dynamics of binaural processing in the mammalian sound localization pathway—the role of GABA(B) receptors. *Hear. Res.* 279, 43–50. doi: 10.1016/j.heares.2011.03.013
- Hall, C., Bell, C., and Zelik, R. (1995). Behavioral evidence of a latency code for stimulus intensity in mormyrid electric fish. *J. Comparative Physiol. A* 177, 29–39. doi: 10.1007/bf00243396
- Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455
- Jin, D. Z., and Seung, H. S. (2002). Fast computation with spikes in a recurrent neural network. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 65(5 Pt 1):051922. doi: 10.1103/PhysRevE.65.051922
- Johansson, R. S., and Birnie, I. (2004). First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Nat. Neurosci.* 7, 170–177. doi: 10.1038/nn1177
- Joris, P., and Yin, T. C. (2007). A matter of time: internal delays in binaural processing. *Trends Neurosci.* 30, 70–78. doi: 10.1016/j.tins.2006.12.004
- Klug, A., Khan, A., Burger, R. M., Bauer, E. E., Hurley, L. M., Yang, L. C., et al. (2000). Latency as a function of intensity in auditory neurons: influences of central processing. *Hear. Res.* 148, 107–123.
- Kurt, S., Deutscher, A., Crook, J. M., Ohl, F. W., Budinger, E., Moeller, C. K., et al. (2008). Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions. *PLoS ONE* 3:e1735. doi: 10.1371/journal.pone.0001735
- Laing, C. R., and Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *J. Comput. Neurosci.* 12, 39–53. doi: 10.1023/A:1014942129705
- Lee, D. K., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* 2, 375–381. doi: 10.1038/7286
- Leibold, C., and van Hemmen, J. L. (2005). Spiking neurons learning phase delays: how mammals may develop auditory time-difference sensitivity. *Phys. Rev. Lett.* 94:168102. doi: 10.1103/PhysRevLett.94.168102
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (1999). The precedence effect. *J. Acoust. Soc. Am.* 106, 1633–1654.
- Lüling, H., Siveke, I., Grothe, B., and Leibold, C. (2011). Frequency-invariant representation of interaural time differences in mammals. *PLoS Comput. Biol.* 7:e1002013. doi: 10.1371/journal.pcbi.1002013
- McAlpine, D. (2005). Creating a sense of auditory space. *J. Physiol.* 566(Pt 1), 21–28. doi: 10.1113/jphysiol.2005.083113
- McGinley, M. J., Liberman, M. C., Bal, R., and Oertel, D. (2012). Generating synchrony from the asynchronous: compensation for cochlear traveling wave delays by the dendrites of individual brainstem neurons. *J. Neurosci.* 32, 9301–9311. doi: 10.1523/JNEUROSCI.0272-12.2012
- Panzeri, S., and Diamond, M. E. (2010). Information carried by population spike times in the whisker sensory cortex can be decoded without knowledge of stimulus time. *Front. Synaptic Neurosci.* 2:17. doi: 10.3389/fnsyn.2010.00017
- Panzeri, S., Petersen, R. S., Schultz, S. R., Lebedev, M., and Diamond, M. E. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron* 29, 769–777. doi: 10.1016/S0896-6273(01)00251-3
- Shamir, M. (2009). The temporal winner-take-all readout. *PLoS Comput. Biol.* 5:e1000286. doi: 10.1371/journal.pcbi.1000286
- Shamir, M. (2014). Emerging principles of population coding: in search for the neural code. *Curr. Opin. Neurobiol.* 25, 140–148. doi: 10.1016/j.conb.2014.01.002
- Shamir, M., Ghitza, O., Epstein, S., and Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Comput. Biol.* 5:e1000370. doi: 10.1371/journal.pcbi.1000370
- Sheliga, B. M., FitzGibbon, E. J., and Miles, F. A. (2007). Human vergence eye movements initiated by competing disparities: evidence for a winner-take-all mechanism. *Vision Res.* 47, 479–500. doi: 10.1016/j.visres.2006.09.027
- Sheliga, B. M., Kodaka, Y., FitzGibbon, E. J., and Miles, F. A. (2006). Human ocular following initiated by competing image motions: evidence for a winner-take-all mechanism. *Vision Res.* 46, 2041–2060. doi: 10.1016/j.visres.2005.11.033

- Shriki, O., Kohn, A., and Shamir, M. (2012). Fast coding of orientation in primary visual cortex. *PLoS Comput. Biol.* 8:e1002536. doi: 10.1371/journal.pcbi.1002536
- Shriki, O., Hansel, D., and Sompolinsky, H. (2003). Rate models for conductance-based cortical neuronal networks. *Neural Comput.* 15, 1809–1841. doi: 10.1162/08997660360675053
- Standage, D. I., Trappenberg, T. P., and Klein, R. M. (2005). Modelling divided visual attention with a winner-take-all network. *Neural Netw.* 18, 620–627. doi: 10.1016/j.neunet.2005.06.015
- Van Rullen, R., and Thorpe, S. J. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* 13, 1255–1283. doi: 10.1162/08997660152002852
- White, J. A., Chow, C. C., Ritt, J., Soto-Treviño, C., and Kopell, N. (1998). Synchronization and oscillatory dynamics in heterogeneous, mutually inhibited neurons. *J. Comput. Neurosci.* 5, 5–16. doi: 10.1023/A:1008841325921
- Zhou, Y., Carney, L. H., and Colburn, H. S. (2005). A model for interaural time difference sensitivity in the medial superior olive: interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology. *J. Neurosci.* 25, 3046–3058. doi: 10.1523/JNEUROSCI.3064-04.2005
- Zohar, O., Shackleton, T. M., Nelken, I., Palmer, A. R., and Shamir, M. (2011). First spike latency code for interaural phase difference discrimination in the guinea pig inferior colliculus. *J. Neurosci.* 31, 9192–9204. doi: 10.1523/JNEUROSCI.6193-10.2011

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2016 Zohar and Shamir. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*