Spiking Neurons Learning Phase Delays: How Mammals May Develop Auditory Time-Difference Sensitivity

Christian Leibold* and J. Leo van Hemmen

Physik Department, Technische Universität München, 85747 Garching bei München, Germany
(Received 29 November 2004; published 26 April 2005)

How is the central auditory pathway set up to handle information that is encoded with a temporal precision of microseconds? An example that has recently lit an increasingly controversial discussion [1] is the representation of interaural time differences (ITDs) in the brainstem of small mammals, in particular, rodents. It has been shown [2] that in Mongolian gerbils neurons in the medial superior olive (MSO) of a specific brain hemisphere have maximal firing rate when a sound arrives first at the ear located at the opposite side of the head, a property called "tuning asymmetry"; see Fig. 1.

There are two intriguing aspects of the above findings [2]. First, the asymmetry is mediated by precisely timed inhibition. Second, the peak of neuronal discharge rate as a function of ITD occurs at a phase of about 0.1 to 0.2 cycles independently of the auditory frequency channel the cell belongs to. The latter is referred to as a phase delay and its physical cause is heavily debated. There is also data from cat [3] and guinea pig [4] midbrains providing further evidence for a systematic asymmetry that is consistent with the notion of phase delay. There is now agreement [1] that mammals, at least partially, rely on a MSO tuning asymmetry as a basis of the neuronal encoding of sound source location. It is thus important to understand how the asymmetry develops in a self-organized and robust fashion. Here we report how the tuning asymmetry can be explained by correlation-driven (Hebbian) [5–8] synaptic plasticity.

In order to set up a model, we first give a brief description of the underlying biological system. Mammalian MSO neurons receive excitatory (+) as well as inhibitory (−) input from both ears (I, C). For frequencies below 1.5 kHz the temporal structure of the stimulus is basically preserved in the discharge times of MSO inputs through phase locking [9,10]. Asymmetric tuning of MSO cells is attributed to the binaural interaction between excitation and inhibition. It is generally believed [2,11] that contralateral inhibition arrives earlier than bilateral excitation, whereas ipsilateral inhibition carries only little temporal information. Furthermore, tuning asymmetry develops during a critical period [12] around the first few days after hearing onset, so that in young animals tuning curves of downstream neurons are initially symmetric. Even more interestingly, if the auditory experience of animals in the critical period is reduced to omnidirectional noise, no development of asymmetric tuning curves is found [12].

The above findings are strong indications for a plastic process taking place at hearing onset and leads to the main idea of the present Letter. We propose there is a synaptic "learning rule" that strengthens inhibitory synapses that have fired a substantial but small enough time interval before a MSO action potential occurs and depresses those that have fired at the same time [13] or later.

More precisely, if we describe the strength of excitatory (+) and inhibitory (−) synapses by scalar variables \( J_n^{(\pm)} \),

```latex
\begin{align*}
\text{MSO Discharge Rates} & \quad \text{Ear} \quad \text{Ear} \\
\text{Contralateral Hemisphere} & \quad \text{Sound Source} \\
\text{ITD<0 Ipsilateral leading} & \quad \text{ITD>0 Contralateral leading}
\end{align*}
```

Fig. 1. Tuning asymmetry of a gerbil. (a) A sound source (large circle) evokes high neuronal discharge rates in a brainstem nucleus MSO in the opposite (contralateral) brain hemisphere and low rates in the ipsilateral MSO at the same side of the source. (b) Corresponding ITD tuning curve from a gerbil; based on Fig. 3(c) of Brand et al. [2]. Tuning maxima are statistically significant [2] at contralateraly leading (positive) ITDs, mostly even outside the physically accessible ITD range (vertical dashed lines).

DOI: 10.1103/PhysRevLett.94.168102 PACS numbers: 87.19.Dd, 43.64.+r, 87.19.La
our learning algorithm is as follows. Each input spike alters the weight $J_n^{(\pm)}$ by a constant $\eta^{(\pm)}$ $w^{(\pm)} N$, each MSO spike induces a change $\eta^{(\pm)} w^{(\pm)}$. In addition, each pair of input and MSO spikes at times $t_{in}$ and $t_{out}$ accounts for an associative weight change that depends on the time difference $\Delta t = t_{in} - t_{out}$ and has the magnitude $\eta^{(\pm)} w^{(\pm)}(\Delta t)$; see Fig. 2(a). The functions $W^{(\pm)}$ are generally referred to as learning windows [5]. In the limit of small learning rates $\eta$ as they naturally occur, the learning rule for the synaptic weights $J^{(\pm)}$ can be approximated [6] by the mean-field dynamics

$$\frac{dJ_n^{(\pm)}}{dt} = \eta^{(\pm)} \left[ \sum_{X=\text{in/out}} w^{(\pm)} X \nu X + \int_{-\infty}^{t} ds W^{(\pm)}(s) C_n(s) \right]$$

(1)

where $\nu^{in/out}$ denote the mean pre- and postsynaptic firing rates and $C_n(s) = T^{-1} \int_0^T dt P_n(t, t + s)$ is the temporal average over the compound probability density $P_n(t, t + s)$ of having a pre- and postsynaptic spike at times $t_{in} = t + s$ and $t_{out} = t$, respectively. The averaging time $T$ thereby has to substantially exceed neuronal time constants, yet it has to be smaller than the time scale of synaptic weight change [6]. Specification of $\nu^{in/out}$ and $C$, however, requires a model of the statistics of input and output spikes.

For the input part we start with a realization of a homogeneous Poisson process with a mean rate of 100 events/s providing a time series $\{t^i\}$ that is supposed to mimic the stochasticity of the auditory environment. We then obtain four classes $\{(I, +), (I, -), (C, +), (C, -)\}$ of input processes described by the homogeneous Poissonian rate densities $P_{i,c}^{(\pm)}(t) = 0.1 \text{kHz}(1 - c_i^{(\pm)} \alpha) + 1 \text{kHz} e_i^{(\pm)} \alpha \sum_{x,c} \exp[-(t-t')/(1 \text{ ms})]$. The first term on the right represents a homogeneous uncorrelated background activity, whereas the second induces temporal correlation on a time scale that is relevant to synaptic dynamics in the auditory system, i.e., roughly 1 ms; cf., Fig. 2(a). The value of the correlation strength $\alpha$ is taken to be 0.5, unless mentioned otherwise. It determines the overall proportion of correlated to background activity. Motivated by morphological studies [10], we also have to postulate a mechanism that induces symmetry breaking, in that the temporal acuity of ipsilateral inhibitory $(I, -)$ MSO input is slightly worse than that of its contralateral counterpart $(C, -)$. Excitatory inputs $\{(I, +), (C, +)\}$ and inhibitory inputs from the contralateral $(C, -)$ side are thus obtained by a factor $c_i^{(\pm)} = c_i^{(+)} = c_i^{(-)} = 1$, whereas ipsilateral inhibitory inputs are generated with $0 < c_i^{(\pm)} < 1$.

The positive constant $c_i^{(+)}$ measures the ratio of temporal correlation strengths between ipsilateral and contralateral inhibitory inputs. The mean input rate is 100 Hz and does not depend on $c_i^{(\pm)}$ and $\alpha$. The input spike trains are then conveyed to MSO neurons by a considerable number of synaptic contacts [14]. From each hemisphere we assume

$$n^{(+)} \text{ excitatory and } n^{(-)} \text{ inhibitory inputs that are delayed by } \Delta^{(+)}_{n,C,n} \text{, } n = 1, \ldots, N^{(+)}; \text{ cf., [15]. Delays are considered as effective latencies from the ear to the MSO neuron that take into account all occurring mechanical, axonal, and synaptic delays. Independently of the neuron model, the linear response of the membrane potential on a synaptic input spike can be described by kernels } e^{(\pm)} \text{ [16]. If one then models spike generation as a Poisson process such that the probability density of firing is linear in the synaptic potential

$$v(t) = \sum_{h=\text{in/in}} \sum_{t=1}^{N_h^{(\pm)}} \sum_{s} f_h(t_s) \int_{-\infty}^{t} ds W^{(\pm)}(s) C_n(s)$$

(2)
Eq. (1) becomes linear in the weights $J_{I,C}^{(z)}$ and eigenspaces determine synaptic structure formation [6].

Introducing a Fourier transform $\hat{f}(\mu) = \int_{-\infty}^{\infty} dt e^{-i \mu t} \lambda \times f(t)$ on the domain $[0, \Lambda]$ of available transmission delays $\Delta^{(z)}_{n}$, we obtain two classes (1l) of eigenvectors $\Phi^{(1l)}(\mu) = [\phi^{(1l)}(\mu)(+)^{(1l)}(\mu)]^{T}$, where $\Phi^{(1l)}(\mu)$ denote the eigenvectors’ excitatory and inhibitory components, respectively. Here $\Phi^{(1l)}(\mu) = [c^{(+)}(\mu)\Phi^{(+)}(\mu), -N^{(+)}(\mu)]^{T}$ has vanishing eigenvalues and hence does not contribute to the synaptic structure. On the other hand, for $\mu \neq 0$, $\Phi^{(1)}(\mu) = [c^{(+)}(\mu)\Phi^{(+)}(\mu), c^{(-)}(\mu)\Phi^{(-)}(\mu)]^{T}$ has eigenvalues

$$\lambda(\mu) = a^2\tilde{C}(\mu) \sum_{\alpha = \pm 1} (c^{(\sigma)}(\mu)\Phi^{(\sigma)}(\mu)\tilde{e}(\sigma)(\mu)$$

with positive real part [6] and explains three major issues.

First, the inhibitory part $\phi^{(-)}(\mu)$ of $\Phi^{(1)}(\mu)$ and, hence, synaptic structure formation, diminishes as the inhibitory correlation $c^{(-)}$ decreases, which is supposed to occur among the ipsilateral afferents (1,-). Second, the eigenvalues of $\Phi^{(1)}$ are proportional to $\alpha^2$. This prevents structure formation for small temporal input correlations, which is putatively the case in noise box experiments. Third, a phase delay between excitation and inhibition, as observed in experiments [2,4], requires excitation $\tilde{e}^{(+)}(\mu)\Phi^{(+)}(\mu)$ to be aptly delayed with respect to inhibition $\tilde{e}^{-}(\mu)\Phi^{-}(\mu)$. Our learning rule is capable of generating a phase delay $\varphi$ if, in the frequency range where $|\tilde{W}^{(+)}(\mu)|$ is not negligibly small, the learning windows are related to each other approximately like $W^{(+)}(\mu)\tilde{e}^{(+)}(\mu) \approx e^{-i \varphi} \arg(\mu) W^{(+)}(\mu)\tilde{e}^{-}(\mu)$.

The theoretical predictions have then been tested by computer simulations. Input spikes evoke synaptic currents $I_{I,C}^{(z)}(t)$ that depend on the $1 \leq n \leq N^{(z)}$ synaptic weights $J_{I,C}^{(z)}$ on synapse-specific delays $\Delta_{I,C}^{(z)}$ from the ear to the MSO neuron, and on input firing times $\{t_{I,C}^{(z)}\}$ stemming from $N^{(z)}$ independent processes with respective densities $p_{I,C}^{(z)}$ so that

$$I_{I,C}^{(z)}(t) = \sum_{n=1}^{N^{(z)}} \sum_{\Delta^{(z)}_{I,C} > 0} e^{-r_{I,C}^{(z)}(t_{I,C}^{(z)} + \Delta_{I,C}^{(z)})/r^{(z)}_{I,C} + \Delta^{(z)}_{I,C} < t}.$$ 

Both the excitatory and inhibitory synaptic time constants have been chosen to be $r^{(z)} = 0.1$ ms [11]. The currents $I_{I,C}^{(z)}$ are then used as an input to a MSO neuron that is modeled as a threshold element whose membrane potential $V$ (resting potential is set to 0) is calculated according to $dv/dt = -v/\tau_{M} - \frac{1}{R}(I_{I,C}^{(z)} + I_{I}^{(z)}) + (I_{C}^{(z)} + I_{I}^{(z)})$. For the sake of simplicity we have assumed that the resting potential equals the inhibitory reversal potential. The effective membrane time constant was chosen to be $\tau_{M} = 0.2$ ms, which is biophysically motivated by a low-threshold outward current [17]. The fraction 2/15 is obtained from a ratio of 2 between inhibitory and excitatory conductance and a reversal potential (60 mV) being 15 times the threshold (4 mV) [18]. Every 100 ms an ITD is randomly drawn from the interval $[-120 \mu s, +120 \mu s]$ and symmetrically added (subtracted) to (from) the ipsi-(contra)-lateral input spike times $t_{I,C}^{(z)}$. Model details have been specified elsewhere [15].

We also assume $N^{(+)} = 180$ excitatory and $N^{(-)} = 120$ inhibitory afferents from each side (i.e., both ipsi- and contralateral). The learning windows $W^{(+)}$ [see Fig. 2(a)] are parametrized as $W(\delta t) = (A - B) \exp([-\delta t - s^*]/\tau_0)$ for $\delta t < s^*$ and $W(\delta t) = A \exp(-[\delta t - s^*]/\tau_1) - B \exp(-[\delta t - s^*]/\tau_2)$ for $\delta t \geq s^*$. They are the same for excitation and inhibition except for the numerical values of the parameters [19]. For the present choice of parameters the phase delay predicted by the mean-field theory is $\varphi = -\arg(W^+/W^-) = 0.25$ c in the frequency range between 0.1 and 1 kHz. The magnitude of $\eta^{(z)}$ sets the time scale of synaptic modification. In our simulations we arrive at stable synaptic configurations after $T = 3000$ seconds of formal time. In the living animal $\eta^{(z)}$ is to be rescaled so that $T$ becomes a few days.

In Fig. 2(b) we demonstrate the development of synaptic delay distributions at a single cell. Initially, the distributions of afferent delays are unstructured; see Fig. 2(b) top. A spike-timing-dependent learning algorithm [5,7,8,13] tunes the distributions in such a way that excitatory delays become restricted to a small interval, about 250 $\mu$s wide. Inhibitory connections are potentiated most if their delays are slightly shorter than the remaining excitatory ones; see Fig. 2(b). After a learning period $T$, strong temporal correlations among the contralateral inhibitory inputs have led to a delay structure where contralateral inhibition arrives at the cell before excitation. At T, the inhibitory current evoked by a contralateral pulse exhibits a width of 0.5 ms at half maximum (not shown). On the other hand, temporal correlation among ipsilateral inhibitory inputs is too weak for delay selection and thus the inhibitory delays remain unstructured.

In the fully developed state almost all synaptic weights have reached their upper or lower bounds. As a result, ITD tuning of the cell is asymmetric (triangles), whereas ITD functions obtained with deactivated inhibition (asterisks) are symmetric, as are those derived from a synaptic structure that is subject to learning with reduced input correlations $\alpha = 0.2$ (squares); see Fig. 2(c). The latter explains data from gerbils demonstrating that a reduction of acoustic correlations during the critical period almost completely eliminates ITD tuning asymmetry of downstream cells [12]. On hearing onset (crosses) initial conditions can be chosen so that ITD tuning is symmetric.

Although the delay distributions of Fig. 2(b) have been obtained through learning with a nonperiodic input process, tuning curves generated by means of periodic input of various frequencies exhibit the same properties as experimentally found in gerbils [2] and guinea pigs [4]; see Fig. 3. The tuning curves in Fig. 3(a) are due to best fits.
of the Gabor function to the tuning data obtained by simulation. The ITD \( \tau_{\text{best}} \) where a tuning function exhibits a maximum is called best ITD. Best ITDs decrease monotonically with stimulation frequency [Fig. 3(b)], though best interaural phase disparities (IPD = ITD \times frequency) are nearly constant [2] at a phase delay of \( \phi_{\text{sim}} = 0.1 \) c. The value of \( \phi_{\text{sim}} \), however, is smaller than that obtained by linearization since threshold models have a tendency to fire at the rise of excitation, hence earlier. Nevertheless, Fig. 3(b) shows that constant best IPDs may well be a result of the filtering of afferent spike trains through appropriately tuned axonal delay distributions [cf., Fig. 2(a)] rather than frequency-dependent preprocessing at earlier stages of the auditory pathway.

In summary, the present study provides an explanation of response properties of cells in the medial superior olive of mammals. In particular, we show how a phase delay, determined by the ratio \( \bar{W}^+ / \bar{W}^- \) between both learning windows, can emerge in a self-organized fashion. The responsible microscopic mechanism is proposed to be spike-time-dependent synaptic plasticity. This hypothesis is supported by anatomical changes of inhibitory synapses in MSO [14] as well as by activity-dependent inhibitory synaptic plasticity of other projections from the same source [20]. Our model predicts that temporal precision of synaptic currents is enhanced during ontogenetic development and, moreover, that in young and noise-box-reared animals, tuning curves of MSO cells are symmetric.

The authors thank Benedikt Grothe and Christoph Kapfer for stimulating discussions. C. L. has been supported by the DFG (Forschergruppe Hörobjecte).

*Present address: Institute for Theoretical Biology, Humboldt Universität zu Berlin, 10115 Berlin, Germany.
[15] After a MSO spike, the synaptic currents and membrane potential are set to 0 and the neuron is put into a refractory state lasting 1 ms. The firing threshold of the neuron equals the amplitude of a voltage response to a single excitatory input spike with weight \( J^{(+)} \) = 1. Excitatory weights are initialized from a Gaussian distribution with mean 0.04 and standard deviation \( \sigma = 0.012 \), inhibitory weights have a mean value of 0.06 and \( \sigma = 0.018 \). Weights are restricted to the interval [0, 0.12]. The distributions of delays \( \Delta_{I,C,a}^{(\pm)} \) are taken to be Gaussian with mean 1 ms and \( \sigma = 0.3 \) ms.
[19] Learning windows are specified by \( A^{(+)}/A^{(-)} = 2/3 \), \( B^{(+)}/B^{(-)} = 0.989 \), \( \eta^{(+)}/\eta^{(-)} = 4 \times 10^{-4} \), \( w_{\text{in}}^{(+)}/w_{\text{in}}^{(-)} = 1/20 \), \( w_{\text{out}}^{(+)}/w_{\text{out}}^{(-)} = -1/5 \), \( \tau_{\text{in}}^{(+)}/\tau_{\text{in}}^{(-)} = 1/10 \) ms, \( \tau_{\text{en}}^{(+)}/\tau_{\text{en}}^{(-)} = 1/20 \) ms, \( \tau_{\text{exc}}^{(+)}/\tau_{\text{exc}}^{(-)} = 4 \) ms, \( s_{\text{in}}^{(+)}/s_{\text{in}}^{(-)} = -1/40 \) ms, \( A^{(+)}/A^{(-)} = 2/3 \), \( B^{(+)}/B^{(-)} = 0.49 \), \( \eta^{(+)} = 2.4 \times 10^{-4} \), \( w_{\text{in}}^{(+)}/w_{\text{in}}^{(-)} = -1/20 \), \( w_{\text{out}}^{(+)}/w_{\text{out}}^{(-)} = 1/4, \tau_{\text{en}}^{(+)}/\tau_{\text{en}}^{(-)} = 1/5 \) ms, \( \tau_{\text{exc}}^{(+)}/\tau_{\text{exc}}^{(-)} = 1/10 \) ms, \( \tau_{\text{exc}}^{(+)}/\tau_{\text{exc}}^{(-)} = 1/2 \) ms, and \( s_{\text{in}}^{(+)}/s_{\text{in}}^{(-)} = -1/5 \) ms.