2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

A Computational Model for Pitch Pattern Perception with the Echo State Network

Miwa Fukino[†], Yuichi Katori[‡ ¶], and Kazuyuki Aihara [¶]

†Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
‡School of Systems Information Science, Future University Hakodate
116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan
¶Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
Email: fukino@sat.t.u-tokyo.ac.jp, katori@fun.ac.jp, and aihara@sat.t.u-tokyo.ac.jp

Abstract—The predictive coding theory assumes that the sensory system of the cortex continuously predicts incoming stimuli and detects residual errors. The mismatch negativity (MMN) is a neural response to a deviance of learned regularity, and is regarded as an error signal in this theory. Here we report a preliminary study on a computational model of the auditory MMN using the Echo State Network which is one of the recurrence neural network models. We trained the network by an oddball task with two pitch patterns. The result shows that our model simulates a qualitatively similar waveforms with the MMN response to deviant pitches.

1. Introduction

The auditory mismatch negativity (MMN) (for review, see [1]) is a neural response to a deviance of learned regularity. It is one of the event related potentials which arises between 100 and 200 ms after the deviant stimulus onset, and can be measured by electroencephalography (EEG) or magnetoencephalogram (MEG). When a deviant magnitude is higher, the MMN amplitude is larger and the latency is smaller. When a deviant probability is lower, the MMN amplitude is larger and there is no effect for the latency.

The MMN has been widely used in clinical and theoretical studies, for example, as a biomarker of psychosis [2, 3], and as a research of brain plasticity in terms of comparing the difference between musicians' and non-musicians' MMN [4, 5]. However, neurophysiological mechanisms of the MMN are still controversial.

The MMN is often regarded as an error signal of the predictive coding model [6] in recent studies [7]. Several mathematical models of the MMN are proposed based on the predictive coding idea (e.g. [8, 9]). Wacongne et al. (2012) proposed a neuronal model of the auditory cortex accounting for the MMN, whose scheme is as follows. There are four components in the model: Thalamic inputs, Prediction error layer, Predictive layer, and Memory trace. An input sound stream is composed of two pitches A and B. Population of neurons in the Predictive layer continuously predicts incoming Thalamic inputs. Population of neurons in the Prediction error layer receives two inputs: inhibitory inputs coming from the Prediction layer and excitatory inputs, or the Thalamic inputs, and then the residual is the error signal, namely MMN. The error signal is transmitted to the Prediction layer to adjust the internal model of the prediction. Memory neurons keep the inputs of past few hundred milliseconds.

Wacongne et al. used the spiking neuron model [10], and implemented precise neuronal behavior in terms of realistic receptors (AMPA, NMDA, and GABA), synapses, and spiking neurons. However, it is difficult to apply this model for processing more complex sound patterns, because it focuses on the precise neuronal behavior of the primary auditory cortex, and it does not consider more complex auditory patterns.

The Echo State Network (ESN) [11, 12, 13] is one of artificial recurrent neural networks, where neurons are sparsely and randomly connected, and only outputs are trained. The scheme of ESN is as follows. $x_i(n)$ is the ith neuron ($i = 1, \ldots, N$) at time n in the dynamical reservoir. $d_j(n)$ and $y_j(n)$ are the jth teacher data and system output ($j = 1, \ldots, L$) at time n, respectively. $\mathbf{W}^{rec}$, $\mathbf{W}^{back}$, and $\mathbf{W}^{out}$ are weight matrices of recurrent connections inside the dynamical reservoir, feedback connections from outputs to reservoir, and output connections from the reservoir to system outputs, respectively.

The activation of internal units is updated according to

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{W}^{rec}\mathbf{x}(n) + \mathbf{W}^{back}\mathbf{y}(n)), \quad (1)$$

$$\mathbf{y}(n+1) = \mathbf{f}^{out}(\mathbf{W}^{out}(\mathbf{x}(n+1), \mathbf{y}(n)), \quad (2)$$

where $\mathbf{f}$ and $\mathbf{f}^{out}$ denotes the individual units' transfer functions. The internal weights $\mathbf{W}^{rec}$ and feedback weights $\mathbf{W}^{back}$ are set to random and sparse at first, and not changed during the training. Only the output weights $\mathbf{W}^{out}$ are trained. The echo state network can learn nonlinear systems, although the structure is simple.

In the present study, we propose the predictive coding model with ESN for processing complex and realistic auditory patterns, and provide a learning procedure for the proposed network model.

2. Model

We propose a computational model for pitch pattern perception based on Jaeger's echo state network model [11] and the predictive coding architecture of Wacongne et al.'s neuronal model [9]. Structure of the proposed model is shown in Fig. 1

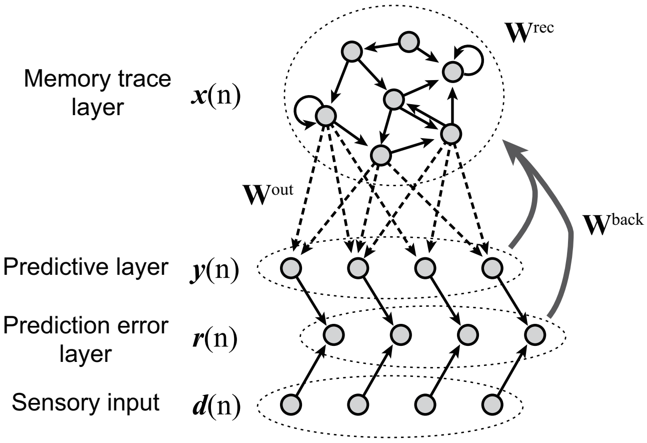

Figure 1: Structure of proposed model

2.1. Overall structure of the network

$\mathbf{x}(n)$ is a status of memory trace layer at time n (N_x dimensions), $\mathbf{y}(n)$ is a status of predictive layer (N_y dimensions), $\mathbf{r}(n)$ is a status of prediction error layer (N_y dimensions), and $\mathbf{d}(n)$ is a status of given sensory inputs (N_y dimensions).

The state of each layer is updated according to

$$\mathbf{x}(n+1) = \mathbf{x}(n) + \frac{\delta}{\tau}\Bigg(-\alpha_0\mathbf{x}(n) + \mathbf{f}_x\Bigg(\mathbf{W}^{rec}\mathbf{x}(n-k_x) + \mathbf{W}^{back}(\mathbf{y}(n-k_y) + \mathbf{r}(n-k_r))\Bigg)\Bigg), \tag{3}$$

$$\mathbf{y}(n) = \mathbf{f}_y(\mathbf{W}^{out}\mathbf{x}(n)), \tag{4}$$

$$\mathbf{r}(n) = \mathbf{f}_r(\mathbf{d}(n) - \mathbf{y}(n)), \tag{5}$$

where $\mathbf{W}^{rec}$ is $N_x \times N_x$ weight matrices of recurrent connections, $\mathbf{W}^{back}$ is $N_x \times N_y$ weight matrices of feedback connections, and k_x, k_y, k_r are delay times. We define $f_x(x) = \tanh x$, $f_y(x) = \tanh x$, and $f_r(x) = \max(x, 0)$. Initial state of $\mathbf{x}(n)$ is randomized as a uniform distribution over $-0.2 \sim 0.2$.

2.2. Configuration of Network

Step1 : Setting $\mathbf{W}^{rec}$ and $\mathbf{W}^{back}$
We define the network of proposed model based on [11].

1. Generate an internal weight matrix $\mathbf{W}_0$. Assign 1 or -1 to randomly selected $\beta_r N_x \times N_x$ components of $\mathbf{W}_0$. Assign 0 to the other units.
2. Normalize $\mathbf{W}_0$ to a matrix $\mathbf{W}_1$ with unit spectral radius by putting $\mathbf{W}_1 = 1/|\lambda_{max}|\ \mathbf{W}_0$, where λ_{max} is maximum eigenvalue of $\mathbf{W}_0$.
3. Scale $\mathbf{W}_1$ to $\mathbf{W}^{rec} = \alpha_r \mathbf{W}_1$.
4. For $\mathbf{W}^{back}$, first, randomly assign $-1 \sim 1$ to $\beta_b N_x \times N_y$ components of $\mathbf{W}_3$ with a uniform distribution. Assign 0 to the other units. Then, normalize the sum of columns of the unit vectors, multiply with coefficient α_b, and set it to $\mathbf{W}^{back}$.

Step2 : Training $\mathbf{W}^{out}$
In our proposed method, the weight values for the output connection $\mathbf{W}^{out}$ are computed by a few iterative epochs.

1. As initial values of $\mathbf{W}^{out,(m)}$, all the components of $\mathbf{W}^{out,(0)}$ are set to 0.
2. Using $\mathbf{W}^{out,(m-1)}$ which is calculated in the $m-1$th epoch, the mth weight matrix $\mathbf{W}^{out,(m)}$ is calculated. Using the sensory input $\mathbf{d}(n)$ with the time ranges $n = 0, ..., T_1$, the model is driven with "teacher-focing", which means that the feedback of output state $\mathbf{y}$(n) is replaced with the teacher signal (the given sensory input) $\mathbf{d}$(n) as following equations.

$$\mathbf{x}(n+1) = \mathbf{x}(n) + \frac{\delta}{\tau}\Bigg(-\alpha_0\mathbf{x}(n) + \mathbf{f}_x\Bigg(\mathbf{W}^{rec}\mathbf{x}(n-k_x) + \mathbf{W}^{back}(\mathbf{d}(n-k_y) + \mathbf{r}(n-k_r)) + \sigma\xi(n)\Bigg)\Bigg), \tag{6}$$

$$\mathbf{y}(n) \quad = \quad \mathbf{f}_y(\mathbf{W}^{out,(m-1)}\mathbf{x}(n)), \tag{7}$$

$$\mathbf{r}(n) \quad = \quad \mathbf{f}_r(\mathbf{d}(n) - \mathbf{y}(n)), \tag{8}$$

where $\xi(n)$ has random values of normal distribution with mean 0 and variance 1. σ specifies the amplitude of random values. $\mathbf{x}(n)$ and $\mathbf{y}(n)$ at time from T_0 to T_1 are used to calculate $\mathbf{W}^{out,(m)}$. Next, input the time series of $\mathbf{x}(n)$ into state collecting matrix M, where M is $(T_1 - T_0 - 1) \times N_x$ matrix. Then, input sigmoid-inverted $\mathbf{d}(n)$, or $\tanh^{-1}\mathbf{d}(n)$ into G, where G is $(T_1 - T_0 - 1) \times N_y$ matrix. $\mathbf{W}^{out,(m)}$ is calculated using ridge regression as

$$(\mathbf{W}^{out,(m)})^T = (M^T M + \lambda E)^{-1} M^T G, \tag{9}$$

where λ is a coefficient for adjusting a sparseness, and E is a unit matrix.

3. After the I iterations of above calculation, we get $\mathbf{W}^{out,(I)}$ as $\mathbf{W}^{out}$ by repeating the learning for several times (I =10 iterations are enough).

3. Results

We use an input data simulating the oddball paradigm which is often used for the stimulus of MMN experiments. The red curves in Fig.2 (b) shows the input data. The lower and upper curve indicate pitch A and B, respectively. Repetition of two pitches A and B make a stream of patterns: the data includes 80% of standard patterns ”AAAB” and 20% of deviant pitches (e.g.,”AAA$\underline{\text{A}}$B”).

We demonstrate our model with the input data simulating the oddball paradigm using following parameter values. $N_x = 200$, $N_y = 2$, $\alpha_r = 0.6$, $\beta_r = 0.1$, $\alpha_b = 0.8$, $\beta_b = 0.1$, $\alpha_0 = 0.7$, $\tau = 2.5$, $\delta = 1$, $k_r = 5$, $k_x = 10$, $k_y = 0$, $\sigma = 10^{-5}$, and $\lambda = 0.1$.

Figure 2 shows the typical response of the proposed model. The red and blue curves of (b) are the teacher data (**d**(n) on the Sensory input of Fig.1) and resulting data (**y**(n) on the Predictive layer) respectively. Many red curves of (a) are time series of all neurons on the dynamical reservoir (**x**(n) on the Memory trace layer). Green curves of (c) are time series of residual between the teacher and resulting data (**r**(n) on the Prediction error layer).

The response on the dynamical reservoir (many red curves of Fig.2 (a)) vary among units, and a set of the whole units represent the resulting data (blue curves in Fig.2 (b)). The time series $\mathbf{r}(n) = f_r(d - y)$ (green curves in Fig.2 (c)) is the residual between the the Sensory input (red curves of Fig.2 (b)) and the prediction, or the resulting data (blue curves in Fig.2 (b)). They are correspond to the error signal of predictive coding, or the MMN.

We can observe that responses on the predictive layer (the blue curves of Fig.2 (b)) are reproducing the sensory input (red curve). Especially, they are almost similar at the timing of peaks, that is, A or B sound is played. On the contrary, when the sounds are omitted or deviant, the response on the prediction layer shows wrong predictions. Note that the deviant sound (AAABAAABAAA$\underline{\text{A}}$B) is played at the timing of a black arrow. There is a small predicted wave arising at the upper blue curve.

Thus, our model simulates qualitatively reproduce waveform of the MMN response to deviant pitches. This results indicate that our proposal model can learn the input pattern, and possibly predict an underlying nonlinear regularity behind the data.

4. Conclusion

We have reported our preliminary study on a computational model of the auditory MMN using the Echo State Network. We trained the network by an oddball task with two pitch patterns. We have showed that our model simulations can qualitatively reproduce waveforms of the MMN response to deviant pitches.

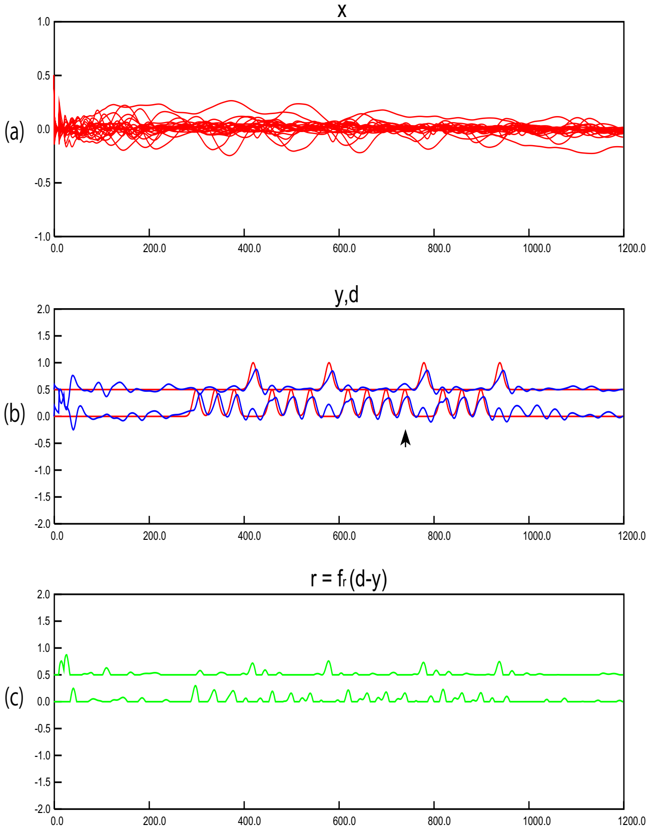

Figure 2: Typical response of the proposed model. The red and blue curves of (b) are the teacher and resulting data respectively. Many red curves of (a) are time series of all **x**(n) in the memory trace layer. Green curves of (c) are time series of **r**(n).

This report is the first step towards a computational model of the MMN. It is an important future problem to analyze this model by changing parameters, and discuss the relation between the computational and implementation levels of analysis in terms of Marr’s levels of analysis.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 16K00246. This work was funded by ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan).

References

[1] R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, & I. Winkler, “‘Primitive intelligence’ in the auditory cortex,” *Trends Neurosci.* vol.24, pp.283-288, 2001.

[2] T. Nagai, M. Tada, K. Kirihara, T. Araki, S. Jinde, & K. Kasai, “Mismatch negativity as a ‘translatable’ brain marker toward early intervention for psychosis: A review,” *Front. Psychiatry*, vol.4, pp.1-10, 2013.

[3] R. Näätänen, T. Shiga, S. Asano, & H. Yabe, “Mismatch negativity (MMN) deficiency: A breakthrough biomarker in predicting psychosis onset,” *Int. J. Psychophysiol.*, vol.95, pp.338-344, 2015.

[4] T. L. van Zuijen, E. Sussman, I. Winkler, R. Näätänen, & M. Tervaniemi, “Auditory organization of sound sequences by a temporal or numerical regularity - a mismatch negativity study comparing musicians and non-musicians,” *Brain Res. Cogn. Brain. Res.*, vol.23, pp.270-276, 2005.

[5] S. C. Herholz, C. Lappe, & C. Pantev, “Looking for a pattern: an MEG study on the abstract mismatch negativity in musicians and nonmusicians,” *BMC Neurosci.*, vol.10, 42, 2009.

[6] R. P. Rao, & D. H. Ballard, “Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects,” *Nat. Neurosci.*, vol.2, pp.79-87, 1999.

[7] I. Winkler, & I. Czigler, “Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations,” *Int. J. Psychophysiol.*, vol.83, pp.132-143, 2011.

[8] F. S. Lieder, K. E. Stephen, J. Daunizeau, M. I. Garrido, & K. J. Friston, “A neurocomputational model of the mismatch negativity,” *PLoS Comput. Biol.*, vol.9, e1003288, 2013.

[9] C. Wacongne, J. P. Changeux, & S. Dehaene, “A neuronal model of predictive coding accounting for the mismatch negativity,” *J. Neurosci.*, vol.32, pp.3665-3678, 2012.

[10] E. M. Izhikevich, “Simple model of spiking neurons,” *IEEE Trans. Neural Netw.*, vol.14, pp.1569-1572, 2003.

[11] H. Jaeger, “A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the ‘echo state network’ approach,” (Revision of H. Jaeger, “Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the ‘echo state network’ approach,” *GMD Report 159*, pp.48, 2002), pp.1-46, 2005.

[12] H. Jaeger, “Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication,” *Science*, vol.304, pp.78-80, 2004.

[13] M.Lukoševičius, & H. Jaeger, “Reservoir computing approaches to recurrent neural network training,” *Comput. Sci. Rev.*, vol.3, pp.127-149, 2009.