# Recognizing Sound Signals Through Spiking Neurons and Spike-timing-dependent Plasticity

Yan Liu
School of Systems Science
Beijing Normal University
Beijing, China
bnuliuyan@bnu.edu.cn

Jiawei Chen
School of Systems Science
Beijing Normal University
Beijing, China
chenjiawei@bnu.edu.cn

Liujun Chen
School of Systems Science
Beijing Normal University
Beijing, China
chenlj@bnu.edu.cn

## ABSTRACT
Spiking Neural Networks (SNNs) are regarded as brain-inspired neural networks. Most SNNs described spiking neurons with the leaky integrate-and-fire model, which does not incorporate biological properties of real neurons. In this paper, a model motivated by the human auditory pathway is proposed to explore the possible sound signals recognition mechanism based on the biological dynamic properties of Hodgkin-Huxley (HH) neurons and the spike-timing-dependent-plasticity (STDP) rule of synapses. The first mechanism is that HH neurons have the property of frequency selective response. They only respond to their characteristic frequencies in burst spike trains, which makes the recognition of sound intensity based on the dynamic neurons become possible. The second mechanism is that according to the STDP rule, a synaptic connection structure is formed, and the frequency and the intensity information of input signals are stored in the synaptic delay times. Finally, the neural networks recognize sound signals with spatiotemporal firing patterns.

## CCS Concepts
**•Applied computing → Life and medical sciences → Computational biology → Biological networks**

## Keywords
Sound Signals Recognition; Spiking Neural Networks; Hodgkin-Huxley Neuron; Spike-timing-dependent Plasticity.

## 1. INTRODUCTION
Spiking Neural Networks (SNNs) are regarded as third generation neural networks because of their power efficiency and competitive capabilities in several cognitive tasks such as object recognition and speech recognition [1][2][3]. The reason SNNs are called brain-inspired neural networks is that neurons and synapses in SNNs are similar with those in the biological neural networks. There are two main differences between SNNs and artificial neural networks (ANNs) [4]. Firstly, the neurons in SNNs communicate with each other through bursting spikes, while the neurons in ANNs communicate by real numbers. Secondly, the training methods of most ANNs are based on the non-local propagation, which is implausible in brain, where neurons can only communicate with others with real connections. In SNNs, neurons are connected by synapses and the training methods are also based on the mechanisms found in brain, such as spike-timing-dependent-plasticity (STDP) [5][6].

For recognizing sound signals, several SNN models have been proposed, which mostly used leaky integrate-and-fire (LIF) model to describe a spiking neuron [5][7][8]. The LIF neuron is a simple neuron model which fires when its membrane potential reaches its threshold, so it is a good model for an integrator. However, many behaviors of the real neurons are mediated by ion channel dynamics, which determine the various dynamic and powerful properties of the real neurons. In some cases when the aim is just to improve the computing efficiency, the simplicity of LIF neuron is an acceptable sacrifice. While in the case of developing brain-inspired neural networks, we need to consider the biological neurons and their dynamic functional roles in brain [9]. Hodgkin-Huxley (HH) model is the first biologically plausible model of spiking neurons, which incorporates the detailed dynamics of the membrane potential and the $Na^+$, $K^+$, and leak ion channels [10]. Later, several biophysical models have been developed to characterize the dynamics of real neurons [11]. Therefore, these biological neuron models should be incorporated in further SNNs to explore the brain-inspired neural networks [12].

Since SNNs aim to build the brain-inspired neural networks, for developing the neural network models to recognize sound signals, the mechanism that we have known about human auditory pathway should be the most important reference. In this paper, a model motivated by the mechanism in human auditory pathway is proposed to simulate the recognition of auditory signals. The neural network is developed based on HH neurons and the STDP rule, which both include important biological properties in the real neural system. The aim of this paper is to explore the possible role of the dynamic mechanisms of neurons and synapses in recognizing sound signals.

## 2. METHODS
### 2.1 Input Encoding
The input encoding method is based on the mechanism in human auditory pathway. Sound signals are firstly transmitted through the auditory receptors, where the sound signals are converted into neural signals, and then transmitted to brain stem, MGN, and auditory cortex for further integrating. An important process of translating the physical sound signals into neural signals is in the cochlea, where the two main features of sounds, intensity and frequency, are represented in a typical way.



*AIPR 2019*, August 16–18, 2019, Beijing, China

ACM ISBN 978-1-4503-7229-9/19/08…$15.00
DOI: https://doi.org/10.1145/3357254.3357264

In human auditory pathway, the sound frequency is coded according to the tonotopic map [13]. Based on the mechanism in the cochlea, the sound signal is first decomposed by frequencies based on tonotopy, and each neuron is most sensitive at its characteristic frequency. Neurons connected to hair cells near the apical basilar membrane have low characteristic frequencies, and those connected to hair cells near the basal basilar membrane have high characteristic frequencies. So, the frequency information of the input signal is coded by the position of neurons. In this modal, a sound signal is firstly Fourier transformed and divided into several frequencies (for simplicity, here divided into 10 frequencies), and then the signal of each frequency is inversed Fourier transformed. The neurons of different positions in the input layer receive the corresponding signals, which models the mechanism of neurons on different positions decomposing sound signals based on frequencies.

In human auditory pathway, the sound intensity is coded in the firing rates of neurons and the number of active neurons. When the intensity is increased, the membrane potential of the activated hair cells will be more depolarized or hyperpolarized, and then the auditory nerve fiber will fire faster. According to this mechanism, in this model, the neurons in the input layer fire burst spike trains. There are two temporal information in a burst spike train, the inter-burst period and the intra-burst period. The intensities of sound signals on a typical frequency are coded by the intra-burst periods. The smaller the intra-burst period, the greater the sound intensity at that moment. The times when each intensity appears are coded by the inter-burst periods.

Thus, the frequency and intensity information of input sound signals are coded by the positions and firing patterns of the neurons in the input layer.

## 2.2 Dynamic Model of Spiking Neurons

There are two types of spiking neurons in the network. The first type of neuron is HH neuron, which is described by the following equations [10],

$$C\dot{V} = k_V[I_{syn} - g_K n^4(V-E_K) - g_{Na}m^3h(V-E_{Na}) - g_L(V-E_L)] \quad (1)$$

$$\dot{x} = k_V[\alpha_x(V)(1-x) - \beta_x(V)x], \qquad x = n,m,h \quad (2)$$

in which $V$ is the membrane potential. $n,m,h$ are the gating variables $\alpha_n(V) = 0.01(10-V)/(\mathrm{e}^{(10-V)/10}-1)$, $\beta_n(V) = 0.125\,\mathrm{e}^{-V/80}$, $\alpha_m(V) = 0.1(25-V)/(\mathrm{e}^{(25-V)/10}-1)$, $\beta_m(V) = 4\,\mathrm{e}^{-V/18}$, $\alpha_h(V) = 0.07\,\mathrm{e}^{-V/20}$, $\beta_h(V) = 1/(\mathrm{e}^{(30-V)/10}+1)$, $E_K = -12\mathrm{mV}$, $E_{Na} = 120\mathrm{mV}$, $E_L = 10.6\mathrm{mV}$, $g_K = 36\mathrm{mS/cm^2}$, $g_{Na} = 120\mathrm{mS/cm^2}$, $g_L = 0.3\mathrm{mS/cm^2}$, $C = 1\mu\mathrm{F/cm^2}$, $I_{syn}$ is the synaptic input current.

HH neurons have the property of frequency selective response. When the input signal to an HH neuron is a burst like current, the response property of the neuron will be different from the DC current. HH neurons respond and fire when the input DC current is higher than a threshold, which is called the property of integrate and fire. But when the input signal is burst spike trains, some HH neurons will not fire even when the frequency of the burst current is quite high. In fact, HH neurons only respond to a typical frequency of a burst spike train. It means that an HH neuron has a characteristic frequency selection from the burst spike train input. The characteristic frequency is dependent on the parameters of the neuron. Thus, a group of HH neurons with different parameters can recognize the burst spike trains with different frequencies and each neuron selects one burst frequency to response. This property is consistent with the tonotopy in human auditory pathways, which makes the recognition of sound intensity based on the dynamic neurons become possible.

The second type of neuron in the network is the continuous spike neuron, which is described as follows [14],

$$\dot{V} = g_{Na}m_\infty(E_{Na}-V) + g_K n(E_K - V) + g_L(E_L - V) + I_{syn} \quad (3)$$

$$\dot{n} = (n_\infty - n)/\tau_n \quad (4)$$

in which $n_\infty = 1/(1+e^{(V_n-V)/k_n})$, $m_\infty = 1/(1+e^{(V_m-V)/k_m})$, $g_{Na} = 20\mathrm{mS/cm^2}$, $g_K = 10\mathrm{mS/cm^2}$, $g_L = 8\mathrm{mS/cm^2}$, $E_{Na} = 60\mathrm{mV}$, $E_K = -90\mathrm{mV}$, $E_L = -80\mathrm{mV}$, $V_m = -20\mathrm{mV}$, $V_n = -25\mathrm{mV}$, $k_m = 15\mathrm{mV}$, $k_n = 5\mathrm{mV}$, $I_{syn}$ is the synaptic input current.

A continuous spike neuron maintains a resting state when there is no external signal input. When there is an action potential input, it will fire continuously.

## 2.3 Unsupervised Learning Method

The STDP rule is an unsupervised learning mechanism found in brain. The magnitude of change of synaptic weight between a pre- and a postsynaptic neuron depends on the timing of spikes: if the presynaptic spike arrives at the postsynaptic neuron before the postsynaptic neuron fires, the synapse is potentiated.

The equation of synaptic transmission is as follows [15],

$$I_{syn}(t) = \bar{g}_{syn} r(t)[V_{syn}(t) - E_{syn}] \quad (5)$$

in which $r(t)$ is the amount of neurotransmitter released from the pre-synapse, which follows $\alpha$ function $r(t) = \alpha\, t\, e^{-t/\tau}$. $t$ is the firing time of the pre-synaptic neuron, $\alpha = 0.015$, $\tau = 2\mathrm{ms}$, $\mathrm{E}_{syn} = 0\mathrm{mV}$. $k_V$ is the time parameter, which is different among neurons. $\bar{g}_{syn}$ is the maximal synaptic conduction, which is also called the synaptic weight in neural networks. In training, $\bar{g}_{syn}$ is changed according to STDP rule as follows [16],

$$\Delta\bar{g}_{syn}(\Delta t) = \begin{cases} A_+ \exp[-\Delta t/\tau_1] & \text{for } \Delta t > 0 \\ A_- \exp[\Delta t/\tau_2] & \text{for } \Delta t < 0 \end{cases} \quad (6)$$

in which $\Delta\bar{g}_{syn}$ is the change of the maximal synaptic conduction, $A_+ = -A_- = 1$, $\tau_1 = 10\mathrm{ms}$, $\tau_2 = 20\mathrm{ms}$.

## 2.4 Structure

The structure of the neural network is shown in Figure 1. The network includes the input layer, the hidden layer, and the output layer. In the input layer, 10 neurons respond to the input signals according to the tonotopy mechanism. The frequencies of input sound signals are coded by the position of the neurons, and the intensities of the input sound signals are coded by the burst spike trains fired by the neurons.

The hidden layer is consisting of two parts. The first part includes 10 HH neurons, which are fully connected with the input layer neurons. The second part includes 10*10 neurons, where the neurons on one row connect with one neuron in the first part. What's more, the neurons of the second part fully connected with each other with dynamic synapses. The synapses follow the STDP rule.

The output layer includes 10*10 continuous firing neurons, which connect with the corresponding neurons in the second part of the hidden layer.

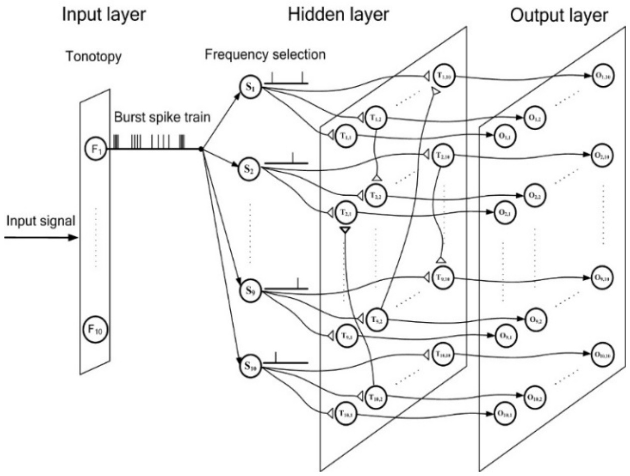


**Figure 1. The structure of the neural network.**

# 3. RESULTS

In this section, we demonstrate the capability of this neural network recognizing the sound signals. The training sound samples are the spoken sound signals of /ba/ and /ka/. The test sound sample is the spoken sound signal of /ga/.

Firstly, the training sound signals of /ba/ and /ka/ are input to the network alternatively. In the input layer, according to the tonotopy, sound signals of different frequencies are responded by the neurons in the way that the neurons on different positions are only sensitive at their characteristic frequencies. The response of a neuron in the input layer is spiking a burst spike train. The intensities of sound signals on a typical frequency are coded by the intra-burst periods in the burst spike train, and the times of each intensity appears are coded by the inter-burst periods in the burst spike train.

Next, the burst spike trains are input to the first part of the hidden layer. The neurons $S_1$-$S_{10}$ have different oscillating periods due to different parameters of $k_V$ and $\bar{g}_{syn}$, as shown in Table 1, so they only respond to their characteristic intra-burst periods. If the input burst spike train includes a burst with one neuron's characteristic frequency, the neuron will fire a spike. The spike times between neurons $S_1$-$S_{10}$ shows the temporal information of the intensities in the input signals.

Thus, the spatiotemporal firing patterns of neurons $S_1$-$S_{10}$ code both the frequency and the intensity information of the input signals.

**Table 1. Oscillating periods due to different parameters**

| | $k_V$ | $\bar{g}_{syn}(nS)$ | period (ms) |
|---|---|---|---|
| $S_1$ | 3.2 | 17.5 | 5 |
| $S_2$ | 2.9 | 18.5 | 5.5 |
| $S_3$ | 2.6 | 20.5 | 6 |
| $S_4$ | 2.4 | 21 | 6.5 |
| $S_5$ | 2.2 | 24 | 7 |
| $S_6$ | 2 | 25 | 7.5 |
| $S_7$ | 1.96 | 26 | 8 |
| $S_8$ | 1.78 | 28.4 | 9 |
| $S_9$ | 1.6 | 30 | 10 |
| $S_{10}$ | 1.52 | 33.2 | 11 |

Then the spike trains transmit to the second part of the hidden layer, where the neurons connect with each other through dynamic synapses. According to the STDP rule, synaptic parameters are adjusted based on intervals of spike times between the input neurons $S_1$-$S_{10}$, and the unadjusted synapses do not function during this signal transmission. After training, a synaptic connection structure will be formed between the neurons and the spike time intervals are stored in the synaptic delay times, as shown in Table 2.

**Table 2. Spike time intervals stored in synaptic delay times**

| $S_3$-$S_7$ | $S_7$-$S_{10}$ | $S_{10}$-$S_5$ | $S_3$-$S_{10}$ | $S_{10}$-$S_7$ | $S_7$-$S_5$ | $S_5$-$S_8$ | $S_8$-$S_1$ |
|---|---|---|---|---|---|---|---|
| 60ms | 62ms | 84ms | - | - | - | 70ms | 80ms |
| - | - | - | 52ms | 80ms | 68ms | 70ms | 80ms |

When the intra-burst periods and the inter-burst periods are coded through the parameters of neurons and synapses, the corresponding neurons in the output layer fire. The output firing pattern is shown in Figure 2.

Then the test sound signal /ga/ is input. When the input signal contains the same parts of burst spike trains as the trained signals, the network extracts the same parts of the signals. As a result, the network recognizes the same vowel /a/, and it activates the same output pattern in the output layer. Thereby the recognition of the same syllable in the sound signals is realized.

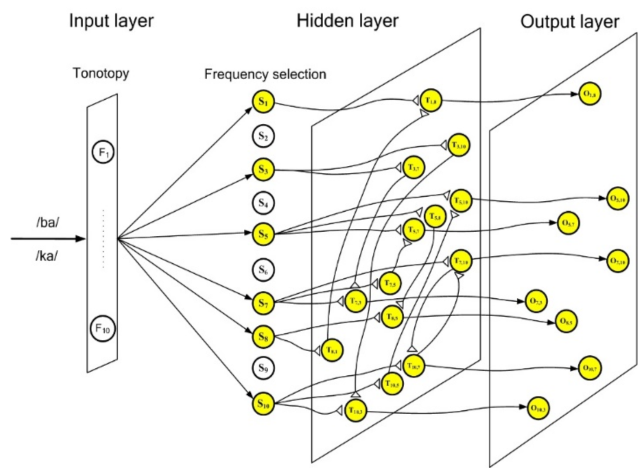


**Figure 2. The firing pattern of neurons in the output layer.**

# 4. CONCLUSION

In this paper, based on the mechanism in human auditory pathway, a neural network with HH neurons and the STDP rule is proposed to simulate the recognition of speech signals. The input encoding method is based on the mechanism in the cochlea that the sound frequency is coded by tonotopy and the sound intensity is coded in the firing rates of neurons. So, the frequency and intensity information of input sound signals are coded by the positions and firing patterns of the neurons in the input layer.

Then the model demonstrates the dynamic roles of HH neurons and the STDP rule in the recognition process. Firstly, HH neurons have the property of frequency selective response that they only respond to their characteristic frequencies in burst spike trains input. This property is consistent with the tonotopy mechanism in human auditory pathways, which makes the recognition of sound intensity based on the dynamic neurons become possible. Secondly, according to the STDP rule, a synaptic connection structure is formed between the neurons and the spike time intervals are stored in the synaptic delay times. Finally, after the frequency and intensity information of the sound signals are coded through the parameters of neurons and synapses, the neurons in the output layer fire corresponding spatiotemporal firing patterns to realize the recognition.

This model explores the possible sound signals recognition mechanism based on the biological dynamic properties of neurons and synapses in SNNs.

## 5. ACKNOWLEDGMENTS

We would like to thank Professor Fukang Fang who provided his knowledge to this work. This work was supported by the National Natural Science Foundation of China (Grant No. 71731002).

## 6. REFERENCES

[1] Maass, W. 1997. Networks of spiking neurons: The third generation of neural network models. *Neural Networks* 10, 9 (Dec. 1997), 1659-1671.

[2] Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., Jackson, B. L., Imam, N., Guo, C., Nakamura, Y., Brezzo, B., Vo, I., Esser, S. K., Appuswamy, R., Taba, B., Amir, A., Flickner, M. D., Risk, W. P., Manohar, R. and Modha, D. S. 2014. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345, 6197 (Aug. 2014), 668-673.

[3] Seung, H. S. 2003. Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron* 40, 6 (Dec. 2003), 1063-1073.

[4] Dong, M., Huang, X. and Xu, B. 2018. Unsupervised speech recognition through spike-timing-dependent plasticity in a convolutional spiking neural network. *Plos One* 13, 11 (Nov. 2018), e204596.

[5] Caporale, N. and Dan, Y. 2008. Spike timing-dependent plasticity: A Hebbian learning rule. *Ann. Rev. Neurosci.* 31 (July 2008), 25-46.

[6] Lee, C., Panda, P., Srinivasan, G. and Roy, K. 2018. Training deep spiking convolutional neural networks with STDP-Based unsupervised pre-training followed by supervised Fine-Tuning. Front Neurosci. 12 (Aug. 2018), 435.

[7] Nager, C., Storck, J. and Deco, G. 2002. Speech recognition with spiking neurons and dynamic synapses: A model motivated by the human auditory pathway. *Neurocomputing* 44-46 (June 2002), 937-942.

[8] Ghosh-Dastidar S. and Adeli H. 2009. Spiking Neural Networks. *Int. J. Neural Syst.* 19, 4 (Aug 2009), 295–308.

[9] Izhikevich, E. M. 2010. *Dynamical systems in neuroscience: The geometry of excitability and bursting.* MIT Press, Cambridge, MA.

[10] Hodgkin A. L. and Huxley A. F. 1952. A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *J. Physiol.* 117, 4 (Aug. 1952), 500-544.

[11] Izhikevich, E. M. 2003. Simple model of spiking neurons. *IEEE Trans. Neural Netw.* 14, 6 (Nov. 2003), 1569-1572.

[12] Nandakumar S. R. , Kulkarni S. R. , Babu A. V. and Rajendran B. 2018. Building Brain-Inspired Computing Systems: Examining the Role of Nanoscale Devices. *IEEE Nanotechnol Mag*. 12, 3 (Aug 2018), 19-35.

[13] Bear M. F., Connors B. W. and Paradiso M. A. 2001 Neuroscience*: Exploring the Brain.* Lippincott Williams & Wilkins Inc., Philadelphia, PA.

[14] Abarbanel, H. and Talathi, S. S. 2006. Neural circuitry for recognizing interspike interval sequences. *Phys. Rev. Lett.* 96, 14 (Apr. 2006), 148104.

[15] Destexhe A., Mainen Z. F. and Sejmowski T. J. 1994. An efficient method for computing synaptic conductances based on a kinetic model of receptor binding. *Neural Comput.* 6 (Apr. 1994), 14-18.

[16] Gerstner W. and Kistler W. M. 2002. *Spiking neuron models: single neurons, populations, plasticity.* Cambridge University Press, New York, NY.