SPIKING NEURON AUDITORY MODEL FOR SPEECH PROCESSING SYSTEMS *

A. V. IVANOV, D. S. LIKHACHOV AND A. A. PETROVSKY

Real-Time Systems Departament at the Bialystok Technical University, Bialystok, Poland; Computer Engineering Department at the Belarusian State University of Informatics and Radioelectronics 6, P.Brovky str. 220027 Minsk, Belarus
E-mail: alexei_v_ivanov@yahoo.com, den2000@tut.by, palex@gw.bsuir.unibel.by

In this paper we present an attempt to develop a novel approach to modeling auditory nerve fibers activity and possibly next auditory processing centers in dorsal and ventral cochlear nuclei based on a simplified model of spiking neuron. Current state of the model already allows talking about good match between real systems and our model at the qualitative level and supports our belief that this model is robust to additive noise.

1. Introduction

Speech processing in noisy and reverberant environments at the moment becomes one of the biggest scientific priorities. A vast variety of approaches are tested and used to solve this task. An illustration of one of such approaches may be found in [1].

Throughout last decade a lot of research effort was dedicated to creation of so-called auditory models, which simulate in considerable detail the core process, which takes place in auditory system, yet are less complicated than strict and detailed biological models. Ensemble Interval Histogram (EIH) technique, presented in [2],[3] may serve as a typical example of this approach.

Representation of the auditory nerve (AN) fibers as an array of level-crossing detectors, which is used in EIH doesn't effectively capture essential properties of the AN.

Indeed, the core controversy between experimental data and model proposed in [2], [3] is reflected by the fact that refractory period of neurons, that form AN (those, having their nuclei contained in cochlear (spiral) ganglion) is of order of magnitude of 1 ms. In other words the maximum frequency of spike generation in axons of AN neurons is only about 1 kHz, while proposed level-crossing model even for 20kHz tone (approximate upper bound of hearing area) will give a sequence of spikes with the frequency of 20 kHz.

Another unsatisfactory property of the model is an inability to capture dependence of the long-time averaged firing rate of individual neurons in AN upon the amplitude of excitation. Such dependence arises only after introduction

* This work is supported in part by the Bialystok Technical University under the grant S/II/2/01.

of multiple levels for level-crossing detectors and summation of the activity at different levels. Thus, as it was noted by Ghitza in [2], [3], individual level-crossing detectors do not represent individual neural fibers, but the array of detectors at different levels represents activity, averaged upon ensemble of neural fibers.

Further, construction of EIH histogram assumes summation of inverted interspike periods across detectors, connected to different filter-bank channels (different Inner Hair Cells). This strategy conflicts with the observation that in reality tonotopical organization of the fibers is maintained up until the highest levels of processing in auditory cortex.

It is also important to note that majority of psychoacoustical dependencies change their behavior around 500 Hz – 1 kHz [4]. In other words, there should exist two principally different ways of sound perception – in the low frequency range (below 1kHz) and high frequency range (above 1kHz). We tend to believe, that this fact is linked to the properties of the AN fibers, in particular, to the refractory period of the AN neurons.

2. Spiking Neuron Auditory Model

The proposed model consists of several modules, each modelling particular processing stage. An outermost Digital Cochlear Model (DCM) represents human cochlear. Elaborate description of the model can be found in [5-7]. Next, a model of Inner Hair Cell (IHC) is employed. IHC is represented by serial connection of half-wave rectifier to the leaky integrator [8]. Furhter on, a Spiking Neuron model is used for simulation of the auditory nerve. Finally layers of Pulse-Coupled Spiking Neurons are used to simulate processes, which take place in cochlear nuclei.

3. Spiking Neuron as a Model of AN fiber

As an alternative to modeling of AN fibers with level-crossing detectors we propose the model of spiking neuron.

This model assumes spike generation as soon as internal activity of the neuron becomes equal to some firing threshold, as opposed to weighted input summation and continuous sigmoid activation function of the classical McCulloch-Pitts neuron. We must note, that response of the classical model represents averaged (over time or ensemble) number of firings of the spiking neuron.

More information about spiking neuron model might be obtained in [9], [10]. In the most general case spiking neuron is described by system of equations (1).

$$F_i = \sum_{\forall j} M_{ji} Y_j(t) \otimes \Phi_{ji}(t) + I_i \ ; \ L_i = \sum_{\forall k} W_{ki} Y_k(t) \otimes \Phi_{ki}(t) + J_k \ ;$$
$$U_i = F_i(1+\beta L_i) \ ; \ Y_i(t) = \begin{cases} 1, if & U_i - \Theta_i > 0 \\ 0, if & U_i - \Theta_i < 0 \end{cases} \ ; \ \Delta\Theta_i = -\alpha_T \Theta(t) + V_T Y_i(t) \ ; \qquad (1)$$

where F_i - the total feeding signal of the neuron; L_i - the total linking signal of the neuron, (linking inputs as a rule have smaller time constant); M_{ji}, W_{ki} - weights, which model synaptic strength; $\Phi_i = e^{-\alpha_{ji} t}$ - response kernel of leaky-integration circuit, which models neuronal input; I_i, J_i - input biases; U_i - total internal activity of the neuron; β - linking depth; Y_i - output signal of the neuron; Θ_i - firing threshold; α_T, V_T - time constant and feedback strength output spike generator.

In our model AN neuron – output impulse generator of the spiking neuron (1), which is driven by IHC membrane potential. While modeling AN fibers conventional spiking model was updated to account for 1 ms refractory period.

AN fiber responses come to dorsal and ventral cochlear nuclei, which in our model are represented as layers of mutually connected spiking neurons, described by full set of equations (1). Their feeding inputs are connected to the outputs of AN fiber models, their output activity is recurrently connected to their linking input. Such model became known as Pulse Coupled Neural Network (PCNN).

During experiments a special attention was paid to the configuration of the PCNN layers in which one of them, with smaller linking depth, sends additional inhibitory feeding input to the other layer. Such configuration is inspired by studies of topology and physiology of the mamalian cochlear nuclei [11],[12].

4. Experiments

Experiments with the model showed qualitative agreement of the AN fiber model attached to the cochlear filter bank to the behavior of the real AN fiber [13], [14]. Dependency of spike count upon stimuus amplitude is sigmoidal and dependency upon stimulus frequency resembles AN fiber tuning curve.

Figure 1 depicts response of the model at the level of AN to a number of test tones. It is very important to observe differences in the responses at low and high frequency. Continuous transition of the curve form from typical low frequency response to the high frequency one can also be seen.

Phase-locked activity of the neurons in the AN fiber model leads to step-like character of the curve at low frequences. Those steps occur at the multiples of the period of the test tone.

Another notable point is that peak response levels form curve which resembles equal-loudness curve of the human ear, this is exclusively a property of the AN fiber model since all cochlear filters in the bank had 0 db attenuation at the central frequency.

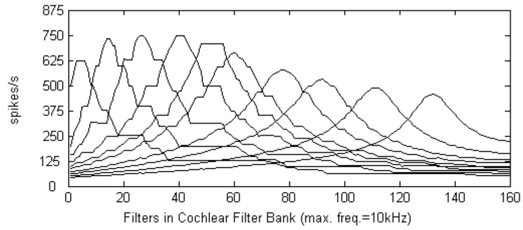
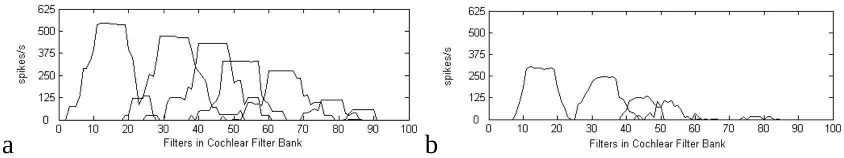


Figure 1. Response of the model at the level of AN to test tones.

Figures 2a and 2b respectively show response of the second of two PCNN layers to the test tones without additional noise and with additive noise at ~6db SNR. It can be seen that "wide-band inhibited" configuration is capable of creation of localised (in frequency) response to the tones and to much lesser extent is sensitive to the noise. Indeed in the presence of noise PCNN generates weaker response to the tone, but this response is not distorted.



a b

Figure 2a & 2b. Response of the second of two PCNN layers to the test tones without additional noise and with additive noise at ~6db SNR respectively.

5. Conclusion

Current state of the model allows us to believe that despite it's simplicity it is capable of capturing behavior of the real auditory nerve fibers activity at the qualitative level.

With the help of two PCNN layers a configuration sensitive to tones and less sensitive to broad-band noise is proposed. This feature is important for speech processing applications but by no means can be considered as a sort of "biologically" grounded model.

Another important property of the described system is the fact that neuron parameters are uniform for all neurons, belonging to a certin layer: audtory nerve layer, first PCNN layer of wide-band inhibitors, second PCNN layer. At this point they were chosen in experimentation. A kind of more justified lerning or adaptation algorithm, which would be capable of creating the system with certain prespecified parameters is needed.

In this paper activity of the spiking neurons was measured as simple time-averaged number of spikes. This procedure is suitable for stationary signals. However, in order to apply proposed model to speech processing of more realistic non-stationary signals a new, more elaborated procedure of the convertion of the outpt neuronal activity to compact representation, suitable for processing is needed.

References

1. J.Baszun, A.A.Petrovsky "Enhancement of speech as a preprocessing for hearing prosthesis by time-varying tunable modulation filters", The proc. 17th international Congress on Acoustics, 2001
2. O.Ghitza "Auditory Nerve Representation as a Basis for Speech Processing" in Advances in Speech Signal Processing, S.Furrui M.M.Sondhi Eds., New York: Marcel-Dekker, 1992
3. O. Ghitza "Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition", IEEE Transactions on Speech and Audio Processing, vol.2, no.1, part II, 1994
4. E.Zwicker, H.Fastl "Psychoacoustics, facts and models" Springer-Verlag,1990
5. A. Petrovsky, "The synthesis of high order digital band-pass filters with tunable center frequency and bandwidth", EUSIPCO 96, pp. 1527–1530
6. W. Wan., A. A. Petrovsky, and C. Fan, "A two-dimensional nonlinear cochlear model for speech processing: response to pure tones", in Proceedings, 6th International Fase -Congress, 1992, pp. 233–236.
7. W. G. Wan, A.A. Petrovsky, C. X. Fan, "A digital cochlear model: transfer function, cochlear map and bandwidth characteristics" (unpublished).
8. F.Rattay, I.C. Gebeshuber "The mammalian auditory hair cell: A simple electric circuit model.", J. Acoust. Soc. Am., 103 (3), March 1998
9. J.L.Johnson, M.L. Padgett "PCNN Models and Applications", IEEE Transactions on Neural Networks, vol.10, no. 3, May, 1999, pp.480-498
10. W. Maas, C.M. Bishop, ed. "Pulsed Neural Networks", MIT Press, 1998
11. M. C. Reed, J.J. Blum "Model calculations of the effects of wide-band inhibitors in the dorsal cochlear nucleus", J. Acoust. Soc. Am. 102(4), 1997
12. M. C. Reed, J.J. Blum "Effects of wide band inhibitors in the dorsal cochlear nucleus. II. Model calculations of the responces to complex sounds", J. Acoust. Soc. Am. 103(4), pp. 2000-2009, 1998
13. J.M. Kates "Two-Tone Suppression in a Cochlear Model", IEEE Transactions on Speech and Audio Processing, vol. 3, no. 5, 1995
14. J.Lasarro, C.Mead "Cirquit models of Sensory Transduction in the Cochlea", http://citeseer.nj.nec.com/cs