Computational Models of Auditory Function
S. Greenberg and M. Slaney (eds.)
IOS Press, 2001

A COMPUTATIONAL MODEL OF AUDITORY SOUND LOCALIZATION

Kazuhito Ito and Masato Akagi

School of Information Science
Japan Advanced Institute of Science and Technology (JAIST)
1-1 Asahidai, Tatsunokuchi, Ishikawa, 923-12 Japan

1. Introduction

Sound localization based on the interaural time difference (ITD) detects sound source locations using the difference in arrival times of the sound waves at the two ears [11] [12]. To understand the process and represent it computationally, we developed a computational model of auditory sound localization based on the ITD.

Sound waves arriving at the ears are decomposed into their frequency components and are changed into impulse trains by the auditory periphery. The impulse trains accurately represent the time intervals between firings because auditory nerve firings tend to be phase-locked or synchronized to the stimulating waveforms (Figure 1) [11][13]. The difference between the temporal information from the two ears is used for sound localization [2]. It is known that humans can perceive an ITD variation of about 10 μs at 900 Hz, corresponding to a minimum audible angle (MAA) of about 1° [11][10].

A nerve impulse is an electrical excitation called an action potential and its duration is over 1 ms. Additional synaptic transmissions in the auditory system extend the duration of the signal [1]. Thus the duration of an auditory firing is long compared with that of the ITD perceived by humans. Such a long-duration signal is problematic for the minute temporal information that should be transmitted and it may obscure temporal information. Although impulses from the auditory nerves are in synchronization with a particular phase of the stimuli, it is known that impulses fluctuate slightly in time [6]. Again, this may obscure temporal information. Given all these conditions, it is amazing that humans can perceive an ITD variation of about 10 μs.

In this study, the signals in the nervous system such as action potentials and synaptic transmission, were modeled computationally and these models were used to detect ITDs. Impulse trains, with fluctuation in time that simulate spikes in the auditory nerve fibers, were

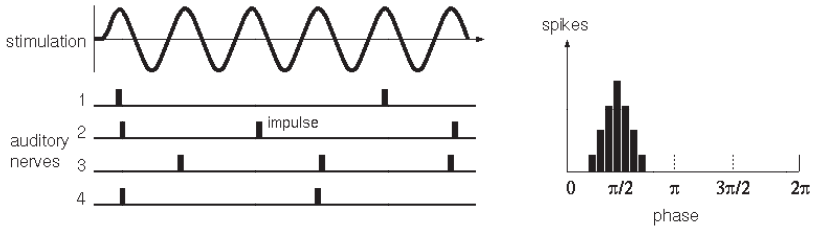

Figure 1 Temporal information and impulse fluctuation.

used as input data to the model. Then processes of sound localization with temporal redundancy and impulse fluctuation were studied through this model.

2. Representation of Signals

To represent the action potential computationally, we use Hodgkin–Huxley type equations [3] [14] [9].

$$C_m \frac{dV(t)}{dt} = -g_{Na}(V(t) - E_{Na}) - g_K(V(t) - E_K) - g_L(V(t) - E_m) \tag{1}$$

Here, $V(t)$ is the membrane potential at time t, C_m is the membrane capacitance, and E_m is the resting potential; E_{Na} and E_{K} are equilibrium potentials for sodium and potassium, respectively; and g_{Na} and g_{K} are the conductances of sodium and potassium. The ion conductances are

$$g_n(t) = a_n(t - t_n)e^{-(t-t_n)/\tau_n} \tag{2}$$

where $g_n(t)$ indicates the conductance for ion n at time t, t_n is the time of the most recent onset of ion n conductance, τ_n is the time constant for the conductance, and a_n is the amplitude constant related to the permeability of ion n. Then, the leakage conductance g_L is represented by

$$g_L(t) = a_L 1 - e^{-\alpha|V(t) - E_m|} \tag{3}$$

where α is the coefficient for the relationship between the leakage conductance and the membrane potential and a_L *is* the maximum conductance.

Synaptic transmission is also represented by the same type of equations. Although applying these equations might not be accurate in this case [1], this is one way to model the temporal redundancy of signals. The equations that describe the behavior of synaptic potentials, (4) and (5), do not model the firing of the post-synaptic cell. These firing thresholds aredefined in equation (6)

$$C_m \frac{dV(t)}{dt} = -G_{Na}(V(t) - E_{Na}) - G_K(V(t) - E_K) - g_L(V(t) - E_m) \tag{4}$$

$$G_n(t) = A_n(t - t_n)e^{-(t-t_n)/T_n} \tag{5}$$

where $V(t)$ is the postsynaptic potential at time t; C_m is the membrane capacitance; E_m is the resting potential; E_{Na} and E_{K} are the equilibrium potentials for sodium and potassium; and G_{Na}, G_{K}, and g_L are the postsynaptic conductances for sodium, potassium, and the leakage, respectively. Again $G_n(t)$ indicates the postsynaptic conductance of ion n at time t. t_n is the time of the most recent onset of ion n conductance. τ_n is the time constant for the postsynaptic conductance, and A_n is the amplitude constant related to the permeability of ion n.

In this model — to represent the effects of other conductances such as early potassium channels, voltage-gated calcium channels, and calcium-activated potassium channels — the firing threshold level is varied according to Eq. (6), which translates the magnitude of the grand postsynaptic potentials into the frequency of firing of action potentials [1].

$$V_{threshold}(t) = \beta \cdot e^{-(t-t_r)/(\tau_r + \tau_a)} + E_{threshold} \tag{6}$$

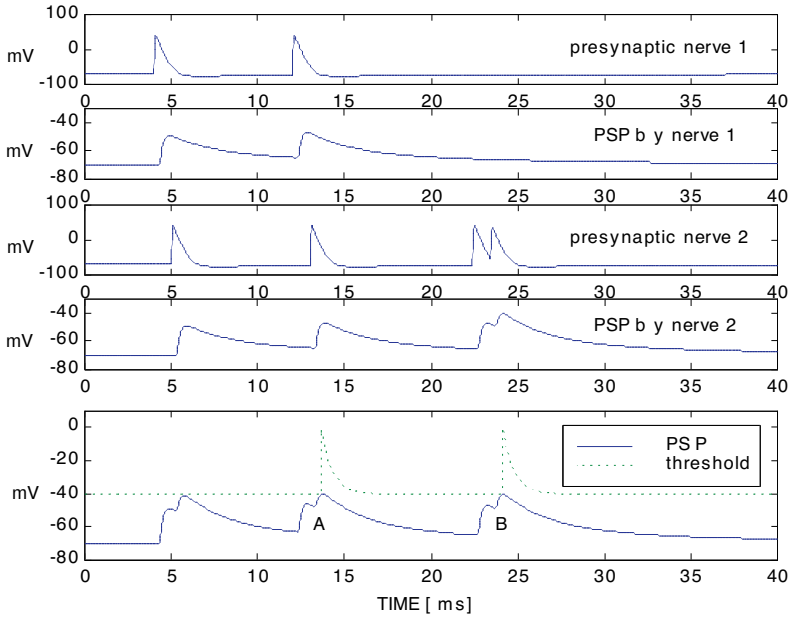

Figure 2 Spatial summation (A) and temporal summation (B)

where $V_{\text{threshold}}(t)$ indicates the threshold level at time t and t_r is the time of the most recent discharge. $E_{\text{threshold}}$ is the basis of threshold level in this function, τ_r is the time constant representing the relative refractory period, and β is the amplitude constant. And to express adaptation to a prolonged stimulation, τ_a is used

$$\tau_a = \tau_{max}\left(1 - \exp\left[-\gamma \int_{t-T}^{t} (V_{threshold}(t) - E_{threshold})dt\right]\right) \tag{7}$$

Adaptation related to the frequency of firing during a certain period of T (hundreds of ms) is represented by extending the length of the relative refractory period τ_a of $V(t)$, and γ is the coefficient for the relationship between the permeability of potassium and the frequency of firing. Hence, if

$$V(t) \geq V_{threshold}(t) \text{ for t-t}_r > \text{absolute refractory period} \tag{8}$$

then the postsynaptic cell fires.

Simulations of these models are illustrated in Figure 2. Presynaptic action potentials are shown in the top panel labeled 'nerve 1' and each action potential arriving at a synapse produces a postsynaptic potential (PSP) on the postsynaptic membrane of a cell, as shown in the second panel down from the top. The other pair are shown in the third and fourth panels, labeled 'nerve.' When both presynaptic nerves innervate the same cell, PSPs produced by both nerves are summed to produce a larger PSP. While spatial summation combines the effects of signals received at different sites on the membrane (at A in the bottom panel), temporal summation combines the effects of signals received at different times (at B in the bottom panel). The firing threshold level is illustrated by the dotted line in the bottom panel.

Reprinted from Computational Models of Auditory Function, Volume 312 NATO Science Series: Life Sciences (Edited by: S. Greenberg and M. Slaney), Kazuhito Ito and Masato Akagi, A computational model of auditory sound localization, 97-111, Copyright 2007, with permission from IOS Press

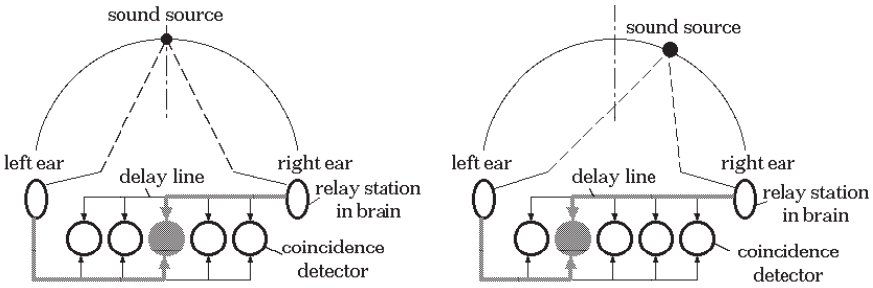

Figure 3 A coincidence detector circuit.

When the summed PSP exceeds a given threshold level, an action potential is generated on the postsynaptic cell. Then, these representations of the signals in the nervous system are used to detect ITDs.

3. Model Circuits For Detecting ITDs

3.1 Cross Correlation

The system for detecting ITDs exists in the medial superior olive (MSO), where the neural pathways from the left and right ears cross for the first time [17] [7]. The Jeffress model is well known as a model for this detection [5]. It is represented as a circuit consisting of an array of coincidence detectors and two nerve fibers from the left and right ears. The coincidence detectors fire most often when impulses from both sides arrive simultaneously. The model calculates ITDs with cross correlation between impulse trains coming from both sides (Figure 3). When a sound source is placed in front of the face, the arrival times from the left and right pathways are the same, because the lengths of time taken by the sound wave to get to the ears and by the impulse trains to get to the circuit are equal. Thus, the center detector in the circuit responds most strongly. The position of the responding detector varies as a sound source moves.

Because auditory nerve firings tend to be phase-locked to the stimulating waveforms, impulse trains contain temporal information. A computational cross-correlation model like the Jeffress model works well to detect ITDs using such impulse trains that synchronize with a particular phase of stimuli. Cross correlation is represented by

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y(t+\tau)dt \tag{9}$$

In this equation, x and y are impulse trains from the right and left ears and τ is the time difference between the two signals. The cross-correlation model outputs the τ which gives the maximum response. The cross correlation is usually implemented at discrete temporal intervals. If the intervals between adjacent coincidence detectors become smaller, the model will detect ITDs more accurately. Hence, the cross correlation model is a basic system for detecting ITDs [18] [16].

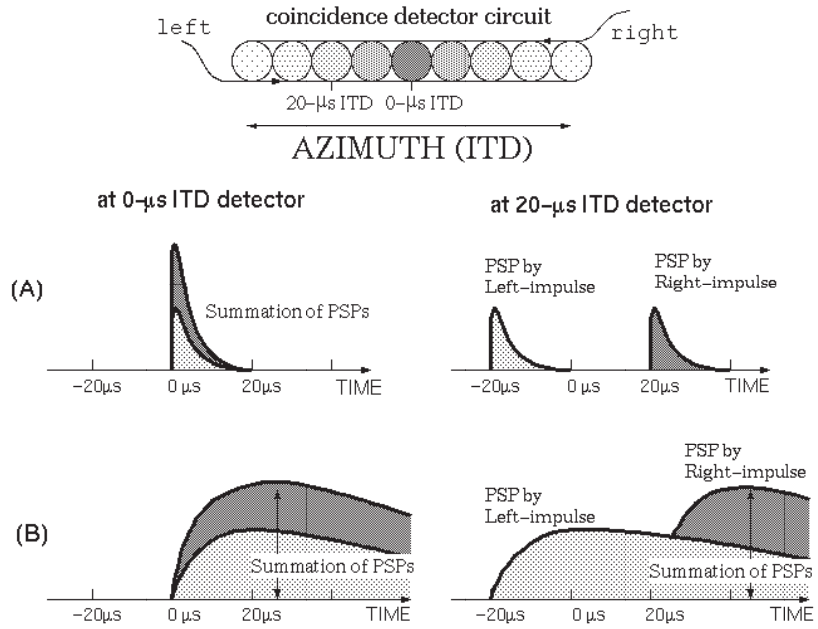

Figure 4 Temporal redundancy for short (A) and normal (B) PSPs.

3.2 Temporal Redundancy

In this study, features of the nervous system were used and the models of nerve impulses and synaptic transmission were applied to a coincidence detector circuit. The duration of the nerve impulse is over 1 millisecond and that of synaptic transmission ranges from several milliseconds to hundreds of milliseconds.

Figure 4 shows two types of temporal transition of PSPs on two coincidence detectors, corresponding to the detection of 0-μs and 20-μs ITDs. On the left side of Figure 4A, an impulse from each side is applied to the circuit without any time difference. For a human to perceive an ITD variation of about 10 μs as the MAA, it is best if the duration of PSP is shorter than 10 μs (Figure 4A). Since impulses arrive at the 0-μs ITD detector at the same time, two PSPs combine together and make a large potential. On the right side of Figure 4B, the arrival times of impulses at the 20-μs ITD detector do not match, PSPs decline without affecting each other. Thus, it is easy to distinguish the difference between the two detectors and determine ITDs.

However, the duration of PSP by synaptic transmission is several milliseconds or more (Figure 4B). At the 0-μs ITD detector, two long PSPs combine and give a large potential. Likewise, at the 20-μs ITD detector, two long PSPs combine and give a large potential, even though the arrival times of the two impulses do not match. This is because the temporal interval between them is much smaller than the duration of PSPs. Thus, the duration of PSP gives an ambiguous result that obscures minute information such as the temporal interval between the arrival times of impulses.

Figure 5 shows the temporal transition of postsynaptic potentials on the coincidence detector cells arranged along the azimuth (the axis of ITDs). An impulse from each side is

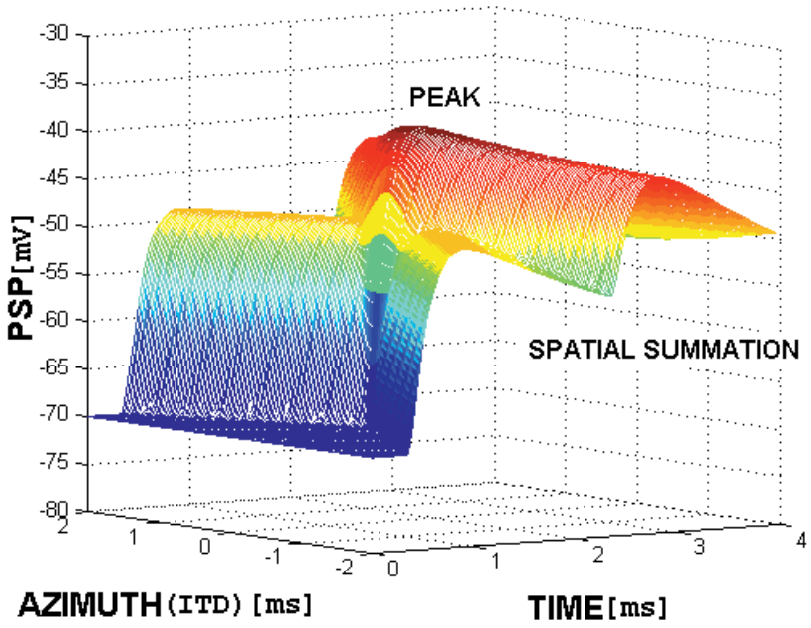

Figure 5 Postsynaptic potentials in a coincidence detector circuit. This graph indicates the temporal transition of PSPs in the model after an impulse is applied to each input of the circuit.

applied to the circuit without any time difference. The impulse from each ear stimulates coincidence detectors sequentially from each side of the circuit and a small postsynaptic potential is generated at every detector. When impulses from both sides meet at the middle detector, the PSPs are summed and a large potential is generated. Then the impulses separate and keep stimulating other detectors on the other side and summations of PSPs are generated on both sides. The envelope with the maximum potential on every detector draws a peak on the axis of ITDs. The peak looks very broad but should indicate the ITD.

Although it is not certain that threshold levels on all detectors in the MSO are the same, we assume that they are the same in this model. When the threshold level is set to the same level as the peak of the potential envelope, its simulation is equivalent to calculating the cross correlation because just one detector fires in this case (Figure 6).

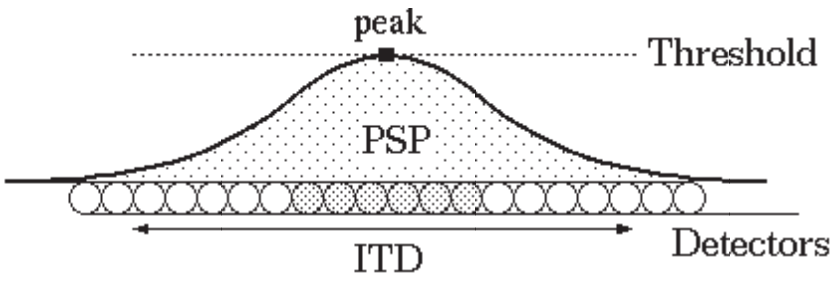

Figure 6 Threshold level at the peak of the potential envelope.

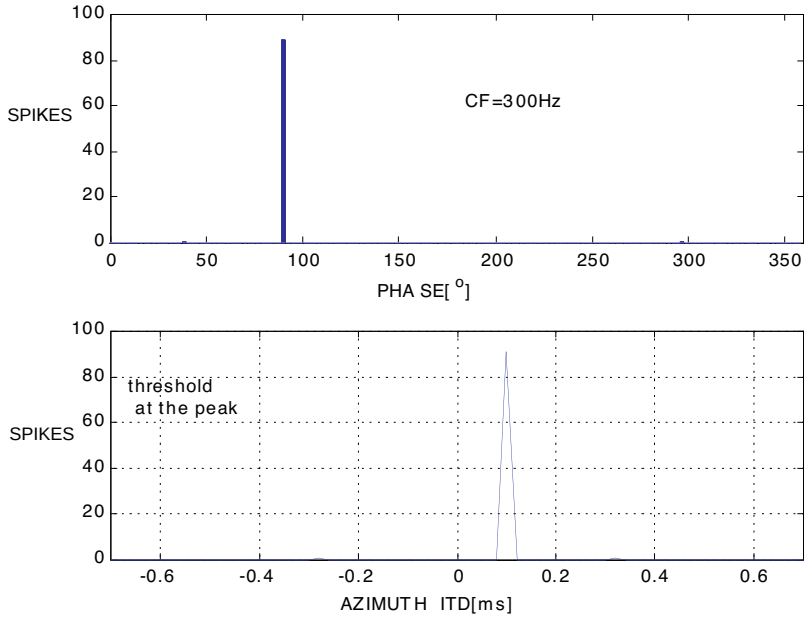

Figure 7 Period histogram of the impulse train firing in synchronization with a certain phase of stimuli and the spike histogram obtained by the simulation indicates the ITD (=100 μs) in azimuth.

Impulse trains firing in synchronization with a particular phase of a stimuli with a frequency of 300 Hz and a 0.3-s duration with an interaural time difference of 100 μs are provided as input to the model. The upper panel in Figure 7 shows the period histogram of the impulse train and the lower panel shows the spike histogram obtained by this simulation. The spikes are concentrated at an ITD of 100 μs in azimuth.

3.3 Nonlinear Output Mechanism

However, it is difficult to set the threshold level precisely at the level of the peak of the potential envelope. It is natural to set it to a level below the peak. In that case, all the detectors whose potential exceeds the threshold level fire and a broad range of firings appear along the azimuth (Figure 8). Accordingly, our model includes a nonlinear output mechanism.

Figure 9 shows the result of a simulation of the nonlinear output mechanism using the same impulse trains as in Figure 7. The upper panel in Figure 9 shows the period histogram of the impulse train and the lower panel shows the spike histogram obtained by this simulation. The nonlinear output mechanism outputs spikes over a broad range along the axis of ITD and the envelope of the spike histogram looks so square that it is difficult to determine the ITD. Although this output mechanism seems inappropriate for detecting ITDs, the output can be improved by using the variability of impulses on auditory nerve fibers.

3.4 Impulse Fluctuation

Auditory nerve fibers do not always fire in synchronization with the same phase of the stimuli; impulse trains from the auditory nerves fluctuate slightly in time (Figure 1) [6]. Since our model uses impulse trains from the auditory peripheral model that fluctuate in

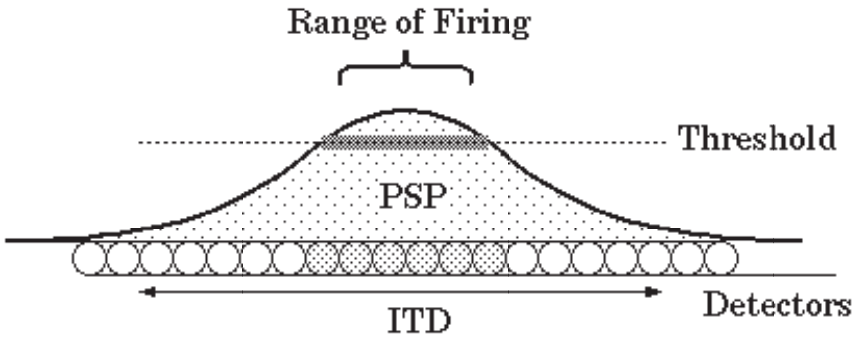

Figure 8 Threshold level below the peak of the potential envelope.

time [9], the model will be affected by the impulse fluctuation. For ITD detection by cross correlation, or the threshold level at the peak of the potential envelope in a coincidence detector circuit in particular, the fluctuation act like noise. Impulse trains having a characteristic frequency of 300 Hz with a large fluctuation in time. duration of 0.3s and with a time difference of 100 μs, are provided as input to the cross correlation model. The upper panel in Figure 10 shows the period histogram of one of those impulse trains with a large fluctuation and the lower panel shows the spike histogram obtained by this simulation. The spike histogram has some peaks but they do not indicate the ITD. Thus, it is difficult to determine the ITD.

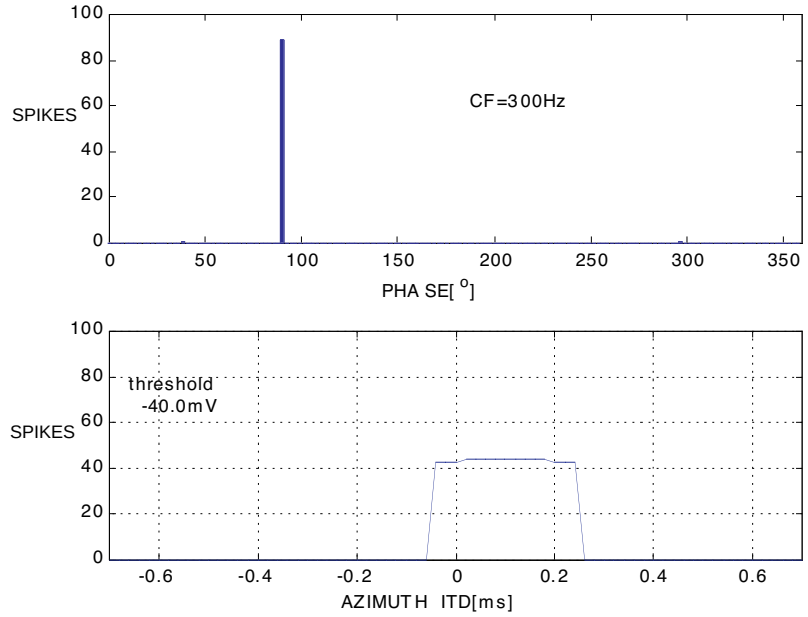

Figure 9 Period histogram of the impulse train firing in synchronization with a certain phase of stimuli and the spike histogram obtained by the nonlinear output mechanism (ITD =100 μs). The envelope of the spike histogram looks so square that it is difficult to determine the ITD

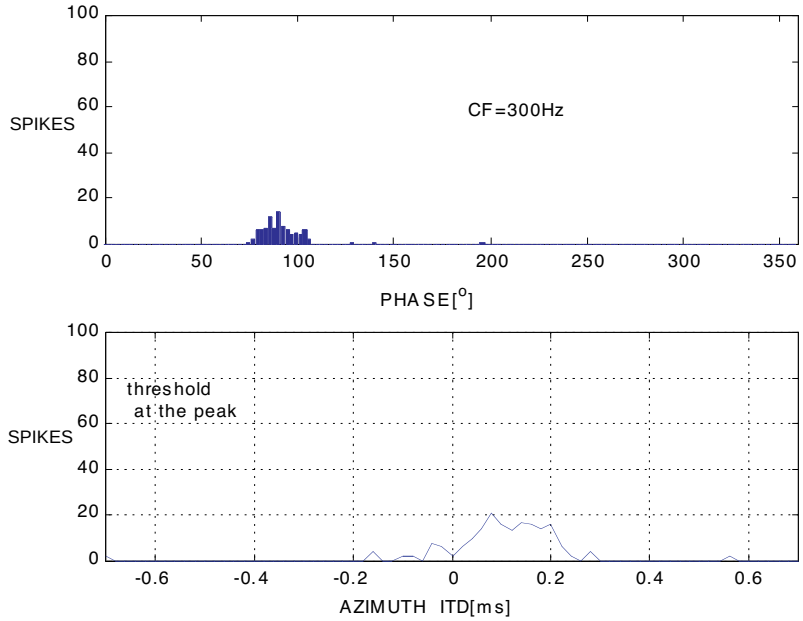

Figure 10 Period histogram of the impulse train with a large fluctuation in time and the spike histogram obtained by cross correlation (ITD =100 μs). The envelope of the spike histogram has several peaks and it is difficult to determine the ITD.

Therefore, since we have impulse trains with a large fluctuation in time as input to our model, we set the firing threshold level below the peak of the potential envelope. Impulse fluctuations affect the location where the peak of the potential envelope appears on a coincidence detector circuit because the detector where impulses from both sides of the circuit encounter each other changes based on the noise. The location of the peak of the potential envelope varies and the firing range also shifts along the axis of ITD (Figure 11A). If the threshold level is set to an appropriate level, the firing ranges will often overlap each other in spite of the variation in peak location, and the detectors in the overlapping area will keep firing. Since impulse fluctuation of an auditory nerve fiber has a normal distribution, it is likely that the actual ITD is included in the response curve.

If a spike histogram is drawn according to the variation in the firing range, it is found that the number of spikes in the overlapping area is greater than in other areas (Figure11B). This means that we use impulse trains fluctuating in time as input data improves the output of the

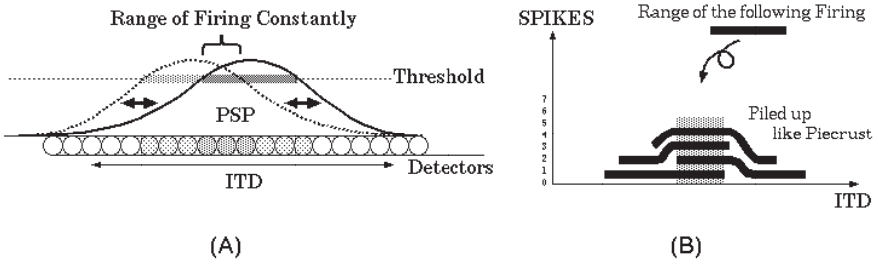

Figure 11 Nonlinear output mechanism.

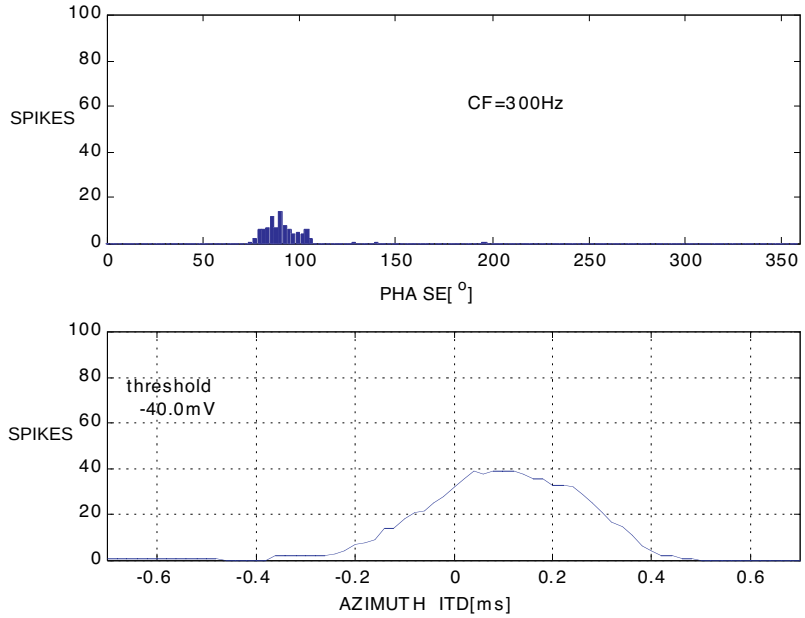

Figure 12 Period histogram of the impulse train with a large fluctuation in time and the spike histogram obtained by the nonlinear output mechanism (ITD =100 μs). The envelope of the spike histogram on the ITD rises.

model compared with using ones with no fluctuation, and the envelopes of spike histograms output from the nonlinear output mechanism tend to have a peak that indicates the ITD or its vicinity. Figure 12 shows the result of a simulation by the nonlinear output mechanism using the same impulse trains as in Figure 10. The upper panel in Figure 12 shows the period histogram of the impulse train with a large fluctuation and the lower panel shows the spike histogram obtained by this simulation. The envelope of the spike histogram as a function of the ITD is rising and beginning to form a peak.

4. Improving the Accuracy

4.1 Emphasizing

It is still difficult to determine the ITD using only one coincidence detector circuit. Actually, there are hundreds of detector circuits for each frequency in the MSO and outputs from these circuits are integrated [2] [17]. Therefore, the model can emphasize the peak indicating the ITD and achieve higher accuracy by having many circuits with different thresholds for each frequency and by integrating all the outputs. To avoid a large increase in the amount of calculation, we used multi-threshold paths for every coincidence detector instead of many circuits. Then, the outputs from all the paths were integrated. This method can emphasize the peaks of the potential to detect the ITD (Figure 13).

We used an inhibition-like model to determine the ITD more effectively. Since the coincidence detector indicating the ITD tends to fire earlier than others in the circuit (Figure 14), the inhibition suppresses the succeeding firings and thus emphasizes the initial firing. Although there are reports about inhibition in the MSO [15] [3], this inhibition-like model is

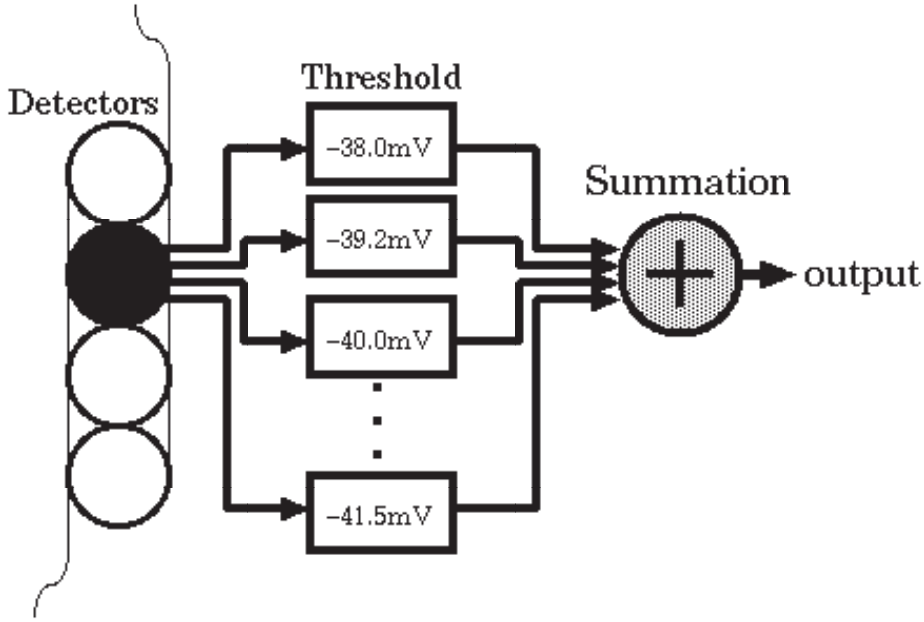

Figure 13 Multi-threshold model.

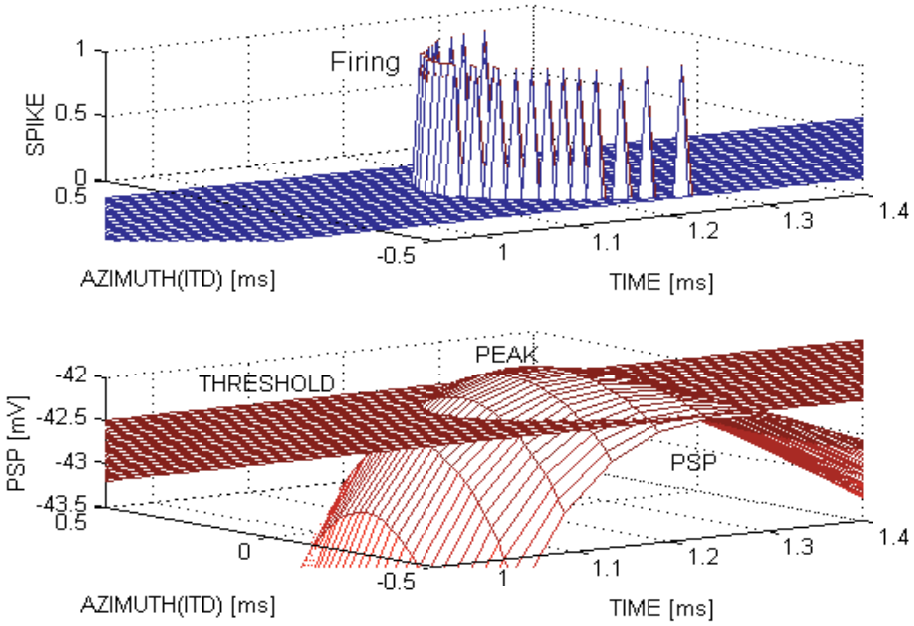

Figure 14 The detector indicating the ITD tends to fire earlier than others in the circuit.

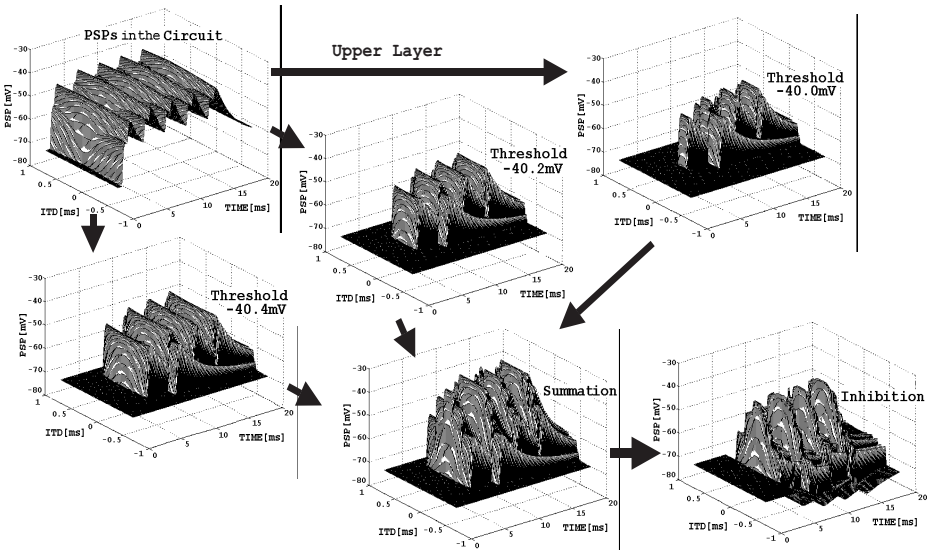

Figure 15 Outputs from the coincidence detector circuit to a upper layer. The potentials are summed by the multi-threshold model and emphasized by the inhibition-like model.

not based on physiological data. Note that this is another way to emphasize the peak and suggests the possibility of a multi-threshold mechanism. Thus, our computational localization model is constructed using models of action potentials and synaptic transmission, the multi-threshold model, and the inhibition-like model. The sound localization model outputs spikes. that indicate the best ITD.

4.2 Simulation Results

Figure 15 shows an example of outputs from the coincidence detector circuit with the multi-threshold and inhibition models. Impulse trains with a small fluctuation — having a characteristic frequency of 300 Hz and about 10-ms duration, or five impulses — were used as input data and the model output the postsynaptic potentials instead of spikes to show the effects of the emphasis. Consequently, the potentials in the detectors indicating the ITD and its vicinity were emphasized by the multi-threshold model and the inhibition-like model.

Next, we examined the potential of the auditory sound localization model to achieve greater accuracy at detecting ITDs. Impulse trains having characteristic frequency of 300 Hz, 0.3-s duration and with a time difference of 100 μs were used as input to the model. Each simulation used three types of impulse trains, which fired in synchronization with a fixed phase of the stimuli, with small and large fluctuations in time. The impulse train with large fluctuation mimics the phase locking of actual auditory nerves [9]. The results of the simulations were as follows; The upper panel in Figure 16 shows the period histogram of the impulse train firing in synchronization with a fixed phase of stimuli. The lower panel shows the spike histogram that result from this simulation. The envelope of the spike histogram has no peak and it is difficult to determine the ITD.

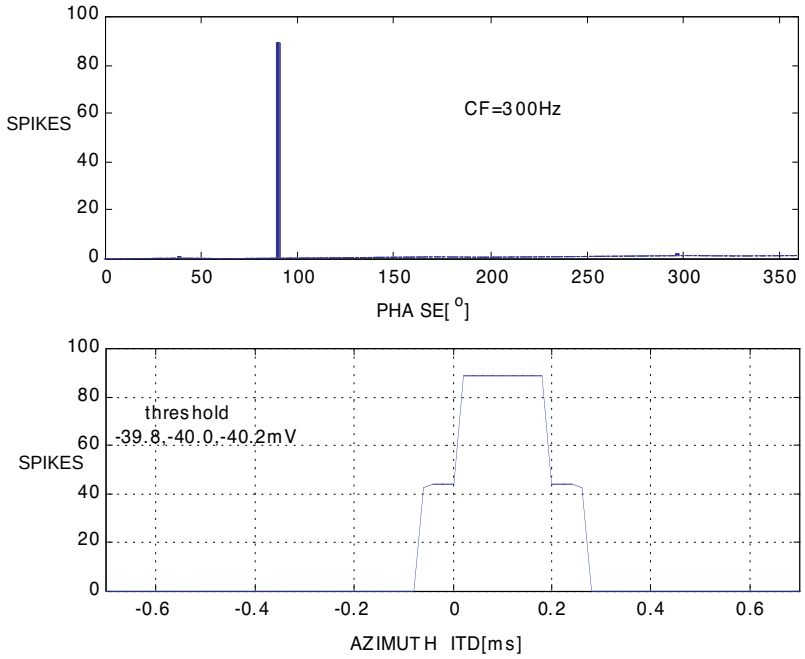

Figure 16 Period histogram of the impulse train firing in synchronization with a certain phase of stimuli and the spike histogram obtained by the simulation (ITD =100 μs). The envelope of the spike histogram has no peak and it is difficult to determine the ITD.

Impulse trains do not always keep firing in synchronization with a fixed phase of the stimuli. The upper panel in Figure 17 shows the period histogram of the impulse train with a small fluctuation in firing time. And the lower panel shows the spike histogram obtained by this simulation. The spikes are distributed around ITD of 100 μs in azimuth. However, the peak indicates the correct ITD.

The upper panel in Figure 18 shows the period histogram of the impulse train with a large fluctuation in firing time. The lower panel shows the spike histogram obtained by this simulation. Even though the spikes are distributed around ITD of 100 μs in azimuth and the envelope is smooth, the peak indicates the correct ITD.

5. Conclusions

A computational model of the auditory sound localization based on the interaural time difference was presented. Nerve impulses and synaptic transmission in the nervous system were modeled computationally and these models were applied to a coincidence detector circuit model to detect ITDs. Impulse trains with fluctuation in time were used as input data and the effects of the impulse fluctuation on the detection of ITDs were investigated.

The simulation results show that the peak indicating the ITD in azimuth obviously sharpens when impulses fluctuating in time are used as input. Using such impulse trains as input data improves the output of the model compared with using ones having no fluctuation. This suggests that impulse fluctuation can contribute to the detection of ITDs in the temporally redundant process and the nonlinear output mechanism.

Reprinted from Computational Models of Auditory Function, Volume 312 NATO Science Series: Life Sciences (Edited by: S. Greenberg and M. Slaney), Kazuhito Ito and Masato Akagi, A computational model of auditory sound localization, 97-111, Copyright 2007, with permission from IOS Press

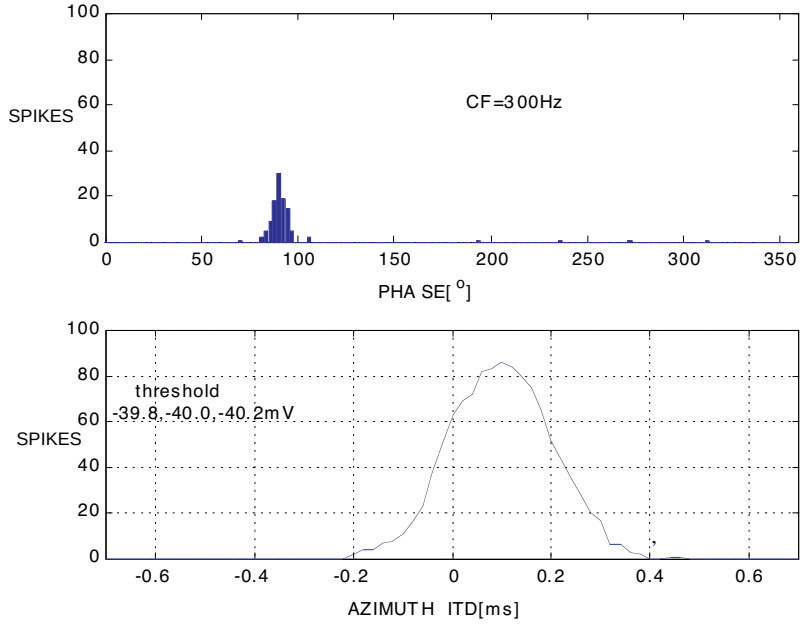

Figure 17 Period histogram of the impulse train with a small fluctuation in time and the spike histogram obtained by the simulation (ITD =100 μs). The peak of the envelope of the spike histogram indicates the ITD.

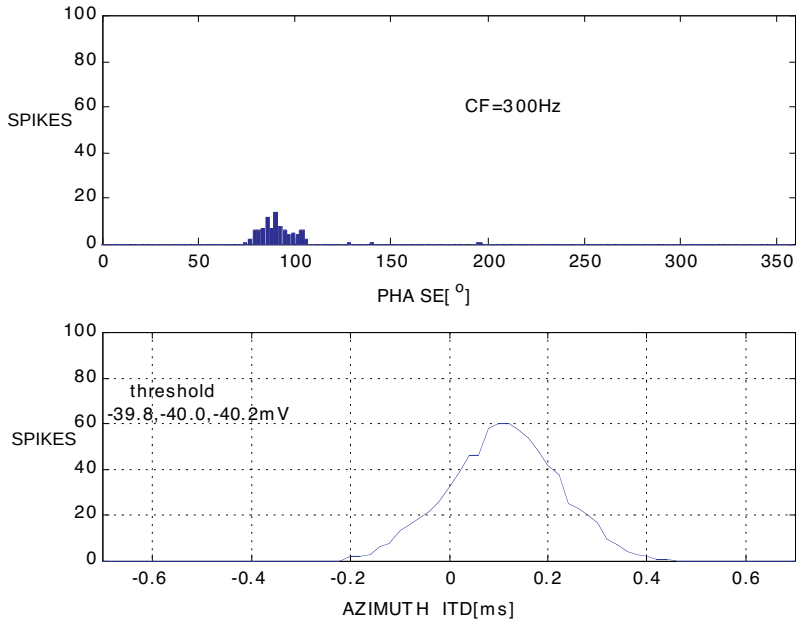

Figure 18 Period histogram of the impulse train with a large fluctuation in time and the spike histogram obtained by the simulation (ITD =100 μs). The peak of the envelope of the spike histogram indicates the ITD.

Acknowledgments

This work was supported by CREST and by Grant-in-Aid for Science Research from Ministry of Education (No.10680374).

References

[1] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. *Molecular Biology of the Cell (3rd ed.)*. New York: Garland, 1994.

[2] Carr, C. E. and Konishi, M. "A circuit for detection of interaural time differences in the brain stem of the barn owl." *J. Neurosci.*, 10: 3227–3246, 1990.

[3] Funabiki, K., Koyano, K., and Ohmori, H. "The role of GABAergic inputs for coincidence detection in the neurons of nucleus laminaris of the chick." *J. Physiol.* 508: 851–869, 1998.

[4] Hodgkin, A. L. and Huxley, A. F. "A quantitative description of membrane current and its application to conduction and excitation in nerve." *J. Physiol.,* 117: 500–544, 1952.

[5] Jeffress, L. A. "A place theory of sound localization." *J. Comp. Physiol. Pychol.*, 45: 35–49, 1948.

[6] Johnson, D. H. "The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones." *J. Acoust. Soc. Am.* 68, pp. 1115–1122, 1980.

[7] Kuwada, S., Batra, R., and Fitzpatrick, D. C. "Neural processing of binaural temporal cues." In *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson (eds.), Hillsdale, NJ: Lawrence Erlbaum, pp. 399–425, 1997.

[8] Konishi, M. "Listening with two ears." *Sci. Am.*, pp. 34–41, April 1993.

[9] Maki, K. and Akagi, M. "A functional model of the auditory peripheral system." *Proc. ASVA97*, Tokyo, pp. 703–710, 1997.

[10] Mills, A. W. "On the minimum audible angle." *J. Acoust. Soc. Am.*, 30: 237–246, 1958.

[11] Moore, B. C. J. *An Introduction to the Psychology of Hearing*. London: Academic Press, 1997.

[12] Palmer, A. R. "Neural signal processing," in *Hearing,* B. C. J. Moore (ed.), San Diego: Academic Press, pp. 75–122, 1995.

[13] Pickles, J. O. *An Introduction to the Physiology of Hearing*. London: Academic Press, 1998.

[14] Rothman, J. S., Young, E. D., and Manis, P. B. "Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: Implications of a computational model." *J. Neurophysiol.* 70: 2562–2583, 1993.

[15] Smith, P. H. "Structural and functional differences distinguish principal from nonprincipal cells in the guniea pig mso slice." *J. Neurophysiol.* 73: 1653–1667, 1995.

[16] Stern, R. M. and Trahiotis, C. "Models of binaural interaction." In *Hearing,* B. C. J. Moore (ed.), London: Academic Press, pp. 347–386, 1995.

[17] Takahashi, T. T. and Konishi, M. "Projections of the cochlear nuclei and nucleus laminaris to the inferior colliculus of the barn owl." *J. Comp. Neurol.,* 274: 190–211, 1988.

[18] Yin, T. C. T., Joris, P. X., Smith, P. H., and Chan, J. C. K. "Neuronal processing for coding interaural disparities." In *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson (eds.), Hillsdale, NJ: Lawrence Erlbaum, pp. 427–445, 1997.