A Biologically Inspired Spiking Neural Network for Sound Localisation by the Inferior Colliculus

Jindong Liu, Harry Erwin, Stefan Wermter, and Mahmoud Elsaid

University of Sunderland, Sunderland, SR6 0DD, United Kingdom,
jindong.liu@sunderland.ac.uk,
WWW home page: http://www.his.sunderland.ac.uk

Abstract. We introduce a biologically inspired azimuthal sound localisation system, which simulates the functional organisation of the human auditory midbrain up to the inferior colliculus (IC). Supported by recent neurophysiological studies on the role of the IC and superior olivary complex (SOC) in sound processing, our system models two ascending pathways of the auditory midbrain: the ITD (Interaural Time Difference) pathway and ILD (Interaural Level Difference) pathway. In our approach to modelling the ITD pathway, we take into account Yin's finding that only a single delay line exists in the ITD processing from cochlea to SOC for the ipsilateral ear while multiple delay lines exists for the contralateral ear. The ILD pathway is modelled without varied delay lines because of neurophysiological evidence that indicates the delays along that pathway are minimal and constant. Level-locking auditory neurons are introduced for the ILD pathway network to encode sound amplitude into spike sequence, that are similar to the phase-locking auditory neurons which encode time information to the ITD pathway. A leaky integrate-and-fire spiking neural model is adapted to simulate the neurons in the SOC that process ITD and ILD. Experimental results show that our model performs sound localisation that approaches biological performance. Our approach brings not only new insight into the brain mechanism of the auditory system, but also demonstrates a practical application of sound localisation for mobile robots.

1 Introduction

The known performance of animals using two ears in the sound localisation task has inspired researchers to work on new computational auditory models to understand the mechanisms of auditory neural networks. During the last twenty-five years, the structure and function of a number of pathways in the auditory brainstem have been well studied and become better understood [1]. For example, multiple spectral representations [2] are known to exist both in the early stages of sound processing, in cochlea and cochlear nucleus, and in later stage, from SOC to IC. Excitatory and inhibitory connections in the neural network of the midbrain sound processing pathways have been clarified [3]. Modelling these networks can help us to understand the brain mechanisms and provide a robust approach of sound understanding to mobile robots.

Binaural sound localisation systems take advantage of two important cues [4] of the arriving sound signals in two ears: (i) interaural time difference (ITD) or interaural phase difference (IPD), and (ii) interaural level difference (ILD). Using these two cues, the sound source position can be estimated in the horizontal or azimuthal plane. ILD is believed to be more efficient at localising low- and mid- frequency sounds (50 Hz $\sim$3 kHz) while ILD is better for mid- and high-frequency sound (>1.2 kHz) [4].

Models of ITD and ILD processing have been developed by several researchers [4][5]. Jeffress [4] originally proposed a widely used model to detect ITDs, in which sound from each ear passes different delay lines before reaching a coincidence neuron which fires with a maximum rate when two specific delay times are present for the sound. Yin [5] improved Jeffress's model in response to biological evidence by introducing a single delay line for the ipsilateral ear while retaining multiple delay lines for the contralateral ear. For ILD, Jeffress suggested a so-called "latency hypothesis" to explain the processing mechanism, in which the latency of inhibitory input is delayed relative to the excitatory input from the ipsilateral ear. Evidence for this idea was provided by Hirsch [6], however, the mechanism of ILD processing remains unclear.

This paper presents a new auditory processing system designed to provide live sound source positions via a spiking neural network (SNN). In this SNN, we implement Yin's ITD model and additionally propose an ILD model. Then ITD and ILD are combined together to simulate the function of IC in the human. It is the first example of applying both ITD and ILD into a spiking neural network to cover a wide frequency range of sound inputs. The synaptic and soma models in the system are treated as independent of sound frequency, ITD or ILD as there is no current evidence for specialisation. This simplifies the system and is a potential advantage for future adaptive auditory systems.

The rest of this paper is organised as follows. Section 2 presents the neurophysiological data of human auditory pathway. In that section, we make an assumption of level locking auditory neurons. Section 3 proposes an pyramid model to calculate ILDs. Section 4 proposes a system model with combines the ITD pathway and the ILD pathway. In Section 5, experimental results are presented to show the feasibility and performance of the sound localisation system. Finally, conclusions and future work are given in Section 6.

2 Biological Fundamentals and Assumptions

When sound arrives at the external ear, it is directed through the ear drum and then propagates to the inner ear, i.e. cochlea. The inner hair cells along the cochlea respond to the sound pressure to generate spikes that are sent through the auditory nerve (AN) up to the central nervous system. The temporal and amplitude information of the sound wave are encoded by the hair cells up to auditory nerve [4]. Two properties of the hair cells are important for this encoding. First, the cochlea is tonotopically organised so that each inner hair cell is maximally excited by stimulation at a characteristic frequency (CF) [7]. In

other words, each hair cell has a specific frequency with a highest sensitivity. Second, the hair cells are polarised so that their movement is excited during one specific phase of the sinusoidal sound wave while inhibited during other phases. This phase locking occurring at frequency of 50 ∼1.2 kHz is the basis for the encoding of timing information for sound.

As the sound pressure level (SPL) increases, auditory nerve fiber increase their rate of discharge with a sigmoidal relationship to the decibel of sound over a relative range of 30 dB. In order to cover a wide SPL range, e.g. 120 dB, the relative range is adaptively changed according to the background sound level. However, the mechanism of this adaptivity is still not clear. In this paper, we assume there are "level locking" auditory nerve fibers which generate spiking signals only if the sound signal level lies in a specific fixed range.

After encoding the temporal and amplitude information, AN fibers pass the spiking sequence through the superior olivary complex (SOC) up to the inferior colliculus (IC) to calculate the ITDs and ILDs. Two separate circuits, the ITD pathway and the ILD pathway, are involved in the calculation. The ITDs [7] have been proved to be processed in the medial superior olive (MSO), which is one part of the SOC, and the ILDs are processed in another part of the SOC, i.e. the lateral superior olive (LSO). The MSO in one side receives excitation from the AVCN (anteroventral cochlear nucleus) from both the ipsilateral and contralateral ears. Besides the excitation, recent neurophysiological studies have revealed that the MSO also receives inhibition shortly following the excitation [7]. After the ITD processing in the MSO, the results are projected to the IC. For the ILDs processing, cells in the LSO are excited by sounds that are greater in level at the ipsilateral ear than the contralateral ear and inhibited by sounds that are greater in level at the contralateral ear [7]. Therefore, the LSO receives excitation from the ipsilateral AVCN, but inhibition from the MNTB (medial nucleus of the trapezoid body) which converts the excitation from the contralateral AVCN to inhibition. Finally, the spiking output of the LSO is also tonotopically projected to IC.

3 Model of ILD processing in LSO

In the LSO, the model of the ILD processing is unclear yet. In this paper, we propose an ILD model based on the neurophysiological evidence of single delay line and the assumption of the existing of level locking AN. This model, called pyramid ILD model, is illustrated in Figure 1. The left slope of the pyramid receives the inhibition inputs from the contralateral MNTB and the level order of the inhibition is arranged in the reversed way of the pyramid layer, i.e. the level number increases from the top to bottom layer. The inhibition of one specific sound level gets into the network along the dot lines in the figure. The right slope of the pyramid takes the excitation inputs from the ipsilateral AVCN and the level order of the excitation is the same way of the pyramid layer. The excitation of one sound level passes down the network along the dash line. The neuron at the cross point of the excitatory dash line and the inhibitory dot line is the ILD

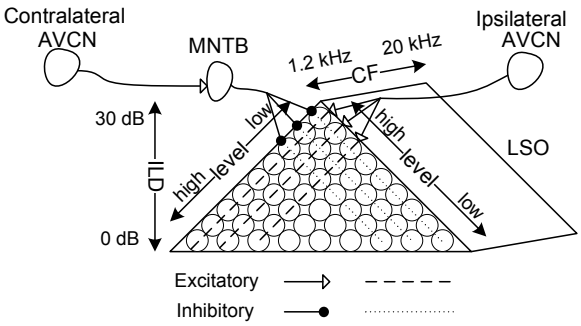


Fig. 1: Schematic diagram the ILD processing of the LSO in the human.

coincidence neuron. It fires only if there are spikes at the excitatory line and no spike at the inhibitory line. The layer where the coincidence neuron belongs indicates the ILD, i.e. the lower layer represents the small level difference and vice versa.

4 System Model of Sound Localisation

Inspired by the neurophysiological data and the proposed ILD pathway assumption presented in Section 3, we designed a system model of sound localisation by using spiking neural networks (SNNs). The sound is first encoded into spikes as inputs to the SNN. The synaptic response $I(t)$ to a spike, occurred at $t = t_s$, is modelled as a constant square current with an amplitude (also called weight) of w_s, a latency l_s relative to the timing of the spike, and a lasting time τ_s. The sign of the the amplitude indicates whether the synapse inhibits (negative) or excitates (positive) the following soma. The soma response to $I(t)$ can be modelled based on the leaky integrate-and-fire model in [8].

$$u(t) = u_r \exp(-\frac{t - t_s}{\tau_m}) + \frac{1}{C}\int_0^{t-t_s} \exp(-\frac{s}{\tau_m}) I(t-s)ds \tag{1}$$

where u_r is the initial membrane potential, and τ_m is a time constant. In this paper, a typical value for τ_m is 1.6 ms. C is the capacitor which is charged by $I(t)$, in order to simulate the procedure of the postsynaptic current charging the soma. The soma model has one more parameter, the action potential φ. When $u(t) = \varphi$, the soma will fire a spike, then $u(t)$ is reset to 0.

In contrast to other sound localisation systems, e.g. [9], which only applied an ITD pathway, our model utilises both the ITD and ILD pathways for both sides of ear. This feature provides a wide-frequency localisation processing ability to the model as we will see in Section 5. Furthermore, the SNN of the model simplifies the model of the synapse, and the parameters of the synapse and soma are independent to sound frequencies in contrast to the model in [9]. This feature enables our system to have a real-time computation ability and potential

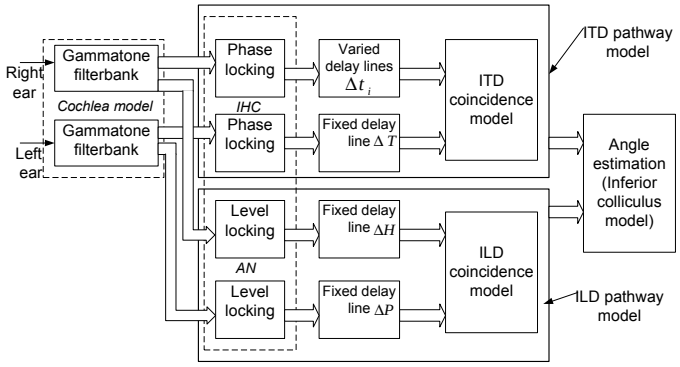


Fig. 2: Schematic structure of biologically inspired sound localisation system. This example assumes the left ear is the ipsilateral ear.

optimisation possibility. Figure 2 shows a schematic structure for the sound localisation procedure in the left ear side. Please note that there is another corresponding mode in the right ear side. In the figure, the tonotopical sound encoding feature of the cochlea is simulated by a bandpass filterbank consisting a series of Gammatone filters [10]. The filterbank decomposes the sound into several frequency channels in a similar way as the cochlea processing the sound.

After the Gammatone filterbank, the temporal information of the sound in each frequency channel is encoded into a spike sequence by the phase locking module in Figure 2, which simulates the phase locking character of the inner hair cell in the cochlea. At every positive zero-crossing of the incoming sound, a spike is triggered. At the same time, the amplitude information is encoded into several spike sequences by the level locking module.

Following the ITD pathway in Yin's model [5], the spike sequence of the contralateral ear, i.e. the right ear in Figure 2, passes variable delay lines Δt_i. We denote the delayed spike sequence as $S_{CP}(\Delta t_i, f_j)$, where C stands for the contralateral, P for phase locking, Δt_i for the delay time, f_j for the frequency channel j. Similarly, $S_{IP}(\Delta T, f_j)$ represents the delayed spike sequence of the ipsilateral ear with a fixed delay time ΔT. $S_{CP}(\Delta t_i, f_j)$ and $S_{IP}(\Delta T, f_j)$ are then input to the ITD coincidence model (please see Figure 3-(a) for details) to calculate the ITD. The calculated output of the ITD coincidence model is a new spike sequence represented as $S_{ITD}((\Delta_T - \Delta t_i), f_j)$. If there are spikes in $S_{ITD}((\Delta_T - \Delta t_i), f_j)$, it means the sound arriving to the ipsilateral ear is earlier than that to the contralateral ear by $ITD = \Delta_T - \Delta t_i$ second. Once the ITD calculation is implemented for all frequency channels, a three dimension ITD distribution map can be drawn, where the x-axis is for ITD, y-axis for the frequency channel and z-axis for the mean spike number in a unit time.

The ILD pathway is modelled in the bottom rectangular box in Figure 2. The level locking spike sequences from the contralateral and ipsilateral sides pass fixed delay lines, Δ_H and Δ_P, respectively. Then they go into the ILD coincidence

model for calculating the level difference. Figure 3-(b) illustrates the coincidence model in detail. In the figure, $S_{CL}(l_i, f_j)$ represents the contralateral level locking spike sequence of frequency channel j and level l_i, while $S_{IL}(l_k, f_j)$ the ipsilateral level locking spike of level l_k. The output of ILD model is $S_{ILD}((l_k - l_i), f_j)$, which indicates the spike sequence when $ILD = l_k - l_i$. Once the ILD calculation is implemented for all frequency channels and for both sides, a ILD distribution map can be drawn in the similar way to the ITD distribution.

The calculation results of the ITD and ILD coincidence model are finally merged together as shown in the last module of Figure 2. Considering the complex head transfer function between the ILDs and source source angles [2], we use the ILD results to identify whether the sound came from the left or right side. Then we use this information to remove the ambiguity in the ITD results. For example, if the sum of the negative ILD spikes is larger than that of positive ILD spikes, we can conclude that the sound came from left side and then all the positive ITD spikes can be ignored in the following angle estimation of the sound source. After correcting ITD spikes, we can choose the significant ITD by using several methods, such as winner-take-all and the weighted mean method. Finally, the sound source azimuth angle can be calculated by:

$$\theta = \arcsin(\mathrm{ITD} \times \mathrm{V}_{sound}/\mathrm{d}_{\mathrm{ear}}) \tag{2}$$

where $\mathrm{V}_{\mathrm{sound}}$ is the sound speed, typically 344 m/s in 20^{o}C. $\mathrm{d}_{\mathrm{ear}}$ is the distance between two ears, i.e. microphones in robot.

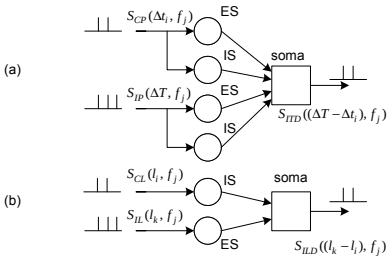


Fig. 3: (a) ITD coincidence model (b) ILD coincidence model. ES stands for excitatory synapse, and IS inhibitory synapse.

5 Experimental Results

To justify the feasibility and performance of our sound localisation model, we designed two groups of experiments: (i) artificial pure tone localisation, and (ii) real pure tone localisation. The sound in the former experiments is generated by a computer and it is manually added time difference and level difference on two sound channels according to the physical parameter of our mobile robot,

MIRA. For example, to simulate a sound coming from the left 90 degree, the right channel sound is added by an extra time $\frac{\sin(90^\circ)\mathrm{d}_{ear}}{\mathrm{V}_{sound}}$ at the beginning and decreased in level by an arbitrary 50%. The ear distance $\mathrm{d}_{\mathrm{ear}}$ is 21 cm. The sound in real pure tone localisation is recorded in a general environment with 30 dB background noise. The distance of the sound source to the robot is 90 cm and the sound pressure at the speaker side is controlled to be 90±5 dB. The sample rate is 22050 Hz and the duration is 100 ms with 10 ms silence at the beginning. Pure tones with frequency 500, 1000 and 2000 Hz were recorded at the positions of -90, -60, -30, 0, 30, 60 and 90 degrees. The parameters of the spiking neural network are listed in Table 1. The Gammatone filterbank is set with 16 channels to cover from 200 to 3000 Hz.

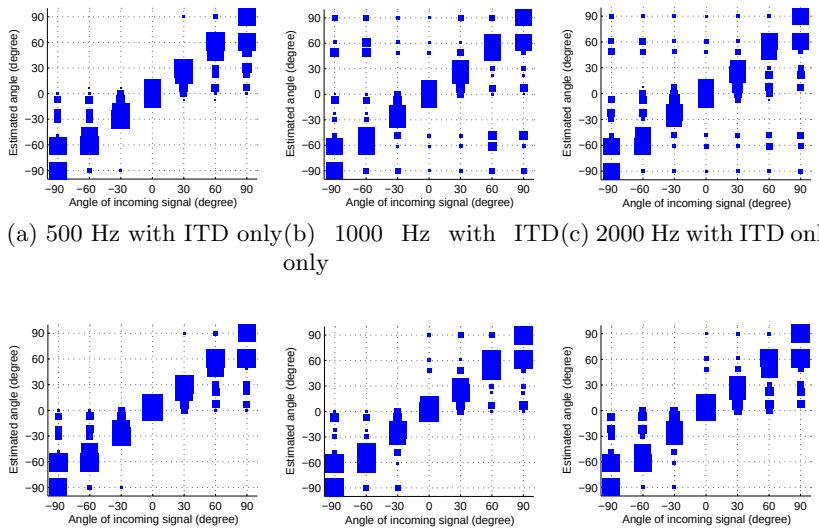
Table 1: Parameters in experiments

	Excitatory Synapse			Inhibitory Synapse			Soma		
	l_s	τ_s	w_s	l_s	τ_s	w_s	φ	t_m	C
ITD	2.1	0.08	0.1	2.28	0.08	-0.1	8e-4	1.6	10
ILD	2.1	0.08	0.1	2.28	0.08	-0.1	8e-4	1.6	10
	channels		channels		Δt_i		ΔT	ΔH	ΔP
ITD	17		n/a		[1.4 2.6]		2	n/a	n/a
ILD	n/a		10		n/a		n/a	2	1.8

*Note: the unit of l_s, τ_s, w_s, t_m,Δ_{t_i}, Δ_T, ΔH and ΔP is ms and the unit of C is mF.

Figure 4 shows the spike distributions in the sound localisation of the artificial signal. In these distribution figures, the x-axis is the expected angle of sound localisation and the y-axis is the estimated angle after calculation. The size of square is proportional to the spiking number in the corresponding angle estimation. For example, in Figure 4a, a big square at (0 0) means that when the artificial signal came from 0 degree the spiking number of estimated angle of 0 degree is the majority of spikes of all estimated angles. The localisation efficiency of the system is defined as a percentage of the spiking number at the correct estimation point, such as (-30 -30) and (60 60), to the total spiking number. In the figure, we compare the performance of two methods, i.e. the localisation using (i) the ITDs only and (ii) the ITDs with the ILDs.

In the top column of Figure 4, the localisation efficiency across all frequencies (500, 1000, 2000 Hz) and angles is about 70%. 85% of the sound signals across all frequencies were recognised correctly between -45 to 45 degree. The highest localisation efficiency occurred at the 0 degree. The efficiency decreases when the frequency goes high or the sound moves to the sides. This result matches the fact [7] that (i) the ITDs cue has the highest efficiency for sound localisation when the sound source is in front of the observer, and (ii) the ITD cue effect on sound localisation fades down over 1.2 kHz.



(a) 500 Hz with ITD only (b) 1000 Hz with ITD only (c) 2000 Hz with ITD only

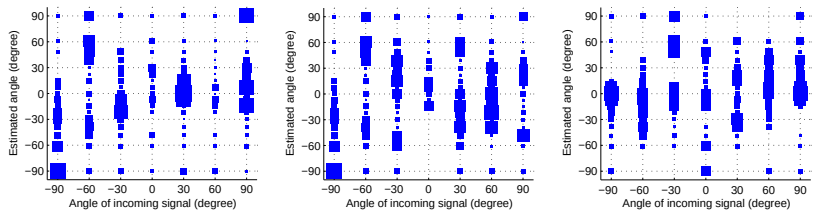
(d) 500 Hz with ITD & ILD (e) 1000 Hz with ITD & ILD (f) 2000 Hz with ITD & ILD

Fig. 4: The artificial sound localisation results for 500, 1000 and 2000 Hz. The size of square represents the spiking proportion in the corresponding angle estimation.

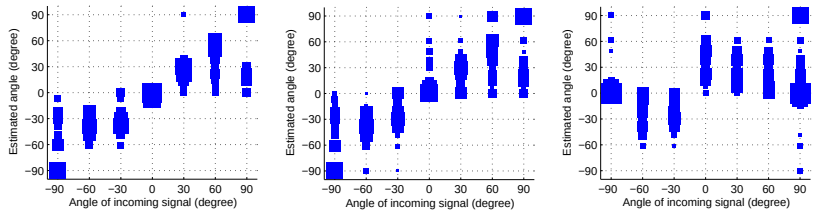
After adding the ILDs into our system, the experiments results shown in the bottom column of Figure 4 demonstrate that the spiking distribution is more concentrated, especially in the figure of 1000 Hz and 2000 Hz, than that of the results by only using the ITDs. This results match the fact [7] that the ILDs is the main cue for the high frequency sound localisation. The overall localisation efficiency is increased to 80% and 95% of the sound signals across all frequencies were recognised correctly between -45 to 45 degree.

Figure 5 shows the spike distributions for the sound localisation of a real pure tone signal. The same experimental methods for the artificial sound localisation were applied. In these figures, the ambiguity of the estimated sound source angle is large when only the ITD cue is used. The overall localisation efficiency dropped down below 50% due to acoustic clatter which affects the ITD calculations because the phase-locking block in our system will generate more irrespective spike sequence in terms of these noise. However, after adding ILDs into the system, the ambiguity in the ITD calculation is improved and the overall localisation efficiency is raised to 65% because the noise level in the signal generally is not different in two microphones and therefore does not affect the ILD calculation much.

The time cost for processing a 100 ms sound signal is less than 50 ms on a 2.6 GHz CPU. Comparing the 9.1 s time cost when using the model in [9], our



(a) 500 Hz with ITD only (b) 1000 Hz with ITD only (c) 2000 Hz with ITD only



(d) 500 Hz with ITD & ILD (e) 1000 Hz with ITD & ILD (f) 2000 Hz with ITD & ILD

Fig. 5: The real sound localisation results for 500, 1000 and 2000 Hz. The size of square represents the spiking proportion in the corresponding angle estimation.

system performs much quicker without losing localisation efficiency. It is feasible to apply our system for real time sound localisation in real world.

6 Conclusion and Future Work

In this paper, we implemented a sound localisation system using spiking neural network inspired by mammalian auditory processing. In this system, both the ITD and ILD pathway were adopted and modelled based on recent neurophysiologic findings. In the ITD pathway, inhibitory inputs to the MSO are added together with traditional excitatory inputs in order to get a shape localisation results. In the ILD pathway, we proposed an assumption of the level-locking auditory nerve and built a pyramid model to calculate ILDs. The parameters of the SNN were set independent to the sound frequency and ITDs (ILDs) in contrast to similar work in [9]. The experimental results showed that our system can localise the sound source from the azimuth -90 to 90 degree. The sound frequency varied from 500 to 2000 Hz. The effect of frequency and sound source position to the localisation efficiency had a high correspondence with neurophysiologic data. It proved the reasonability of the proposed system.

In the future, active sound localisation, which can specify the feature frequencies of an interesting object, will be the next step of our research. In addition, an adaptive relative range of the level-locking encoding by using the feedback from

SOC will increase the localisation efficiency of our system in a cluttered environment. For the application of our system to a mobile robot, we are planning to implement a self-calibration sound localisation system which can adaptively adjust the synapse and soma parameters according to the environment or electrical hardware change.

Acknowledgment

This work is supported by EPSRC (EP/D055466). Our thanks go to Dr. Adrian Rees and Dr. David Perez Gonzalez at University of Newcastle for their contribution on neurophysiologic support toward the project. This project, MiCRAM, is a cooperative project with them and they are mainly specialising in the neurophysiologic structure on the inferior colliculus for sound processing. We would also like to thank Chris Rowan for helping with building the MIRA robot head.

References

1. Oertel, D., Fay, R., Popper, A., eds.: Integrative Functions in the Mammalian Auditory Pathway. Springer: New York (2002)
2. Young, E., Davis, K.: Circuitry and Function of the Dorsal Cochlear Nucleus. In: Integrative Functions in the Mammalian Auditory Pathway. Springer: New York (2002) 160–206
3. Fitzpatrick, D., Kuwada, S., Batra, R.: Transformations in processing interaural time differences between the superior olivary complex and inferior colliculus: Beyond the jeffress model. Hear. Res. **168**(1-2) (2002) 79–89
4. Jeffress, L.: A place theory of sound localization. J. Comp. Physiol. Psychol. **61** (1948) 468–486
5. Smith, P., Joris, P., Yin, T.: Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: evidence for delay lines to the medial suprior olive. J. Comp. Neurol. **331** (1993) 245–260
6. Hirsch, J.A., Chan, J.C., Yin, T.C.: Responses of neurons in the cat's superior colliculus to acoustic stimuli. i. monaural and binaural response properties. J. Neurophysiol. **53** (1985) 726–745
7. Yin, T.: Neural mechanisms of encoding binaural localization cues in the auditory brainstem. Integrative Functions in the Mammalian Auditory Pathway (2002) 99–159
8. Gerstner, W., Kistler, W.M.: Spiking Neuron Models, Single Neurons, Populations, Plasticity. Cambridge Unviersity Press (2002)
9. Voutsas, K., Adamy, J.: A biologically inspired spiking neural network for sound source lateralization. IEEE Trans Neural Networks **18**(6) (2007) 1785–1799
10. Meddis, R., Hewitt, M., Shackleton, T.: Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse. J. Acoust. Soc. Am. **87**(4) (1990) 1813–1816