

A Spiking Neural Network Implementation of Sound Localization

J. Wall, L. J. McDaid, L. P. Maguire and T. M. McGinnity

*School of Computing and Intelligent Systems
University of Ulster, Magee Campus
Northland Road, Derry, BT48 7JL
{wall-j, lj.mcdaid, lp.maguire, tm.mcginny}@ulster.ac.uk*

Abstract – The focus of this paper is the implementation of a spiking neural network to achieve sound localization; the model is based on the influential short paper by Jeffress in 1948. The SNN has a two-layer topology which can accommodate a limited number of angles in the azimuthal plane. The model accommodates multiple inter-neuron connections with associated delays, and a supervised STDP algorithm is applied to select the optimal pathway for sound localization. Also an analysis of previous relevant work in the area of auditory modelling supports this research.

Keyword – sound localization, STDP, spiking neural networks

I. INTRODUCTION

Of all the organs in the body, there are very few that can compare to the ear with regards to the degree of functionality it contains within such a small and compressed space. Sound localization is one function that the ears perform, defined as determining where a sound signal is generated in relation to the position of the human head. It is a very powerful aspect of mammalian perception, allowing an awareness of the environment and permitting mammals to locate prey, potential mates and to determine from where a predator is advancing [1].

Mammalian sound localization can be determined with interaural time difference (ITD); a sound at one side of the body will arrive at one ear before the other and the ITD is the very small difference in their arrival times [2]. From this the brain can calculate the angle of the sound source in relation to the head [3, 4]. ITD is very sensitive and can differentiate between angles of 1-2°; however when the frequency of the sound is greater than 900Hz, ITD becomes much less reliable and interaural intensity difference (IID) is used as confirmation [2]. The auditory pathway (see Figure 1) for ITD begins at the cochlear nucleus of both ears. The sound signal travels to the medial superior olive (MSO) of the superior olivary complex (SOC) and it is here that the ITD is calculated [5]. This information travels further up the pathway to the inferior colliculus (IC) where the azimuth (angle from a certain direction) is worked out. The cortex,

with the known angle, now has awareness of the sound source and the superior colliculus enables reflex movement of the head and eye to move toward the sound source [5].

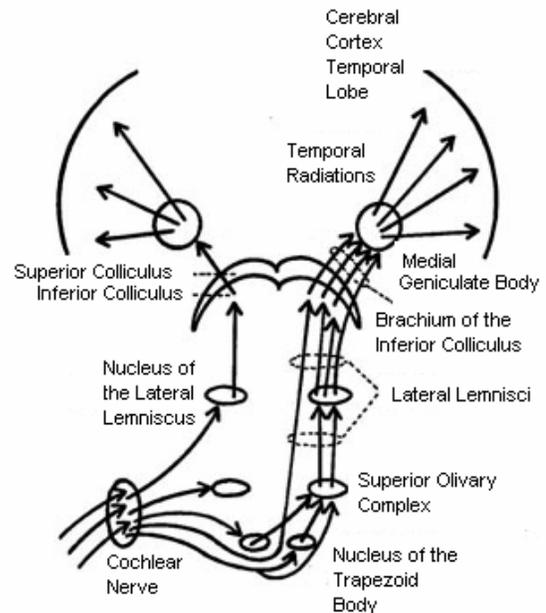


Figure 1: Auditory Pathway

This processing is achieved in real time as the brain uses parallel processing using many neurons to simultaneously transmit the information up through

the auditory pathway; the number of neurons varies from six to forty for each one-third-octave frequency band [6].

In this paper, a spiking neural network will implement the Jeffress sound localization technique for a limited number of angles in the azimuthal plane. The ITD will be encoded in the two inputs to the network and the output neurons will determine the angle of location using both axon delay lines and coincidence functions. Section 2 provides a review of the related research conducted in the field. Section 3 outlines the model proposed in the paper while section 4 outlines the experimental results found. Finally, section 5 concludes the paper and outlines some possible future work.

II. LITERATURE REVIEW

Research in the area of post-cochlear perception is based on the work carried out on modelling of the cochlea itself. This work assists in the understanding of how the brain utilizes these post-cochlear signals for sound localization. Implementation of the cochlea has been an area of research and one focus is in embedded systems. Prominent researchers in this field have modelled a cochlea on silicon [7]. Lyon and Mead introduced the analog electronic cochlea; built in CMOS VLSI technology it involved a cascade of second-order filter stages to mimic the travelling-wave system of fluids in the cochlea, test results showed that it matched both previous theories and observations of real cochleas [8]. In the mid 90s Kuszta outlined two main methodologies for designing artificial cochleas based on Mead's description of Very Large Scale Integration (VLSI) systems containing electronic analog circuits that imitate neuro-biological architectures existing in the nervous system [9]. Other research in this area involved Van Schaik's models of the cochlea as a three-tier design that included the artificial cochlea, the inner hair cell (IHC) model and a spiking neuron circuit on one chip [10, 11]. On the circuit, thirty-two neurons could be combined together to create a small and simple network that can reproduce the spiking behaviour of neurons in the auditory system [12]. Integrate-and-fire neurons were modelled on the silicon chip, but it is unclear what training was applied [13].

Software modelling of the biological cochlea has not received as much attention. However, neural networks have been used to some degree to model the cochlear nuclei in the brain. Just as electronic engineering created a spiking neuron circuit, software neural networks have been created to model the output of the IHCs of the cochlea. Sheikhzadeh and Deng created a three layer feed-forward neural network model of the dorsal cochlear nucleus (DCN)

[14]. They first created the basilar membrane (BM), IHC and the action potential generator. The BM model used the biophysical mechanisms behind the BM vibration. This gave a dynamic nonlinear BM filter function as opposed to previous models which used simple linear digital filters. The spike generator produced random sequences of auditory nerve action potentials as inputs to the DCN model. The model was tested with both synthetic and natural speech sounds and processed these sounds in a similar way to a true biological auditory model.

In 1948, Jeffress created a computational model to show how ITD works in mammals to determine the angle of origin of a sound signal [3, 15]. His model involved time or phase locked inputs; a set of delay lines to vary the axonal path lengths arriving at the neuron; and an array of coincidence detector neurons which will only fire when presented with simultaneous inputs from both ears [1, 3, 4, 5]. Coincident inputs only occur when the ITD is exactly compensated for by the delay lines. The fundamental importance of Jeffress' model and why it has become the prevailing model of binaural sound localization is its ability to depict auditory space with a neural representation in the form of a topological map, even though Jeffress himself acknowledged the simplicity of his model [1].

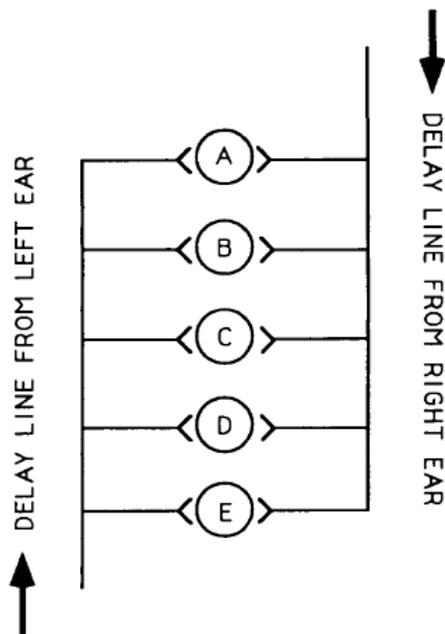


Figure 2: The Jeffress (1948) Model

Schauer et al. have based their work extensively on the Jeffress sound localization model. Their initial research involved a biologically inspired model of binaural sound localization again by means of ITD; using a spike response model for implementation in analog VLSI. Slight modifications

to the Jeffress model were made including a digital delay line with AND gates. Data recorded in an open environment was used in testing which was carried out offline; results showed that the model was proficient at localizing single sound sources for sixty-five azimuthal angles [16]. Schauer and Gross extended this work to discriminate between sound sources of different orientations. However, this was achieved in a biologically implausible way. The authors simply specified one microphone for the front and another for the back. Differences in the sound colour of the binaural signals, calculated using a short-term Fast Fourier Transform (FFT), determined from which direction the sound approached. Again positive results were achieved during testing in open environments, including a lecture hall and a shopping centre [17]. Currently, Schauer and Gross are working on a computational model for early auditory-visual integration for the aim of performing a robust multimodal attention-mechanism in artificial systems. Using their auditory model from [17], they combined this with a visual and bimodal map; with the visual map based on spatio-temporal intensity differences. To test their model they combined recordings of real-world situations and off-line simulations. The authors perceive their model as a bench-mark for future research in audio-visual integration [18].

The development of a learning technique that can detect the location of sound is an area of research which has produced only a modest amount of work. Ultimately research needs to address this issue leading to the development of a bio-inspired system that can locate a sound source in the presence of background noise. The model presented here represents the initial state in the development of such a system, whereby a learning capability provides the basis from which to build a sound localization technique.

III. SNN SOUND LOCALIZATION

This paper proposes an implementation of the Jeffress sound localization model using SNNs by considering a sound source at five distinct angles (θ) on the horizontal azimuthal plane (180°): 0° , 45° , 90° , 135° , and 180° .

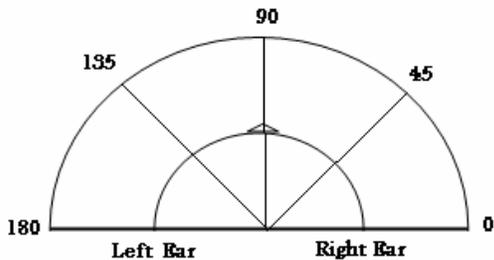


Figure 3: Angles of the Horizontal Plane

The SNN models the coincident-detection neurons of the medial superior olive (MSO). The MSO is the largest of the nuclei in the superior olivary complex (SOC) containing between 10,000 – 11,000 cells; it has a tonotopic response pattern which favours low frequencies. Its cells work as coincidence detectors to identify the ITD of a sound signal and thus recognition of the sound source angle. The MSO is very important in the localization process, in particular as cell types here are coincidence detectors [19].

The topology of the model, shown in figure 4, includes a SNN network consisting of five processing neurons implemented using the leaky integrate and fire (LIF) model [20]. The inputs t_1 and t_2 (chosen arbitrarily) correspond to the length of time taken for the sound to reach both cochleas, and these inputs are passed to the processing neurons via the cochlear nodes. The synapse on each pathway encompasses the multiple delay structure as shown in figure 5.

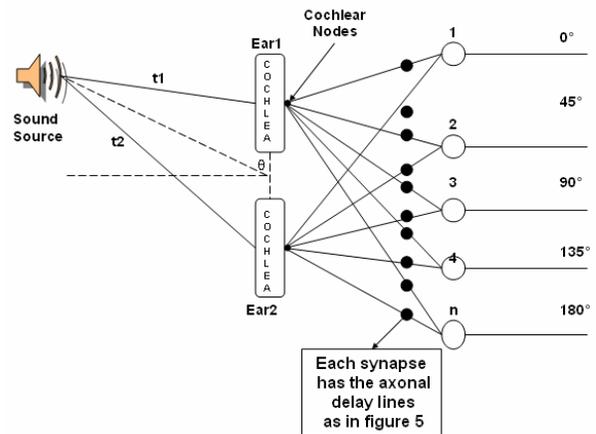


Figure 4: Network Topology

Figure 5 shows how delay lines are used in this model, where t_{pre} is the presynaptic spike time; d_i are the axonal delays; w_i are the weights; and t_{post} is the postsynaptic spike time [21]. The output from neuron A is converted to m number of outputs where $m=5$, each with their own weight w_i .

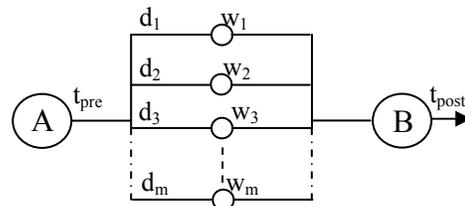


Figure 5: Pre and Post Synaptic Neurons with Delay Lines [21]

Spike Timing Dependent Plasticity

STDP occurs naturally in neurons and is a form of synaptic plasticity, i.e. the capacity for the synapse connecting two neurons to change strength [22]. It is a form of Hebbian learning which strengthens the weights of the synapses that are activated before the post synaptic spike and weakens those synaptic weights that occur after the post synaptic spike [23, 24]. The weight updates are potentiated according to

$$\delta w_i = A_1 e^{-\frac{\delta t_i}{\tau_1}} \quad (1)$$

and depressed according to

$$\delta w_i = -A_2 e^{-\frac{\delta t_i}{\tau_2}} \quad (2)$$

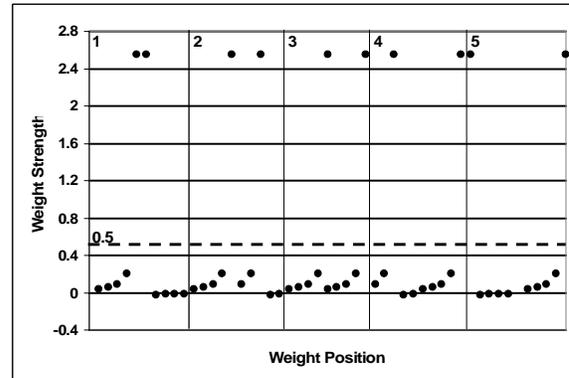
where δw_i is the weight change, A_1 is the maximum value of the weight potentiation, A_2 is the maximum value of weight depression, δt_i is the output spike time minus the input spike time, τ_1 is the width of the window for long term potentiation and τ_2 is the width of the window for long term depression.

Supervised training is used in this work, thus the post synaptic spike time for each neuron is known. The input sets are passed to the network and the weight values for each neuron are calculated. For the delay lines which caused coincidence at the neuron, STDP increased their weights according to the above learning rule in (1) and the weights of the other delay lines are decreased according to (2). For instance, the neurons corresponding to each of the five angles (0° , 45° , 90° , 135° and 180°) will be passed inputs t_1 and t_2 , and after training the classifying neuron for each angle will only fire when presented with their unique inputs. After a period of training the ITD (encoded by t_1 and t_2) is compensated for by the delay lines, two inputs will coincide at the neuron and only in that case will the neuron fire. The other sets of inputs will also have reached the neuron but due to the training procedure their combined post synaptic potentials (PSP) will not have caused the neuron to fire.

IV. EXPERIMENTAL RESULTS

To evaluate the SNN model, the network was trained by passing the inputs (t_1 and t_2), encoded as single spikes, to the processing neurons. Dependent on the azimuthal angle, the output neuron was supervised to fire at a pre-determined time; thus allowing the STDP rule to select the best pathway to facilitate coincidence. The graph in figure 6 shows the weight distribution on the vertical axis of the post trained SNN where each window 1 to 5 represents the

classifying neurons and their associated weights: the horizontal axis is the spatial distribution of synapses across the network. The dotted line at a weight value of 0.5 represents the pre-trained weight distribution. Note a bimodal weight distribution which is characteristic of the STDP process. Potentiated weights at approximately 2.5 are associated with pathways that have been selected by the STDP training rule because their delays cause coincidence at the appropriate classifying neuron.



VI. REFERENCES

- [1] McAlpine, D. and Grothe, B. (2003) 'Sound Localization and Delay Lines – Do Mammals Fit the Model?', *Trends in Neurosciences*, 26(7):347 – 350.
- [2] Lewicki, M. S. (2006) 'Sound Localization I', Computational Perception and Scene Analysis Module, Carnegie Mellon University, <http://www.cs.cmu.edu/~lewicki/cpsa/sound-localization1.pdf>.
- [3] Grothe, B. (2003) 'New Roles for Synaptic Inhibition in Sound Localization', *Nature Reviews Neuroscience*, 4(7):540 – 550.
- [4] Carr, C. E. (1993) 'Delay Line Models of Sound Localization in the Barn Owl', *American Zoology*, 33:79 – 85.
- [5] Alexander, M. 'Sound Localization of Humans', UCSD Vision Group, <http://psy.ucsd.edu/~dmacleod/159/presentations05/MichaelSound%20Localization%20of%20Humans.pdf>.
- [6] Hartmann, W. M. (1999) 'How We Localize Sound', *Physics Today on the Web*, <http://www.aip.org/pt/nov99/locsound.html>.
- [7] Lyon, R. F. (1982) 'A Computational Model of Filtering, Detection, and Compression in the Cochlea', *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Paris, France, 1282 – 1285.
- [8] Lyon, R. F. and Mead, C. (1988) 'An Analog Electronic Cochlea', *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7):1119-1134.
- [9] Kuszta, B. (1998) 'Silicon Cochlea', *Proc. of the Western Electronic Show and Convention*, Anaheim, California, 282 – 285.
- [10] Chan, V., Van Schaik, A. and Liu, S. C (2006), 'Spike Response Properties of an AER EAR', *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS)*, Island of Kos, Greece, 859 – 862.
- [11] Liu, S. C. and Van Schaik, A. (2005) 'An AER Ear', *The Neuromorphic Engineer*, 2(1):11-12.
- [12] Van Schaik, A., Fragniere, E. and Vittoz, E. (1996) 'An Analogue Electronic Model of Ventral Cochlear Nucleus Neurons', *Proc. of the Fifth International Conference on Microelectronics for Neural Networks and Fuzzy Systems (MicroNeuro)*, Lausanne, Switzerland, 52 – 59.
- [13] Van Schaik, A. (2003) 'The Electronic Auditory Pathway', University of Sweden, <http://www.eelab.usyd.edu.au/andre/eap/>.
- [14] Sheikhzadeh, H. and Deng, L. (1999) 'A Layered Neural Network Interfaced with a Cochlear Model for the study of Speech Encoding in the Auditory System', *Computer Speech and Language*, 13(1):39-64
- [15] Jeffress, L. A. (1948) 'A Place Theory of Sound Localization', *Journal of Comparative Physiology and Psychology*, 41:35 – 39.
- [16] Schauer, C., Zahn, T. et al. (2000) 'Binaural Sound Localization in an Artificial Neural Network', *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2:865 – 868.
- [17] Schauer, C. and Gross, H. M. (2001) 'Model and Application of a Binaural 360° Sound Localization System', *Proc. of the International Joint Conference on Neural Networks*, 2:1132 – 1137.
- [18] Schauer, C. and Gross, H. M. (2003) 'A Computational Model of Early Auditory-Visual Integration', *Proc. of the 25th Pattern Recognition Symposium (DAGM)*, Magdeburg, 362 – 369.
- [19] Delgutte, B., and Oxenham, A. (2005) Hearing and the Auditory System: Overview. Harvard-MIT Division of Health Sciences and Technology, <http://ocw.mit.edu/OcwWeb/Health-Sciences-and-Technology/HST-725Spring2004/LectureNotes/index.htm>.
- [20] Ibarz, B., Cao, H. and Sanjuan, M. A. F. (2007) 'Map-Based Neuron Networks', *Cooperative Behaviour in Neural Systems: Ninth Granada Lectures*, 887:69 – 76.
- [21] Strain, T. J., McDaid, L. J., Maguire, L. P. and McGinnity, T. M. (2001) 'A Novel Supervised STDP Based Training Approach for Spiking Neural Network Classification', *Intelligent Systems Engineering Laboratory*.
- [22] Martin, S. J., Grimwood, P. D. and Morris, G. M. (2000) 'Synaptic Plasticity and Memory: An Evaluation of the Hypothesis', *Annual Review of Neuroscience*, 23:649 – 711.
- [23] Vogelstein, R. J., Tenore, F., Philipp, R., Adlerstein, M. S., Goldberg, D. H. and Cauwenberghs, G. (2002) 'Spike Timing-Dependent Plasticity in the Address Domain', *Advances in Neural Information Processing Systems 15*.
- [24] Izhikevich, E. M. and Desai, N. S. (2003) 'Relating STDP to BCM', *Neural Computation*, 15:1511 – 1523.