A Comparison of Sound Localisation Techniques using Cross-Correlation and Spiking Neural Networks for Mobile Robotics

Julie A. Wall, Thomas M. McGinnity, Liam P. Maguire

Abstract—**This paper outlines the development of a cross-correlation algorithm and a spiking neural network (SNN) for sound localisation based on real sound recorded in a noisy and dynamic environment by a mobile robot. The SNN architecture aims to simulate the sound localisation ability of the mammalian auditory pathways by exploiting the binaural cue of interaural time difference (ITD). The medial superior olive was the inspiration for the SNN architecture which required the integration of an encoding layer which produced biologically realistic spike trains, a model of the bushy cells found in the cochlear nucleus and a supervised learning algorithm. The experimental results demonstrate that biologically inspired sound localisation achieved using a SNN can compare favourably to the more classical technique of cross-correlation.**

I. Introduction

One of the key functions that the ears and auditory pathways perform is the ability to determine the point of origin of a sound source. It is a powerful aspect of mammalian perception, allowing an awareness of the environment and permitting mammals to locate prey, potential mates and predators [1]. The neural components of sound localisation are complicated, as the location of a stimulus can only be determined by combining input from both ears [2].

Mammalian sound localisation is determined with a combination of binaural cues; ITDs, which are processed in the medial superior olive (MSO) for low frequency sound-signals and interaural intensity differences (IID), which are processed in the LSO for high frequency sounds (> 2 kHz) [3]. Both the LSO and MSO are located within an area of the auditory system called the superior olivary complex [3]. The combination of ITD and IID processing is better known as the "duplex theory of sound localization" and was first devised by Thompson and Rayleigh [4-5]. In this paper there is a focus on sound localisation by means of ITD, defined as the different points in time at which a sound from a single location arrives at each individual ear [6]. From this time difference, the brain can calculate the angle of the sound source in relation to the head [7].

The ITD cue works most effectively for sounds greater than ~200 Hz to about 1.5 kHz in humans since the sound wavelengths are wide and sound intensity is not discernibly weakened by the size of the head [8]. Low frequency sound waves have a wavelength that is greater than the diameter of the head; therefore each ear receives the sound wave at a different point in time. For example, if a sound signal originates to the extreme left of the head, it will reach the left ear first and after a time delay which is specific to the azimuthal angle of the sound source it will then reach the right ear, generating the ITD. ITDs occur at both the onset of the sound and throughout the duration of the sound, known as onset ITDs and ongoing ITDs respectively [9].

The ITDs in continuous and periodic sounds produce interaural phase differences (IPD), i.e. differences in the phase of the sound wave that approach each ear. The fibers of the auditory nerve which respond best to low frequencies produce spike trains which are time locked to the signal's sine curve, meaning that the intervals between spikes is a period of the curve or a multiple of that period. This feature of the auditory nerve is called phase-locking and is important in sound localisation for extracting the ITD from the sound arriving at each ear; it also occurs in bushy cells of the cochlear nucleus and can only occur at low frequencies, [2].

In 1948, Jeffress created a theoretical computational model to show how ITD works in mammals to determine the angle of origin of a sound signal [10]. This is one of the earliest and most durable models of binaural hearing developed and is used to this day as a basis for binaural hearing research. It was quite remarkable considering how little was known at the time about the structure of the auditory system. The model involved three distinct theories:

1. The inputs to the binaural cells are phase-locked and thus retain accurate timing information.
2. A set of delay lines vary the axonal path lengths arriving at the neuron.
3. An array of coincidence detector neurons fire maximally when presented with coincidental inputs from both ears; these 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 localisation 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.

One of the earliest studies of the MSO was that of Goldberg and Brown in 1969 which showed that MSO neurons were most responsive to low frequencies and extremely sensitive to ITDs [8]. Also of significance was their finding that the spike output of MSO neurons varied with ITD, affirming them to be one of the most temporally sensitive neurons in the nervous system. They also showed

Manuscript received January 5, 2011. This research is supported under the Centre of Excellence in Intelligent Systems (CoEIS) project, funded by the Northern Ireland Integrated Development Fund and InvestNI.

The authors are with the Intelligent Systems Research Centre, School of Computing and Intelligent Systems, University of Ulster, Magee Campus. (phone: +44 (0)28 71675166; email: j.wall@ulster.ac.uk).

that differing neurons of the MSO were most sensitive to a particular ITD, called their "best ITD", which depended on the time delay of their inputs, i.e. neurons fired maximally only when their inputs passed a delay which allowed their inputs to arrive in coincidence at the neuron. Consequently, Goldberg and Brown gave weight to Jeffress' simple model for processing ITDs over twenty years later. Many other researchers continued in this vein, producing findings which supported and augmented Goldberg and Brown's work [8, 11-14].

The research presented in this paper builds on earlier work in biologically inspired sound localisation where the input consisted of experimentally derived HRTF data from an adult domestic cat [15-16]. We feel that it is beneficial to extend this research to be applicable to the area of mobile robotics, and the research presented in this paper is a proof of concept for this overall aim which establishes that a biologically inspired SNN can compare favourably in its sound localisation ability to the more classical methodology of cross-correlation. Mobile robotics provides the ideal platform for the development of a human-like auditory system which can operate in a dynamic and noisy environment. In this paper, two different approaches for sound localisation are presented and their ability to perform accurate sound localisation compared; a classical method in the form of cross-correlation and a biologically inspired method using SNNs influenced by the Jeffress model. Both methodologies record sounds in a dynamic and noisy environment using a mobile robot.

The paper is organised as follows. Section II outlines the experimental setup and the two methodologies used, cross-correlation and SNNs. Section III presents the results obtained from each methodology, discusses and compares these results and also draws a comparison to the state of the art. Finally, Section IV presents the conclusions.

II. Methodology

A. Robotic Framework

A Pioneer 3-DX mobile robot with a pair of stereo omni-directional microphones placed 30cm apart was used in the robotics arena of the Intelligent Systems Research Centre at the University of Ulster. A Vicon motion tracking system was used to model both the robot and the sound source with accurate positional data using reflective markers and high speed cameras. These models were used to determine the actual angle of the sound source in relation to the mobile robot using coordinate geometry with the inverse of the Cosine rule, see Figure 1:

$$\theta_2 = \cos^{-1}\left(\frac{a^2 + c^2 - b^2}{2ac}\right) \quad (1)$$

Knowing the actual angle allows the accuracy of both the cross-correlation algorithm and the SNN to be determined.

The sound source was placed at a distance of 1.5m from the robot and the robot was rotated to record a low frequency (400 Hz) two second pure tone at seven different angles in the range of ±60° in steps of 20°. The sound samples were recorded at a sampling rate of 88.2 kHz and the maximum rise time for each individual sound sample was 50 ms. This limited range of angles and single pure tone frequency were used as the work presented in this paper is a proof of concept towards the development of a human-like auditory system implemented on a mobile robot. Ten recordings were made at each angle to produce an input dataset for both the cross-correlation algorithm and the SNN. Figure 2 shows an example of a pure tone sound recorded by both the left and right microphones, where the sound originates from a -60° angle. These recordings resulting from a simple pure tone sound source are extremely noisy and this demonstrates the difficulty their processing with either the cross-correlation algorithm or the SNN will be.

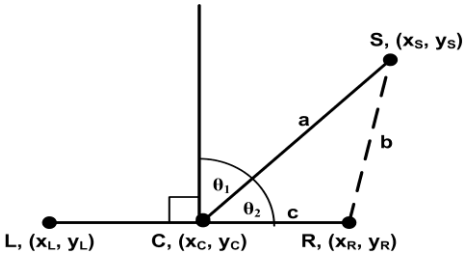

Fig. 1: Use the inverse of the Cosine rule to determine angle θ_2 which enables you to determine angle θ_1, where L is the position of the left microphone; R is the position of the right microphone; C is the centre point between the two microphones and S is the sound source.

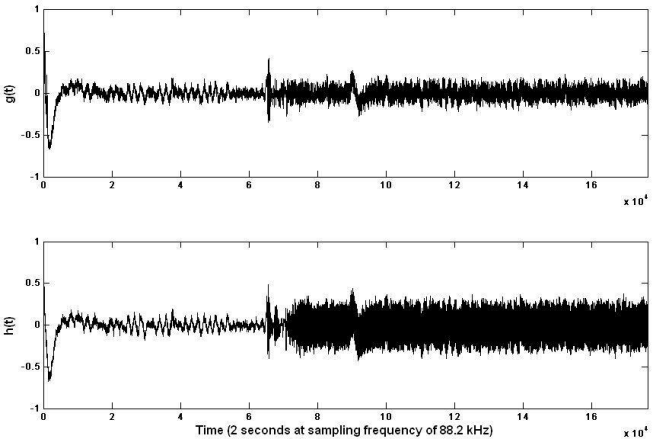

Fig. 2: Two second recording of sound signal at sampling frequency of 88.2 kHz for each microphone, *g(t)* and *h(t)*, at -60°.

B. Sound Localisation by Cross-Correlation

Both the cross-correlation algorithm and the SNN were developed in Matlab. The recorded signal was read into Matlab producing a waveform for each microphone, *g(t)* and *h(t)*, and the sampling frequency (*f=88.2 kHz*). The sampling frequency provides the time interval Δt for each sample within the waveforms:

$$\Delta t = \frac{1}{f} \quad (1)$$

The cross-correlation function is then applied to the two waveforms producing an offset σ which corresponds to the number of samples within either of the waveforms, *g(t)* and

h(t), which will cause full correlation, i.e. cause the two wave-forms to be in phase with one another. Cross-correlation is applied to the two waveforms, *g(t)* and *h(t)*:

$$R_{gh}(\sigma) = \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} g(t)h(t+\sigma)dt \quad (2)$$

where T is the length of the sample. Figure 3 plots the cross-correlation function for an angle of -40°, showing the offset σ of 59 samples. The offset σ corresponds to the maximum value in the cross-correlation function. From this offset σ, the ITD can be determined:

$$ITD = \Delta t * \sigma \quad (3)$$

and thus the azimuthal angle θ of the original sound source:

$$\theta = \sin^{-1} \frac{C_{air} * ITD}{c} \quad (4)$$

where C_{air} is the speed of sound, 343.477m/s, and c is the distance between the two microphones in metres. It should be noted that the ITD is calculated across the entire sound signal, and not just at the onset of the signal.

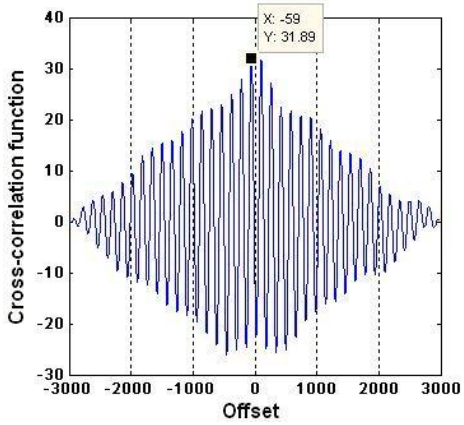

Fig. 3: Cross-correlation function for -40° and the resulting offset σ of 59 samples

The accuracy of the estimated angle produced by the cross-correlation algorithm can vary depending on the sampling rate used when recording the sound signals. The ITD is calculated based on the offset σ which corresponds to the number of samples within either of the waveforms which will cause full correlation; and the time interval Δt of each sample within the recorded signal vectors. The range of offsets σ produced by the cross-correlation function varies greatly in regards to differing sampling rates used. When one second of data was recorded using three differing sampling rates, 8 kHz, 44.1 kHz and 88.2 kHz, the number of estimated angles that could be calculated differed greatly.

Figure 4 demonstrates that only six different angles can be estimated when the sampling rate of 8 kHz is used; 26 different angles with a sampling rate of 44.1 kHz; and 53 angles with a sampling rate of 88. 2 kHz. It is for this reason that all recordings made in these experiments used a sampling rate of 88.2 kHz in order to achieve the greatest level of granularity.

C. *Sound Localisation by Spiking Neural Networks*

In contrast to the cross-correlation method, using networks of spiking neurons to generate the azimuthal angle is more biologically inspired as they are based on the modelling of

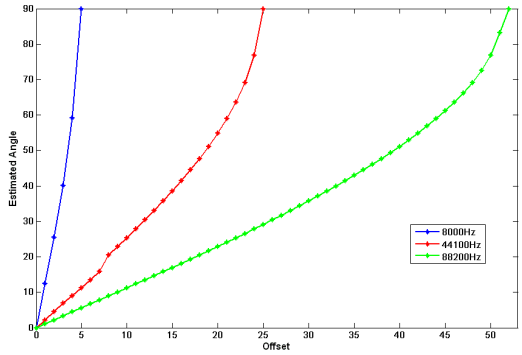

Fig. 4: Range of estimated angles is dependent on the sampling rate used when recording the sound signals.

the interconnecting system of neurons in the auditory pathway and they take individual spikes as input which allows for spatio-temporal information to be included in the computation [17]. The SNNs outlined in this paper use the same waveforms, *g(t)* and *h(t)*, as the cross-correlation algorithm and use a learning algorithm to classify these inputs to angles of location.

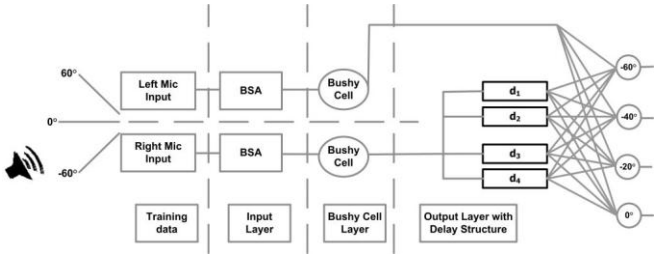

Fig. 5: SNN architecture for angles originating from the range -60 to 0° (left network)

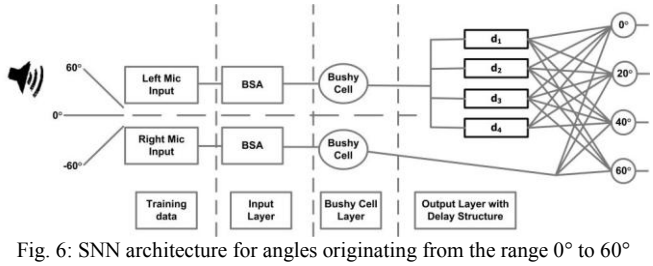

Fig. 6: SNN architecture for angles originating from the range 0° to 60° (right network)

Reflecting the bilateral symmetry of the nervous system, there are two SNNs; the topology of the left network corresponding to angles in the range of -60° to 0° can be seen in Figure 5; and the right network relating to the angles 0° to 60° in Figure 6. For the duration of this section, processing of the left network will be described as both networks have identical processing; they can be considered as mirror images of one another. There is a unique ITD for each angle in the positive or negative range which is dependent on the distance between the two ears/microphones

and the speed of sound [8], i.e. the ITD for each angle ± is the same. This is the reason why angles in the range of -60° to 0° are not processed together with angles in the range of 0° to +60°, as to do so would cause confusion when the SNN is being trained to produce an estimated angle as output. It is also the reason why an output neuron for 0° can be placed in both networks.

The input waveforms, *g(t)* and *h(t)*, from the left and right microphones pass through the input layer which consists of Ben's Spiker Algorithm (BSA); a spike encoding methodology which uses a convolution filter optimised for encoding by a genetic algorithm [18-19]. BSA converts the sound signals into biologically realistic phase-locked spike trains, which are then routed through a bushy cell neuron. This can be seen in Figure 7, where the output of the BSA algorithm and the output of the bushy cell layer for a portion of the waveform *g(t)* are plotted.

Knowledge of the bushy cells in biology is limited, however it is known that the main function of these cells is to maintain the phase-locked signal and to minimise noise. In the network, spike trains such as those in the centre panel of Figure 7 proved difficult to train due to their bursting nature and erroneous spikes which are not phase-locked to the waveform. Therefore, the role of the bushy cell layer in this network is to remove any erroneous spikes in the spike train (i.e. to remove noise), and to transform the phase-locked bursts to single spike instances.

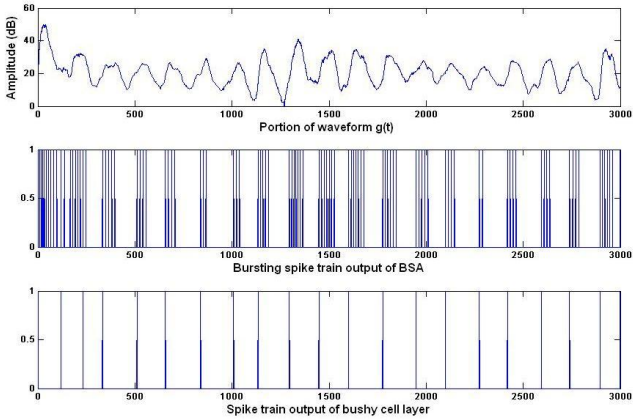

Fig. 7: The waveform *g(t)* as it is encoded into a spike train by the BSA algorithm and then routed through the bushy cell layer

This processing was implemented using a LIF neuron. All LIF neurons in the network are modelled by [20]:

$$\tau_{mem} \frac{dv}{dt} = -v + R_{in} I_{syn}(t) \qquad (5)$$

where τ_{mem} refers to the membrane time constant of the neuron, v is the membrane potential and R_{in} is the membrane resistance, driven by a synaptic current $I_{syn}(t)$. The phase-locked single spike output in place of a burst was achieved through selection of an appropriate neuron threshold and refractory period. The parameters are fixed for every bushy cell in the network, i.e. the same parameters are used for every angle with which the network was trained and tested. It is not the time of the first spike in each resulting spike train that is important when the two spike trains synchronise at the output neurons; each individual spike is necessary for the sound localisation process to be achieved, i.e. the ITD is not extracted just at the onset of the stimulus but across the length of the stimulus. The delay structure causes the two spike trains to become coincident at the output layer, each spike in the left spike train will then be in coincidence with a spike in the right spike train, causing maximum output.

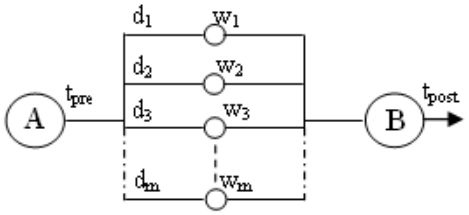

Fig. 8: Pre and postsynaptic neurons with interconnecting delay lines d_1 to d_m, and weights w_1 to w_m, from [21].

Figure 8 shows how the multiple delay structure, similar to the graded series of delays found in the biological MSO, is 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. The output spike from neuron A is passed to i interneuron connecting pathways, each with their own weight w_i, where ($i = 1, m$).

Each delay line is connected to every output neuron, also modelled by LIF neurons, producing sixteen synaptic connections in the output layer; the synaptic weights are identical before training begins. Therefore, every output neuron will receive both in-phase and out-of-phase inputs for every angle. The objective of the trained network is to associate each delay to a particular output neuron. For example, the first delay will become associated with the first output neuron, and so on. To do this the post-trained weight on the connection between the associated delay and output neuron must be larger than the weights on any of the other connections also providing stimulus to that output neuron. Ultimately, the association of particular delay lines within the delay structure to particular output neurons is specified by the training algorithm alone, and is not a matter of network design.

A multiplicative form of Supervised Hebbian Learning (SHL) using STDP windows was employed [22]; the multiplicative form was found to produce more stability during the training period. During training the following behaviour occurs:

1. Determine whether the current output neuron is being supervised or not.
2. If it is supervised, the positive part of the STDP window is used to increase the weights on the synapses between the supervised output neuron and the appropriate delay lines providing the current input.
3. If it is not supervised, the negative part of the STDP window is employed to decrease the weights on the synapses between the non-supervised output neuron and any delay lines providing the current input.

This training algorithm proved successful in producing the desired output of the network, i.e. the appropriate output neuron has the highest firing frequency when its associated input data is routed through the network.

The final weights on the synapses between the delay line connections and each output neuron at the end of training are bimodal, i.e. the delay which is associated with an output neuron after training has the largest weight and the weights associated with any other delays are much lower. This distribution of weights allows the fully connected SNN with a generic delay structure to produce the desired angles with the accuracies outlined in the next section.

III. Results

A. Simulated Data

An initial experiment was carried out to test the cross-correlation algorithm using a set of simulated waveforms from the MIT dataset [29]. These are a set of HRTF measurements from a KEMAR dummy head microphone which represent the left and right ear impulse responses from a Realistic Optimus Pro 7 loudspeaker mounted 1.4 metres from the KEMAR. Impulse responses are available for 710 different azimuthal positions with elevations between -40° and +90°. A left impulse response for -90° can be seen in Figure 9.

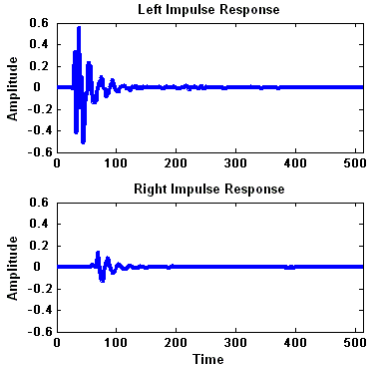

Fig. 9: Left and right impulse responses from a sound source at -90°

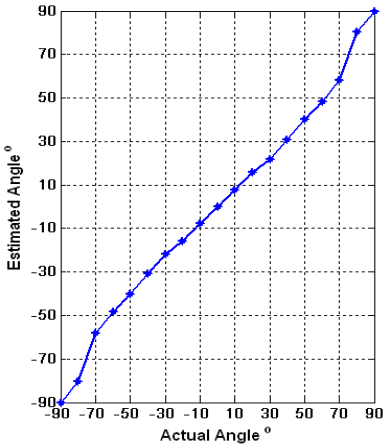

Fig.10: Results of cross-correlation algorithm when presented with HRTF impulse responses convoluted with a waveform

These impulse responses were convoluted with a single channel waveform to produce a set of left and right simulated data in the range of ±90° in steps of 10° at an elevation of 0°. Convolution is a mathematical way of combining two waveforms to generate a third waveform:

$$w(k) = \sum_j u(j)v(k+1-j) \qquad (6)$$

where *w(k)* is the new waveform, *u(j)* is the impulse response and *v(k)* is the original waveform. The results of this experiment can be seen in Figure 10 and show that the cross-correlation algorithm is very successful at producing the azimuthal angle.

B. Experimental data

The next experiments involved utilising the waveforms recorded by the mobile robot in a noisy and dynamic environment. Initially, the cross-correlation algorithm was tested with the entire two second low frequency (400 Hz) waveforms. Ten waveforms for each angle in the range of ±60° in steps of 20° were presented to the algorithm. Table I outlines the resulting mean and standard deviation of the estimated angles. It is clear to see, that this classical methodology is proficient in its ability to perform sound localisation on real sounds, achieving a mean error of ±3.3°.

TABLE I
MEAN AND STANDARD DEVIATION RESULTS OF CROSS-CORRELATION ALGORITHM

Cross-Correlation		
Actual Angle	Mean Estimated Angle	Standard Deviation
-60°	-58.95°	±0.00°
-40°	-47.07°	±0.57°
-20°	-20.52°	±0.53°
0°	+1.34°	±0.47°
+20°	+17.14°	±0.38°
+40°	+35.11°	±0.86°
+60°	+54.54°	±1.41°

For computational efficiency, a sample of each two second low frequency waveform was presented to the SNN. This sampling related to approximately 70 ms of the sound stimulus; the sample was taken from the half way point of each waveform. Again, ten waveforms for each angle in the range of ±60° in steps of 20° were presented. In order to directly compare the two methodologies, the ten waveforms for each angle were also presented to the cross-correlation algorithm. Table II outlines the resulting mean and standard deviation of the estimated angles for both methodologies.

TABLE II
MEAN AND STANDARD DEVIATION RESULTS OF CROSS-CORRELATION ALGORITHM AND SNN

	Cross-Correlation		Spiking Neural Network	
Actual Angle	Mean Estimated Angle	Standard Deviation	Mean Estimated Angle	Standard Deviation
-60°	-67.01°	±15.88	-59°	±3.06°
-40°	-55.32°	±18.33°	-43°	±17.67°
-20°	-20.57°	±2.06°	-23°	±6.74°
0°	+1.04°	±1.06°	0°	0°
+20°	+17.69°	±0.84°	+20	0°
+40°	+37.15°	±1.92°	+60°	0°
+60°	+55.82°	±1.95°	+60°	0°

TABLE III
COMPARISON OF ROBOTIC SOUND LOCALISATION SYSTEMS AS REPORTED IN THE RESEARCH LITERATURE, THE RESULTS PRESENTED IN THIS PAPER ARE INDICATED IN THE FINAL ROW

No. of Microphones	Cues	Angular Sweep	Angular Resolution	Method to Determine Direction of Sound Source	Localisation Accuracy	Ref.
2	ITD, IID, Spectral Notches Vision	N/A	132 positions	Broyden update rule & Visual Servoing Loop	Error < 0.1 radians = ~5°	[23]
2	ITD, IID, IPD Relative IID	±80	10°	Parameter-Less Self-Organizing Map with Reinforcement Learning	10°	[24]
2	ITD, IID	0° to 360°	5°	Fuzzy Neural Network	Mean rms azimuth error: ±5.97° Mean rms elevation error: ±3.9°	[25]
12	TDOA	~360°	N/A	Genetic Algorithm	Mean azimuth error: 2.9° Mean elevation error: 1°	[26]
2	ITD, IID	±90°	30°	Spiking Neural Network with Conditional Probability	N/A	[27]
2	ITD	±90°	5°	Recurrent Neural Networks with Back Propagation	Error between ±1.5° and ±7.5°	[28]
2	ITD	±60°	20°	**Full stimulus:** Cross Correlation **Sampled stimulus:** Cross Correlation Spiking Neural Networks	 ±3.3° ±4.7° ±3.8°	-

The results show a mean error of ±4.75° for the cross-correlation algorithm and ±3.8° for the SNN. When presented with a limited sample of each waveform (70 ms), the SNN performs with a higher accuracy. Furthermore, considering that all ten samples for +40° classified incorrectly for the SNN, the fact that the mean error is still low suggests that the SNN methodology has considerable potential for performing both accurate and biologically inspired sound localisation. Establishing that a biologically inspired SNN can not only compare favourably but improve on the classical methodology of cross-correlation for limited samples of the sound stimulus is important for the overall aim of this research, i.e. incorporating a human-like auditory system within a mobile robot.

In order to carry out this research, an important step was to know and understand the state of the art. With the knowledge of these existing techniques, it was possible to formulate ideas which would advance said techniques in a more biologically inspired way. Table III outlines and compares these techniques. It is difficult to make direct comparisons between the results achieved in this research and the work outlined by other researchers. This is due to the many different methods that are used for sound localisation modelling. However, based on the combination of the classification results, the biologically inspired SNN based architecture and the use of a learning algorithm; we feel that this research advances the work in this field.

IV. Conclusion

This paper presents a comparison between two methodologies for sound localisation; a classical cross-correlation method and a biologically inspired SNN. Real sounds recorded by a mobile robot in a noisy and dynamic environment are used as input to both models. The cross-correlation method produces satisfactory results with a mean error of ±3.3° when presented with the entire waveform. However, when a limited sample of the stimulus was presented to both the cross-correlation algorithm and the SNN, mean errors of ±4.75° and ±3.8° were achieved respectively. This suggests that the SNN is more adept at producing accurate angles of location with a more limited input stimulus. Furthermore, the SNN technique also compares favourably to the state of the art.

Future work is planned which will involve extending the SNN to take different types of sound as input, from pure tones of different frequencies to complex sounds such as speech. A control system for the mobile robot based on the azimuthal outputs of the SNN will be developed. Additionally, issues associated with the encoding of spike trains and the noise reduction of the original waveforms will also be investigated. Future experiments will also outline how the SNN accuracy is affected by different environments; and the comparison between the cross-correlation algorithm and the SNN will be discussed against many more factors rather than just performance accuracy. Such factors will include reverberation levels, robustness to noise and the computational complexity of the two methodologies.

V. References

[1] D. McAlpine and B. Grothe, "Sound localization and delay lines - do mammals fit the model?" *Trends Neurosci*, vol. 26, no. 7, pp. 347-350, 2003.

[2] T. C. T. Yin, *Integrative Functions in the Mammalian Auditory Pathway*. Springer-Verlag, 2002, ch. Neural mechanisms of encoding binaural localization cues in the auditory brainstem, pp. 99-159.

[3] D. J. Tollin, "The lateral superior olive: A functional role in sound source localization," *Neuroscientist*, vol. 9, no. 2, pp. 127-143, 2003.

[4] S. P. Thompson "On the function of the two ears in the perception of space," *Philos Mag*, vol. 13, no. 83, pp. 406-416, 1882.

[5] L. Rayleigh, "On our perception of sound direction," *Philos Mag*, vol. 13, no. 74, pp. 214-232, 1907.

[6] M. S. Lewicki. (2006) Sound localization 1. [Online]. Available: http://www.cs.cmu.edu/~lewicki/cpsa/sound-localization1.pdf

[7] B. Grothe, "New roles for synaptic inhibition in sound localization," *Nature Rev Neurosci*, vol. 4, no. 7, pp. 540-550, 2003.

[8] R. M. Burger and E. W. Rubel, "Encoding of interaural timing for binaural hearing," *The Senses: A Comprehensive Reference*, vol. 3, pp. 613-630, 2008.

[9] P. X. Joris and T. C. T. Yin, "A matter of time: Internal delays in binaural processing," *Trends Neurosci*, vol. 30, no. 2, pp. 70-78, 2007.

[10] L. A. Jeffress, "A place theory of sound localization," *J. Comparative Physiological Psychology*, vol. 41, no. 1, pp. 35-39, 1948.

[11] J. K. Moore, "Organization of the human superior olivary complex," *Microsc Res Tech*, vol. 51, no. 4, pp. 403-412, 2000.

[12] D. C. Fitzpatrick, S. Kuwada and R. Batra, "Transformations in processing interaural time differences between the superior olivary complex and inferior colliculus: beyond the Jeffress model," *Hearing Research*, vol. 168, no. 1-2, pp. 79-89, 2002.

[13] I. Bazwinsky, H. Hilbig, H. J. Bidmon and R. Ruebsamen, "Characterization of the human superior olivary complex by calcium binding proteins and neurofilament H (SMI-32)," *J. Comparative Neurology*, vol. 456, no. 3, pp. 292-303, 2003.

[14] R. J. Kulesza, "Cytoarchitecture of the human superior olivary complex: Medial and lateral superior olive," *Hearing Research*, vol. 225, no. 1-2, pp. 80-90, 2007.

[15] J. A. Wall, L. J. McDaid, L. P. Maguire and T. M. McGinnity, "Spiking neuron models of the medial and lateral superior olive for sound localisation," in *IEEE Int. Joint Conf. Neural Networks (IJCNN) (IEEE World Congr. Computational Intelligence)*, 2008, pp. 2641-2647.

[16] B. Glackin, J. A. Wall, T. M. McGinnity, L. P. Maguire and L. J. McDaid, "A spiking neural network model of the medial superior olive using spike timing dependent plasticity for sound localisation," *Front. Comput. Neurosci*, vol. 4, pp. 1-16, 2010.

[17] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659-1671, 1997.

[18] H. de Garis, N. E. Nawa, M. Hough and M. Korkin, "Evolving an optimal de/convolution function for the neural net modules of ATR's artificial brain project," in *Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN)*, 1999, vol. 1, pp. 438-443.

[19] B. Schrauwen and J. Van Campenhout, "BSA, a fast and accurate spike train encoding scheme," in *Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN)*, 2003, vol. 4, pp. 2825-2830.

[20] W. Gerstner and W. M. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press, 2002.

[21] S. M. Bohte, J. N. Kok and H. La Poutre, "Spike-prop: Error-backpropagation for networks of spiking neurons," in *Proc. European Symp. Artificial Neural Networks (ESANN)*, 2000.

[22] R. Legenstein, C. Naeger and W. Maass, "What can a neuron learn with spike-timing-dependent plasticity?" *Neural Computation*, vol. 17, no. 11, pp. 2337-2382, 2005.

[23] J. Hörnstein, M. Lopes, J. Santos-Victor and F. Lacerda, "Sound localization for humanoid robots – building audio-motor maps based on the HRTF," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2006, pp. 1170-1176.

[24] E. Berglund and J. Sitte, "Sound source localisation through active audition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2005, pp. 653-658.

[25] F. Keyrouz and K. Diepold, "A novel biologically inspired neural network solution for robotic 3D sound source sensing," *Soft Comput*., vol. 12, no. 7, pp. 721-729, 2008.

[26] N. M. Kwok, J. Buchholz, G. Fang and J. Gal, "Sound source localization: Microphone array design and evolutionary estimation," in *Proc. IEEE Int. Conf. Ind. Technology (ICIT)*, 2006, pp. 281-286.

[27] J. Liu, D. Perez-Gonzalez, A. Rees, H. Erwin and S. Wermter, "A biomimetic spiking neural network of the auditory midbrain for mobile robot sound localisation in reverberant environments," in *Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN)*, 2009, pp. 1855-1862.

[28] J. C. Murray, H. R. Erwin and S. Wermter, "Robotic sound-source localisation architecture using cross-correlation and recurrent neural networks," *Neural Networks*, vol. 22, no. 2, pp. 173-189, 2009.

[29] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *J Acoust Soc Am*, vol. 97, pp. 3907-3908, 1995.