Using the Interaural Time Difference and Cross-Correlation to Localise Short-Term Complex Noises

Julie Wall[1] , Martin McGinnity[1] and Liam Maguire[1]

[1] Intelligent Systems Research Centre, University of Ulster, Derry, Northern Ireland
{j.wall, lp.maguire, tm.mcginnity}@ulster.ac.uk

Abstract. The mammalian binaural cue of interaural time difference (ITD) and cross-correlation have long been used to determine the point of origin of a sound source. The ITD can be defined as the different points in time at which a sound from a single location arrives at each individual ear [1]. From this time difference, the brain can calculate the angle of the sound source in relation to the head [2]. Cross-correlation compares the similarity of each channel of a binaural waveform producing the time lag or offset required for both channels to be in phase with one another. This offset corresponds to the maximum value produced by the cross-correlation function and can be used to determine the ITD and thus the azimuthal angle θ of the original sound source. However, in indoor environments, cross-correlation has been known to have problems with both sound reflections and reverberations. Additionally, cross-correlation has difficulties with localising short-term complex noises when they occur during a longer duration waveform, i.e. in the presence of background noise. The cross-correlation algorithm processes the entire waveform and the short-term complex noise can be ignored. This paper presents a technique using thresholding which enables higher-localisation abilities for short-term complex sounds in the midst of background noise. To determine the success of this thresholding technique, twenty-five sounds were recorded in a dynamic and echoic environment. The twenty-five sounds consist of hand-claps, finger-clicks and speech. The proposed technique was compared to the regular cross-correlation function for the same waveforms, and an average of the azimuthal angles determined for each individual sample. The sound localisation ability for all twenty-five sound samples is as follows: average of the sampled angles using cross-correlation: 44%; cross-correlation technique with thresholding: 84%. From these results, it is clear that this proposed technique is very successful for the localisation of short-term complex sounds in the midst of background noise and in a dynamic and echoic indoor environment.

Keywords: Sound Localisation, Interaural Time Difference, Cross-Correlation.

References

1. Lewicki, M.S.: Sound Localization 1, http://www.cs.cmu.edu/ lewicki/cpsa
2. Grothe, B.: New roles for synaptic inhibition in sound localization. Nature Reviews Neuroscience. 4, 540-550 (2003)

Intelligent Systems Research Centre

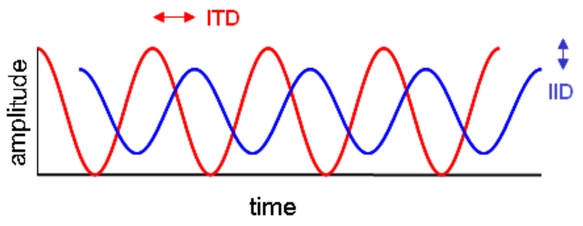
University of Ulster
Magee Campus

Using the Interaural Time Difference and Cross-Correlation to Localise Short-Term Complex Noises

Julie Wall, Martin McGinnity and Liam Maguire
Intelligent Systems Research Centre (ISRC), University of Ulster, Derry

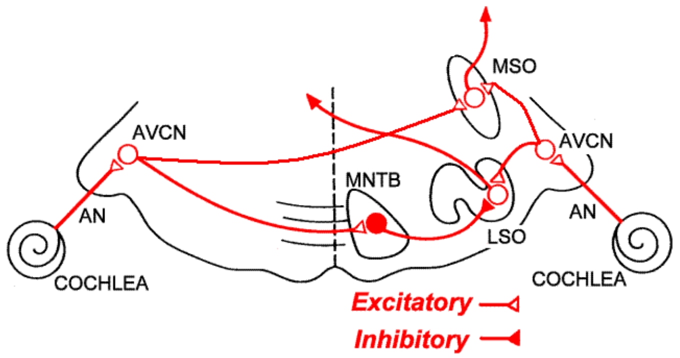
Biological Inspiration

- Sound localisation can be defined as determining where a sound signal is generated in relation to the position of the human head.
- Uses binaural cues of interaural time difference (ITD) and interaural intensity difference (IID).



Time for ITD and Amplitude for IID

- ITD is the time difference between the arrival time of a sound to the two ears. ITD is calculated in the medial superior olive of the auditory system and works on low frequency sound signals.
- IID is the difference in sound pressure levels of the sound signal between each ear. It is computed in the lateral superior olive (LSO) of the auditory system and works on high frequency sound signals.



ITD and IID pathway of the biological auditory system

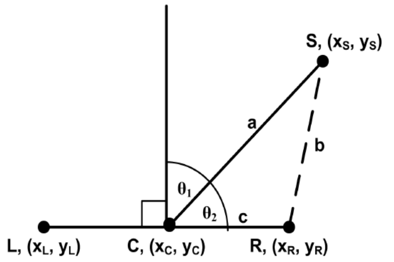
General Research Aim

- This research builds on earlier work in biologically inspired sound localisation (Glackin et al. 2010), (Wall et al. 2010), (Wall et al. 2011).
- Spiking neural network model of the auditory pathways which operates on the output signals of the cochlea to achieve sound localisation.
- Input consisted of experimentally derived HRTF data from an adult domestic cat and a biologically plausible learning algorithm was used to classify the HRTF data to azimuthal angles.
- Extend this research to the area of mobile robotics.
- Provides ideal platform for the development of a human-like auditory system which can operate in a dynamic and noisy environment.

Robotic Framework

- Pioneer 3-DX mobile robot with a pair of stereo omni-directional microphones placed 30cm apart.
- Vicon motion tracking system modelled both the robot and the sound source, 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.

Pioneer 3-DX



Inverse of the Cosine rule determines angle θ2 which enables you to determine angle θ1; L and R are the positions of the left and right microphones; C is the centre point between the two microphones and S is the sound source.

Sound Localisation by Cross-Correlation

1. Cross-correlate left and right waveforms, *g(t)* and *h(t)*, to produce offset σ:

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

2. Determine sampling frequency Δt of waveforms:

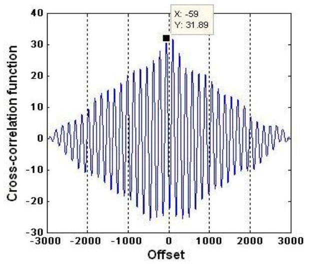
$$\Delta t = \frac{1}{f}$$

3. Calculate the ITD:

$$ITD = \Delta t * \sigma$$

4. Thus, the azimuthal angle is:

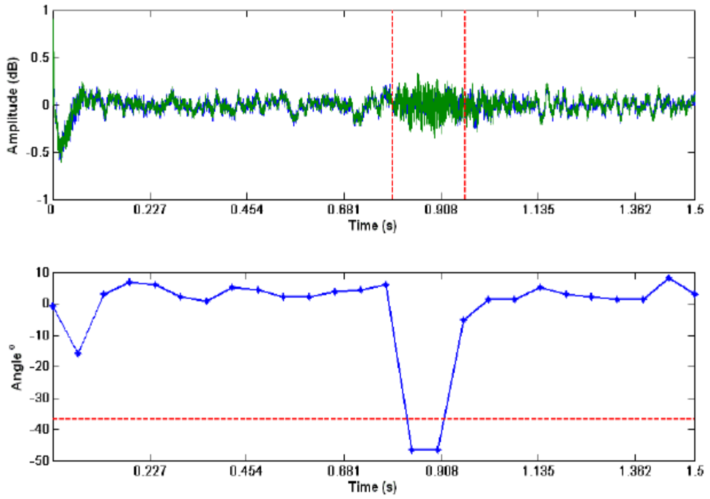
$$\theta = \sin^{-1} \frac{C_{air} * ITD}{c}$$



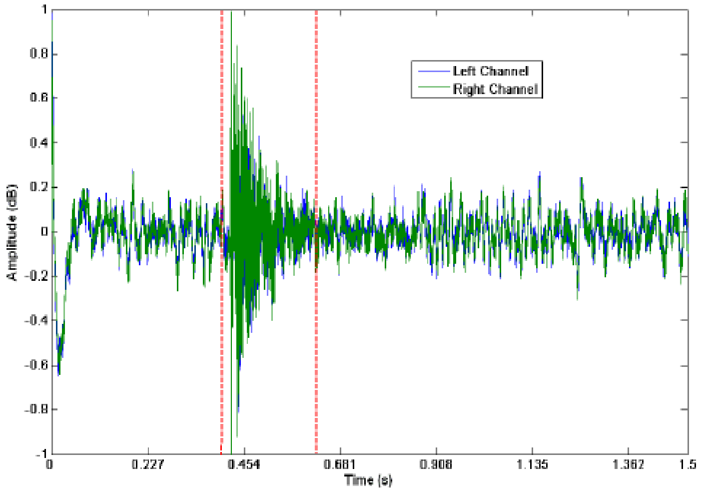
Cross-correlation function for -40° sound source, showing offset σ of 59 samples.

Proposed Thresholding Technique

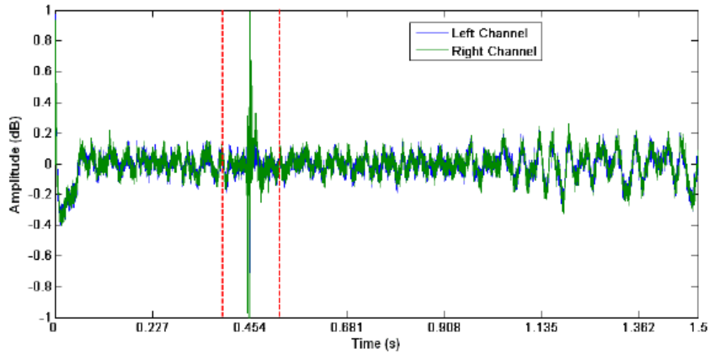
- Cross-correlation has difficulties with localising short-term complex noises when they occur during a longer duration waveform.
- Cross-correlation processes entire waveform and the short-term complex noise can be ignored.
- Thresholding enables greater localisation abilities for short-term complex sounds in the midst of background noise.
- Considered a two-pass cross-correlation strategy.
- Divide waveform into smaller samples, ~50 ms each.
- Perform cross-correlation on every individual sample.
- Threshold is automatically generated from the range of azimuthal angles generated.
- Threshold extracts the portion of the sound signal which is important and removes the background noise.



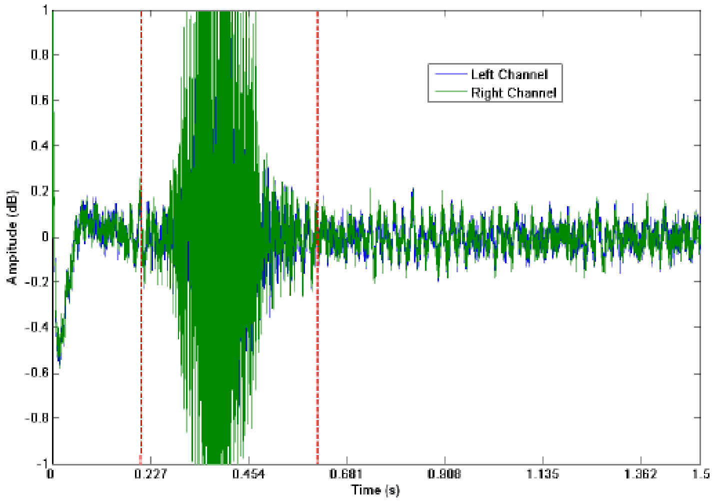
Original waveform and range of azimuthal angle outputs from the cross-correlation of each sample



Thresholding technique extracting **“clap”**



Thresholding technique extracting **“finger click”**



Thresholding technique extracting **“speech sample”**

Initial Results

- Twenty-five sounds recorded in a dynamic and echoic environment.
- Hand-claps, fingers-clicks and speech (isolated words)
- Average sound localisation accuracy
  - **Regular cross-correlation: 44%**
  - **Cross-correlation with Thresholding: 84%**
- Proposed technique very successful for localisation of short-term complex sounds in the midst of background noise and in a dynamic and echoic indoor environment.

References

- Glackin, B., Wall, J. A., McGinnity, T.M., Maguire, L. P. & McDaid, L. J. "A Spiking Neural Network Model of the Medial Superior Olive using Spike Timing Dependent Plasticity for Sound Localisation", Frontiers in Computational Neuroscience, vol. 4, pp. 1-16, 2010.
- Wall, J. A., McDaid, L. J., Maguire, L. P. & McGinnity, T.M. "Spiking neural network model of sound localisation using the interaural intensity difference", IEEE Transactions on Neural Networks. (Submitted 2010, Accepted with Minor Revision).
- Wall, J. A., McGinnity, T. M. & Maguire, L. P., “A comparison of sound localisation techniques using cross-correlation and spiking neural networks for mobile robotics,” IEEE International Joint Conference on Neural Networks (IJCNN), 2011.