Work-in-Progress: A Neuromorphic Approach of the Sound Source Localization Task in Real-Time Embedded Systems

Daniel Gutierrez-Galan
dgutierrez@atc.us.es
University of Seville
Seville, Spain

Juan Pedro Dominguez-Morales
University of Seville
Seville, Spain

Angel Jimenez-Fernandez
University of Seville
Seville, Spain

Ricardo Tapiador-Morales
University of Seville
Seville, Spain

Antonio Rios-Navarro
University of Seville
Seville, Spain

Alejandro Linares-Barranco
University of Seville
Seville, Spain

ABSTRACT

Autonomous robots have become a very popular topic within the artificial intelligence field. These systems are able to perform difficult or risky tasks that could be dangerous when done by humans or trained animals. Vision is commonly considered the most relevant input sensor for autonomous robots and tracking systems. However, auditory information is also important in some specific situations where vision cannot provide any useful information when navigating. In this work, a spike-based model of the medial superior olive of the inner ear has been implemented in reconfigurable hardware for performing sound source localization in real time. Future works will focus on integrating this information with vision in order to achieve a fully bio-inspired autonomous tracking system.

CCS CONCEPTS

• **Applied computing → Event-driven architectures**.

KEYWORDS

Neuromorphic auditory sensor, FPGA, binaural sound localization, real-time audio processing, spike-based system

ACM Reference Format:
Daniel Gutierrez-Galan, Juan Pedro Dominguez-Morales, Angel Jimenez-Fernandez, Ricardo Tapiador-Morales, Antonio Rios-Navarro, and Alejandro Linares-Barranco. 2019. Work-in-Progress: A Neuromorphic Approach of the Sound Source Localization Task in Real-Time Embedded Systems. In *2019 International Conference on Embedded Software Companion (EMSOFT'19 Companion), October 13–18, 2019, New York, NY, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3349568.3351544

1 INTRODUCTION

Robotics have become a highly active research topic within artificial intelligence. This, combined with the new trends on autonomous systems, have made this field very relevant for carrying out tasks that are difficult or risky to perform by humans and trained animals. Autonomous systems are guided by a set of sensors in which the most important one is vision, since it provides useful and sufficient information for navigating under standard conditions. However, when tracking targets based on vision, the object that the system is following could go outside of the field of view, leaving the system with no clue on where the target is. Animals solve this problem by integrating the information from both the vision and the hearing. An audio sensor attached to the robot could make it head to the direction where the target is, using then the vision to move towards and avoid obstacles. State-of-the-art studies have experimented with using bio-inspired auditory cues based on the interaural level difference (ILD) and the interaural time difference (ITD) to perform the sound localization task [4]. Researchers have tried to use current state-of-the-art neuromorphic platforms as SpiNNaker in the past to implement both ITD and ILD with Spiking Neural Networks (SNNs) in hardware. These experiments have not been fruitful since the minimum timestep that SpiNNaker uses to simulate the network is 1ms, which is not precise enough for ITD. For this purpose, in this work we present a spike-based implementation of ITD based on the Jeffress model [1] in an Field-Programmable Gate Array (FPGA) that directly takes the output spikes from a Neuromorphic Auditory Sensor (NAS) [3] as input.

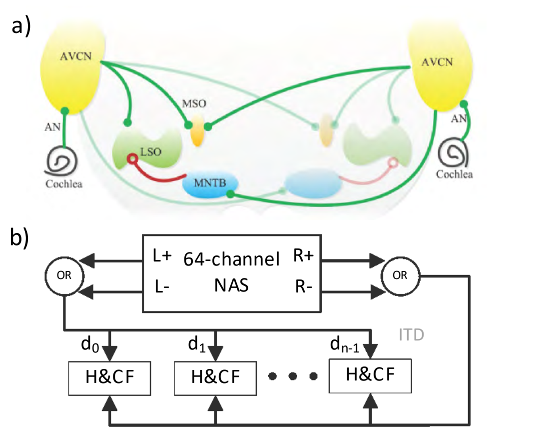

Figure 1: a) SOC biological model (from [4]) b) Neuromorphic approach of the Jeffress model.

EMSOFT'19 Companion, October 13–18, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6924-4/19/10...$15.00
https://doi.org/10.1145/3349568.3351544

2 AUDITORY CUES: A NEUROMORPHIC APPROACH

2.1 Bio-inspired approach

The sound source localization in mammals is based in two physical phenomena of the sound. First, the position of a sound source can be estimated by computing the intensity differences between the incoming sound waves that arrive to each ear. This cue is known as ILD, and it achieves better results for high frequency sounds. The localization can be also determined by detecting the phase or time difference between the sound waves received at both ears. This cue is known as ITD. In this case, the more lateral the source is placed, the higher the time difference is. This mechanism performs better with low frequency sounds. Regarding Fig1a, in the human brain the Medial Superior Olive (MSO) has excitatory projections from both the ipsilateral and the contralateral cochlear nucleus. A model proposed in [1] defines the MSO as a coincidence detector array which carries out a spike-based correlation operation.

2.2 Spike-based models

Instead of implementing mathematical algorithms to solve this task, we proposed a spike-based model of the MSO. The advantage of working in the spikes domain instead of the time domain lies on the simplification of the mathematical operations [2], and thus, the speed when processing the information. Furthermore, the time constraints are not a problem in the spike-based processing blocks, since time bins are not needed. For implementing the model proposed in [1], the Hold & Coincidence Fire (H&CF) block was used. The inhibitory input was converted to excitatory, and an output spike is fired when two excitatory input spikes arrive during the detection period. In order to have different delay lines, they were implemented as independent timers with variable resolution. The result is an array of spike-based coincidence detectors as it is showed in Fig 1b. In this work, sixteen H&CF blocks were used.

3 MODEL VERIFICATION

Some preliminary experiments were performed in order to verify the behavior of the proposed model. A testbench was implemented for the behavioral verification of the coincidence detectors. Virtual input stimuli were generated with an incremental time difference of 10 µs from zero to 700 µs, which is the maximum time difference that humans are able to detect [1]. Results are shown in Fig2. Then,

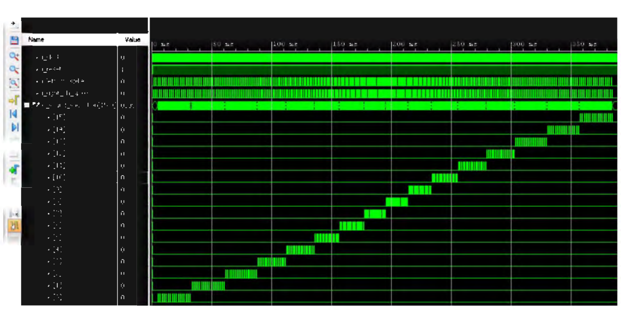

Figure 2: Results from the behavioral simulation using a simulated sound sweep as input.

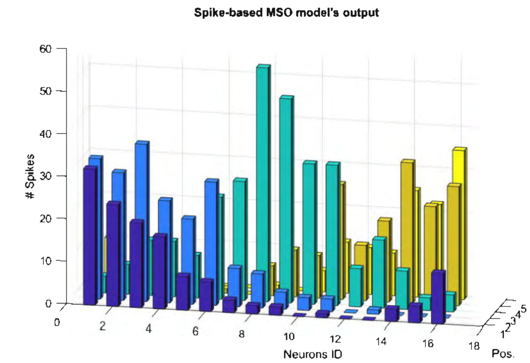

Figure 3: Results of the real-time experiment. Numbers 1 to 5 correspond to the positions -90, -45, 0, 45 and 90 degrees.

we created a test scenario in which a receptor was placed in a virtual room, and five sound sources playing a 500 Hz pure tone were placed equidistant to the receptor, with a 2 meters distance in a semicircular shape, generating five sound files. We used a 64-channels NAS [3] which was modified by adding the proposed MSO models. This sensor is a spike-based digital sensor that can be implemented in an FPGA. Hence, we used this hardware setup for a real-time test. The auditory information was sent in real time through a line-in connector to the NAS, and the output of the MSO was recorded using jAER. This test configuration allowed us to check the model response to the same input stimulus from different positions. Results obtained from the recordings are shown in Fig 3.

4 CONCLUSIONS AND FUTURE WORKS

In this work we have presented a bio-inspired digital implementation of the Jeffress model in FPGA. Two test scenarios were proposed in order to verify the expected behavior of the model. In future works, the authors will integrate this spike-based sound localization model together with a neuromorphic silicon retina in order to make a robot able to navigate and track objects autonomously. Further research will focus on studying the feasibility of these bio-inspired implementations for performing more difficult or risky tasks like rescue tasks in dangerous environments.

ACKNOWLEDGMENTS

This work was supported by the Spanish grant (with support from the European Regional Development Fund) COFNET (TEC2016-77785-P). The work of D. Gutierrez-Galan was supported by a "Formación de Personal Investigador" Scholarship from the Spanish Ministry of Education, Culture and Sport.

REFERENCES

[1] Lloyd A Jeffress. 1948. A place theory of sound localization. *Journal of comparative and physiological psychology* 41, 1 (1948), 35.

[2] Angel Jimenez-Fernandez et al. 2010. Building blocks for spikes signals processing. In *Neural Networks (IJCNN), The 2010 International Joint Conference on.* IEEE, 1–8.

[3] Angel Jiménez-Fernández et al. 2017. A Binaural Neuromorphic Auditory Sensor for FPGA: A Spike Signal Processing Approach. *IEEE Trans. Neural Netw. Learning Syst.* 28, 4 (2017), 804–818.

[4] Qian Liu et al. 2013. Modeling populations of spiking neurons for fine timing sound localization. In *The 2013 International Joint Conference on Neural Networks (IJCNN).* IEEE, 1–8.