Neurocomputing 000 (2018) 1–11

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Real-time neuro-inspired sound source localization and tracking architecture applied to a robotic platform

Elena Cerezuela Escudero[a,*], Fernando Pérez Peña[b], Rafael Paz Vicente[a], Angel Jimenez-Fernandez[a], Gabriel Jimenez Moreno[a], Arturo Morgado-Estevez[b]

[a] Robotics and Computer Technology Lab (RTC), Universidad de Sevilla, ETSI Informática, Avd. Reina Mercedes s/n, Seville 41012, Spain
[b] Applied Robotics Research Lab, Universidad de Cádiz, Faculty of Engineering, Avda. Universidad, 10, Puerto Real, Cadiz 11519, Spain

ARTICLE INFO

Article history:
Received 3 April 2017
Revised 8 November 2017
Accepted 19 December 2017
Available online xxx

Communicated by Wei Wu

Keywords:
Sound localization
Interaural intensity difference
Spike signal processing
Neuromorphic auditory sensor
Neurorobotics
FPGA

ABSTRACT

This paper proposes a real-time sound source localization and tracking architecture based on the ability of the mammalian auditory system using the interaural intensity difference. We used an innovative binaural Neuromorphic Auditory Sensor to obtain spike rates similar to those generated by the inner hair cells of the human auditory system. The design of the component that obtains the interaural intensity difference is inspired by the lateral superior olive. The spike stream that represents the IID is used to turn a robotic platform towards the sound source direction. The architecture was implemented on FPGA devices using general purpose FPGA resources and was tested with pure tones (1-kHz, 2.5-kHz and 5-kHz sounds) with an average error of 2.32°. Our architecture demonstrates a potential practical application of sound localization for robots, and can be used to test paradigms for sound localization in the mammalian brain.

1. Introduction

Sound localization is a function that the ears, auditory pathways and auditory cortex of the brain perform together to determine the source of a sound. It is a powerful feature of mammalian perception that allows the animal to be aware of the environment and to locate prey and predators. This has inspired researchers to develop new computational models of the auditory pathways and biological mechanisms that underlie sound localization in the brain.

The ability to model the ways in which mammals locate a sound source can improve the perception and navigation of mobile robots, allow the development of better virtual realities, improve teleconferencing, provide surveillance systems with omnidirectional sensitivity, and enhance hearing aids.

During the last decades, the structure and function of pathways in the auditory brainstem for sound localization have been extensively studied [1–4]. The direction of a sound in the horizontal plane is determined by a combination of binaural cues derived from the incident acoustic waves arriving at the ear from different angles: interaural time difference (ITD) and interaural intensity, or level, difference (IID or ILD, respectively). Sounds that do not generate directly in front of or behind the receptor arrive earlier at one ear than at the other, creating an ITD. For wavelengths roughly equal to, or shorter than, the diameter of the head, a shadowing effect is produced at the ear that is further away from the source, creating an IID [1,2].

For example, in general terms, if a pure tone sound source is positioned on the left side, the sound signal at the left ear is represented by the equation:

$$Left_{signal} = a \times \sin(2\pi ft) \tag{1}$$

where a is the sound amplitude, f the sound frequency and t the time. The sound at the right ear is represented by the equation:

$$Right_{signal} = (a/\Delta a) \times \sin(2\pi f(t - \Delta t)) \tag{2}$$

where Δa and Δt are related to, respectively, the intensity difference (IID) caused by the shadowing effect of the head, and the additional time (ITD) required for the sound wave to travel the further distance to the right ear.

Due to the head size, the ITD cue in humans is effective for locating low frequency sounds (20 Hz - 1 kHz). However, the information it provides becomes ambiguous for frequencies above 1 kHz. In contrast, the IID cue is not useful for locating sounds below 1 kHz, but it is more efficient than the ITD cue for mid-and high-frequency (<1 kHz) sound localization.

* Corresponding author.
E-mail addresses: ecerezuela@us.es, ecerezuela@atc.us.es (E.C. Escudero).

https://doi.org/10.1016/j.neucom.2017.12.041
0925-2312/

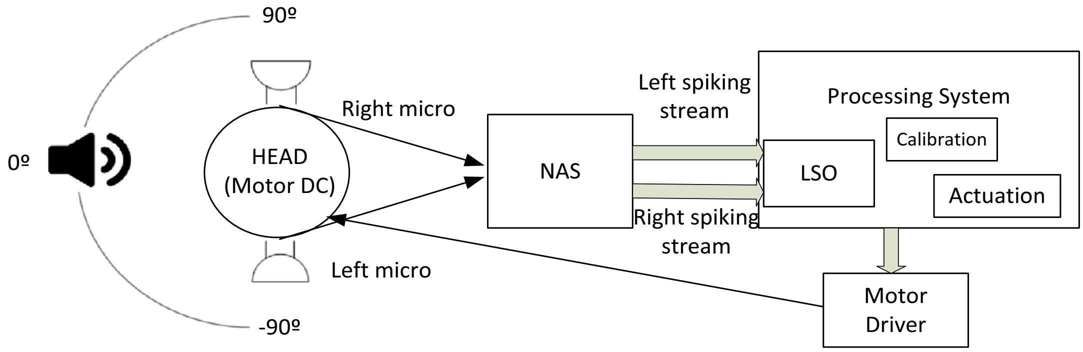

Fig. 1. Tracking Sound system architecture.

The ITD and IID cues are extracted in the medial and lateral superior olive, respectively (MSO and LSO) which are located within an area of the auditory system called the superior olivary complex [2]. The LSO has a tonotopical organization: high frequencies are represented in the middle of the LSO and continually decreasing frequencies to both sides [5]. LSO neurons are inhibited by sounds to the contralateral ear and excited by sounds to the ipsilateral ear, resulting in a neural form of subtraction [2].

There is an increasing demand for the development of real-time and low-power sound localization techniques in the industry of hearing aid [6–8] and robotic applications [9–11]. Currently, various digital processing techniques based on Fast Fourier Transform (FFT) have been proposed to determine the source of a sound signal [12,13]. However, these techniques require high power consuming devices, such as Digital Signal Processors (DSP), and memory to perform such complex signal processing. Traditional Digital Signal Processing (DSP) techniques commonly apply Multiply-Accumulate (MAC) operations over a collection of discrete samples codified as fixed or floating point representations. MAC operations often require dedicated and complex resources, i.e. float-point multipliers, which are available in FPGAs as dedicated expensive resources in relatively small quantities. Therefore, applying a sequence of MAC operations over a dataset with these units requires multiplexing them in time. So they are reused with different input data and output results, which are stored in a global memory. It often requires high frequency clock signals to achieve a competitive data throughput. Furthermore, large memory depths to store intermediate data and results are needed. These facts are reflected in the power consumption and circuitry complexity. On the other hand, Spike Signal Processing (SSP) implements the basic operations that commonly are performed in DSP, but over spike rate coded signals [14]. Thus, operations are performed directly over spike streams, being equivalent to simply adding or removing spikes at the right moment (although it is not evident which). The circuits that implement SSP operations use general purpose FPGA resources, as counters, comparators and logic gates. This allows the building of large scale dedicated systems in hardware, which process spike coded signals in real time using low frequency clocks in a fully parallel way for (low cost) FPGAs, for example, the auditory sensor used in this work demands 29.7 mW for 64 channels in stereo operation [15].

The ability to replicate the ways in which mammals locate a sound source could allow the development of better virtual realities. In addition, the performance of robotics with lower power consumption will be increased. Furthermore, hearing aids could be enhanced by improving the localization of individual sounds. These improvements, which are enabled by the ability to understand and mimic mammalian sound localization, are the main reasons for the research carried out in this paper.

The aim of this research involves the development of a spike-based system that processes and extracts the binaural cue of IID with a topology inspired by the mammalian auditory pathways, specifically the LSO. Using the IID cue, the system performs the task of tracking a sound in real time, in a biologically inspired way.

In this paper, to obtain spike rates similar to those generated by the inner hair cells of the human auditory system, we used an innovative binaural Neuromorphic Auditory Sensor (NAS). This decomposes an audio signal into different frequency bands where the audio information is encoded in the spike rates [15]. Using the out coming spike rates from the NAS as the stimulus to the LSO model we propose, the whole architecture deals with biologically inspired data. The NAS, the LSO model and the actuation system have been implemented on FPGA devices using general purpose FPGA resources. These models are developed using SSP techniques [14].

There are previous works that propose audio localization systems inspired by the mammalian auditory system: the papers by [8] and [11] reported that a neuromorphic silicon cochlea can be used for spatial audition and auditory scene analysis; both papers were based on ITD. In [8], the sound localization circuit was devised by mimicking the neuronal organization of barn owl's auditory pathway to obtain ITD.

The works of [16] and [17] proposed a Spiking Neural Network (SNN) to partially simulate the superior olivary complex, but they did not use a neuromorphic device to obtain the spike streams that represent sound. In the system proposed in [16], the input sound passes through a Gammatone filterbank and is then encoded into phase-locked spikes using a model of the half-wave rectified receptor potential of inner hair cells. ITD processing uses a series of delays and a leaky integrate-and-fire neuron model; the ITD is calculated for all frequency channels to form a full map of ITD processing. IID processing does not use a neuron model; instead, a logarithmic ratio computes the intensity difference. The model classifies the sound source between 7 discrete azimuthal angles (from -90° to 90° in steps of 30°). The model was tested using a robotic head on broadband sounds, both noise and speech, and it achieved overall localization accuracies of 80%. The paper by [17] presented a SNN architecture to simulate the sound localization ability of the mammalian auditory pathways using the IID. To train and validate the localization ability of the architecture, experimentally derived head-related transfer function acoustical data from adult domestic cats were employed; the supervised learning algorithm known as "remote supervision method" was used for the training to determine the azimuthal angles. The SNN classified the sound source between 13 discrete azimuthal angles (from -60° to 60° in steps of 10°). The experimental results using the same sound frequency used for the training were 52% for 5-kHz sounds, 83% for 15-kHz sounds and 40% for 25-kHz sounds. Reference [18]