Neurocomputing 74 (2010) 129–139

Contents lists available at ScienceDirect

Neurocomputing

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

A biologically inspired spiking neural network model of the auditory midbrain for sound source localisation

Jindong Liu [a,*], David Perez-Gonzalez [b], Adrian Rees [b], Harry Erwin [a], Stefan Wermter [c]

[a] School of Computing and Technology, University of Sunderland, Sunderland SR6 0DD, UK
[b] Institute of Neuroscience, The Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH, UK
[c] Department of Informatics, University of Hamburg, 22527 Hamburg, Germany

ARTICLE INFO

Available online 22 June 2010

Keywords:
Spiking neural network
Sound localisation
Inferior colliculus
Interaural time difference
Interaural level difference
Intelligent robotics

ABSTRACT

This paper proposes a spiking neural network (SNN) of the mammalian subcortical auditory pathway to achieve binaural sound source localisation. The network is inspired by neurophysiological studies on the organisation of binaural processing in the medial superior olive (MSO), lateral superior olive (LSO) and the inferior colliculus (IC) to achieve a sharp azimuthal localisation of a sound source over a wide frequency range. Three groups of artificial neurons are constructed to represent the neurons in the MSO, LSO and IC that are sensitive to interaural time difference (ITD), interaural level difference (ILD) and azimuth angle (θ), respectively. The neurons in each group are tonotopically arranged to take into account the frequency organisation of the auditory pathway. To reflect the biological organisation, only ITD information extracted by the MSO is used for localisation of low frequency (<1 kHz) sounds; for sound frequencies between 1 and 4 kHz the model also uses ILD information extracted by the LSO. This information is combined in the IC model where we assume that the strengths of the inputs from the MSO and LSO are proportional to the conditional probability of $P(\theta|ITD)$ or $P(\theta|ILD)$ calculated based on the Bayes theorem. The experimental results show that the addition of ILD information significantly increases sound localisation performance at frequencies above 1 kHz. Our model can be used to test different paradigms for sound localisation in the mammalian brain, and demonstrates a potential practical application of sound localisation for robots.

1. Introduction

Humans and other animals show a remarkable ability to localise sound sources using the disparities in the sound waves received by the ears. This has inspired researchers to develop new computational auditory models to help understand the biological mechanisms that underlie sound localisation in the brain. The project described in this paper discusses sound processing in the mammalian brain and aims to build a computational model that can be tested on biomimetic mobile robots to validate and refine models for focused hearing.

During the last decades, the structure and function of pathways in the auditory brainstem for sound localisation have been extensively studied [9]. Binaural sound localisation systems take advantage of two important cues [6] derived from the sound signals arriving at the ears: (i) interaural time difference (ITD), and (ii) interaural level difference (ILD). Assuming a pure tone sound source is positioned on the left side, the sound signal at the left ear is represented by $a\sin 2\pi ft$ (where a is the sound amplitude and f the sound frequency) while the sound at the right ear is represented by $(a-\Delta a)\sin 2\pi f(t+\Delta t)$, where Δa and Δt represent, respectively, the level difference (ILD) 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. Using these two cues sound source direction can be estimated in the horizontal or azimuthal plane.

The ranges over which these cues operate depend on head size. In humans the ITD cue is effective for localising low frequency sounds (20 Hz $\sim$1 kHz) [7], however, the information it provides becomes ambiguous for frequencies above $\sim$1 kHz. In contrast, the ILD cue has limited utility for localising sounds below 1 kHz, but is more efficient than the ITD cue for mid- and high-frequency (>1 kHz) sound localisation [7]. The ITD and ILD cues are extracted in the medial and lateral nuclei of the superior olivary complex (MSO and LSO), which project to the inferior colliculus (IC) in the midbrain. In the IC these cues are combined to produce an estimation of the azimuth of the sound [15].

Several hypotheses for ITD and ILD processing have been proposed [6,12,2], with one of the most influential being a model

---

[*] Corresponding author.
E-mail addresses: jindong.liu@gmail.com (J. Liu), davidpg@usal.es (D. Perez-Gonzalez), adrian.rees@ncl.ac.uk (A. Rees), harry.erwin@sunderland.ac.uk (H. Erwin), wermter@informatik.uni-hamburg.de (S. Wermter).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.10.030

advanced by Jeffress [6]. In his model, ITDs are extracted by a mechanism in which neural activity elicited by sound from each ear travels through a number of parallel delay lines, each one of which introduces a different delay into the signal and connects with a particular MSO cell. One of these delays compensates for the interaural delay of the sound waves, thus causing the signal from both ears to arrive coincidentally at a neuron that fires maximally when it receives simultaneous inputs. Smith et al. [12] provided partial evidence for Jeffress's model in the cat with the description of axons that resemble delay lines for the signal arriving at the MSO from the contralateral ear, but they found no evidence for delay lines for the MSO input from the ipsilateral side. For ILDs, physiological evidence suggests this cue is encoded in the neuronal firing that results from the interaction of an excitatory input from the side ipsilateral to the LSO, and an inhibitory input driven by the sound reaching the contralateral side. Thus, as the sound moves from one side to the other, the firing rate of the neurons decreases in one LSO and increases in the other.

Modellers have taken different approaches to represent this system. In an engineering study, Bhadkamkar [1] proposed a system to process ITDs using a CMOS circuit, while Willert [14] built a probabilistic model which separately measures ITDs and ILDs at a number of frequencies for binaural sound localisation. Recently, Voutsas and Adamy [13] realised a multi delay-line model using spiking neural networks (SNN) which incorporate realistic neuronal models. This model only takes into account ITDs and while it gives good results for low frequency sounds, it is not effective for frequencies greater than 1 kHz. Some models seek to incorporate the contribution of the inferior colliculus. In an engineering model, Rodemann [10] applied three cues for sound localisation, however, it did not take advantage of the biological connection between the superior olivary complex (SOC) and the IC. Willert [14] and Nix [8] implemented a probabilistic model to estimate the position of the sound sources, which includes models of the MSO, LSO and IC and uses the Bayesian theorem to calculate the connections between them. However, the model did not use SNN to simulate realistic neuronal processing.

This paper presents a model designed to identify sound source direction by means of a SNN. It is the first to employ an SNN that combines both ITD and ILD cues derived from the SOC in a model of the IC to cover a wide frequency range. To simulate the biological connection between the MSO/LSO and the IC, we propose a model which applies Bayes probability theorem to calculate the synaptic strength of the connection between cells in these nuclei. This model incorporates biological evidence on the inputs from the MSO and LSO to the IC, and is able to build a sharp spatial representation of a sound source.

The rest of this paper is organised as follows. Section 2 presents the neurophysiological organisation of the mammalian auditory pathway as derived mainly from cat and guinea pig experimental data. It also presents an IC model which takes account of the projection from MSO and LSO. Section 3 proposes a system model which simulates the mammalian auditory pathway from the cochlea up to the IC. In Section 4, experimental results are presented to show the feasibility and performance of the sound localisation system. Finally, conclusions and future work are considered in Section 5.

2. Biological fundamentals and assumptions

When sound waves arrive at the external ear, they enter the auditory meatus and vibrate the tympanic membrane, or ear drum, to then be propagated through the auditory ossicles in the middle ear to the cochlea of the inner ear. There the vibrations generate a travelling wave of displacement that propagates along the basilar membrane inside the cochlea, such that the point of maximum displacement is dependent on the frequency of the sound, thus leading to a spatial separation of the frequencies in the stimulus. The motion of the basilar membrane activates the inner hair cells arrayed along its length, which in turn trigger action potentials in auditory nerve (AN) fibres that transmit the spike encoded information to the central nervous system. Each auditory nerve fibre is maximally sensitive at a characteristic frequency (CF) which is determined by the location of the inner hair from which it receives its input [15]. This tonotopic representation of frequency is maintained in subsequent nuclei of the ascending auditory pathway.

In addition to this tonotopic representation, the AN fibres also encode temporal information about the sound waveform. The hair cells act as halfwave rectifiers so that the probability of AN fibre excitation is maximal during the peak phase of the sound waveform. This phase locking occurs at frequencies of 20 Hz $\sim$5 kHz, and is an essential step in the extraction of ITDs, because it represents the basis for comparing the relative timing of the waveforms at the ears. Fig. 1 shows an example of spikes phase-locked to the peaks of the sound waveform (t_1^l, t_1^r, t_2^l and t_2^r).

As the sound pressure level (SPL) increases, the discharge rate of most AN fibres increases sigmoidally over a relative range of $\sim$30 dB. In order to cover the wide range of SPL to which we are sensitive, e.g. 120 dB, the relative operating range changes adaptively according to the background sound levels. There is also a smaller population of AN fibres that have a higher threshold and a wider dynamic range. For simplicity, in this paper we do not model the biological details of the encoding of sound amplitude, but rather we use the measured SPL (e.g. p_i^l and p_i^r in Fig. 1) in the first stages of ILD processing.

After encoding the temporal and amplitude information, the spike-encoded information from each ear is transmitted via the cochlear nuclei to the SOC to extract ITDs and ILDs in the MSO and LSO, respectively [15] (Fig. 2). The MSO on one side receives excitatory inputs from the anteroventral cochlear nucleus (AVCN) from both the ipsilateral and contralateral sides. An ITD-sensitive cell in the MSO fires when the contralateral excitatory input leads the ipsilateral by a specific time difference. According to hypotheses based on Jeffress's original model, activation of these coincidence detectors is thought to occur when the contralateral delay line network compensates for the time delay of the sound in the ipsilateral ear, i.e. ITD. These ITD-sensitive cells in the MSO can be idealised as a coincidence cell array where each cell receives a delay-line input, and they are assumed to be

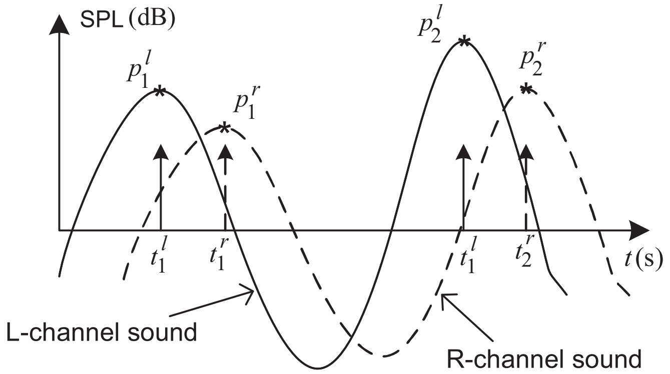

Fig. 1. An example of sound signals arriving at both ears (left, continuous line; right, dashed line), and the phase-locked spikes (t_1^l, t_1^r, t_2^l and t_2^r) triggered by them. The signal corresponding to the right ear is delayed and has a smaller amplitude than the left one, indicating that the origin of the sound was on the left side of the head. $p_i^{l/r}$ is the sound pressure level at which the spikes are generated.