Neural Networks 23 (2010) 819–835

Contents lists available at ScienceDirect

Neural Networks

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

Evolving spiking neural networks for audiovisual information processing

Simei Gomes Wysoski [a,*], Lubica Benuskova [a,b], Nikola Kasabov [a]

[a] *Knowledge Engineering and Discovery Research Institute,*[1] *Auckland University of Technology, 1051 Auckland, New Zealand*

[b] *Department of Computer Science, University of Otago, Dunedin, New Zealand*

ARTICLE INFO

Article history:
Received 7 July 2008
Received in revised form 26 January 2010
Accepted 27 April 2010

Keywords:
Spiking neural network
Audio and visual pattern recognition
Face recognition
Speaker authentication
Online classification

ABSTRACT

This paper presents a new modular and integrative sensory information system inspired by the way the brain performs information processing, in particular, pattern recognition. Spiking neural networks are used to model human-like visual and auditory pathways. This bimodal system is trained to perform the specific task of person authentication. The two unimodal systems are individually tuned and trained to recognize faces and speech signals from spoken utterances, respectively. New learning procedures are designed to operate in an online evolvable and adaptive way. Several ways of modelling sensory integration using spiking neural network architectures are suggested and evaluated in computer experiments.

1. Introduction

A number of systems have used the terms *biologically realistic* or *brain-like* to define a new generation of neural networks that attempt to process information in a way similar to the human brain. What mainly motivates researchers in this direction is that artificial information processing systems, despite enormous effort, still struggle to deliver general and reliable solutions. The majority of the attempts were on modelling the visual and auditory systems, perhaps because their inputs and outputs are better defined than for somatosensory or olfactory/gustatory systems and also because there is a strong interest in more intelligent visual and acoustic computer systems in a wide variety of industrial sectors (e.g., car manufacturing, aerospace, medicine, etc.).

Several models of visual systems used the Hubel and Wiesel model of the primary visual cortex with contrast, directionally selective and complex cells placed in a hierarchical pathway (Hubel & Wiesel, 1962) for the purpose of pattern recognition (Fukushima & Miyake, 1982; Mel, 1998; Riesenhuber & Poggio, 1999). Examples of brain-like auditory models can be found in Ghitza (1988) and Shamma, Chadwick, Wilbur, Morrish, and Rinzel (1986). Also under the *biologically realistic* label, many approaches showed how artificial systems could adapt and evolve in an intelligent and autonomous way. In this direction, networks of processing units learn what is the best structural configuration based on a few soft constraints and self-growing/shrinking procedures (see Gallant, 1995; Kasabov, 2007, for extensive reviews on adaptive methods and procedures).

Thus, up to this point, there are *brain-like* models of network structures and *brain-like* ways to perform network connectivity and reconfiguration. Recently another factor has added to the momentum. The principle that neurons can perform pattern recognition using spike timings is another addition to *biologically realistic* information processing (Hopfield, 1995). In Gerstner and Kistler (2002) this concept is properly clarified, stating that, in order to avoid any prior assumptions on neural computation, neurons need to process and exchange information at the level of spikes. Thus, spiking neurons and spiking neural networks (SNNs), historically used as a tool for neuroscientists to study the dynamics of single or ensembles of neuronal units, emerged as a new generation of neural network models for pattern recognition.

Although SNNs are mathematically more complex than traditional artificial neural networks, they are potentially better suited for hardware implementation due to the "integrate-and-fire" nature of spiking neurons (Tikovic, Vöros, & Durackova, 2001). There are several advantages of the hardware implementations of spiking neurons, e.g. no multiplications as in traditional models, pulse processing can be implemented using shifts and adds, interconnections transmit only a single bit instead of real numbers. Sparse and asynchronous communication can also be easily implemented. However, it is important to note that this prospective advantage does not manifest itself yet when implementing SNNs in a general purpose computer platform.

As the theory of spiking neurons is currently most accepted to describe brain-like way of processing (Gerstner & Kistler, 2002), and is the most promising with respect to the future super-fast and reliable hardware implementations, it is the basis for all

* Corresponding author. Tel.: +64 9 526 4486.
E-mail address: wysoski@hotmail.com (S.G. Wysoski).

[1] http://www.kedri.info.

0893-6080/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2010.04.009

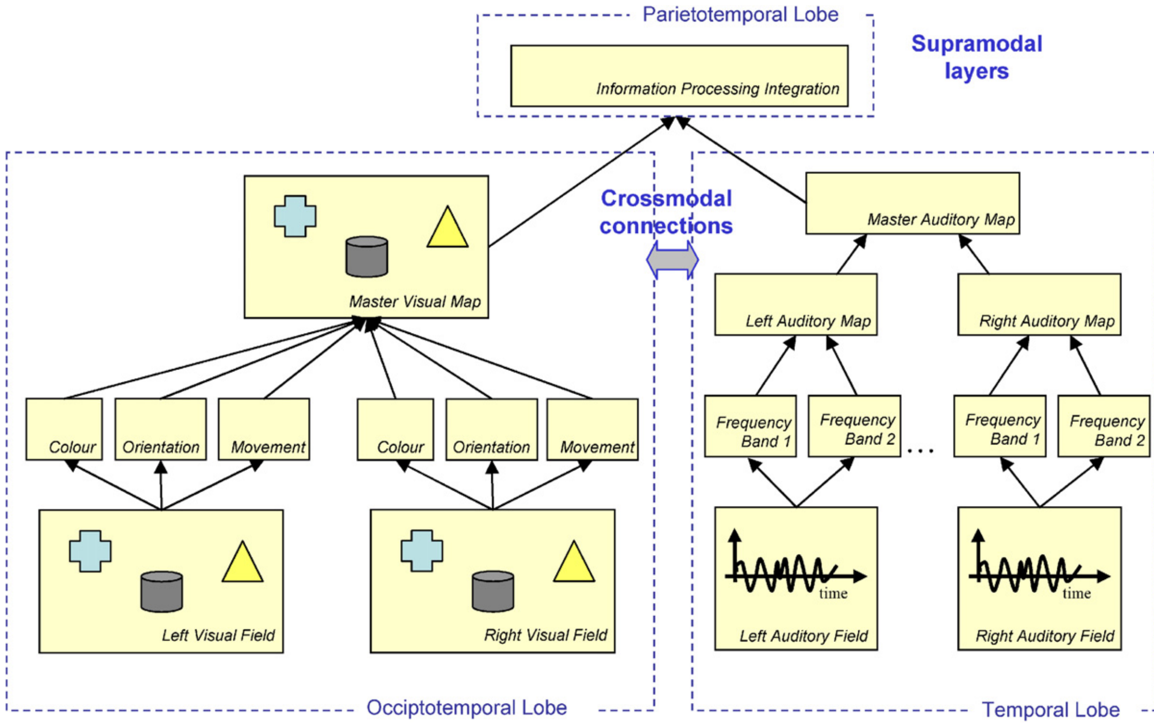

Fig. 1. Integration of sensory modalities of the visual system (left side), the auditory system (right side) and the subsequent integration of modalities (above).

new designs presented in this paper. At the systemic level, the behaviour of ensembles of neurons and the information processing pathways are also evaluated under the biological perspective. Of particular relevance to this research are the auditory and visual systems that are discussed separately in Sections 2 and 3. The auditory and visual pathways are considered in a new integrative audiovisual pattern recognition approach. The learning theories and the corresponding algorithms to implement them are discussed from the perspective of computation with spiking neurons.

The main original contributions of this paper are:

(a) Design of a new spiking neural network architecture to perform person authentication through the processing of signals from auditory and visual modalities. The integrative architecture combines opinions from individual modalities within a supramodal layer, which contains neurons sensitive to multiple sensory inputs. An additional feature that increases biological relevance is the crossmodal coupling of modalities, which effectively enables a given sensory modality to exert direct influence upon the processing areas typically related to other modalities.
(b) Extension of adaptive online learning procedure to audiovisual pattern recognition. An online learning procedure that enables the system to change its structure by creating and/or merging neuronal maps of spiking neurons is presented and evaluated.
(c) Experimental evaluation of a new architecture that integrates sensory modalities on a person authentication problem, and comparison with traditional approaches.

A schematic illustration of the complexity of integrating the auditory and the visual senses is shown in Fig. 1. Each sensory modality has distinct pathways where information is processed. Within a sensory modality, information is decomposed, e.g., in the visual system, the information is divided into submodalities (colour, shape, motion, etc.) that are independently processed in different pathways. In the auditory system, the ventral cochlear nucleus with mainly tonotopical organization of cells and dorsal cochlear nucleus (mainly non-tonotopical) also define different pathways. In different modalities and submodalities, it is reasonable to think that the speed of transduction and the speed of information propagation in different pathways is not the same. If this is true, afferent stimuli from different sensory modalities arrive at the cerebral cortex at different times. The separation and integration of pathways within a modality as well as the integration of pathways from different modalities (and all the synchronizations implied in it) constitute a complex network that cannot be in principle accurately described and reproduced in a computational model. Therefore, many simplifications and abstractions need to be introduced to the design of sensory integration models.

The visual and the auditory models, and the novel online learning procedure that were described in Wysoski, Benuskova, and Kasabov (2006, 2007, 2008a) are evaluated in Sections 2 and 3 of this paper. Section 4 explores the integration of modalities (with preliminary description presented in Wysoski, Benuskova, & Kasabov, 2008b), which is followed by experimental evaluation of the adaptive properties of the system. Section 5 concludes the paper and points to further directions to explore in order to achieve more biologically realistic and reliable pattern recognition systems.

2. The visual model

2.1. Short literature review

In a pioneering attempt to create a network in which the information is processed through several areas resembling the visual system, Fukushima and Miyake proposed the Neocognitron, which processes information with rate-based neural units (Fukushima & Miyake, 1982). A new type of model for object recognition based on computational properties found in the brain cortex was described by Riesenhuber and Poggio (1999). This model uses hierarchical layers similar to the Neocognitron and processing units based on MAX-like operation, to define the postsynaptic response, which results in relative position and scale invariant features. This