Stochastic correlative firing for figure-ground segregation

Zhe Chen
Adaptive Systems Lab, McMaster University, Hamilton, Canada L8S 4K1

Received: 2 April 2004 / Accepted: 10 January 2005 / Published online: 5 March 2005

Abstract. Segregation of sensory inputs into separate objects is a central aspect of perception and arises in all sensory modalities. The figure-ground segregation problem requires identifying an object of interest in a complex scene, in many cases given binaural auditory or binocular visual observations. The computations required for visual and auditory figure-ground segregation share many common features and can be cast within a unified framework. Sensory perception can be viewed as a problem of optimizing information transmission. Here we suggest a stochastic correlative firing mechanism and an associative learning rule for figure-ground segregation in several classic sensory perception tasks, including the cocktail party problem in binaural hearing, binocular fusion of stereo images, and Gestalt grouping in motion perception.

1 Introduction

Segregation and streaming are important and fundamental tasks for sensory perception in both the auditory and visual domains (Livingstone and Hubel 1988). However, understanding the mechanisms underlying the brain’s remarkable ability to segregate an object of interest in a complex auditory or visual scene remains a major challenge. Computational neuroscientists often view or simplify the development of perceptual systems as an unsupervised learning problem, where learning refers to the process of synapse adaptation and the learning rule refers to the procedure for adjusting the synapse. Mathematically, learning can be cast as an optimization process. In formulating the learning process, we must derive a procedure to find an optimal solution that achieves a minimum or maximum value of a predesigned cost function. Understanding how the brain solves such optimization problems with remarkable ease and efficiency is of theoretical and technical importance. This enigma, despite many efforts, remains unsolved. We conjecture that the candidate learning rule to the solution should be simple, robust, biologically plausible, and amenable to parallel and distributed implementation.

In a biological context, the correlated firing mechanism of synapse was first postulated by Hebb (1949). Hebbian-like learning is temporally local since it only depends on pre- and postsynaptic firing rates and the present state of the synapse. Correlation-based learning is indeed a special case of the Hebbian rule, and therefore is particularly attractive to serve as a neurobiological model of learning. Correlation theory also has roots in associative memory (see, e.g., Arbib 1995; Eggermont 1990), perception (Cook 1991), coincidence detection (Eggermont 1990), sensory segmentation (von der Malsburg and Schneider 1986; von der Malsburg 1999), and optimization and learning (Harth and Tzanakou 1974; Tzanakou et al. 1979; Harth et al. 1987; Unnikrishnan and Venugopal 1994). Here we illustrate, for the first time, that a simple yet powerful correlation learning rule is sufficient to accomplish several figure-ground segregation and perceptual vision tasks. Formulated within a similar framework, the correlative firing mechanism and learning procedure work equally well for auditory and visual processing tasks, despite the drastic differences between auditory and visual sensory inputs.

2 Stochastic correlative firing mechanism

The suggested stochastic correlative learning rule follows closely ALOPEX, a gradient-free optimization procedure originally developed in vision research (Harth and Tzanakou 1974; Tzanakou et al. 1979). Let \(\mathbf{w}\) denote a vector of some synaptic weights or unknown parameters, and let \(E\) denote a predefined cost function induced by \(\mathbf{w}\); the correlative learning rule has the following form:

\[
\mathbf{w}(t+1) = \mathbf{w}(t) + \eta \Delta \mathbf{w}(t) \Delta E(t) + \gamma \mathbf{r}(t),
\]

where \(\Delta \mathbf{w}(t) = \mathbf{w}(t) - \mathbf{w}(t-1), \Delta E(t) = E(t) - E(t-1), \mathbf{r} \) is an additive (uniformly distributed) random noise vector, \(\eta\) is a learning-rate parameter, and \(\gamma\) is another scalar that controls the amount of the noise. Equation (1), being a generalized form of Hebbian rule, says that the synaptic adaptation \(\Delta \mathbf{w}(t+1)\) at time \(t+1\) is proportional to the...
cross-correlation term $\Delta w(t) \Delta E(t)$, a product of previous synaptic modulation (presynaptic activity), and the resulting cost modification (postsynaptic activity) at time $t$. The algebraic sign (positive/negative) depends on the form of the cost function (to be maximized/minimized). It is noteworthy to highlight some features of the learning rule (1) here:

- **Statistical learning perspective:** The freedom of designing $E$ allows flexibility to incorporate higher-order correlations in the learning process. The main role of the noise is to introduce certain randomness into the optimization process to help escape local maxima or minima. In common with many other stochastic optimization procedures (Ballard et al. 1983; Kirkpatrick et al. 1983; Geman and Geman 1984), this correlative learning rule relies heavily on the effect of noise. We have observed that the influence of noise is critical to the optimization process in all of the tasks undertaken here.

- **Biological computation perspective:** The correlative mechanism described in (1) postulates a model of temporally asymmetric Hebbian-like synapse plasticity and reinforcement-reward-like causality: the synaptic modification $\Delta w$ induces the delayed reward $\Delta E$. This can be viewed as a top-down projection effect: the synaptic strength can be influenced by electrochemical substance released via high-level cognitive activities. In addition, (1) characterizes a *synchronous* firing mechanism in the sense that all the synapse connections in $w$ are modified at the same time. The existence of the noise is consistent with the fact that ubiquitous stochastic variability exists in the firing pattern of neurons at the sensory cortices. Another attractive feature of (1) is that it is model independent and therefore applicable for arbitrary cost functions and neural architectures that have a modular, hierarchical, or feedback structure.

3 Figure-ground segregation

3.1 Binaural hearing at cocktail party

The cocktail party effect demonstrates the capability of the human brain to recognize the voice of an attended speaker in a noisy room environment, despite the existence of competitive sound sources or noises (Cherry 1953). Segregating the attended speech is an essential task in auditory perception (Bregman 1990). Recently, the cocktail party effect has been tackled in the framework of independent component analysis (ICA) (Bell and Sejnowski 1995; Brown et al. 2001; Makeig et al. 1997; Hyvärinen et al. 2001). Let $S$ denote a source matrix that contains the original signals, where each source is represented by a row vector and each vector contains the samples of one source. It is assumed that the sensors (the haircells of the inner ear) receive the signals subject to a linear mixing process described by $X = AS$, where $A$ is called the mixing matrix. In the auditory cortex, the brain attempts to reconstruct the original auditory scene, which requires sound identification and localization, analogous to the “what” and “where” tasks of the visual system. The recovered signals are written in a matrix form $Y = WX$, where $W$ is called the demixing matrix. We envision a distributed parallel architecture for auditory perception, as illustrated in Fig. 1.

The goal of figure-ground segregation is to recover the speech signal of interest (“figure”) from the auditory scene. We speculate that in different parts of the auditory cortex there exist distributed modules that are responsible for processing the mixing signals $X = [x_1, x_2]^T$ received from the two ears after sophisticated preprocessing by the cochlea. Different modules perform distinct segregation tasks in the independent circuits, with individual one attending one source. Each module runs a stochastic correlative rule as in (1) and produces two coherent outputs $Y = [y_1, y_2]^T$, that is, $y_1$ is as close to $y_2$ as possible. The computations in the modules are performed in parallel; the “figure” of interest is chosen via selective attention, modeled by a gating network through a thalamocortical loop. The switch of attention can be influenced by the top-down expectation as well as the relative intensity difference between the left-ear and right-ear inputs, which might modulate the factor with which sound coming from either (left or right) direction is forwarded. It should be stressed that the learning rule makes no attempt to perform a complete separation of all possible sources as in conventional ICA. Here the outcomes from each module are coherent, hence only one stream is attained in each module (Fig. 1). This is consistent with human perceptual processing, where at a particular moment we only pay special attention to one or another source, but not both. For example, in the segregation task, we are only interested in recovering the “figure” rather than decomposing the scene. Therefore, in each module the sound or voice in the foreground should be perceived synchronously. This is a key observation we discovered in the experiments using (1). Another nice feature of this learning rule in a computational context is that it is insensitive to the sampling frequency of the source signal and is therefore numerically robust.

Designing an appropriate cost function for the learning procedure is the next step in applying a correlative learning rule to perception tasks. Information-theoretic measures have often been used to characterize sensory coding (Attick 1992; Becker and Hinton 1992). The most informative features in sensory data are those that are highly non-Gaussian (Field 1994; Olshausen and Field 1996). We used the maximum kurtosis (absolute value) criterion in order