

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Learning distributed selective attention strategies with the Sigma-if neural network

Maciej Huk
Wroclaw University of Technology
Poland

1. Introduction

Selective attention systems are found to be very interesting as well from theoretical point of view, as also as useful tools for many practical applications, such as analysis of large data sets, real time route planning for autonomic robots in dynamical environments, and dispersed sensor networks control. However up to date models of biological selective attention systems and other methods that realise similar functionalities still have some drawbacks. Decision trees are easy to analyse by humans, but once created are hard to adapt to the changes in the system environment. They are also not resistant to damages in their rigid logical structure as well as in the hardware that executes them. The next example - statistical feature selection algorithms create solutions that use globally suboptimal feature sets, but in general they are not optimal from the point of view of single input vectors, and those algorithms are hard to apply in fast changing environments. This is why biologically founded neuronal selective attention systems seem to offer better properties for many possible applications. They have distributed nature and in the effect are more damage resistant than centralised models. They are also easy to adapt to the changing environment conditions, and were tested and selected by nature as simple, cheap and scalable solutions. The only problem with neuronal selective attention systems is that their existing models implement artificial, centralised attention directing algorithms and in the effect they do not offer functionalities observed in the nature. This is why there is still a need for analysing properties of natural selective attention mechanisms, as well as for further exploration of possible models of biological neuronal selective attention systems.

2. Sources and properties of natural selective attention systems

In the nature selective attention is a mechanism that gives a living organism the possibility to extract from the incoming information a part that is most important at a given moment and that should be processed in detail (Broadbent, 1982; Treisman, 1960). This mechanism is necessary to avoid wrong or delayed decisions and reactions, due to limited processing capabilities of the typical nerve system which does not allow rapid analysis of the whole visual and other senses scene (Tsotsos et al., 2001; VanRullen & Koch, 2003). So selective

attention can be viewed as a strategy of dynamical input space selection for gaining predefined goals by an organism interacting with a very complicated environment.

It is easy to search for selective attention manifestations during animal life-threatening situations. When fast conditioned and unconditioned reflex start to rule over animal behavior, often some important signals from one given sense cause that other signals or signals from other senses are left almost completely unprocessed. A good example here is a very hungry sparrow that observe something to eat and suddenly hears sound of an eagle – in most cases at this very moment signals about food in front of the eyes as well as signals about hunger automatically became not important for the little bird and do not affect its decisions – it starts to run away. What is also interesting, such behaviour happens sometimes even if the little bird is young and has never heard the eagle before. This is because some very important strategies of senses and signals prioritisation can be those that were developed by evolution and in such cases are already hard-coded in the nervous system of the animal (Tinbergen, 1951; Dole, 2008).

Above simple example suggests basic properties of selective attention mechanisms. They are simple, process information very fast and most probably have direct access to signals from sensors. What's more – they seem to function as independent subsystems and can have great influence on decisions and behaviour of an animal. It can also be seen, that once developed, selective attention strategy can help its user to reduce time from signal arrival to target decision or reaction, to reduce information acquisition and processing costs, e.g. in terms of energy. But there is more – it can help to gain better decisions – especially in multi-criteria optimisation problems, by reducing them to simpler ones with use of historical knowledge encoded in used selective attention strategy.

It would be a great advantage to use such systems as tools in science and industry. Applications range here is very wide – from data mining and picture recognition, through robot control and advisory systems (e.g. medical, financial) to strategic planning and autonomous target selection and tracking. It would be also a new opportunity to study psychophysiology of human perception, development of artificial senses and artificial agents that mimic human behaviour. In theory the key to such an interesting and very useful functionality seems to be simple – a generation and use of a set of rules that for every possible situation define parts of perception space (fields of attention) that should be observed and processed with highest, medium or lowest priority (Tsal, 1983, Tsotsos et al., 2001, Claus et al., 2004). But in practice to develop such systems and to see selective attention phenomenon in greater detail one should familiarize himself with a key observations of selective attention mechanisms in humans and other primates.

Among observations of selective attention in humans the most widely known are early experiments of Cherry, Gray and Treisman on such called “cocktail party problem”. They have shown that even in crowded and noisy party room, people can effectively discuss in pairs by passing over the global noise, and – in the same time – can react on some keywords (e.g. their name, surname) even if those words were told in the opposite side of the room (Cherry, 1953; Gray & Wedderburn, 1960; Treisman, 1960). This gives a suggestion that we use some kind of automated processes that are specialised in filtering out unimportant information. Properties of those processes were further examined by Deutsch, Lewis and Spelke in experiments on dichotic listening (Deutsch & Deutsch, 1963; Lewis, 1970; Spelke et al., 1976). They have shown that all signals reaching our perceptual system, at first undergo non-conscious, automatic semantic preprocessing.

Nature of this semantic preprocessing was later partially uncovered in experiments on human analysis of words (Neely, 1977; Grosjean, 1980; Marslen-Wilson & Tyler, 1980) and sentences (Swinney, 1979; Swinney, 1982; Pynte et al., 1984). Measurements of the time of understanding and reaction on heard words have shown that the process of understanding of their meaning lasts for about 200ms from the beginning of signals arrival. In this time usually are received only two first phonemes of the word, but it allows for significant automatic reduction of the number of word's potential meanings (in English language first phoneme in average reduces that number to 1033 and the second phoneme to 87 meanings (Kucera & Francis, 1967)). It should be also noticed that in such a short time the signals can be processed only by two or three subsequent layers of human brain neurons. Thus such structures most probably act as information context selection subsystems, and quickly activate synaptic paths to parts of the human neural system, that are semantically connected with just processed signals and do the final part of those signals interpretation (LaBerge, 1990). Those findings coincide with previous results of Spelke and Shiffrin, that show that automated processing is very flexible, and can be significantly modified by training in a short period of time (about six weeks) (Spelke et al., 1976; Shiffrin & Schneider, 1977). Their experiments concerned a task of writing some phrases and reading other phrases in the same time. While before the training it was hard enough to disturb processing of other signals, that after the training, doing the same was easy and almost automatic, and saved resources could be used to realise other tasks.

But one may ask - what is the main, most crucial physical or statistical effect that makes selective attention to appear and function so well? The answer is not simple and most probably we currently know only a half of the truth. What we know, is that there is no central, specialised part of a brain that govern our attention (Allport et al., 1972). Also simple biological neural networks that somehow learn how to direct our attention, seem not to have any special neurons or uncommon architectures of interconnections between them. Thus basic mechanism of selective attention emerge invisibly somewhere between levels of single neurons and simple neural network. While it is hard to analyse the way neural network process information, nature gives us a unique chance to observe the way we direct our visual attention by moving our eyes. The most important achievement in this field is an identification of the role of saccadic eyes movements in picture recognition and objects tracking (Noton & Stark, 1971; Findlay, 1982; Chelazzi et al., 1993; Schall & Hanes, 1993). For every scene eyes control subsystem learns to quickly direct gaze to the most important scene elements, by following a unique directed scan path, that defines number, locations and sequence of picture points to be analysed. This seem to be the central mechanism of natural selective attention systems, because an element of automatic, cyclic selection of regions of attention, guided by previously learned scan paths, was found to be common for almost all input channels of human brain (Tsal, 1983; Kastner et al., 1998). Unfortunately we still don't know exactly how biological neural networks develop and realise selective attention mechanisms and how to mimic this functionality.

3. Towards neuronal models of selective attention

Promising prospects of the use of human-like artificial selective attention systems made many authors to look for precise theoretical description of this phenomenon. But it took almost seventy years from the moment of writing down the first definition of selective

attention by sir William James in 1890, to the formulation of the first models. Unfortunately those first attempts to model selective attention, such as Broadbent's bottleneck theory (Broadbent, 1958), Treisman's two step filtering (Gray & Wedderburn, 1960; Treisman, 1964) and Kahneman's centralised attention resource model (Kahneman, 1973) were too general and too imprecise (not to say completely not true) to stand the tests of further experiments. After those first steps, proposed models evolved in the direction of more and more specialised ones, such as Allport's multiple resource capacity, that captured distributed nature of selective attention processes, and Posner's late selection model, that described the role of non conscious, automatic semantic preprocessing observed by Deutsch, Neely and Swinney.

But the first groups of models that could have been used in practice emerged as the realisations of Noton and Stark scan path theory — those were so called neuronal hierarchical routing and shifting circuits models. Their authors assumed that to realise dynamic changes of region of interest, neural networks should be governed by specialised external element that uses predefined algorithm of selection of inputs to activate (Koch & Ullman, 1985; Anderson & Van Essen, 1987; Olshausen et al., 1993) or various artificial mechanisms for dynamic network architecture changes (Niebur et al., 1993; Tsotsos et al., 1995; Houghton & Tipper, 1996; Pelc, 1998), which in fact also were realisations of some constant embedded strategies of selective attention. This was not correct in the light of neurobiological observations and earlier experimental proofs for distributed nature of selective attention, but it allowed producing specialised tools that were useful in tasks such as face recognition, objects tracking or even pointing interesting objects for observation by cosmic probes (Yamada & Cottrell, 1995; Rybak et al., 1998; Hager & Toyama, 1999; Privitera et al., 2000).

As those models were not universal solutions of the problem of building neural selective attention systems, were very complicated and neurobiologically unfounded (Privitera & Stark, 2000; Eckstein et al., 2001; Clauss et al., 2004), this gave a strong impulse to further searching for basic attention mechanisms on the level of single synapses and neurons (Lee, 2000; VanRullen, 2003; Renninger, 2004). The basic paradigm of this direction of research is that selective attention at higher levels of brain structure organization emerges as an effect of synergy between elementary structures at lower levels. Thus the attention of researchers was concentrated on exploring new artificial neurons models — especially those, that allow interactions or other dependencies between signals arriving to neuron's inputs.

To achieve that, one had to consider input signals aggregation methods other than simple weighted sum of input values known from perceptron neuron model. This is due to the fact, that aggregation function is the only one part of neuron transmit function that has full access to all information that is available through neuron's inputs — and after the aggregation process most of the information about the sources of particular signals and dependencies between them is lost. It would also provide a way to search not only for models that take into consideration dependencies between signals from distinct inputs, but also that allow in some situations the reduction of the number of signals read in for processing. For those reasons, a set of appropriate neuron models was proposed, along with definitions of two basic types of aggregation functions. First family of those aggregation functions incorporate so called higher order neuron models, and the second family — nonlinear neuron models.

3.1 Higher order polynomial aggregation functions

Higher order neuron models extend simple weighted sum aggregation schema by adding into it also higher order terms. In general such solutions are variations of polynomial aggregation function:

$$\varphi_{Poly}(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^N w_i x_i + \sum_{i,j=1}^N w_{ij} x_i x_j + \dots + \sum_{i,j,\dots,k=1}^N w_{ij\dots k} x_i x_j \dots x_k \quad (1)$$

in which \mathbf{x} and \mathbf{w} are input and weights vectors (respectively of dimension N and $N+1$). They include all or only chosen terms of the right side of this expression. Such extension of aggregation function form, not only increases information capacity of resulting neural networks (Cover, 1965; Venkatesh & Baldi 1991), but also their ability of learning of geometrically invariant properties of input patterns (Giles & Maxwell, 1987; Perantonis & Lisboa, 1992). Unfortunately with the increasing number of neuron's inputs, this leads to the phenomenon of exponential explosion of the number of higher order terms. But in practice, when it is necessary to implement some particular function, usually only few high order terms are required (Redding et al., 1993). This can prevent exponential explosion of aggregation function complexity, and is the key to the effectiveness and practical use, of such standard examples of the higher order aggregation functions as the Sigma-Pi (sum of products) and the Clusteron (Mel, 1990; Mel, 1992):

$$\phi_{Sigma-Pi}(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^d \left(w_i \prod_{k \in R_i} x_k \right). \quad (2)$$

It can be seen that aggregation function of Sigma-Pi neuron defines a set of d input connections clusters R_i within which connections collectively build up an importance of incoming signals group by multiplying them with each other. In this model, low value of only one input signal in the group can block signals from all other inputs in a cluster. This simple conditional behavior of aggregation function generates interesting properties of Sigma-Pi neural networks (Neville & Eldridge, 2002; Weber & Wermter, 2007), but in general it is hard to optimally define inputs clusters and collective blocking of the whole clusters seems to be too rigorous. This is why Clusteron aggregation function was made to almost automatically define clusters of neighbouring inputs (one has only to specify radius r of all the clusters), and independently consider mutual dependencies between signals from each pair of inputs in a given cluster:

$$\phi_{Clusteron}(\mathbf{x}, \mathbf{w}, r) = w_0 + \sum_{i=1}^N \left(w_i x_i \sum_{k=i-r}^{i+r} w_k x_k \right). \quad (3)$$

Nevertheless, those improvements don't outdate the question, if chosen or predefined clusters of connections are in the given situation optimal or even suboptimal. In the effect in the literature exist also a set of slightly different solutions that do not use the concept of input connections clusters. In their case the only mechanism that allows taking into account interconnection dependencies is defined by chosen, simple higher order polynomial terms. Good examples of those are Compensatory aggregation function (Sinha et al., 2001):

$$\varphi_{Compensatory}(\mathbf{x}, \mathbf{w}, \mathbf{w}') = w_0 + \sum_{i=1}^N w_i x_i + 0.5 \sum_{j=1}^M \sum_{\substack{i=1 \\ i \neq j}}^N w'_i w'_j x_i x_j, \quad (4)$$

as well as aggregation functions of the Quadratic neuron:

$$\varphi_{Quadratic}(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^N \sum_{j=i}^N x_i x_j w_{ij}, \quad (5)$$

and the Cubic neuron (Bukovsky et al., 2007):

$$\varphi_{Cubic}(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{i=1}^N \sum_{j=i}^N \sum_{k=j}^N x_i x_j x_k w_{ijk}. \quad (6)$$

Their simplicity makes them easier to apply than Sigma-Pi or Clusteron function, and they still mimic some aspects of selective attention (Gupta, 2008). But in practice it is unanswered fundamental question if chosen subset of higher order terms is a best one for solving a given problem.

3.2 Nonlinear neural aggregation functions

As polynomial aggregation functions do not solve all the problems with modeling low-level selective attention, there also exists a parallel research on different, nonlinear aggregation methods. The main idea of those solutions is to model nonlinear aspects of input signals accumulation observed in biological neurons (Anderson et al., 1985; Karlholm, 1993; Stuart & Spruston, 1998; Larkum et al., 1999; Körding & König, 2001), with the use of operations other than signals multiplication. In the effect selective attention behaviour could be generated also in the situations when all neuron's input signals values are definitely non zero.

The first widely known nonlinear aggregation function was proposed with a neuron model called the product unit (Durbin & Rumelhart, 1990):

$$\varphi_{PU}(\mathbf{x}, \mathbf{w}) = w_0 + \prod_{i=1}^N w_i^{x_i}. \quad (7)$$

It uses nonlinearity of the exponential function, but still the dependencies between input signals are generated through multiplication of the terms connected with particular input connections. Many authors reported usefulness and interesting theoretical properties of the product unit model and it is still an object of a research (Leernik, 1995; Schmidt, 2002; Martínez-Estudillo et al., 2008).

Another essential example and the first attempt to build a strictly non multiplicative, nonlinear aggregation function is the Spratling-Hayes function (Spratling & Hayes, 2000):

$$\varphi_{SH}(\mathbf{x}, \mathbf{w}, \mathbf{c}, \kappa) = w_0 + \sum_{i=1}^N w_i \min \left(\frac{x_1 + \kappa}{c_{i,1}}, \frac{x_2 + \kappa}{c_{i,2}}, \dots, \frac{x_{i-1} + \kappa}{c_{i,i-1}}, x_i, \frac{x_{i+1} + \kappa}{c_{i,i+1}}, \dots, \frac{x_N + \kappa}{c_{i,N}} \right). \quad (8)$$

In its case nonlinearity as well as dependencies between neuron's input connections are induced by the minimum function. This solution uses also a matrix c , which elements define a ratio of mutual dependence between each two input connections, and a parameter κ that determines a threshold of signals values difference, below which dependence between two given connections is not considered.

As it is hard to choose most representative subset of aggregation functions proposed up to date, let that the last presented nonlinear solution will be very interesting aggregation function of generalised-mean neuron model (Yadav et al., 2004):

$$\varphi_{GMN}(\mathbf{x}, \mathbf{w}) = \left(w_0 + \sum_{i=1}^N w_i x_i^r \right)^{1/r}. \quad (9)$$

The behaviour of this function is highly dependent on the value of single continuous parameter r . When r is equal one, resulting function is identical to the perceptron's aggregation function. When r is greater than one, no connection influences the other, and when r is greater than zero and less than one, GMN function can include various higher order terms (e.g. for $r=1/2$). Such flexibility makes that function applicable in many problems (Yadav et al., 2006), but there is no explicit rule describing how to properly choose value of the r parameter for a given problem.

3.3 Remarks and further directions of research

Above examples of polynomial and nonlinear aggregation functions present proposed up to date basic schemes of mimicking biological low-level selective attention systems. Resulting models of artificial neurons during input signals processing can take advantage of additional information about mutual dependencies between signals from different input connections. Additional nonlinearities, after proper selection of model's parameters values, allow also effective directing of an attention to the most important elements of input data. But they all also have one common fundamental disadvantage – in every situation they require reading in of all input signals, regardless of their importance. Thus their selective attention abilities are very limited when compared to properties of natural systems.

In theory for above functions can exist additional mechanism that in some situations would prevent reading in unimportant information. If the aggregation function would include higher order terms and input signals from different inputs would be read in and analysed in many steps, similarly as in the Stark's scan path model, a zero valued signal could be information that signals from some other input connections can be left unread. But in practice results of such a solution would strongly depend on the order of input signals analysis. It is interesting that similar effects are detected in nature – observations of neurobiologists show that biological neurons are asynchronous in nature, and that their behaviour depend on the order in which input signals reach the neuron. An example of biologically founded asynchronous neurons models are integrate and fire neurons, that process information coded in the time domain (e.g. with use of frequency modulation) –

their most important property is ability to mimic capacitive character of signals accumulation process (Stein, 1967; Maass, 1997; Burkitt & Clark, 1999; Quiles et al., 2008).

Unfortunately artificial asynchronous neuron networks also do not show attentional properties observed in the nature. This is due to the fact, that asynchronous, continuous accumulation of input signals, neglect the information about that, from which neuron's inputs come the signals which influenced neuron's actual activation potential. This make it impossible to realise selective attention in the form other than making neuron's output value dependant from only most intense stimuli and regardless of their meaning.

As both synchronous and asynchronous neuron signals aggregation schemes don't provide general and precise models of selective attention, it seem to be plausible to search for hybrid solutions that are both synchronous, and read in and analyse signals in a step by step manner. This research direction finds strong support in recent observations that link selective attention of biological systems with synchronisation processes in asynchronous neural networks (Wróbel, 2000; Niebur et al., 2002; Grammont & Riehle, 2003; Gross et al., 2004; Usher, 2006). Further parts of this chapter will describe a proposition of artificial neuron model that try to fulfil mentioned requirements, and realise low-level selective attention by conforming to the Noton & Stark scan path theory.

4. Synchronous, conditional signals aggregation as a key to distributed selective attention

Analysis of the properties of aggregation functions presented in the previous subchapters leads to the conclusion that up to date neuronal models of low-level selective attention can be enriched by using synchronous data processing along with input connections grouping and stepwise conditional input signals accumulation. As even simple perceptron neuron model is synchronous in nature – all its input signals are treated like if were received in the same time – it should be possible to extend its attention directing abilities by dividing its input connections into groups, defining proper sequence in which signals from different groups should be read in and processed, and specifying a condition that stops signals aggregation when is met. In general this is a very difficult task, especially finding input's grouping and sequence of input signals processing proper for a given task. But, as it will be shown in further sections, with additional assumptions on the inputs grouping algorithm and the form of scan path used to analyse them, the problem can be essentially reduced and effectively solved. In the effect, proposed synchronous conditional aggregation strategy and resulting neuron models can be easily used as one-to-one nodes replacement in widely used fully connected, multilayer perceptron neural networks. A basic but legible example of such a solution is Sigma-if neuron along with Sigma-if neural network (Huk, 2004; Huk, 2007).

4.1 The Sigma-if neuron

The Sigma-if neuron, in contrast to a typical perceptron neuron, aggregates input signals for the given data vector $\mathbf{X}=[x_1, x_2, \dots, x_N]$ not in one but in a series of given K steps according to the corresponding state graph. Its input connections are divided into K discrete subsets during the training process. Subsequently, when the neuron's aggregation function value is computed, in every k -th step of this process, the subset of input signals X_k is taken from the environment and processed to determine the current value of the partial activation level

φ_k of the neuron. The process continues until the value of φ_k exceeds a given aggregation threshold φ^* . When that condition is met, signals which were not analysed are ignored, and φ_k is considered the input value for the neuron's activation function F . As a result, signal level information coding and even the use of a non-local activation function (e.g. sigmoid) do not degrade the neuron's selective attention abilities. A sample scheme of a state graph for a Sigma-if neuron is presented in Fig. 1.

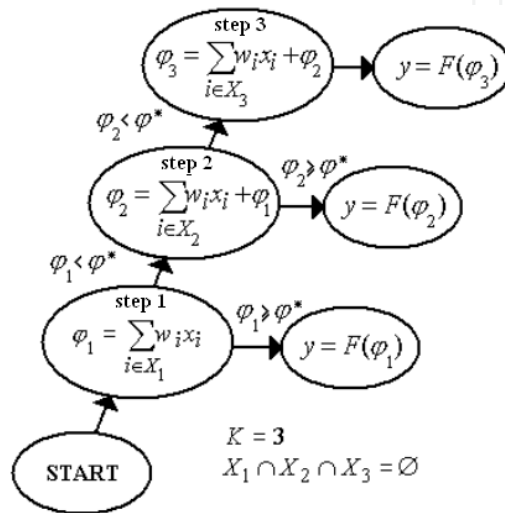


Fig. 1. Sample process of a three-step input signal aggregation in the Sigma-if neuron

Formally speaking, N dendrites of the Sigma-if neuron are divided into K distinct groups, by complementing each i -th input connection with an additional integer parameter $\theta_i \in \{0, \dots, K-1\}$, determining membership in one of the groups. This allows us to divide the process of signal accumulation into K steps, where K is a function of the neuron's grouping vector $\theta^T = [\theta_1, \theta_2, \dots, \theta_N]$:

$$K(\theta) = \max_{i=1}^N \theta_i. \quad (10)$$

During each step k (from 0 to $K-1$), the neuron accumulates data belonging to one selected group, such that

$$\theta_i = k. \quad (11)$$

Within each k -th group, partial activation $\Delta \varphi_k$ is determined as a weighted sum of input signals and the appropriate Kronecker's delta:

$$\Delta\varphi_k(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta}) = \sum_{i=1}^N w_i x_i \delta(k, \theta_i), \quad (12)$$

where w_i and x_i are coefficients of the neuron's weight vector \mathbf{w} and an input vector \mathbf{x} . This process is repeated until the actual activation φ_k derived from respective inputs groups exceeds a preselected aggregation threshold φ^* . It can be described by the following recursive formula (vectors \mathbf{w} , \mathbf{x} and $\boldsymbol{\theta}$ are omitted for clarity):

$$\varphi_k = \begin{cases} \Delta\varphi_k \cdot H(\varphi^* - \varphi_{k-1}) + \varphi_{k-1} & : k \geq 0 \\ 0 & : k < 0' \end{cases} \quad (13)$$

where H is Heaviside step function. This sum is then treated as the neuronal activation value. The input from remaining (heretofore unconsidered) groups is neglected. Thus, the proposed form of the aggregation function $\varphi_{\text{Sigma-if}}$ is:

$$\varphi_{\text{Sigma-if}}(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta}) = \varphi_K(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta}). \quad (14)$$

In the final stages of determining the output value Y of the neuron, function (14) serves as a parameter of the nonlinear threshold (e.g. sigmoidal) function F :

$$Y(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta}) = F(\varphi_{\text{Sigma-if}}(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta})). \quad (15)$$

The described model assumes that the state graph used during signal aggregation is always a simple directed path of non-terminal nodes corresponding with the neural activation accumulation procedure. In a general case, the Sigma-if neuron, besides the weights vector \mathbf{w} , includes one continuous valued parameter for aggregation threshold φ^* , and an additional connections grouping vector $\boldsymbol{\theta}$ with only one nominal valued coefficient for each neuronal input connection.

It is easy to find, that if all coefficients of grouping vector have the same value (what means that all input connections belong to the same group), the functionality of Sigma-if neuron is equivalent to the functioning of the perceptron. Similar effect appears also in the case when there are many different groups and after aggregation of all $\Delta\varphi$ the total activation is not greater than threshold value φ^* , or is greater than φ^* just after adding $\Delta\varphi_K$. But it is much more interesting how Sigma-if neurons function, when for given input values:

$$\exists_{k^* < K} : \varphi_{k^*}(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta}) \geq \varphi^*. \quad (16)$$

This corresponds to the case, when due to suitable selection of \mathbf{w} and $\boldsymbol{\theta}$ vectors, for a given input vector Sigma-if neuron reads in and analyse only a part of available input signals. The value k^* defines the number of inputs groups, which are used to determine the activation of the neuron.

It also is important to notice, that assumption that introduced value k^* is known for every input vector, helps to avoid recursive formula (13) and to write that:

$$\varphi_{\text{Sigma-if}}(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^{k^*} \Delta \varphi_k(\mathbf{w}, \mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^{k^*} \sum_{i=1}^N w_i x_i \delta(k, \theta_i). \quad (17)$$

Above form of proposed aggregation function reflects actual simplicity of the conditional inputs accumulation method, and is more suitable to use in practice and in theoretical considerations than its recursive equivalent. However it should be remembered, that in practice k^* value remains unknown, until the end of calculation of Sigma-if neuron activation.

4.2 Sigma-if network architecture and training

Proposed Sigma-if neurons can be used to build a simple model called Sigma-if neural network, which possess selective attention abilities. The proposed neural network is a generalisation of synchronous, three layer, fully connected perceptron neural network (MLP), in which hidden perceptron neurons were changed to the Sigma-if neurons. Such a neural network does not need separate centralised attention guidance module. This is because its ability to realise selective attention functionality emerges as an effect of synergy between its hidden Sigma-if neurons.

In this place it is good to ask, how described properties of Sigma-if network can be achieved? It is easy to see, that in comparison to MLP neural network training, searching for an optimal set of Sigma-if network parameters would be very computationally challenging, due to the non-continuous character of Sigma-if neuron grouping vector. The answer to this general question is that while we don't know quick and effective method for global searching for optimal Sigma-if network weights and grouping vectors, the problem should be reduced to a simpler one.

The proposed solution assumes that at each Sigma-if neuron, coefficients of the grouping vector $\boldsymbol{\theta}$ are in fact direct functions of weight vector \mathbf{w} . In this work, the grouping vector computation procedure (as well as the predefined value of the aggregation threshold φ^*) is common for all Sigma-if neurons. It simply divides input connections into a given number of groups of similar sizes, according to "the greater the connection weight, the smaller the connection group number" principle. Such a search problem reduction leads to very interesting results, and allows practical elimination of the additional $\boldsymbol{\theta}$ vectors after the end of neural network training process. However, during further analysis of the general Sigma-if model, it is still very helpful to use the grouping vector concept.

As a result, network connection weights are established by the well known error backpropagation algorithm, but for every ω training epoch, actual grouping vectors are computed. This reflects the application of the self-consistency idea widely used in physics. According to this idea, two sets of mutually dependent parameters of a system converge to the optimum, during repeated forcing known directions of improvement at parameters values of at least one of those sets (Kohn & Sham, 1965; Mannheim, 1975; Noyh et al., 1991; Fonseca et al., 1998; Raczkowski et al., 2001). During the training, this method allows also regulation of ratio of mutual dependency between neurons weights and grouping vectors, by setting the number of training epochs ω , after which actualisation of the grouping

vector takes place. It is interesting, when ω tends to infinity, described training algorithm tends to be identical to original backpropagation method.

The last element needed to carry out described backpropagation process, is an information about that how proposed conditional aggregation function of the Sigma-if neuron changes methods of calculating errors and delta rule values for each layer of Sigma-if neural network. Fortunately it is easy to show, that when the Sigma-if aggregation method φ is given by expression (17), the output error $\delta_j^{m\mu}$ of j -th neuron in m -th hidden layer of Sigma-if neurons for the input vector of number μ is given by the following equation:

$$\delta_j^{m\mu} = F'(\varphi_j^{m\mu}) \sum_{l=1}^{n_{m+1}} \delta_l^{(m+1)\mu} w_{l,j}^{m+1} H(k_l^{*(m+1)\mu} - \theta_{l,j}^{m+1}), \quad (18)$$

where n_{m+1} is the number of neurons in $m+1$ neurons layer, F is neuron's activation function and H is a two-valued Heaviside unit step function. Presented equation is different from analogous expression for the perceptron network only by Heaviside step function. This change causes that when some of neuron's input connections are not used during calculation of its activation, they do not influence neuron's output error, even when their weights and input signals are nonzero.

In turn, by using equation (18), the value of general delta rule $\Delta w_{ji}^{m\mu}$ used to update weight of the connection, between j -th neuron in m -th neuron layer and i -th neuron in $m-1$ layer, can be defined as:

$$\Delta w_{ji}^{m\mu} = \eta \delta_j^{m\mu} u_i^{(m-1)\mu} H(k_j^{*m\mu} - \theta_{j,i}^m), \quad (19)$$

where η is a learning coefficient and $u_i^{(m-1)\mu}$ is an output value of i -th neuron in $m-1$ network layer for given input vector μ . Repeated appearance of Heaviside function is here natural for backpropagation algorithm — while connections not active during input signals aggregation do not influence neuron's error, thus their weights should not be updated.

At the end of this section it is also important to notice, that the procedure of grouping vector modification according to weights values is run only after the phase of connections weights update. Thus, while we do not want to determine Sigma-if neural network behaviour just after random initialisation of neurons connections weights, at the beginning of the training process Sigma-if neurons are reduced to perceptron neurons by setting all coefficients of their grouping vectors to 1.

5. Properties of the Sigma-if model

The conditional aggregation method of the Sigma-if neuron can be considered also from the perspective of its decision space characteristics. In particular it is interesting how its form evolves during single input values aggregation. Before outputting a selected value, Sigma-if neuron can consider multiple hypotheses involving its input signals. Speaking in formal terms, a trained neuron before output value calculation can do many attempts to partition

the data space with hypersurfaces using an increasing number of dimensions, where the maximal number of attempts is determined by the number of its input groups K . In the effect, the final dimensionality of Sigma-if neuron decision space is known just after determination of its total activation value, and effective form of its decision borders is a composition of individual hypersurfaces in appropriate variables value ranges, that with respect to the order of signals aggregation. This capability significantly extends the classification potential of neural processing units, and introduces neuron's selective attention behaviours.

Experience dictates that a single Sigma-if neuron, unlike the perceptron neuron, is capable of solving linearly non-separable problems, despite using a linear or sigmoid non-local threshold function. The simplest example of such an approach is a two-input Sigma-if neuron, which, provided both dendrites belong to different input groups, can implement a function defined by expression (20):

$$D_{LNS}(x_1, x_2) = \begin{cases} 0 : (x_1, x_2) \in \{(0,0), (0.8,0.8)\} \\ 1 : (x_1, x_2) \in \{(0,1), (1,0)\} \end{cases} . \quad (20)$$

Let's now assume that components of the grouping vector θ_1 and θ_2 connected adequately with input connections x_1 and x_2 are not equal – e.g. that θ_1 is less than θ_2 . Then neuron's input x_2 is read in only when partial neuron activation $\Delta\varphi_1$ (equal to the product of w_1 and x_1) is lesser than activation threshold φ^* . This allows a division of input signals classification process into two stages. In the first phase, aggregation method considers only input space dimension connected with variable x_1 , and all input vectors are perceived by neuron as adequate projections to a one dimensional subspace defined by axis X_1 . Adequately, connection weight w_1 and aggregation threshold φ^* define a boundary line that separates region of the full decision space, which analysis is based only on signals from input x_1 , from the region for which neuron's output value is calculated in the second aggregation phase with use of all input signals. However in the one dimension this border is reduced to the point, in the given example defined as:

$$c^* = \frac{\varphi^*}{w_1}, \quad (21)$$

that after considering also the second input variable – that is in the full decision space – it has the form of a straight line that passes through the point c^* , and is perpendicular to the axis X_1 (in general case this is a hyperplane).

With a fixed value of threshold φ^* , proper selection of w_1 allows such division of decision space that breaks linearly non-separable problem into two subproblems that are linearly separable. Since both partial problems are being solved in subspaces of different number of dimensions, it should be remembered that each of the subproblems should have to be linearly separable only in the subspace it is associated with. In the given example, the partial problem that is solved only with use of variable x_1 thus must be linearly separable in one dimensional space. As it is shown on the Fig. 2, in a solved problem this condition can be met.

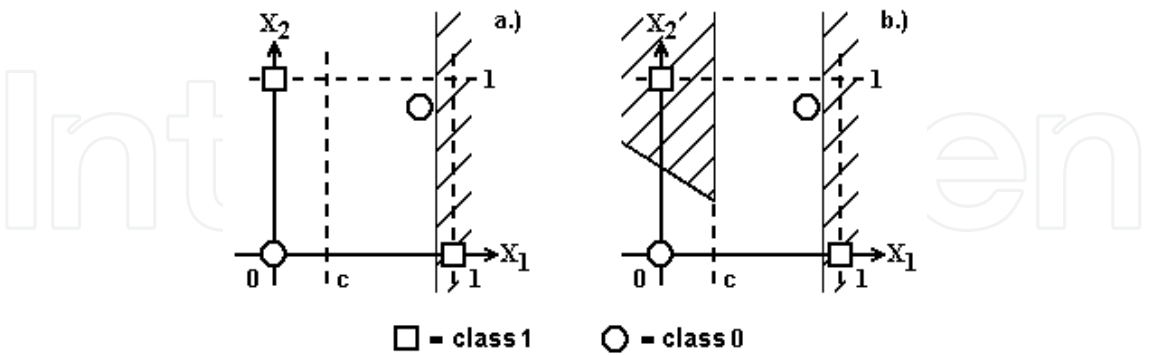


Fig. 2. Phases of two-input Sigma-if neuron decision space evolution, during step-by-step accumulation of input signals ($\theta^T=[1; 2]$, $w^T=[1,3; 2,5]$, $\varphi^*=0,4$). a.) division into two subproblems and solution of the one dimensional problem, b.) composition of decision regions of the first and the second phase

During analysis of the above illustration it should be remembered, that in case of Sigma-if neuron, the final shape of decision borders is also determined by bipolar, sigmoidal activation function and necessity of neuron output values discretization. In the given example, neuron output values that were less or equal 0.5 were treated as class 0, and other values were associated with class 1. Results for chosen input vectors are presented in table 1.

Nr	Input vector (x_1, x_2)	Number of used input values	Total neuron activation	Output value Y	Class
1	(0; 0)	2	0,00	0,00	0
2	(0; 1)	2	2,50	0,85	1
3	(1; 0)	1	1,30	0,57	1
4	(0,8; 0,8)	1	1,04	0,48	0

Table 1. Numerical quantities depicting classification of chosen input vectors by Sigma-if neuron ($\theta^T=[1; 2]$, $w^T=[1,3; 2,5]$, $\varphi^*=0,4$)

On the basis of above analysis, it is easy to point an example of the problem, which is not solvable by single Sigma-if neuron. While sometimes it is not possible to fulfil the criteria of linear separability for all partial problems produced by step-by-step input signals aggregation, proposed solution fails for the original XOR problem. In this case each straight line perpendicular to one of the dimensions of the data space and passing through a selected point of input data belonging to class C, also necessarily contains a point from a class different than C, making it impossible to discern between both points through the use of straight lines perpendicular to data space versors. Such problems may, however, be attacked through rotating the entire coordinate system by a given acute angle - a transformation, which enables the Sigma-if neuron to properly classify points defining the XOR function. It is also worth to mention, that Sigma-if neuron is also capable of solving AND and OR problems (Huk, 2006).

6. Experimental verification of Sigma-if model properties

Above theoretical analysis of the Sigma-if neuron and its conditional aggregation function, was extended and verified by examination of the whole Sigma-if network properties in practical applications, on the example tasks of classification of the UCI Machine Learning benchmark problems. During examination, simulated Sigma-if neural networks were compared to MLP networks with the same architectures. All networks were fully connected and had one hidden layer with the number of neurons for which the MLP network in preliminary tests gained highest test data classification accuracy. Numbers of inputs groups K and aggregation threshold φ^* values of all hidden Sigma-if neurons in the given network were equal. As the sigmoidal perceptron is a special case of a Sigma-if neuron, MLP networks were in fact simulated by Sigma-if networks with the number of inputs groups K of all Sigma-if neurons set to one. In all cases, standard input signal coding was used, and answers of the network were computed in the winner-takes-all manner.

Along with classification accuracies u for training and γ for test data, properties such as neural network data processing time τ , as well as hidden connections and network input activity (designated hca and nia respectively) were considered. The data processing time τ for all trained networks was measured to check relative data processing costs for MLP and Sigma-if networks. Regardless of the very precise time measurement procedure used, actual timings on other hardware setups may vary considerably. Hidden connection activity hca and network input activity nia were used to representing the percentage ratio of the number of hidden and input connections used during data processing, compared to all of the network's hidden and input connections respectively. These parameters allowed check if hidden Sigma-if neurons use their selective attention ability in practice. For the completeness of analysis, for each given problem and trained network, the percent of all inputs used to classify all test vectors niu was calculated. This procedure was important in order to determine if selective attention functionality is also realised at the level of the whole Sigma-if network. All measured values were calculated as averaged outcomes of ten independent 10-fold cross validations.

The experiments were divided into two groups: examination of Sigma-if network classification properties and verification of the hypotheses that Sigma-if networks realise selective attention at the level of single neurons and of the whole network. In both cases all of the considered properties were analysed in relation to the number of Sigma-if neuron inputs groups K . This was because the K value has the highest influence on the properties of the proposed network. Other parameters, such as the aggregation threshold φ^* and the grouping vector actualisation interval ω , were set to 0.6 and 25 respectively, following preliminary tests.

The obtained results indicate that increasing the number of Sigma-if neuron input groups K to more than one, results in an increase of test data classification accuracy γ as well as in simultaneous decrease of data processing time τ . The drawback here is a decrease of training data classification accuracy u . Typical examples of such a dependencies can be observed for the HeartC and Wine problems, which are presented in Fig. 3 and Fig. 4 (as the number of input connections groups K is discrete, values on presented figures are connected only to ease analysis of the results).

The observed decrease in training data classification accuracy u for the HeartC problem was caused by the fact that it is harder to learn when the neuron's input space is changed

every ω epoch. However, and more importantly, the obtained increase in test data classification accuracy γ is a result of rejecting redundant or noisy signals from processed data and the consequence of effective reduction of problem dimensionality. This thesis is confirmed by the observed simultaneous decrease in data processing time τ . Reduction of τ can only be caused by reduction of the network's hidden and input connections activities. However, regardless of the reasons, these results show that the Sigma-if neural network can have better classification properties than MLP.

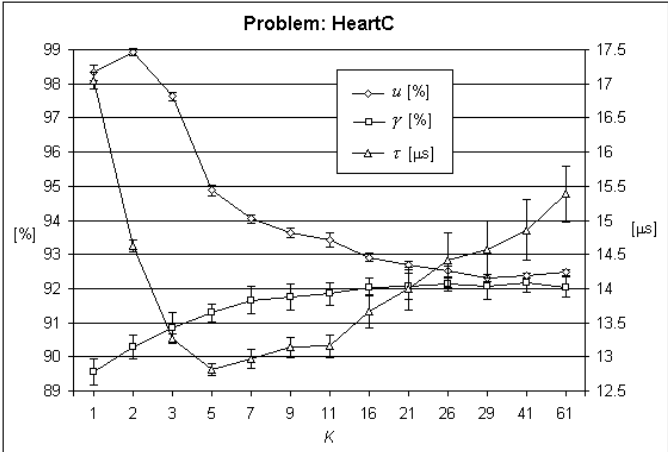


Fig. 3. The time of Sigma-if network output signal generation τ , the classification accuracy of training u and test γ data for the HeartC problem versus the number of hidden neuron input connections groups K (networks architecture: 28 inputs, 10 hidden neurons, 5 outputs)

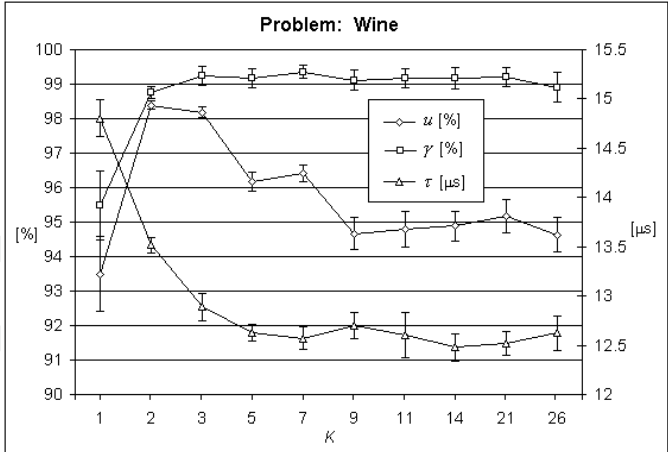


Fig. 4. The time of Sigma-if network output signal generation τ , the classification accuracy of training u and test γ data for the Wine problem versus the number of hidden neuron input connections groups K (networks architecture: 13 inputs, 10 hidden neurons, 3 outputs)

The visible increase of HeartC data processing time τ for K greater than 7 inputs groups is the effect of a linear increase of time cost, connected with the existence of additional instructions for grouping vector θ information processing. This factor can be easily seen for the number of groups K greater than the given number of network inputs. Without it, data processing time would semi-logarithmically decrease with rising K , similarly as in the case of Wine problem, for which low number of network inputs and connections between hidden and input layer, make this effect negligible. This indicates the character of the changes of Sigma-if network hidden hca and input connection activities nia as a function of K , which can be observed in Fig. 5 (for the Sonar problem) and in Fig. 6 (for the Votes problem).

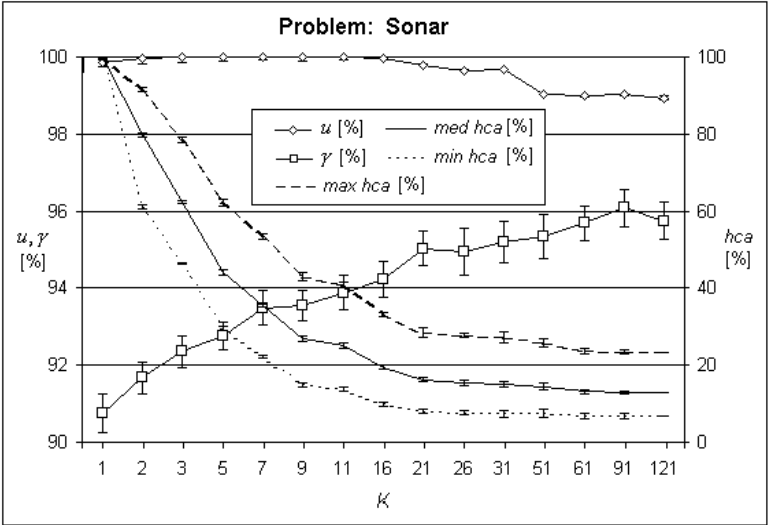


Fig. 5. The Sigma-if network hidden connection activity hca , the classification accuracy of training u and test γ data for the Sonar problem versus the number of hidden neuron input connections groups K (network architecture: 60 inputs, 30 hidden neurons, 2 outputs)

It can be easily seen that in the case of the Sonar problem the increase in K causes an increase of the test data classification accuracy, with a corresponding slight decrease of the training data classification accuracy. These changes are accompanied with a much stronger reduction of hidden connection activities. The shape of the $hca(K)$ function confirms earlier conclusions that the data processing time reduction is connected with Sigma-if neurons' selective attention abilities. All this is clear evidence that Sigma-if neurons use selective attention, and that this can reduce the generalization error level as well as data processing costs. But results for the Wine problem show that these savings are not always gained for the price of lowering training data accuracy (Fig. 4).

The last example concerns how selective attention abilities manifest themselves on the level of the whole Sigma-if network. The analysis of results for the Votes problem (Fig. 6) shows that when a significant decrease of network input activity nia occurs, one can expect a simultaneous reduction in the number of Sigma-if network inputs used to classify data (niu) without a notable decrease of classification accuracy in comparison to the MLP network.

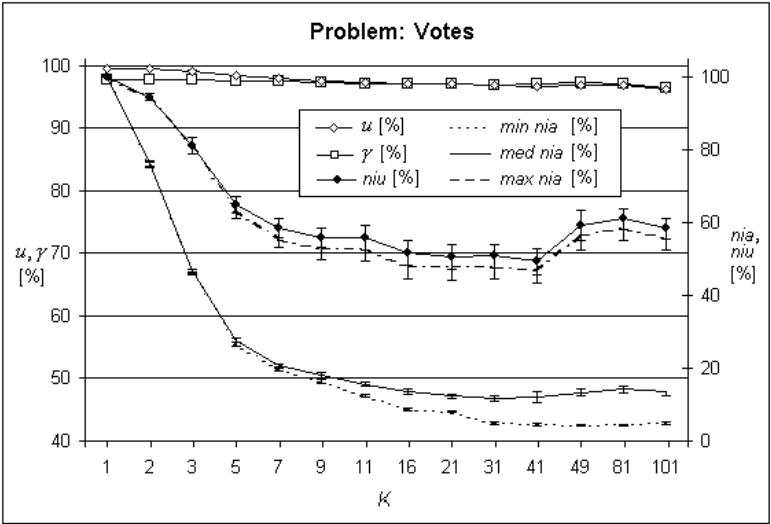


Fig. 6. The Sigma-if network input activity nia , the number of network input used niu , and the classification accuracy of training u and test γ data for the Votes problem versus the number of hidden neuron input connections groups K (network architecture: 48 inputs, 2 hidden neurons, 2 outputs)

While the Sigma-if network has no specialised or separate attention guiding unit, all such activities can emerge only as the effect of synergy between individual neurons. Thus, the observed selective attention behaviour of the proposed network, treated as a black box, is a significant indication that the Sigma-if model effectively mimics basic aspects of low level attentional processes observed in nature. This can, in turn, make the model an interesting tool for feature selection and other data processing purposes.

7. Conclusion

In this chapter, a novel computational model of distributed neuronal selective attention, based on the scan-path theory and up-to-date findings of neurobiologists, has been presented. In described solution, attention is conceived by new, synchronous, sequential and conditional input signals aggregation function of an artificial neuron. While the resulting Sigma-if network has no specialised or separate attention guiding unit, selective attention emerges at the level of the whole neuronal network as an effect of synergy among the network's hidden neurons.

Introduced idea allowed construction of the whole attentional classification neuronal system along with efficient Sigma-if network training method that is combination of gradient backpropagation algorithm and self-consistency paradigm. Thanks to this, a series of experiments have been conducted with the use of benchmark problems, to test basic properties of Sigma-if network. The model's selective attention ability for medium-size test problems manifests itself in an increase of classification accuracy and in a simultaneous decrease of data processing costs. Observed reduction of the number of network inputs used

to classify data shows the possibility of further reduction of data acquisition costs, during as well as after the network training process.

The Sigma-if selective attention feature also introduces new possibilities in the area of analysing the network decision process via its inputs activity interpretation. This can point at features of given data sets that are most important for classification, and help to identify features that are irrelevant, redundant or contaminated by noise. All this makes the Sigma-if neural network a very useful tool for the data acquisition and analysis domain. As proposed Sigma-if network training method allow automatic construction of selective attention strategies, proposed model can be also a very promising solution for applications such as remote sensing in dispersed sensor networks as well as automatic robot navigation and control.

In fact the Sigma-if neuron is very simple and differs in much details from biological neurons. But from theoretical point of view it can be also treated as an interesting model of low-level human selective attention. In this light it is astonishing that for many of mid-sized benchmark classification problems Sigma-if network gained the best and the quickest classification results, when number of input connections groups was set between 3 and 11. This is because it seems to correlate with real life observations of the maximum number of seven information groups (e.g. displayed on a poster) that average human brain finds easy to analyse. But this should be treated only as a kind of intuition not a strict result.

Due to a very interesting theoretical and practical properties, proposed Sigma-if model should be further tested as well on benchmark as also on real-life data. Also the whole idea of synchronised conditional signals aggregation should be further explored, as other aggregation functions than the one presented in this chapter can be proposed. All this makes a wide and promising direction of research on the neuronal selective attention models, and in this time it is a subject of continuous investigation.

8. References

- Allport, D.A.; Antonis, B. & Reynolds, P. (1972). On the division of attention: a disproof of the single channel hypothesis, *Quarterly Journal of Experimental Psychology*, Vol. 24, pp. 225-235
- Anderson, R.A.; Essich, G.K. & Siegal, R.M. (1985). Encoding of spatial location by posterior parietal neurons, *Science*, Vol. 230, pp. 456-458
- Anderson, C. & Van Essen, D. (1987). Shifter Circuits: a computational strategy for dynamic aspects of visual processing, *Proceedings of National Academy of Sciences, USA* 84, pp. 6297-6301
- Baldassi, S. & Verghese, P. (2002). Comparing integration rules in visual search, *Journal of Vision*, Vol. 2, pp. 559-570
- Broadbent, D.E. (1982). Task combination and selective intake of information. *Acta Psychologica*, Vol. 50, pp. 253-290, 0001-6918
- Bryden, M.P. (1971). Attentional strategies and short-term memory in dichotic listening, *Cognitive Psychology*, Vol. 2, pp. 99-116.
- Bukovsky, I.; Zeng-Guang Hou; Bila, J. & Gupta, M.M. (2007) Foundation of Notation and Classification of Nonconventional Static and Dynamic Neural Units, *Proceedings of 6th IEEE International Conference on Cognitive Informatics*, pp. 401-407, 978-1-4244-1328-7

- Burkitt, A.N. & Clark, G.M. (1999). Analysis of integrate-and-fire neurons: synchronization of synaptic input and spike output, *Neural Computation*, Vol. 11, No. 4, pp. 871-901, 0899-7667
- Chelazzi, L.; Miller, E.; Duncan, J. & Desimone, R. (1993). A Neural Basis for Visual Search in Inferior Temporal Cortex, *Nature*, Vol. 363, pp. 345-347, 0028-0836
- Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears, *Journal of the Acoustical Society of America*, Vol. 25, No. 5, pp. 975-979, 0001-4966
- Clauss, M.; Bayerl, P. & Neumann, H. (2004). Evaluation of regions-of-interest based attention algorithms using a probabilistic measure, *Proceedings of the 5th. Workshop Dynamic Perception 2004*, pp. 227-232, 3-89838-059-9, Germany, IOS Press, Tübingen
- Cover, T.M. (1965). Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition, *IEEE Transactions on Electronic Computers*, Vol. EC-14, No. 3, June 1965, pp. 326-334, 0367-7508
- Deutsch, J. & Deutsch, D. (1963). Attention: Some theoretical considerations, *Psychological Review*, Vol. 70, No.1, pp. 80-90
- Dole, G.H. (2008). The Philosophy of Creation, Elsevier Science Publishers B.V., 0559702159, Amsterdam
- Durbin, R. & Rumelhart, D.E. (1990). Product units: A computationally powerful and biologically plausible extension to backpropagation networks, *Neural Computation*, Vol. 1, No. 1, pp. 133-142, 0899-7667
- Eckstein, M.P.; Beutter, B.R. & Stone, L.S. (2001). Quantifying the performance limits of human saccadic targeting in visual search, *Perception*, Vol. 30, pp. 1389-1401, 0301-0066
- Findlay, J.M. (1982). Global visual processing for saccadic eye movements, *Vision Research*, Vol. 22, pp. 1033-1046, 0042-6989
- Fonseca, L.R.C.; Jimenez, J.L.; Leburton, J.P. & Martin R.M. (1998). Self-consistent calculation of the electronic structure and electron-electron interaction in self-assembled InAs-GaAs quantum dot structures, *Physical Review B*, Vol. 57, pp. 4017-4026
- Giles, C.L. & Maxwell, T. (1994). Learning, Invariance, and Generalization in High Order Neural Networks, *Applied Optics*, Vol. 26, No. 23, pp. 4972-4978, 0003-6935
- Grammont, F. & Riehle, A. (2003). Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time, *Biological cybernetics*, Vol. 88, No. 5, pp. 360-373, 0340-1200
- Gray, J. & Wedderburn A. (1960). Grouping strategies with simultaneous stimuli, *Quarterly Journal of Experimental Psychology*, Vol. 12, pp. 180-184, 1747-0218
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm, *Perception and Psychophysics*, Vol. 28, No. 4, pp. 267-283, 0031-5117
- Gross, J.; Schmitz, F.; Schnitzler, I.; Kessler, K.; Shapiro, K. & Hommel, B. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, No. 35, August 2004, pp. 13050-13055, 0027-8424
- Gupta, M.M. (2008). Correlative type higher-order neural units with applications, *Proceedings of the IEEE International Conference on Automation and Logistics 2008*, September 2008, pp. 15-718, 978-1-4244-2502-0, Quingdao, China
- Hager, G. & Toyama, K. (1999). Incremental focus of attention for robust visual tracking, *International Journal of Computer Vision*, Vol. 35, No. 1, pp. 45-63

- Houghton, G. & Tipper, S.P. (1996). Inhibitory Mechanisms of Neural and Cognitive Control: Applications to Selective Attention and Sequential Action, *Brain and Cognition*, Vol. 30, pp. 20-43
- Huk, M. (2004). The Sigma-if neural network as a method of dynamic selection of decision subspaces for medical reasoning systems, *Journal of Medical Informatics and Technologies*, Vol. 7, October 2004, pp. KB-65-73, 1642-6037
- Huk, M. (2006). Sigma-if neural network as a use of selective attention technique in classification and knowledge discovery problems solving, *Annales UMCS Informatica AI*, Szczygieł R., Mikołajczak P., Budzyński M., Kamiński W.A., Sielanko J., Złotkiewicz E. (Ed.), Vol 5., No. 2, pp. 121-131, 1732-1360
- Huk, M. (2007). Manifestation of selective attention in Sigma-if neural network, *Proceedings of the International Multiconference on Computer Science and Information Technology IMCSIT/AAIA'07, 2nd International Symposium Advances in Artificial Intelligence and Applications*, Vol. 2, October 2007, pp. 225-236, 1896-7094
- Kahneman, D. (1973). Attention and Effort, Englewood Cliffs, New Jersey, Prentice-Hall
- Karlholm, J.M. (1993) Associative memories with short-range, higher order couplings, *Neural Networks*, Vol. 6, pp. 409-421
- Kastner, S.; De Weerd, P.; Desimone, R. & Ungerleider, L. (1998). Mechanisms of Directed Attention in the Human Extrastriate Cortex as Revealed by Functional MRI, *Science*, Vol. 282, pp. 108-111, 0036-8075
- Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry, *Human Neurobiology*, Vol. 4, pp. 219-227
- Kohn W. & Sham LJ. (1965). Self-Consistent Equations Including Exchange and Correlation Effects, *Physical Review*, Vol. 140, pp. A1133-A1138
- Körding, K.P. & Köniq, P. (2001) Neurons with two sites of synaptic integration learn invariant representations, *Neural Computation*, Vol. 13, No. 12, pp. 2823-2849, 0899-7667
- Kucera, H. & Francis, W.M. (1967). *Computational analysis of present-day American English*, Brown University Press
- LaBerge, D. (1990). Thalamic and cortical mechanisms of attention suggested by recent positron emission tomographic experiments, *Journal of Cognitive Neuroscience*, Vol. 2, pp. 358-372, 0898-929X
- Larkum, M.E.; Zhu, J.J. & Sakmann, B. (1999) A new cellular mechanism for coupling inputs arriving at different cortical layers, *Nature*, Vol. 398, pp. 338-41, 0028-0836
- Lee, Y. (2000). An information-theoretic framework for understanding saccadic behaviors. *Advances in Neural Processing Systems*, Vol. 12, pp. 834-840
- Leerink, L.R.; Giles, C.L.; Horne, B.G. & Jabri, M.A. (1995). Learning with Product Units, *Advances in Neural Information Processing Systems 7*, Tesauero, G.; Touretzky, D.; Leen T. (Ed.), pp. 537-544, MIT Press
- Lewis, J.L. (1970). Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, Vol. 85, pp. 220-227
- Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models, *Neural Networks*, Vol. 10, No. 9, pp. 1659-1671
- Marslen-Wilson, W.D. & Tyler, L.K. (1980). The temporal structure of spoken language understanding, *Cognition*, Vol. 8, pp. 1-71

- Martínez-Estudillo, C.; Hervás-Martínez, P.A.; Gutiérrez, & A.C. Martínez-Estudillo, Evolutionary product-unit neural networks classifiers, *Neurocomputing, Life System Modelling, Simulation, and Bio-inspired Computing (LSMS 2007)*, Vol. 72, No. 1-3, December 2008, pp. 548-561,
- Mel, B.W. (1999). Why have dendrites? A computational perspective, *Dendrites*, Stuart G., Hausser M., pp. 271-289, Oxford University Press
- Mel, B. W. (1992). The clusteron: toward a simple abstraction for a complex neuron, *Advances in Neural Information Processing Systems*, Vol. 4, pp. 35-42, Morgan Kaufmann
- Mel, B. W. (1990). The sigma-pi column: a model of associative learning in cerebral cortex. *Technical Report CNS Memo 6, Computation and Neural Systems Program*, California Institute of Technology
- Mannheim, P.D. (1975). Dynamical symmetry breaking as a bootstrap, *Physical Review D*, Vol. 12, p. 1772-1793
- Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention, *Journal of Experimental Psychology: General*, Vol. 106, pp. 226-254, 0096-3445
- Neville, R. S.; S. (2002). Eldridge, Transformations of sigma-pi nets: obtaining reflected functions by reflecting weight matrices, *Neural Networks*, Vol. 15, No. 3, pp.: 375-393, 0893-6080, Elsevier Science Ltd. Oxford, UK
- Niebur, E.; Hsiao, S.S. & Johnson, K.O. (2002). Synchrony: a neuronal mechanism for attentional selection?, *Current Opinion in Neurobiology*, Vol. 12, No. 2, April 2002, pp. 190-194, 0959-4388.
- Niebur E.; Koch C. & Rosin C. (1993). An oscillation-based model for the neural basis of attention. *Vision Research*, Vol. 33, pp. 2789-2802
- Noh, T.W.; Song, P.H. & Sievers, A.J. (1991). Self-consistency conditions for the effective-medium approximation in composite materials, *Physical Review B*, Vol. 44, No. 11, pp. 5459-5464,
- Noton D. & Stark L. (1971). Scanpaths in Saccadic Eye Movements While Viewing and Recognizing Patterns, *Vision Research*, Vol. 11, pp. 929-942, 00426989
- Olshausen B.; Anderson C. & Van Essen, D. (1993). A Neurobiological Model of Visual Attention and Invariant Pattern Recognition based on Dynamic Routing of Information, *The Journal of Neuroscience*, Vol. 13, pp. 4700-4719
- Pelc, T. (1998). A formal model of an artificial neural network used to store and recognize the semantics of some sentences of natural language, *Proceedings of the Fifth International Conference on Neural Information Processing ICONIP'98*, pp. 21-23, 9789051994636, October 1998, IOA Press
- Perantonis, S.J. & Lisboa, P.J. (1992). Translation, rotation, and scale invariant pattern recognition by high-order neural networks and moment classifiers, *IEEE Trans. Neural Networks*, Vol. 3, pp. 241-251
- Privitera, C.M.; Azzariti, M. & Stark, L.W. (2000). Locating regions-of-interest for the Mars Rover expedition, *Journal of Remote Sensing*, Vol. 21, No. 17, pp. 3327-3347
- Privitera, C. & Stark, L.W. (2000). Algorithms for Defining Visual Region-of-Interest: Comparison with Eye Fixations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 9, pp. 970-982

- Pynte, J.; Do, P. & Scampa, P. (1984). Lexical decisions during the reading of sentences containing polysemous words, In: *Preparatory States and Processes*, S. Kornblum, J. Requin (Ed.), Hillsdale NJ, Erlbaum
- Quiles, M.G., Breve F., Romero R.A.F., Zhao L. (2008). Visual Selection with Feature Contrast-Based Inhibition in a Network of Integrate and Fire Neurons, Fourth International Conference on Natural Computation, ICNC 2008, Vol. 3, pp. 601-605, 978-0-7695-3304-9
- Raczkowski, D.; Canning, A. & Wang, L.W. (2001). Thomas-Fermi charge mixing for obtaining self-consistency in density functional calculations, *Physical Review B*, Vol. 64, No. 12, pp. 121101-121105
- Redding, N.J.; Kowalczyk, A. & Downs, T. (1993). Constructive higher order network algorithm that is polynomial time, *Neural Networks*, Vol. 6, pp. 997-1010
- Renninger, M. (2004). Sequential information maximization can explain eye movements in an object learning task, *Journal of Vision*, Vol. 4, No. 8, pp. 744a
- Rybak, I.A.; Gusakova, V.I.; Golovan, A.V.; Podladchikova, L.N. & Shevtsova, N.A. (1998). A model of attention-guided visual perception and recognition, *Vision Research*, Vol. 38, pp. 2387-2400
- Schall, J. & Hanes, D. (1993). Neural Basis of Saccade Target Selection in Frontal Eye Field during Visual Search, *Nature*, Vol. 366, pp. 467-469, 0028-0836
- Schmitt, M. (2002). On the Complexity of Computing and Learning with Multiplicative Neural Networks, *Neural Computation*, Vol. 14, No. 2, pp. 241-301
- Schmitt, M. (2000). VC dimension bounds for product unit networks, *Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, Vol. 4, pp. 165-170
- Shiffrin, R.M. & Schneider, W. (1997). Controlled and automatic human information processing: Perceptual learning, automatic attending and a general theory, *Psychological Review*, Vol. 84, pp. 127-190
- Sinha, M.; Gupta, M.M. & Nikiforuk, P.N. (2001). A compensatory wavelet neuron model, *Proceedings of IFSA World Congress and 20th NAFIPS International Conference*, Vol. 3, July 2001, pp. 1372-1377
- Spelke, E.; Hirst, W. & Neisser, U. (1976). Skills of divided attention, *Cognition*, Vol. 4, pp. 215-230
- Sperling, G.A. (1984). A unified theory of attention and signal detection, In *Varieties of attention*, Parasuraman R., Davies D.R. (Ed.), pp. 103-181. New York: Academic Press
- Spratling, M.W. & Hayes, G. (2000). Learning Synaptic Clusters for Nonlinear Dendritic Processing, *Neural Processing Letters*, Vol. 11, No. 1, pp. 17-27
- Stein R.B. (1967). The information capacity of nerve cells using a frequency code, *Biophysical Journal*, Vol. 7, pp. 797-826
- Stuart, G.J. & Spruston, N. (1998) Determinants of voltage attenuation in neocortical pyramidal neuron dendrites, *The Journal of Neuroscience*, Vol. 18, pp. 3501-3510
- Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects, *Journal of Verbal Learning and Verbal Behaviour*, Vol. 18, pp. 645-659
- Swinney, D.A. (1982). The structure and time-course of information interaction during speech comprehension: Lexical segmentation, access, and interpretation, In *Perspectives on Mental Representation*, J. Mehler, E.C.T. Walker & M. Garrett (Ed.), Hillsdale, NJ: Erlbaum

- Tinbergen, N. (1951). *The Study of Instinct*, Oxford: Clarendon Press, Oxford
- Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, Vol. 20, pp. 12-16
- Treisman, A.M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, Vol. 12, pp. 242-248, 1037-1054.
- Tsal, Y. (1983). Movements of attention across the visual field, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 9, No 4., pp. 523-530, 0096-1523
- Tsotsos, J.K.; Culhane, S. & Cutzu, F. (2001). From foundational principles to a hierarchical selection circuit for attention, In: *Visual Attention and Cortical Circuits*, Braun J, Koch C, Davis J. (Ed.), pp. 285-306, MIT Press, 02620-24934, Cambridge, MA,
- Tsotsos, J.K.; Culhane, S.; Wai, W.; Lai, Y.; Davis, N.& Nuflo, F. (1995). Modeling visual attention via selective tuning, *Artificial Intelligence*, Vol. 78, No. 1-2, pp. 507-547
- Usher, M. (2006). What has been learned from computational models of attention, *Neural Networks, Special Issue: Brain and Attention*, Vol.19, No.9, November 2006, pp. 1440-1442, 0893-6080
- van den Bergh, F. & Engelbrecht, A.P. (2001). Training product unit networks using cooperative particle swarm optimisers, *Proceedings of International Joint Conference on Neural Networks*, Vol. 1, pp. 126-131, Washington DC, USA
- VanRullen, R. & Koch, C. (2003). Visual Selective Behavior can be Triggered by a Feedforward Process, *Journal of Cognitive Neuroscience*, Vol. 15, No. 2, pp. 209-217, 0898-929X
- VanRullen, R.; Reddy, L. & Koch, C. (2004). Visual search and dual-tasks reveal two distinct attentional resources, *Journal of Cognitive Neuroscience*, Vol. 16, No. 1, pp. 4-14
- Venkatesh, S.S. & Baldi, P. (1991). Programmed interactions in higher order neural networks: maximal capacity, *Journal of Complexity*, Vol. 7, pp. 316-337
- Weber, C. & Wermter, S. (2007), A self-organizing map of sigma-pi units, *Neurocomputing*, Vol. 70, No. 13-15, August 2007, pp. 2552-2560, 0925-2312
- Wright, R.D., & Ward, L.M. (2008). *Orienting of Attention*, Oxford University Press
- Wróbel, A. (2000). Beta activity: a carrier for visual attention, *Acta neurobiologiae experimentalis*, Vol. 60, No. 2, pp. 247-260, 0065-1400, Warsaw
- Yadav, R.N.; Kalra, P.K. & John, J. (2006). Neural network learning with generalized-mean based neuron model, *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, Vol. 10, No. 3, February 2006, pp. 257-263, 1432-7643, Springer Berlin / Heidelberg
- Yadav, R.N.; Kumar, N.; Kalra, P.K. & John, J. (2004). Multi-layer neural networks using generalized-mean neuron model, *Proceedings of IEEE International Symposium on Communications and Information Technology ISCIT 2004*, Vol. 1, Issue, October 2004, pp. 93 - 97
- Yang, H. & Guest, C.C. (1990). High order neural networks with reduced numbers of interconnectionweights, *Proceedings of IJCNN International Joint Conference on Neural Networks*, Vol. 3, June 1990, pp. 281-286
- Yamada, K. & Cottrell, G.W. (1995). A model of scan paths applied to face recognition, *Proceedings of 17th Ann. Cognitive Science Conference*, pp.55-60, Pittsburgh



Advances in Computer Science and IT

Edited by D M Akbar Hussain

ISBN 978-953-7619-51-0

Hard cover, 420 pages

Publisher InTech

Published online 01, December, 2009

Published in print edition December, 2009

The book presents some very interesting and excellent articles for this divergent title. The 22 chapters presented here cover core topics of computer science such as visualization of large databases, security, ontology, user interface, graphs, object oriented software developments, and on the engineering side filtering, motion dynamics, adaptive fuzzy logic, and hyper static mechanical systems. It also covers topics which are combination of computer science and engineering such as meta computing, future mobiles, colour image analysis, relative representation and recognition, and neural networks. The book will serve a unique purpose through these multi-disciplined topics to share different but interesting views on each of these topics.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Maciej Huk (2009). Learning Distributed Selective Attention Strategies with the Sigma-if Neural Network, *Advances in Computer Science and IT*, D M Akbar Hussain (Ed.), ISBN: 978-953-7619-51-0, InTech, Available from: <http://www.intechopen.com/books/advances-in-computer-science-and-it/learning-distributed-selective-attention-strategies-with-the-sigma-if-neural-network>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen