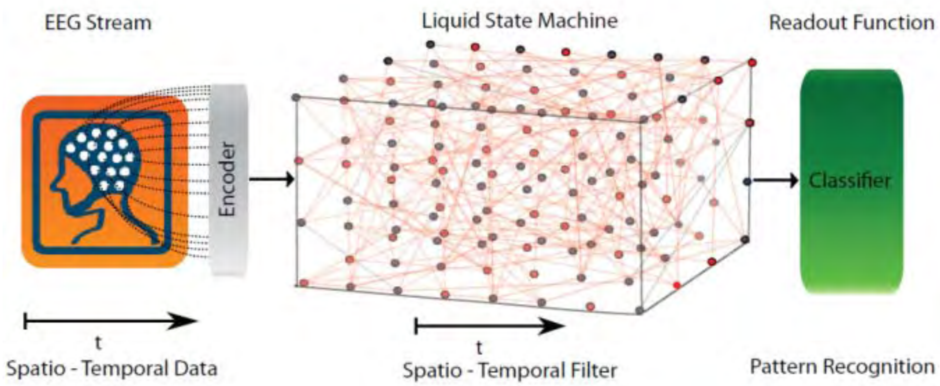
Natural Intelligence

ISSN 2164-8522

The INNS Magazine

Volume 1, Issue 2, Winter 2012



NI Can Do Better Compressive Sensing

Evolving, Probabilistic Spiking Neural Networks and Neurogenetic Systems

Biologically Motivated Selective Attention

CASA: Biologically Inspired approaches for Auditory Scene Analysis

INTERNATIONAL NEURAL NETWORK SOCIETY

Retail: US$10.00

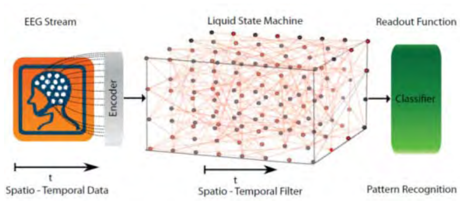
Natural Intelligence: the INNS Magazine Editorial Board

ISSN 2164-8522

Natural Intelligence

The INNS Magazine

Volume 1, Issue 2, Winter 2012



Regular Papers

Columns

News

Call for Papers

A Message from the INNS Vice President for Membership

New Excitements in 2012

Irwin King
INNS Vice-President for Membership

As the Vice-President for Membership, I would like to update you on a few exciting and important items related to INNS members.

First, I am thrilled to share with you once again that we now have our first new Autonomous Machine Learning (AML) Section as it was introduced in our inaugural Natural Intelligence issue. Being elevated from a Special Interest Group (SIG), AML Section enjoys additional benefits for her members in the form of a special track in IJCNN (when organized by IJCNN) and a special issue/section in the new INNS magazine. With this, I look forward to more SIGs be interested in forming new Sections in the years to come.

Second, I would like to share with you some new activities on SIGS and Regional Chapters (RCs). The Spiking Neural Networks SIG led by David Olmsted and Women in Science and Engineering led by Karla Figueiredo are our newly formed SIGs. Based on geographical region, we have a new India RC led by Suash Deb. In additional to the above, we have several communities that are under consideration. They are Biological Neural Networks, Biomedical Applications , Brain Modeling & Neuroscience, and Embedded and Cognitive Robotics. Furthermore, we have China, Korea, and Thailand under consideration to form RCs to further promote INNS in their respective geographical areas. These are exciting news as we continue to expand INNS' reach to emerging topics and new regions.

Third, in recognition of our members' contributions our Senior Membership application process for 2012 will begin soon. If you have been actively participating in INNS events as a member for at least the recent five consecutive years, you could be eligible to be nominated for the Senior Member status. The Call will be sent out shortly and we plan to have the senior membership approved by IJCNN 2012 in Brisbane, Australia this June so polish up your CV and send in your application accordingly.

Lastly, we are looking for enthusiastic volunteers to help INNS with social media projects that aim to promote the society and her members. We are investigating ways to improve our website and complement it with other social media sites such as Facebook, Twitter, etc. If you are interested in getting involved in any of these projects, contact Bruce Wheeler or myself for more information.

Please feel free to send me your inquiries and/or suggestions. I look forward to an existing year working together with you to improve further the benefits for INNS members. ■

NI Can Do Better Compressive Sensing

Harold Szu

Catholic University of America
szuharoldh@gmail.com

Abstract

Based on Natural Intelligence (NI) knowledge, our goal is to improve smartphone imaging and communication capabilities to resemble human sensory systems. We propose adding an enhanced night imaging capability on the same focal plane array (FPA), thus extending the spectral range. Compressive sensing (CS) technology reduces exposure to medical imaging and helps spot a face in a social network. Since Candes, Romberg, Tao, and Donoho (CRT&D) publications in 2007, 300 more contributions have been published in IEEE. What NI does differently is mimic the human visual system (HVS) in both its day-night and selective attention communication capabilities. We consider two killer apps exemplars: Software: generating video Cliff Notes; Hardware: designing day-night spectral camera. NI can do better CS, because of connectionist build-in attributes: fault tolerance filling missing parts; subspace generalization discovering new; unsupervised learning improving itself iteratively.

Keywords: Unsupervised Learning, Compressive Sensing, HVS, Smartphone, Fault Tolerance, Subspace Generalization, Medical Image, Face Identification.

1. Introduction to Compressive Sensing and Natural Intelligence Technologies

Compressive Sensing: Compressive Sensing (CS) technology is motivated by reducing the unneeded exposure of medical imaging, and finding a sparse representation to spot a face in social nets. A large community of CS has formed in the last 5 years, working actively on different applications and implementation technologies. The experience of the International Neural Network Society (INNS) working on traditional computational systems in the last decades developing the unique capabilities of unsupervised learning, cf. Sect. 2, can be beneficial to a larger community, if a concise treatise of learning is made available. Likewise, INNS can benefit from the mathematical insights and engineering implementations of CS.

Face Spotting App: To find a friend, one may turn on a smartphone app for spotting a friendly face among the crowd, or simply surf in Facebook. Such a spotting app may be built upon a massive parallel ANN System On Chip for face detection (SOC-FD), which detects all faces (by color hue pre-processing) in real time and simultaneously places all faces in boxes in 0.04 seconds and identifies whom is smiling and who is not and closed eyes, focusing on the smiling one (by the fuzzy logic post-processing). Each high resolution image on a smartphone has mega pixels on target (pot). Each face has a smaller pot denoted by $N \cong 10^4$. Since the FD-SOC can cut equal-size facial pictures $\{\vec{x}_t, t = 1,2,3,4\}$ at different poses, likewise, the other person $\{\vec{y}_t, t = 1,2,3,4\}$, etc., if so wishes, forms a database $[A]_{N,m} = [\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \vec{y}_1, ., \vec{z}_1, .]_{N,m}$ with *m= 3x4* faces. The app CS algorithm matches an incoming face $\vec{Y}_N$ with the closest face in the database $[A]_{N,m}$. Mathematically speaking, since the database $[A]_{N,m}$ is over-sampled, finding a match is equivalent to finding a sparse representation $\vec{X}_m$ of $\vec{Y}_N$, i.e. $\vec{X}_m$ has few ones (=yes) matched, among many mismatched zeros (=no).

$$\vec{Y}_N = [A]_{N,m}\vec{X}_m \tag{1}$$

Yi Ma *et al.* of UIUC have further applied a down-sampling sparse matrix $[\Phi]_{\mathrm{m,N}}$ to the database $[A]_{N,m}$ for the linear dimensionality reduction for real-time ID in IEEE/PAMI 2007.

Medical Imaging App: Emmanuel Candes of Stanford (formerly at Caltech), Justin Romberg of GIT, and Terrence Tao of UCLA [1,2] as well as David Donoho of Stanford [3] (CRT&D) jointly introduced the Compressive Sensing (CS) sparseness theorem in 2007 IEEE/IT, in order to save patients from unneeded exposure to medical imaging with a smaller number of m views of even smaller number k of radiation exposing pixels as all-pass filter representing ones among zeros. Thus, CS is not post-processing image compression; because otherwise, the patients have already suffered the radiation exposure; thus, CS happened at the sensing measurement. They adopted a purely random sparse sampling mask $[\Phi]_{\mathrm{m,N}}$ consisting of $m \cong k$ number of one's (passing the radiation) among seas of zeros (blocking the radiation). The goal of multiplying such a long horizontal rectangular sampling matrix $[\Phi]_{m,N}$ is to achieve the linear dimensionality reduction from N to m (m<<N), and the reduced square matrix follows:

$$\vec{y}_m = [B]_{m,m}\vec{X}_m, \tag{2}$$

where $\vec{y}_m \equiv [\Phi]_{m,N}\vec{Y}_N$ and $[B]_{m,m} \equiv [\Phi]_{m,N}[A]_{N,m}$. Remarkably, given a set of sparse orthogonal measurements $\vec{y}_m$, they reproduced the original resolution medical image $\vec{X}_N$. CRT&D used an *iterative hard* *threshold (IHT)* of the largest guesstimated entries, known as linear programming, based on the $min.|\vec{X}|_1$ subject to $E=|\vec{y}_m-[B]_{m,m}\vec{X}_m|_2^2\le\varepsilon$, where l_p-norm is defined $|\vec{X}|_p\equiv(\sum_{n=1}^{N}|x_n|^p)^{1/p}$.

ANN supervised learning adopts the same LMS errors energy between the desired outputs $\vec{v}'_i=\vec{y}_m$ & the actual weighted output $v_i=\sigma(u_i)=\sigma(\sum_{j=1}^{m}[W_{i,j}]u_j)$. Neurodynamics I/O is given $du_i/dt=-\partial E/\partial v_i$; Lyapunov convergence theorem $dE/dt\le 0$ is proved for the monotonic sigmoid logic $d\sigma/du_i\ge 0$ in Sect.2. ANN does not use the Manhattan distance, or going around a city-block l_1 distance at $p=1$ because it is known in ANN learning to be too sensitive to outliers. Mathematically speaking, the true sparseness measure is not the l_1-norm but the l_0-norm: $|\vec{X}|_0\equiv(\sum_{n=1}^{N}|x_n|^0)^{1/0}=k$, counting the number of non-zero elements because only non-zero entry raised to zero power is equal to 1. Nevertheless, a practical lesson from CRT&D is that the $min.\ |\vec{X}|_1$ subject to $min.|\vec{X}-\vec{Y}|_2^2$ is sufficient to avoid the computationally intractable $min.\ |\vec{X}|_0$. In fact, without the constraint of minimization l_1 norm, LMS is blind to all possible direction cosines within the hyper-sphere giving rise to the *Penrose's inverse* $[A]^{-1}\cong[A]^T([A][A]^T)^{-1}$; or $[A]^{-1}\cong([A]^T[A])^{-1}[A]^T$ (simplified by A =QR decomposition). To be sure, CRT&D proved a Restricted Isometry Property (*RIP*) Theorem, stating that a limited bound on a purely random sparse sampling: $\|[\Phi]_{m,N}\vec{X}\|/\|\vec{X}\|\cong O(1\mp\delta_k)$, $m\cong 1.3k\ll N$. As a result, the $min.l_1$-norm is equivalent to the $min.l_0$-norm at the same random sparseness. Such an equivalent linear programming algorithm takes a manageable polynomial time. However, it is not fast enough for a certain video imaging.

The following **subtitles** may help those who wish to selectively speed read through ANN, NI and CS. The universal language between man and machine will be the mathematics (apology to those who seeks a popular science reading).

NI Definition: NI may be defined by unsupervised learning algorithms running iteratively on connectionist architectures, naturally satisfying fault tolerance (FT), and subspace generation (SG).

Hebb Learning Rule: If blinking traffic lights at all street intersections have built-in data storage from all the transceivers, then traffic lights function like neurons with synaptic junctions. They send and receive a frequency modulation Morse code ranged from 30 Hz to 100 Hz firing rates. Physiologist Donald Hebb observed the plasticity of synaptic junction learning. The Hebbian rule describes how to modify the traffic light blinking rate to indicate the degree of traffic jam at street intersections. The plasticity of synapse matrix $[W_{i,j}]$ is adjusted in proportion to the inputs of u_i from the i-th street weighted by the output change, Δv_j at the *j*-th street as the vector product code:

$$\text{Do 10:}\quad \Delta W_{j,i}\approx\Delta v_j\,u_i\ . \tag{3a}$$

$$\text{10:}\quad [W_{j,i}]'=[W_{j,i}]+\Delta W_{j,i};\ \text{Return.} \tag{3b}$$

An event is represented in the m-D subspace of a total $N=10^{10}$ D space, supported by 10 billion neurons in our brain. The synergic blinking patterns of m communicating neurons/traffic lights constitute the subspace. The volume of m-D subspace may be estimated by the vector outer products called associative memory [AM] matrix inside the hippocampus of our central brain (Fig. 4c). Even if a local neuron or traffic light broke down, the distributed associative memory (AM) can be retrieved. This is the FT as the nearest neighbor classifier in a finite solid angle cone around each orthogonal axis of the subspace; then, the subspace generalization (SG) is going along a new direction that is orthogonal to the full m-D subspace.

Unsupervised lesson learned: Supervised learning stops when the algorithm has achieved a desired output. Without knowing the desired output, an unsupervised learning algorithm doesn't know when to stop. Since the input data already has some energy in its representation; the measurement principle should not bias the input energy for firing sensory system reports accordingly. Thus, the magnitude of output's firing rates should not be changed by the learning weight. Note that in physics the photon energy field is the quadratic displacement of oscillators. Thus, the constraint of unsupervised learning requires adjusting the unit weight vector on the surface of hyper-sphere of R^m. In fact, the main lesson of Bell-Sejnowski-Amari-Oja (BSAO) unsupervised learning algorithm is this natural stopping criterion for the given set of input vectors $\{\vec{x}\}\in R^N$. The BSAO projection pursuit algorithm is merely a rotation within a Hyper-sphere. It stops when the weight vectors $[W_{i,j}]=[\vec{w}_i,\vec{w}_{i\prime},..]$ becomes parallel in time to the input vectors of any magnitude $\vec{w}_i||\vec{x}_i$. The following stopping criterion of an unsupervised learning will be discovered thrice in Sect.2

$$[\boldsymbol{\delta_{\alpha,\beta}}-\boldsymbol{x_\alpha x^T{}_\beta}]\vec{\boldsymbol{x}}_\alpha\vec{\boldsymbol{x}}_\beta=\mathbf{0},$$

$$\Delta\vec{\boldsymbol{w}}\equiv\vec{\mathrm{w}}'-\vec{w}=\boldsymbol{\epsilon}[\boldsymbol{\delta_{\alpha,\beta}}-\vec{\boldsymbol{w}}'_\alpha\vec{\boldsymbol{w}}'^T{}_\beta]\vec{\boldsymbol{x}}_\alpha\frac{dK(\vec{u}_\beta)}{d\vec{w}^i}\ , \tag{3c}$$

where K is a reasonable contrast function for the source separation of the weighted input $\vec{u}_i=[W_{i,\gamma}]\vec{x}_\gamma$. The contrast function could be (i) the maximum a-posteriori entropy (filtered output entropy) used in Bell & Senowski algorithm in 1996; (ii) the fixed point algorithm of 4-th order cumulant Kurtosis (Fast ICA) adopted in Hyvarinen & Oja in 1997; (iii) the isothermal equilibrium of minimum thermodynamic Helmholtz free energy ($Brain\ T_o=37^o\ C$) known as Lagrange Constraint Neural Net, in terms of $min.H=E-T_o\,max.S$ (maximum a-priori source entropy) by Szu & Hsu, 1997.

When we were young, unsupervised learning guided us extracting sparse orthogonal neuronal representations in an effortless fashion defined at the minimum isothermal free energy. Subsequently, the expert systems at school supervised learning come in handy with these unsupervised mental representations. As we get older, our unsupervised ability for creative emotional side *e*-Brain is inevitably eroded and outweighed by the expert systems matured mainly at the logical and left-side *l*-Brain.

In this paper, we assume that the CRT&D RIP theorem works for both the purely random sparse $[\Phi]_{m,N}$ and the organized sparse $[\Phi_s]_{m,N}$. A sketch of proof is given using the exchange entropy of Brandt & Pompe [14] (successfully used recently in NI magazine V1, No 1 by Morabito et al. to the EEG data for Alzheimer patients) for the complexity of organized sparseness at the end of Sect. 7. Intuitively speaking, we do not change the number of ones within $[\Phi_s]_{m,N}$; only we endow a feature meaning to the ones' locations, beyond the original meaningless all-pass filtering. In other words, we found that the admissible ANN storage demands the orthogonal sparse ***moments of*** spotting ***dramatic orthogonal changes of salient features.*** Consequently, we will not alter the value of unknown image vector $\| \vec{X} \|$ more than δ_k. In fact, for real-time video, we have bypassed random CS coding and image recovery algorithms, and chose instantaneous retrieval by MPD Hetro-Associative Memory (HAM) storage defined in Sections 2 and 4.

After the introduction of the goals and the approaches of CS, we review ANN in Section 2; Neuroscience 101 with an emphasis on the orthogonal sparseness representations of HVS in Section 3; the AM storage in Section 4; Software simulation results in Sect. 5; Unsupervised Spatial-Spectral CS Theory in Section 6; and Hardware design of camera in Section 7. The following, Sect 2, might provide the simplest theory of learning from supervised ANN to unsupervised ANN.

2. Reviews of Artificial Neural Networks

Traditional ANN performs supervised learning, known as a soft lookup table, or merely as 'monkey sees, monkey do'. Marvin Minskey and Seymour Papert commented on ANN and Artificial Intelligence (AI) as well as extending multi-layer finite state machine to do "Ex OR" in 1988. This was about the year when INNS was incepted by 17 interdisciplinary Governors. Notably, Stephen Grossberg, Teuvo Kohonen, Shun-ichi Amari serve as editors in chief of INNS, which was published quarterly by Pergamum Press and subsequently, monthly by Elsevier Publisher. In the last two decades, both INNS and Computational Intelligence Society of IEEE have accomplished a lot (recorded in IJCNN proceedings). Being the founding Secretary and Treasurer, a former President and Governor of INNS, the author apologizes for any unintentional bias. Some opinions belong to all shade of grey; taking binary or spicy story approach is one of the great pedagogical techniques that our teachers George Uhlenbeck and Mark Kac often did at the Rockefeller University. For example, 'nothing wrong with the supervised learning exemplars approach using the lookup table, the curse is only at the 'static' or closed lookup table.' Also, 'this limitation of supervised learning is not due to the connectionist concept, rather, due to the deeply entrenched "near equilibrium" concept'; Norbert Wiener developed near equilibrium Cybernetics in *1948 & 1961*. 'What's missing is the ability to create a new class far away from equilibrium.' 'INNS took the out of the box, interdisciplinary approach to learn from the Neurosciences how to develop unsupervised learning paradigm from Neurobiology.' 'This is an important leg of NI tripod. The other two legs are the fault tolerance and the subspace generalization.' I used in teaching but have deleted in writing.

2.1 Fault Tolerance and Subspace Generalization

Fault Tolerance (FT): The read out of m neuronal representation satisfies the fault tolerance. This is due to the geometry of a circular cone spanned in 45^o solid angle around the m-D vector axis. This central axis is defined as the memory state and the cone around it is its family of turf vectors. Rather than precisely pointing in the same vector direction of m-D, anything within the turf family is recognized as the original axis. This is the reason that the read out is fault tolerant. Thus, [AM] matrix storage can enjoy a soft failure in a graceful degradation fashion, if and only if (iff) all storage state vectors are mutually orthogonal within the subspace; and going completely outside the subspace in a new orthogonal direction to all is the subspace generalization (SG).

Subspace Generalization (SG): We introduce the inner product BraCKet notation $< Bra\,|Ket >= C$, in the dual spaces of $< Bra|$ and $|ket >$, while the outer product matrix is conveniently in the reverse order $\left[w_{j,i}\right] = |v_j >< u_i|$ introduced by physicist P. Dirac. We prove the 'traceless outer product' matrix storage allows SG from the m-D subspace to one bigger m+1-D subspace. Defined, the Ortho-Normal (ON) basis is $< n'|n >= \delta_{n',n}$; $n, n' = 1, .. m.$ Then, SG is the Trace-less ON $[AM]_{m,m} = \sum_{n=1}^{m} |n >< n| - Tr[AM]_{m,m}$. Trace operator Tr: summing all diagonal elements is the projection operator defined $Tr^2 = Tr$.

SG Theorem: Without supervision, a traceless matrix storage of ON sub-space can self-determine admitting $|x > = |m+1 >$, iff $< m+1|n >= \delta_{m+1,n}$ satisfying the fixed point of cycle 2 rule: $[AM]_{m,m}{}^2 |x >= |x >$, then

$[AM]_{m+1,m+1} = \sum_{n=1}^{m+1} |n >< n| - Tr[AM]_{m+1,m+1}$.

Q.E.D.

AM is MPD computing, more than the nearest neighbor Fisher classifier. These FT & SG are trademarks of connectionist, which will be our basis of CS approach. Unsupervised learning is a dynamic trademark of NI. New learning capability comes from two concepts, (i) engineering filtering concept and (ii) physics-physiology isothermal equilibrium concept.

Semantic Generalization: Semantic generalization is slightly different than the subspace generalization, because it involves a higher level of cognition derived from both sides of the brain. Such an e-Brain and *l*-Brain combination processes thinking within two boxes of brain related by a set of independent degrees of freedom. Thus, this semantic generalization is the different side of the same coin, in Sect 4. We are ready to set up the math language leading to the modern unsupervised learning as follows:

2.2 Wiener Auto Regression

Norbert Wiener invented the *near equilibrium control* as follows. He demonstrated a negative feedback loop for the missile trajectory guidance. He introduced a moving average Auto Regression (AR) with LMS error:

$$min. E =< (u_{(m)} - y)^2 >$$

where the scalar input $u_{(m)} = \vec{w}_m{}^T \vec{x}_m \equiv < \vec{x}_m >$ has weighted average of the past m data vector

$$\vec{x}_m = (x_t, x_{t-1}, x_{t-2}, \dots, x_{t-(m-1)})^T$$

to predict the future as a desired output $y = x_{m+1}$.

A simple near equilibrium analytical filter solution is derived at the fixed point dynamics

$$\frac{\partial E}{\partial \vec{w}} = 2 < (\vec{w}^T \vec{x}_m - x_{m+1})\vec{x}_m >= 0, \text{ e.g. } m=3$$

$$\begin{bmatrix} c_o & c_1 & c_2 \\ c_1 & c_o & c_1 \\ c_2 & c_1 & c_o \end{bmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix};$$

$$c_{t-t'} \equiv < x_t x_{t'} >; \; c_1 =< x_t x_{t-1} >; c_2 =< x_t x_{t-2} >; \dots$$

Solving the Teoplitz matrix, Wiener derived the filter weights. Auto Regression (AR) was extended by Kalman for a vector time series with nonlinear Riccati equations for extended Kalman filtering. In image processing: $\vec{X} = [A]\vec{S} + \vec{N}$ where additive noisy images become a vector $\vec{X}$ represented by a lexicographic row-by-row order over 2-D space $\vec{x}$. Wiener image filter is derived using AR fixed point (f.p.) algorithm in the Fourier transform domain:

$$\hat{X}(\vec{k}) = \iint d^2\vec{x} \exp(j\vec{k}\cdot\vec{x}) \vec{X}(\vec{x}); j = \sqrt{-1}$$

Using Fourier de-convolution theorem, $\exp(j\vec{k}\cdot\vec{x}) = exp(j\vec{k}\cdot(\vec{x}-\vec{y}))\exp(j\vec{k}\cdot\vec{y})$, gives a linear image equation in the product form in Fourier domain, as noisy speech de-mixing in Fourier domain:

$$\hat{X} = \hat{A}\hat{S} + \hat{N}$$

Wiener sought $\hat{S} = \widehat{W}\hat{X}$ to minimize the LMS errors

$$E =< (\widehat{W}\hat{X} - \hat{S}')^* \cdot (\widehat{W}\hat{X} - \widehat{S'}) >;$$

$$\because f.p. \; \frac{\partial E}{\partial \widehat{W}^*} =< 2\hat{X}^* \cdot (\widehat{W}\hat{X} - \hat{S}') > = 0;$$

Interesting is the termination: $\overrightarrow{data} \perp \overrightarrow{error} \to 0: \hat{S}' \to \hat{S}$

$$\therefore \widehat{W} =< \hat{X}^* \cdot \widehat{S'} > [< \hat{X}^* \cdot \hat{X} >]^{-1} \cong \hat{A}^{-1}[1+\varepsilon]^{-1},$$

where noise to signal ratio is $\varepsilon \equiv < \hat{N}^*\hat{N} > / |\hat{A}|^2 |\hat{S}|^2$.

Wiener filtering is the inverse filtering $\widehat{W} = \hat{A}^{-1}$ at strong signals, and becomes Vander Lugt filtering $\widehat{W} = \hat{A}^* \frac{|\hat{S}|^2}{<|\hat{N}|^2>}$ for weak signals. A mini-max filtering is given by Szu (Appl. Opt. V. 24, pp.1426-1431, 1985).

Such a near equilibrium adjustment influenced generations of scientists. While F. Rosenblatt of Cornell U. pioneered the 'perceptron' concept for OCR, B. Widrow of Stanford took a leap of faith forward with 'multiple layers perceptrons.' Hyvarinen & Oja developed the Fast ICA algorithm. The author was fortunate to learn from Widrow; co-taught with him a short UCLA course on ANN, and continued teaching for a decade after 1988 (thanks to W. Goodin).

2.3 ANN generalize AR

Pedagogically speaking, ANN generalizes Wiener's AR approach with 4 none-principles: (i) non-linear threshold, (ii) non-local memory, (iii) non-stationary dynamics and (iv) non-supervision learning, respectively Equations (4a,b,c,d).

2.3.1 Non-linear Threshold: **Neuron model**

McCullouch & Pitts proposed in 1959 a sigmoid model of threshold logic: mapping of neuronal input $u_i(-\infty,\infty)$ to the unary output $v_i[0,1]$ asymptotically by solving Ricati nonlinear $\frac{dv_i}{du_i} = v_i(1-v_i) = 0$, at 'no or yes' limits $v_i = 0$; $v_i = 1$. Exact solution is:

$$v_i = \sigma(u_i) \equiv [1+\exp(-u_i)]^{-1}$$
$$= \exp\left(\tfrac{u_i}{2}\right)[\exp\left(\tfrac{u_i}{2}\right) + \exp\left(-\tfrac{u_i}{2}\right)]^{-1}, \quad (4a).$$

Three interdisciplinary interpretations are given:

Thermodynamics, this is a two state equilibrium solution expressed in firing or not, the canonical ensemble of the brain at the equilibrium temperature T, and the Boltzmann's constant K_B, as well as an arbitrary threshold θ:

$$y = \sigma_T(x-\theta) = \left[1+\exp\left(-\frac{x-\theta}{K_B T}\right)\right]^{-1}.$$

Neurophysiology, this model can contribute to the binary limit of a low temperature and high threshold value a single ***grandmother*** neuron firing in a family tree subspace (1,0,0,0,0..) as a ***sparse network representation.***

Computer Science, an overall cooling limit, $K_B T \Longrightarrow 0$, the sigmoid logic is reduced to the binary logic used by John von Neumann for the digital computer: $1 \geqq \sigma_o(x \geq \theta) \geqq 0$.

2.3.2 Nonlocal memory: D. Hebb learning rule of the communication is efficiently proportional to what goes in and what comes out the channel by $W_{i,j} \propto v_i\, u_j$ measuring the weight matrix of inter-neuron synaptic gap junction. A weight summation of $\vec{x}_i$ given by **Compressive Sensing** rise to a potential sparse input $\vec{u}_i = [W_{i,\alpha}]\vec{x}_\alpha$ Eq(4b).

2.3.3 Non-stationary dynamics is insured by Laponov control theory: $\frac{du_i}{dt} = -\frac{\partial E}{\partial v_i}$. Eq(4c)

2.3.4 Non-supervised learning is based on nonconvex energy landscape: $E \cong H(open/no\ exemplars)$ Eq(4d)

2.4 Energy Landscape Supervised Algorithm

Physicist John Hopfield broadened the near-equilibrium Wiener notion and introduced a non-convex energy landscape E(v_i) at the output v_i space to accommodate the (neurophysiologic) associative memory storage. He introduced Newtonian dynamics $du_i/dt = -\partial E/\partial v_i$ as a generalization of the fixed point LMS Wiener dynamics. He proved a simple Lyapunov style convergence insured by a square of any real function which is always positive:

$$\frac{dE}{dt} = \sum_i \frac{\partial E}{\partial v_i}\frac{dv_i}{du_i}\frac{du_i}{dt} = -\sigma'(\frac{\partial E}{\partial v_i})^2 \text{ , Q.E.D.}$$

independent of energy landscapes, as long as a real monotonic positive logic $dv_i/du_i \equiv \sigma' \geq o$, in terms of $(\text{in}, \text{out}) = (u_i, v_i)$ defined by

$$v_i = \sigma(u_i);\ \&\ u_i = \sum_j W_{i,j}v_j\ ; E = -\frac{1}{2}\sum_{i,j} W_{i,j}v_iv_j.$$

Physicist E. R. Caianiello is considered a thinking machine beyond Wiener's AR. He used causality physics principles to generalize the instantaneous McCullough & Pitts neuron model building in the replenishing time delay in 1961.

Psychologist James Anderson, in 1968, developed a correlation memory while **Christopher von der Malsburg,** 1976, developed a self-organization concept. They described a *brain in a box* concept, inspired by the binary number predictor box built by **K. Steinbuch & E. Schmidt** and based on a *learning matrix* as the beginning of *Associative Memory* (AM) storage in biocybernetics in Avionics 1967. **Kaoru Nakano,** 1972, and **Karl Pribram,** 1974, enhanced this distributive AM concept with a fault tolerance (FT) for a partial pattern completion (inspired by Gabor hologram).

Engineer Bernard Widrow took multiple layers perceptrons as adaptive learning neural networks. For computing reasons, the middle layer neurons took the cool limit $T \to 0$ of the sigmoid threshold as non-differentiable bipolar logic, and achieved a limited adaptation.
From the connectionist viewpoints, **Shun-ichi Amari** indicated in 1980 that the binary logic approach might suffer a premature locking in the corners of hyper-cubes topology.

2.5. Backprop Algorithm

It took a team of scientists known as the Cambridge PDP group (Neuropsychologists David Rumelhart, James McClelland, Geoffrey Hinton, R. J. Williams, Michael Jordan, Terrence Sejnowski, Francis Crick, and graduate students) to determine the backprop algorithm. They improved Wiener's LMS error $E = \sum_i |v_i - v_i^*|^2$ with a parallel distributed processing (PDP) double decker hamburger architecture, consisting of 2 layers of beef (uplink $w_{k,j}$ & downlink $w'_{j,i}$) sandwiched between in 3 layers of buns made of neurons. The sigmoid logic: $\sigma' \equiv d\sigma/du_k < \infty$ is analytic, they unlocked the bipolar 'bang-bang' control from Widrow's corners of hypercubes. They have analytically derived the 'Backprop' algorithm. Namely, passing boss error to that of the hidden layer; and, in turns, to the bottom layer which has exemplar inputs.

$$\frac{\partial w_{j,i}}{\partial t} \cong \frac{\Delta w_{j,i}}{\Delta t} = -\frac{\partial E}{\partial w_{j,i}} \tag{5a}$$

The Hebb learning rule of uplink weight is obtained by the chain rule:

$$\Delta w_{k,j} = -\frac{\partial E}{\partial w_{k,j}}\Delta t \cong -\sum_n \frac{\partial E}{\partial u_n}\frac{\partial u_n}{\partial u_k}\frac{\partial u_k}{\partial w_{k,j}}\Delta t$$
$$= \sum_n \delta_n \delta_{n,k} v'_j \Delta t = \delta_k v'_j \Delta t, \tag{5b}$$

Kronecker $\delta_{n,k} \equiv \frac{\partial u_n}{\partial u_k}$ selects top layer post-synaptic δ_j (error energy slope) and hidden layer pre-synaptic v'_i:

$$\delta_k \equiv -\frac{\partial E}{\partial u_k} = -\frac{\partial E}{\partial v_k}\frac{\partial v_k}{\partial u_k} = -(v_k - v_k^*)\sigma^{(\prime)}. \tag{5c}$$

The sigmoid slope $\sigma^{(\prime)}$ is another Gaussian-like window function. The PDP group assumed Hebb's rule $\Delta w_{k,j} \approx \delta_k v'_j$ holds true universally, and cleverly computed the hidden share of blaming δ'_j from fan-in boss errors δ_k

$$\delta'_j \equiv -\frac{\partial E}{\partial u'_j} = -\sum_k \frac{\partial E}{\partial u_k}\frac{\partial u_k}{\partial v'_j}\frac{\partial v'_j}{\partial u'_j} \equiv \sum_k \delta_k w_{k,j}\sigma^{(\prime)}. \tag{5d}$$

Each layer's I/O firing rates are denoted in the alphabetic order as (input, output) = (u,v) respectively; the top, hidden, and bottom layers are labeled accordingly:

$$(v_k, u_k) \leftarrow w_{k,j} \leftarrow (v'_j, u'_j) \leftarrow w'_{j,i} \leftarrow (v"_i, u"_i),$$

where $v_k = \sigma(u_k) \equiv \sigma(\sum_j w_{k,j}\, v'_j)$; $v'_j = \sigma(u'_j \equiv \sum_i w'_{j,i}\, v''_i)$. Hebbian rule turns out to be similar at every layers, e.g., $\delta''_j \equiv -\frac{\partial E}{\partial u''_j} = \sum_k \delta'_k w'_{k,j}\sigma^{(\prime)}$, etc. Such a self-similar chain relationship is known as backprop.

2.6 Bio-control

Independently, Paul Werbos took a different viewpoint, assigning both the adaptive credit and the adaptive blame to the performance metric at different locations of the feedback loops in real world financial-like applications (IEEE Handbook Learning & Approx. Dyn. Prog., 2004). As if this were a 'carrot and stick' model controlling a donkey, to be effective, these feedback controls must be applied at different parts of the donkey. Thus, this kind of bio-control goes beyond the near-equilibrium negative feedback control. Such broad sense reinforcement learning, e.g. sought after a clear reception of a smartphone by moving around without exemplars, began a flourishing era, notably, Andrew Barto, Jennie Si, George Lendaries, Kumpati Narendra, et al. produced heuristic dynamic programming, stochastic, chaos, multi-plants, multi-scales, etc., bio-control theories.

2.7 Self-Organization Map (SOM)

Teuvo Kohonen computed the batched centroid update rule sequentially:

$$< \vec{x} >_{N+1} = < \vec{x} >_N \left(\frac{N+1-1}{N+1}\right) + \frac{1}{N+1}\vec{x}_{N+1}$$
$$= < \vec{x} >_N + \rho(\vec{x}_{N+1} - < \vec{x} >_N), \tag{6}$$

replacing the uniform update weight with adaptive learning $\rho = \frac{1}{N+1} < 1$. SOM has significantly contributed to database applications with annual world-wide meetings, e.g. US PTO Patent search, discovery of hidden linkage among companies, genome coding, etc.

2.8 NP Complete Problems

David Tank and John Hopfield (T-H) solved a class of computationally intractable problems (classified as the nondeterministic polynomial (NP), e.g. the Travelling Salesman Problem, Job scheduling, etc.) The possible tours are combinatorial explosive in the factorial sense: $N!/2N$, where the denominator is due to the TSP having no home base, and the clockwise and counter-clockwise tours having an equal distance. T-H solved this by using the powerful MPD computing capability of ANN (Cybernetics, & Sci. Am. Mag).

$$E = \sum_{\vec{c}=1}^{N} \sum_{\vec{t}=1}^{N} v_{\vec{c}} \left[W_{\vec{c},\vec{t}}\right] v_{\vec{t}} + \text{Constraints.}$$

Their contribution is similar to DNA computing for cryptology RSA de-coding. Both deserve the honor of Turing Prizes. Unfortunately, the T-H constraints of the permutation matrix $\left[W_{\vec{c},\vec{t}}\right]$ are not readily translatable to all the other classes of the NP complete problems:

$$\left[W_{\vec{c},\vec{t}}\right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

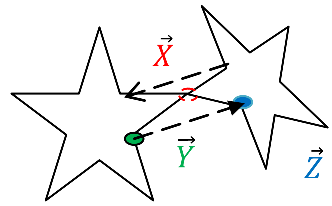
(city #1 visits tour No.1; #2 for 2nd, #3 for 3rd etc., and returned to No. 1; T-H labeled each neuron with 2-D vector index for convenience in both city & tour indices. Meanwhile, Y. Takefuji & others mapped the TSP to many other applications including genome sequencing (Science Mag.).

Divide & Conquer (D&C) Theorem: Szu & Foo solved a quadratic NP complete computing by D&C, using orthogonal decompositions $\vec{A} = \vec{B} + \vec{C}$, l_2-norm:

$$min.|\vec{A}|_2^2 = min.|\vec{B}|_2^2 + min.|\vec{C}|_2^2;\ iff\ \vec{B}\cdot\vec{C} = 0. \qquad (7)$$

Unfortunately, searching boundary tours for divisions could be time consuming. Moreover, Simon Foo and I could not solve the TSP based on the original constraints of row-sum and column-sum and row products and column products of the matrix, generating a Mexican standoff dual like in Hollywood western movies.

Improved TSP with D&C Algorithm: A necessary and sufficient constraint turns out to be *sparse orthogonal sampling Matrix* $[\Phi_s]$ which is equivalent to a permutation mix up of the identity matrix. Iff each row and column is added up to one, similar to N queens constraint (in chess, queens can kill each other, unless only one queen occupies one row and column). Furthermore, a large scale democratic "Go game," is defined by a unlimited number of rank-identical black or white stones for two competing groups roaming over the same no-man land, square lattice space. To win the territory is forming a cowboy lasso loop fashion surrounding the other color stones territory with one's own color stones; but one stone is put down at the intersection of the empty lattice at one step at one time, including any of the same color stones putting down early by foresights. Thus, the winning goal is to gain the maximum possible territory on the square common lattice board (a simplified form has been solved by ANN by Don Wunsch et al.). The strategy to win is usually put down one's own color stone in the center of the board about a half size, and this is the basis of our divide and conquer theorem. We create a surrogate or ghost city at the mid point $\vec{X}$.



Without the need of a boundary search among cities, adding a ghost city $\vec{X}$ finds two neighborhood cities $\vec{Y}$ and $\vec{Z}$ with two vector distances: $\vec{B} = \vec{Y} - \vec{X}$, $\vec{C} = \vec{X} - \vec{Z}$. Iff $\vec{B} * \vec{C} = 0$ satisfies the D&C theorem, we accept $\vec{X}$. Then we conceptually solve two separate TSP problems in parallel. Afterwards, we can remove the ghost city and modify the tour sequences indicated by dotted lines. According to the triangle inequality, $|\vec{a}| + |\vec{b}| \geq |\vec{c}|$, the vector $\vec{c}$ represents a dotted line having a shorter tour path distance than the original tour involving the ghost city. **Q.E.D.**

This strategy should be executed from the smallest doable regions to bigger ones; each time one can reduce the computational complexity by half. In other words, solving the total N=18 cities by two halves N/2=9; one continues the procedure without stopping solving them, further dividing 9 by 4 and 5 halves, until one can carry out TSP in smaller units 4 and 5, each has de-ghosted afterward. Then, we go to 9 and 9 cities, de-ghost in a reverse order.

2.9 ART

Gail Carpenter and Stephen Grossberg implemented the biological vigilance concept in terms of two layers of analog neurons architecture. They proved a convergence theorem about short and long term traces $Z_{i,j}$ in 1987 App. Op. Their two layer architecture could be thought of as if the third layer structure were flipping down to touch the bottom layer using two phone lines to speak to one another top-down or bottom up differently. Besides the PDP 3 layers buns sandwiched 2 layers of weights have the original number of degrees of freedom, they created a new degree of freedom called the vigilance defined by $\rho = (\overline{m}_{t+1}, < \vec{w}_t >) = cos(\leq \pi/4) \geq 0.7$. This parameter can either accept the newcomer and updating the leader's class Centroid with the newcomer vector; or rejecting the newcomer letting it be a new leader creating a new class. Without the need of supervision, they implemented a self-organizing and robust MPD computing who follows the leader called (respectively binary, or analog, or fuzzy) the adaptive resonance theory (ART I, II, III). ART yields many applications by Boston NSF Center of Learning Excellence. Notably, M. Cader, et. al. at the World Bank implemented ART expert systems for typing seeds choice for saving the costly diagnosis needs by means of PCR amplification in order to build up enough DNA samples (pico grams); a decision prediction system based on the past Federal Reserve open forum reports (Neural Network Financial Expert Systems, G. J. Deboeck and M. Cader, Wiley 1994).

2.10 Fuzzy Membership Function

Lotfi Zadeh introduced an open set for imprecise linguistic concept represented by a 'possibilities membership function', e.g. beauty, young, etc. This open set triangular shape function is not the probability measure which must be normalized to the unity. Nevertheless, an intersection of two fuzzy sets, e.g. young and beautiful, becomes sharper in the joint concept. Such an electronic computing system for the union and the intersection of these triangle membership functions is useful, and has been implemented by Takeshi Yamakawa in the fuzzy logic chips. Whenever a precise engineering tool meets an imprecise application, the fuzzy logic chip may be useful, e.g. automobile transmission box and household comfort control device. Such a fuzzy logic technology becomes exceedingly useful as documented in a decade of soft computing conferences sponsored by The Japan Fuzzy Logic Industry Association.

ANN Modeling of Fuzzy Memberships: Monotonic sigmoid logic is crucial for John Hopfield's convergence proof. If the neuron had a piecewise negative response in the shape of a scripted N-letter: $v_i = \sigma_N(u_i)$, then, like the logistic map, it has a single hump height adjustable by the λ-knot (by M. Feigenbaum for the tuning of period doubling bifurcation cascade). If we represent each pixels by the sick neuron model $v_i = \sigma_N(u_i)$, then recursively we produce the nonlinear Baker transform of image mixing. Such a Chaotic NN is useful for the modeling of drug-induced hallucinating images, olfactory ball smell dynamics of Walter Freeman, and learnable fuzzy membership functions of Lotfi Zadeh.

2.11 Fast Simulated Annealing

Szu and Hartley have published in Phys Lett. and IEEE Proc. 1986, the Fast Simulated Annealing (FSA) approach. It combines the increasing numbers of local Gaussian random walks at a high temperature T, with an unbounded Levy flights at a low temperature in the combined Cauchy probability density of noise. A speed up cooling schedule is proved to be inversely linear time step $T_C = \frac{T_o}{1+t}$, for any initial temperature T_o that guarantees the reaching of the equilibrium ground state at the minimum energy. Given a sufficient low temperature $\widetilde{T_o}$ Geman and Geman proved in 1984 the converging to the minimum energy ground state by an inversely logarithmic time step: $T_G = \frac{\widetilde{T_o}}{1+log(1+t)}$. Sejnowski & Hinton used the Gaussian random walks in the Boltzmann's machine for a simulated annealing learning algorithm emulating a baby learning the talk, called Net-talk, or Boltzmann Machine.

Cauchy Machine: Y. Takefuj & Szu designed an electronic implementation of such a set of stochastic Langevin equations. Stochastic neurons are coupled through the synapse AM learning rule and recursively driven by Levy flights and Brown motions governed by the Cauchy pdf. The set of Cauchy-Langevin dynamics enjoys the faster inversely linear cooling schedule to reach the equilibrium state.

Do 10 t'=t'+1; $T_C(t') = \frac{T(t')}{1+t'}$

$\Delta x = T_c(t') \tan[(2\theta[0,1] - 1)\ \pi/2]$;

x(t')=$x_{t'}$+Δx; $E(t') = \sum_{i=1}^{m} \frac{1}{2} k(x(t') - x_i)^2$;

$\Delta E = E(t') - E(t'-1)$;

If $\Delta E \le 0$; accept x(t'), Go To 10; or

compute Metropolitan exp($-\Delta E / K_B T_C(t')$) $> \varepsilon_o$;

accept x(t') or not.

10: Return.

Optical version of a Cauchy machine is done by Kim Scheff and Joseph Landa. The Cauchy noise is optically generated by the random reflection of the mirror displacement x of the optical ray from a uniformly random spinning mirror angle $\theta(-\frac{\pi}{2}, \frac{\pi}{2})$. The temperature T is the distance parameter between the mirror and the plate generates the Cauchy probability density function (pdf) (Kac said as a French counter example to the British Gauss Central Limiting Theorem). This pdf is much faster for search than Gaussian random walks:

$$\rho_G(\Delta x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\Delta x^2}{T}\right) \cong \frac{1}{\sqrt{2\pi}}\left(1 - \frac{\Delta x^2}{T} + ..\right).$$

$$\rho_C(\Delta x) = \frac{1}{\pi}\left(1 + \frac{\Delta x^2}{T}\right)^{-1} = \frac{1}{\pi}\left(1 - \frac{\Delta x^2}{T} + \cdots\right).$$

Proof: ***Since*** $x = T\, tan\theta$; *then* $\frac{dx}{d\theta} = T(1 + tan\theta^2)$;

$$\pi = \int d\theta = \int \frac{d(\frac{x}{T})}{1 + tan\theta^2};$$

$$1 = \int \rho_c(x) dx = \frac{1}{\pi} \int \frac{1}{1+\left(\frac{x}{T}\right)^2} d\left(\frac{x}{T}\right).$$ ***Q.E.D.***

Global Levy Flights $< \Delta x^2 >_{\rho_C} = \infty$

Local Brownian motion $< \Delta x^2 >_{\rho_C} \cong T(t)$

2.12 NI Expert System

Szu and John Caulfield published an optical expert system in 1987, generalized the AI Lisp programming the pointer linkage map from 1-D vector arrays of $\vec{f} = (A,\ O, V)^T$ to $\vec{f}' = (A',\ O',\ V')^T$, etc. The color, "A attribute," of apple, "O object," is red, "V value". We represent the Lisp map with the MPD $[HAM] = \sum \vec{f}\, \vec{f}'^T$ storage which has demonstrated both the FT and the Generalization capabilities. This FT & SG of AM NI Expert System is a key driving engine for Video image Cliff Notes.

2.13 Unsupervised Learning of *l*-Brain

In order to make sure nothing but the desired independent sources coming out of the filter, C. Jutten and J. Herault adjusts the weights of inverse filtering to undo the unknown mixing by combining the inverse and forward operation as the unity operator (Snowbird IOP conf.; Sig. Proc. 1991). Since J.-F. Cardoso has systematically investigated the blind de-convolution of unknown impulse response function; he called a matrix form as Blinded Sources Separation (BSS) by non-Gaussian higher order statistics (HOS), or the information maximum output. His work since 1989 did not generate the excitement as it should be in the ANN community. It was not until Antony J. Bell and Terry J. Sejnowski (BS), et al. [10] have systematically formulated an *unsupervised learning of ANN algorithm,* solving both the unknown mixing weight matrix and the unknown sources. Their solutions are subject to the constraints of maximum filtered entropy $H(y_i)$ of the output $y_i = [W_{i,\alpha}]x_\alpha$, where $x_j = [A_{j,\alpha}?]s_\alpha?$, and the repeated Greek indices represent the summation. ANN model uses a robust saturation of linear filtering in terms of a nonlinear sigmoid output $y_i = \sigma(x_i) = \{1 + \exp(-[W_{i,\alpha}]x_\alpha)\}^{-1}$. Since a single neuron learning rule turns out to be isomorphic to that of N neurons in tensor notions, for simplicity sake we derive a single neuron learning rule to point out why the engineering filter does not follow the Hebb's synaptic weight updates. Again, a bona fide unsupervised learning does not need to prescribe the desirable outputs for exemplars inputs. For ICA, BS chose to maximize the Shannon output entropy H(y) indicating that the inverse filtering has blindly de-convoluted and found the unknown independent sources without knowing the impulse response function or mixing matrix. Thus, the filter weight adjustment is defined in the following and the BS result is derived as follows:

$$\frac{\delta w}{\delta t} = \frac{\partial H(y)}{\partial w}, \&, \ H(y) = -\int f(y)\log f(y)dy \Rightarrow$$

$$\delta w = \frac{\partial H(y)}{\partial w}\delta t = \{|w|^{-1} + (1-2y)x\}\delta t. \qquad (8a)$$

Derivation: From the normalized probability definitions:

$$\int f(y)dy = \int g(x)dx = 1\,; f(y) = \frac{g(x)}{\left|\frac{dy}{dx}\right|};$$

$$H(y) \equiv -< \log f(y) >_f \ ,$$

we express the output pdf in terms of the input pdf with changing Jacobian variables. We exchange the orders of operation of the ensemble average brackets and the derivatives to compute

$$\frac{\partial H(y)}{\partial w} = \frac{\partial < \log\left|\frac{dy}{dx}\right| >_f}{\partial w} \cong \left|\frac{dy}{dx}\right|^{-1}\frac{\partial\left|\frac{dy}{dx}\right|}{\partial w};$$

Ricati sigmoid: $y = [1 + \exp(-wx)]^{-1}$;

$$\frac{dy}{d(wx)} = y(1-y); \frac{dy}{dx} = wy(1-y) \& \frac{dy}{dw} = xy(1-y).$$

Substituting these differential results into the unsupervised learning rule yields the result. Q.E.D.

Note that the second term of Eq(8a) satisfies the Hebbian product rule between output y and input x, but the first term computing the inverse matrix $|w|^{-1}$ is not scalable with increasing N nodes. This non-Hebbian learning enters through the logarithmic derivative of Jacobian giving $\left|\frac{dy}{dx}\right|^{-1}$. To improve the computing speed, S. Amari et al. assumed identity $[\delta_{i,k}] = [W_{i,j}][W_{j,k}]^{-1}$ and multiplied it to the BS algorithm

$$\frac{dH}{dW_{i,j}}[\delta_{i,k}] = \{[\delta_{i,j}] - (2\vec{y} - 1)\vec{y}^T]\}[W_{i,j}]^{-1},$$

where use is made of $y_i = [W_{i,\alpha}]x_\alpha$ to change the input x_j to the synaptic gap by its weighted output y_i. In information geometry, Amari et al. derived the natural gradient ascend BSA algorithm:

$$\frac{dH}{dW_{i,j}}[W_{i,j}] = \{[\delta_{i,j}] - (2\vec{y} - 1)\vec{y}^T]\}, \qquad (8b)$$

which is not in the direction of original $\frac{dH}{dW_{i,j}}$ and enjoys a faster update without computing the BS inverse .

Fast ICA: Erkki Oja began his ANN learning of nonlinear PCA for pattern recognition in his Ph D study 1982.

$$< \vec{x}\vec{x}^T > \hat{e} \ = \lambda\hat{e};$$

$$w' - w = \vec{x}\sigma(\vec{x}^T\vec{w}) \cong < \vec{x}\vec{x}^T > \vec{w}; \qquad (8c)$$

$$\frac{d\vec{w}}{dt} = < \vec{x}\vec{x}^T > \vec{w} \cong \sigma(\vec{x}^T\vec{w})\vec{x} \cong \frac{dK(u_i)}{du_i}\frac{du_i}{dw_i} \equiv k(\vec{x}^T\vec{w})\vec{x};$$

where Oja changed the unary logic to bipolar hyperbolic tangent logic $v_i = \sigma(u_i) \approx u_i - \frac{2}{3}u_i^3 \cong \frac{dK(u_i)}{du_i};\ u_i = w_{i,\alpha}x_\alpha$. It becomes similar to a Kurtosis slope, which suggested to Oja a new contrast function K. The following is the geometric basis of a stopping criterion of unsupervised learning. Taylor expansion of the normalization, Eq(8c) and set $|\vec{w}|^2 = 1$:

$$|\vec{w}'|^{-1} = [(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x}))^T\left(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x})\right)]^{-\frac{1}{2}}$$

$$= \ 1 - \frac{\epsilon}{2}\,k(\vec{w}^T\vec{x})(\vec{x}^T\vec{w} + \vec{w}^T\vec{x}) + O(\epsilon^2).$$

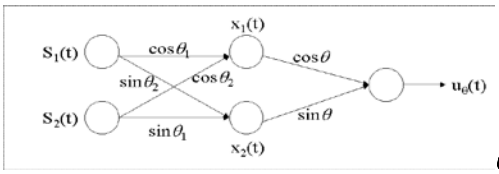
$$\vec{w}'' \equiv \vec{w}'|\vec{w}'|^{-1}$$

$$= \left(\vec{w} + \epsilon\vec{x}k(\vec{w}^T\vec{x})\right)\left(1 - \frac{\epsilon}{2}\,k(\vec{w}^T\vec{x})(\vec{x}^T\vec{w} + \vec{w}^T\vec{x})\right)$$

$$\Delta\boldsymbol{\vec{w}''} = \vec{w}'' - \vec{w} = \boldsymbol{\epsilon[\delta_{\alpha,\beta} - w''_\alpha w''^T_\beta]\vec{x}_\alpha}\frac{dK(u_\beta)}{d\vec{w''}_i} \qquad (8d)$$

This kind of derivation is therefore referred to as **BSAO unsupervised learning** collectively Eqs(8b, 8d).

Fast ICA Example: A. Hyvarinen and Oja demonstrated Fast ICA in 1996, as the fixed point analytical solution of a cubic root: $\frac{dK(u_i)}{dw_i} = 0$, of a specific contrast function named Kurtosis. Rather than maximizing an arbitrary contrast function, or the BS filtered output entropy, they considered the 4th order cummulant Kurtosis $K(y_i) = < y_i^4 > -3 < y_i^2 >^2$ which vanishes for a Gaussian average. K>0 for super-Gaussian, e.g. an image histogram that is broader than Gaussian, and K<0 for sub-Gaussian, e.g. a speech amplitude Laplacian histogram that is narrower than Gaussian. Every faces and voices have different fixed value of Kurtosis to set them apart. At the bottom of a fixed point, they set the slope of Kurtosis to zero and *efficiently* and *analytically solved* its cubic roots. This is called (Fast) I CA, a s c oined by P eter C omo (S ig. P roc., circa ‘90).



$\theta \rightarrow \varphi$

$$\vec{x}_i = \vec{a}(\theta_i)_\alpha \vec{s}_\alpha = cos\theta_i s_1 + \sin\theta_i s_2; i = 1,2$$

$$\vec{u}_j = \vec{k}^T(\varphi_j)_\alpha \vec{x}_\alpha = cos\varphi_j x_1 + sin\varphi_j x_2; j = 1,2$$

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{bmatrix} \cos\varphi_1 & -\sin\varphi_1 \\ \sin\varphi_2 & \cos\varphi_2 \end{bmatrix} \begin{bmatrix} \cos\theta_1 & \sin\theta_2 \\ -\sin\theta_1 & \cos\theta_2 \end{bmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

$$= \begin{bmatrix} \cos(\varphi_1 - \theta_1) & \sin(\varphi_1 - \theta_2) \\ \sin(\varphi_2 - \theta_1) & \cos(\varphi_2 - \theta_2) \end{bmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix},$$

Oja’s rule of independent sources x & y:

$$K(ax + by) = a^4 K(x) + b^4 K(y)$$

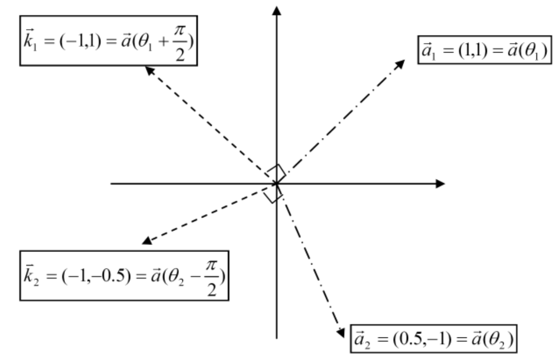
$$K(u_1) = \cos(\varphi_1 - \theta_1)^4 K(s_1) + \sin(\varphi_1 - \theta_2)^4 K(s_2)$$

$$K(u_2) = \sin(\varphi_2 - \theta_1)^4 K(s_1) + \cos(\varphi_2 - \theta_2)^4 K(s_2)$$

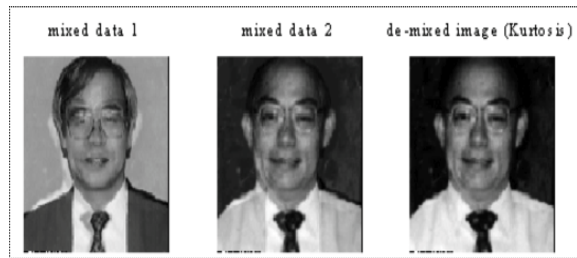
Szu’s ru le: Iff $\varphi_j = \theta_i \pm \frac{\pi}{2}$; then $K(u_1) = K(s_2); K(u_2) = K(s_1)$, verifing Fast ICA $\frac{\partial K}{\partial k_j} = 0$. Given arbitrary unknown θ_i, not ne cessarily or thogonal t o e ach other, the k illing weight $\vec{k}(\varphi_j)$ can eliminate a mixing vector $\vec{a}_i(\theta_i)$.

2.14 Sparse ICA

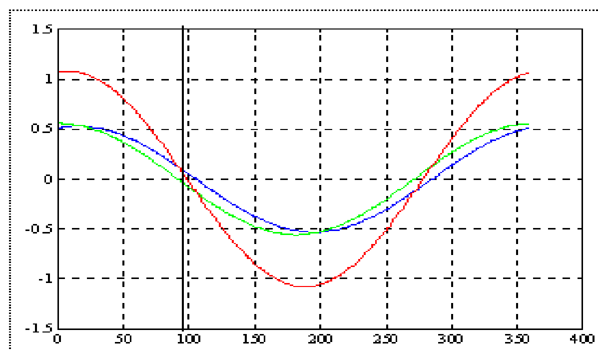
New application is applying a sparse constraint of non-negative matrix f actorization (NMF), which i s useful for image learning of parts: eyes, noses, mouths, (D. D. Lee and H . S . Seung, N ature 401(6755):788–791, 19 99); following a s parse n eural co de f or n atural i mages (B. A. Olshausen and D. J. Field, Nature, 381:607–609, 1996). P. Hoyer provided M atlab c ode t o run sparse N MF $[X] = [A][S]$, (2004). $min.|[A]|_1; min.|[S]|_1$ subject to $E = |[X] - [A][S]|_2^2$. The projection op erator i s derived f rom the Grand-Schmidt decomposition $\vec{B} = \vec{B}_\perp + \vec{B}_\parallel$; where $\vec{B}_\parallel \equiv (\vec{B} \star \vec{A})\vec{A}/|\vec{A}|^2$, and $\vec{B}_\perp \equiv \vec{B} - \vec{B}_\parallel$. Alternative gradient de scend solutions be tween 2 u nknown matrices $\{[A] or\ [S]\}$: n ew $[Z]' = [Z] - \frac{\partial E(|[X]-[A][S]|^2)}{\partial [Z]} = o; min.|[Z]|_1$, where a lternatively s ubstituting $[Z]\ with\ [A]\ or\ [S]\}$. (Q. Du, I. Ko priva & H. S zu, “ICA hyperspectral r emotes s ensing,” O pt E ng.V45, 2006;“Constrained Matrix Factorization for Hyperspectral,” IEEE IGARS 2005). Recently, T-W. L ee and S oo-Young Lee, et al. at KAIST have solved the source permutation challenge of ICA speech sources in the Fourier domain by de-mixing for Officemate automation. They grouped similar Fourier components into a linear combination in a vector unit, and reduced the number of independent vectors in the sense of sparse measurements solving the vector dependent component analysis (DCA) [11].



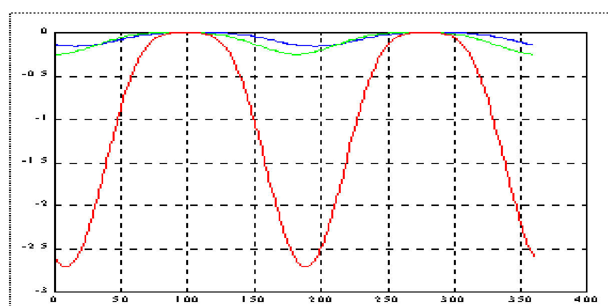
(a)



(b)



(c)



(d)

Figure 1: (a) De-mixing by killing vector(Yamakawa & Szu); (b) a sources image (Szu) (not shown Ymakawa) de-mixed by one of the killing vectors; (c)(left) The vertical axis indicates the blue source of Szu face vector, the green source of Yamakawa face vector, and the red is the Kurtosis value plotted against the killing weight vector. (d) (right) The Kurtosis is plotted against the source angle, where the max of Kurtosis happens at two source angles (Ref: H. Szu, C. Hsu, T. Yamakawa, SPIE V.3391, 1998; “Adv. NN for Visual image Com,” Int’l Conf Soft Computing, Iizuka, Japan, Oct. 1, 1998).

2.14 Effortless Learning Equilibrium Algorithm

An effortless thought process that emulates how the e-Brain intuitive idea process works. Such an effortless thinking may possibly reproduce an intuitive solution, which belong to the local isothermal equilibrium at brain's temperature $K_B T_o$ (K_B is Boltzmann constant, $T_o = 273 + 37^o C = 310^o$ Kelvin). Therefore, the thermodynamic physics gives an inverse solution that must satisfy the minimum Helmholtz free energy: $min.\, \mathrm{E}(s_i) = E - T_o S$. The unknown internal brain energy is consistently determined by the Lagrange multiplier methodology. Thus, we call our m-component 'min-energy max a-priori source entropy' as Lagrange Constraint Neural Network (LCNN) in 1997. Taylor expansion of the internal energy introduced Lagrange parameter μ as variable energy slope.

$$E = E_o^* + \sum_{i=1}^{m} \frac{\partial E}{\partial s_i}(s_i - s_i^*) + O(\Delta^2) = E_o^* + \vec{\mu} \cdot ([W]\vec{X} - \vec{S}^*) + O(\Delta^2)$$

The Lagrange slope variable $\vec{\mu}$ were parallel and proportional to the error itself $\vec{\mu} \approx ([W]\vec{X} - \vec{S})$, our LCNN is reduced to Wiener LMS supervised learning $E \cong E_o^* + |[W]\vec{X} - \vec{S}^*|^2$ of the expected output $\vec{S}^*$ from the actual output $[W]\vec{X}$. Given the Boltzmann entropy formula: $S = -K_B \sum_i^k s_i \log s_i$, of independent s_i sources, the m-components general ANN formulism requires matrix algebra not shown here [9]. In order to appreciate the possibility of blind sources separation (BSS) of individual pixel, we prove the exact solution of LCNN for 2 independent sources per pixel as follows.

Exact Solution of LCNN: Theorem The analytical solution of LCNN of two sources is

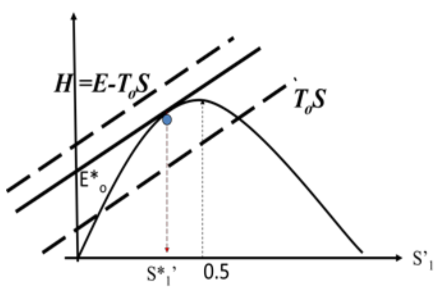


Figure 2: Exact LCNN pixel by pixel Solution

$$s_1^* = 1 - \exp(-\frac{E_o^*}{K_B T_o})$$

Derivation: Convert discret Boltzmann-Shannon entrpy to a single variable s_1 by normalization $s_1 + s_2 = 1$.

$$\frac{S(s_1)}{K_B} = -s_1 \log s_1 - s_2 \log s_2$$

$$= -s_1 \log s_1 - (1 - s_1)\log(1 - s_1),$$

We consider the fixed point solution:

$$Min.H = E - T_o S = 0;$$

so that

$$E = T_o\, S = -K_B T_o [s_1 \log s_1 + (1 - s_1)\log(1 - s_1)]$$

The linear vector geometry predicts another equation:

$$E = intercept + slope\ s_1 = E_o^* + \frac{dE}{ds_1}(s_1 - 0)$$

Consequently, the fixed point slope is computed

$$\frac{dE}{ds_1} = T_o \frac{dS}{ds_1} = T_o K_B (\log(1 - s_1) - \log s_1).$$

$$E = E_o^* + T_o \frac{dS}{ds_1} s_1 = E_o^* + T_o K_B (\log(1 - s_1) - \log s_1) s_1$$

Two formulas must be equal to each other at $s_1 = s_1^*$ yields

$$\frac{E_o^*}{K_B T_o} = -\log(1 - s_1^*). \quad \text{Q.E.D.}$$

The convergence proof of LCNN has been given by Dr. Miao's thesis using Nonlinear LCNN based on Kuhn-Tucker augmented Lagrange methodology. (IEEE IP V.16, pp1008-1021, 2007).

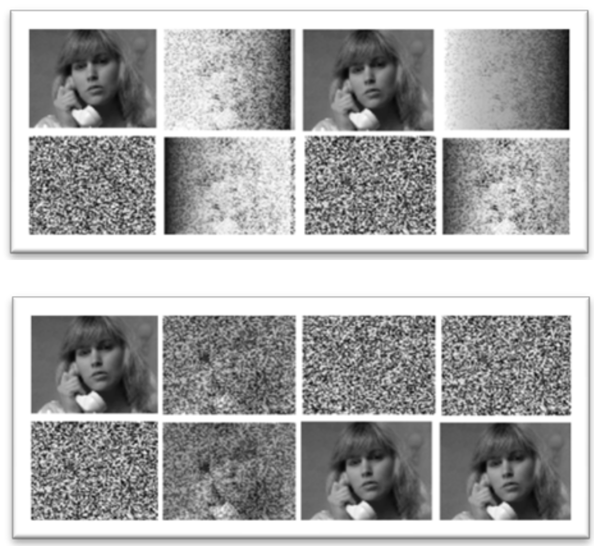


Figure 3: A face picture and a normal noise are mixed by a point nonlinearly (top panel of 8 images; LHS sources) or linearly (bottom panel of 8 images; LHS sources). The top panel has furthermore a column-wise changing mixing matrix (space-variant NL mixing), while the bottom panel has a uniform or identical mixing matrix (space-invariant linear mxing). Since LCNN is a pixel-by-pixel solution, it is designed for a massive and parallel implementation (SIMD, a la Flynn taxonomy), LCNN can solve both the top and the bottom panel at the 3rd column. However, BSAO info-max algorithm cannot solve the top 4th column based on a NL space-variant mixing; only the bottom panel 4th column at a linear and identical mixing matrix for the batch ensemble average.

2.15 Interdisciplinary Contributions

Besides the aforementioned, the author is aware of the interdisciplinary contributions made by mathematicians (e.g. V. Cherkassky, Jose Principe, Lei Xu; Asim Roy);

physicists (e.g. John Taylor, David Brown, Lyon Cooper); and biologists (e.g. Ishikawa Masumi, Mitsuo Kawato, Rolf Eckmiller, Shio Usui); as well as engineers (e.g. Kunihico Fukushima, K.S. Narendra, Robert Hecht-Nielsen, Bart Kosko, George Lendaris, Nikola Kassabov, Jacez Zurada, Cheng-Yuan Liou Of Taiwan U, You-Shu Wu of Tsing Hwa U, Huiseng Chi of Peking U, Toshio Fukuda, Hideo Aiso of 5^{th} Gen Computing, et al.). The author apologized that he could not cover theirs and others younger contributors in this short survey.

Combining ICA and CS linear algebras produced the Feature Organized Sparseness (FOS) theorem in Sect. 7. Contrary to purely random sparseness, the bio-inspired turns out to have the additional meaning, i.e., the locations indicating dramatic moment of changes at salient spatial pixel features. The measure of significance is quantified by the orthogonal degree among admissible states of AM storage [12]. Thus, the author reviewed only the math leading to ICA unsupervised learning to enlighten the Compressive Sensing (CS) community. Traditional ANNs learning are based on pairs of exemplars with desired outputs in the LMS errors, l_2-norm performance. A decade later, modern ANN has evolved close to the full potential of unsupervised connectionists (but still lack the architecture of self-organizing capability). We emphasize in this review that the HVS is a real time processing at a replenishing firing rate of about 1/17 seconds; HVS operates a smart AM tracking of those selected images in order to be retrieved instantly by those orthogonal salient features stored in the Hippocampus.

3. Neuroscience Organized Sparseness

We wish to help interdisciplinary readership appreciate the organization principle of spatiotemporal sparse orthogonal representation for Compressive Sensing. The purpose of sparse orthogonal representation is help increase the chance of pattern hits. A sparse orthogonal representation in computer science is $\{e_i\} = \{(1,0,0,0..), (0,1,0,0,..), .; \ i = 1,2,.)\}$ known as a finite state machine. Human Visual System (HVS) has taken Hubel Wiesel oriented edge map wavelet $[\Psi_1, ..]$ that transform a sparse orthogonal basis to another a sparse feature representation,

$$[\vec{f}_1, ..] = [\Psi_1, ..]^T [\vec{e}_1, ...]$$

which is not a mathematically "Purely Random Sparseness," rather a biologically " Feature-orthogonal Organized Sparseness (FOS)". We shall briefly review Neuroscience 101 of HVS.

3.1 Human Visual Systems (HVS)

Physiologically speaking, the HVS has a uniformly distributed fovea composed of 4 million RG color vision cones for high resolution spot size. 2 millions B cones are distributed in the peripheral out side the central fovea in order to receive the high blue sky and the low blue lake water at 0.4μ wavelengths. This is understood by a simple geometrical ray inversion of our orbital lens when the optical axis is mainly focused on the horizon of green forest and bushes.

Based on Einstein's wavelength-specific photoelectric effect, the G cones, which have Rhodopsin pigment molecules sensitive to green wavelengths shorter than 0.5μ, can perceive trees, grasses, and bushes. Some primates whose G cone's Rhodopsin suffered DNA mutation of (M, L)-genes, developed a remarkable capability of spotting ripe red fruits hidden among green bushes with higher fructose content at a longer wavelength of about 0.7μ. These primates could feed more offspring, and more offsprings inherited the same trait, who then had more offspring, and so on, so forth. Abundant offspring eventually became tri-color Homo sapiens (whose natural intelligence may be endowed by God). Owing to the mutations, it is not surprising that the retina is examined by means of RGB functional stained florescence.

Millions of RG cones were found under a microscope arranged in a seemingly *random sparse* pattern among housekeeping glial (Muller) cells and B cones in the peripheral of the fovea. What is the biological mechanism for organizing sparse representation? If too many of them directly and insatiately send their responses to the brain whenever activated by the incoming light, the brain will be saturated, habituated, and complacent. That's perhaps why HVS developed a summing layer consisting of millions of Ganglions (gang of lions). Massive photo-sensors cones are located at the second layer, shielded behind a somewhat translucent ganglion layer. The Ganglions act as gatekeeper traffic cops between the eyes and the Cortex. A ganglion fires only if the synaptic gap membrane potentials surpass a certain threshold. This synaptic junction gap impedance can serve as a threshold suppressing random thermal fluctuations. It then fires in the Amplitude Modulation mode of membrane potential for a short distance, or Frequency Modulation mode of firing rates for a long distance.

3.2 Novelty Detection

Our ancestors paid attention to novelty detection defined by orthogonal property among local center of gravity changes. Otherwise, our visual cortex may become complacent from too many routine stimuli. Our ancestors further demanded a simple and rapid explanation of observed phenomena with paramount consequences (coded in AM of e-Brain). Thus, we developed a paranoid bias toward unknown events. For example, miracles must have messages, rather than accidents; or 'rustling bushes must be a crouching tiger about to jump out,' rather than 'blowing winds'. Thus, this meaning interpretation has been hard wired and stored in the AM of Hippocampus of e-brain. That's why in biomedical experiments, care must be given whenever rounding off decimals. A double-blind protocol (to the analyst and volunteer participants) with a (negative) control is often demanded, in order to suppress the bias of [AM] interpretation toward False Positive Rate. "In God we trust, all the rest show data." NIH Motto. We now know even given the data set, it's not enough unless there is a sufficient sampling in a double-blind with a control protocol (blind to the patients and the researchers who gets the drugs or the placebo, mixed with a control of no disease). Dr. Naoyuki Nakao thought in his e-brain about how to avoid potential kidney failure for high blood pressure patients who loseproteins. He could have been advocating an intuitive thought about the dual therapy of hypertension drugs; that both ACE inhibitor upon a certain hormone and ARB acting in a different way on the same hormone should be cooperatively administrated together. The paper appeared in the Lancet J. and since Jan. 2003, became the top 2 cited index in a decade. Swiss Dr. Regina Kurz discovered, "the data is too perfect to be true for small sample size of 366 patients." As a result, this dual drug therapy has affected 140K patients, causing a Tsunami of paper retractions at a 7 folds increase.

3.3 Single Photon Detection

The photons, when detected by cones or rods made of multiple stacks of disks, converts the hydrocarbon chain of Rhodopsin pigment molecules from the *cis* configuration to *trans* configuration. As a result, the alternating single and double carbon bonds of trans carbon chains are switching continuously in a domino effect until it reaches and polarizes the surface membrane potential. Then, the top disk has a 'trans state' and will not be recovered until it is taken care of at the mirror reflection layer, and converted back to the 'cis state' upward from the cone or rod base.
A single signal photon at the physiological temperature can be seen at night. No camera can do that without cryogenic cooling (except semiconductor Carbon Nano-Tube (CNT) IR sensor, Xi Ning & H. Szu). To detect a single moonlight photon, we must combat against thermal fluctuations at $300^o\ K \cong \frac{1}{40} eV$. How the thermal noise is cancelled without cooling operated at physiology temperatures. This is accomplished by synaptic gap junctions of a single ganglion integrating 100 rods. These 100 Rod bundles can sense a single moonlight photon $(1\mu\sim 1\ eV)$ because there exist a 'dark light' current when there is no light, as discovered by Hagin [4]. Our eyes supplies the electric current energy necessary to generate the 'dark currents.' It is an ion current made of Potassium inside the Rod and Sodium outside the Rod, circulating around each Rods. (i) Nature separates the signal processing energy from the signal information energy, because (ii) a single night vision photon does not have enough energy to drive the signal current to the back of brain; but may be enough (iii) to de-polarize the membrane potential to switch off the no signal 'dark current,' by 'negate the converse' logic. Any rod of the bundle of 100 rods receives a single photon that can change the rod's membrane potential to detour the 'Hagins dark currents' away from the Rod. Consequently, it changes the ganglion pre-synaptic junction membrane potential. As a result, the incoming photon changes the membrane potential and the ganglion fires at 100Hz using different reservoir energy budget for reporting the information [4]. A single ganglion synaptic junction gap integrating over these 100 rods bundle provides a larger size of the bundle to overcomes (v) the spatial uncertainty principle of a single photon wave mechanics. These (i~v) are lesson learned from biosensors. Another biosensor lesson is MPD computing by the architecture as follows.

3.4 Scale Invariance by Architecture

The pupil size has nothing to do with the architecture of the rod density distribution. The density drops off outside the fovea, along the polar radial direction in an exponential fashion. Thereby, the peripheral night vision can achieve a graceful degradation of imaging object size. This fan-in architecture allows the HVS to achieve scale invariance mathematically, as follows. These 1.4 millions night vision ganglion axon fibers are squeezed uniformly through the fovea channel, which closely packs them uniformly toward the LGN and visual cortex 17 in the back of head. The densities of Rods' and B-cones increase first and drop gently along the radial direction, in an exponential increase and decrease fashion:

$$Input\ locations = \exp(\pm Output\ uniform\ location),$$

which can therefore achieve a graceful degradation of the size variances by means of a mathematical logarithmic transformation in a MPD fashion without computing, just flow through with the fan-in architecture. This is because of the inverse $Output = \pm\log(Input) \cong \text{Output}'$, when Input = 2 x Input' because log(2) is negligible. This size invariance allows our ancestor to run in the moonlight while chasing after a significant other to integrate the intensity rapidly and continuously over the time without computational slow down. For photon-rich day vision, the high density fovea ganglions require 100 Hz firing rate, which might require a sharing of the common pool of resources, before replenishing because the molecular kinetics produces a natural supply delay. As a result, the ganglions who use up the resource will inhibit neighborhood ganglions firing rates, producing the lateral inhibition on-center-off-surround, the so-called Hubel and Wiesel oriented edge wavelet feature map $[\psi_n]$.[5]

3.5 Division of Labor

It's natural to divide our large brain into left and right hemispheres corresponding to our symmetric body limbs reversely. Neurophysiologic speaking, we shall divide our 'learning/MPD storing/thinking' process into a balanced slow and fast process. In fact, Nobel Laureate Prof. Daniel Kahneman wrote about the decision making by slow and fast thinking in his recent book published in 2011. We may explain the quick thinking in terms of intuitive thinking of the emotional side of right hemisphere (in short 'e-Brain') & the logical slow thinking at the left hemisphere, '*l*-Brain'. In fact, Eckhard Hess conducted experiments demonstrating pupil dynamics (as the window of brains) which is relaxed in a dilation state during a hard mental task which uses up mental energy and contracted iris to fit the intensity needed once the computing task is complete. We wish to differentiate by designing different tasks which part of the brain (*l*-brain, *e*-brain) is doing the task. This way we may find the true time scale of each hemisphere. For example, putting together a jigsaw puzzle depicting a picture of your mother or a boring geometry pattern may involve the *e*-Brain or *l*-Brain. How fast can our e-brain or *l*-brain do the job? In the cortex center, there are pairs of MPD storages called the hippocampus, which are closer to each other in female than male.

The female might be more advanced than male for a better lateralization and environmental-stress survivability. The faster learning of speech, when a female is young or the female has a better chance of recovery when one side of the brain was injured. Such a division of labors connected by the lateralization seems to be natural balance to build in ourselves as a self-correction mechanism.

3.6 Lateralization between e-Brain & *l*-Brain

According to F. Crick & C. Koch in 2005, the consciousness layer is a wide & thin layer, called Claustrum, located underneath the center brain and above the lower part lizard brain. The Claustrum acts like a music conductor of brain sensory orchestra, tuning at a certain C note for all sensory instruments (by the winner-take-all masking effect).The existence of a conscious toning remains to be experimentally confirmed (e.g. studying an anesthesia awakening might be good idea). It could be above the normal EEG brain waves types known as alpha, beta, theta, etc., and underneath the decision making neuron firing rate waves at 100 Hz. This pair of hippocampus requires the connection mediated by the Claustrum known as the Lateralization. According to the equilibrium minimum of thermodynamic Helmholtz free energy, the sensory processing indeed happens effortlessly at the balance between minimum energy and maximum entropy, we are operating at.

The sparse orthogonal is necessary for HVS, but also natural for brain neuronal representation. We have 10 billion neurons and 100 billion synapses with some replenishment and regeneration, the synapses could last over 125 years. An other reason the sparse orthogonal representation is not loaded up with all the degree of freedoms and no longer has a free will for generalization. In other words, unlike a robot having a limited memory capacity and computing capability, we prefer to keep our brain degrees of freedom as sparse as possible, about 10 ~15% level (so-called the least developed place on the Earth) about 10% x $10^{20} \supseteq$ *encyclopedia Brittainca*. Todd and Marols in Nature 2004 [6] summarized the capacity limit of visual short-term memory in human Posterior Parietal Cortex (PPC) where sparsely arranged neuronal population called grandmother neurons fires intensely for 1 second without disturbing others, supporting our independence concept yielding our orthogonality attribute. The 'grandmother neuron(s)' may be activated by other stimulus and memories, but is the *sole* representation of 'grandmother' for the individual. To substantiate the *electric brain response* as a *differential response* of visual *event related potentials,* Pazo-Alvarez et al in Neuroscience 2004 [7] reviewed various modalities of brain imaging methodologies, and confirmed the biological base of feature organized sparseness (FOS) to be based on automatic *comparison–selection* of changes. "How many views or frames does a monkey need in order to tell a good zookeeper from a bad one?" Monkeys select 3 distinctive views, which we refer to as *m frames*: frontal, side and a 45^o view [8]. Interestingly, humans need only m = 2 views when constructing a 3-D building from architectural blueprints, or for visualizing a human head. These kind of questions, posed by Tom Poggio et al. in 2003 [8], can be related to an important medical imaging application.

4. Orthogonal Sparse States of Associative Memory

Since the semiconductor storage technology has become inexpensive or 'silicon dirt cheap,' we can apparently afford wasteful 2-D MPD AM storage for 1-D vectors. Here, we illustrate how MPD AM can replace a current digital disk drive storage, a -pigeon-a-hole, without suffering recall confusion and search delays. The necessary and sufficient condition of such AM admissible states requires that rank-1 vector outer product is orthogonal as depicted in Fig.4. Thus, we recapitulate the essential attributes, sparseness and orthogonality as follows.

4.1 Connectionist Storage

Given facial images $\vec{X}_{N,t}$, three possible significant or salient features such as the *eyes, nose,* and *mouth* can be extracted in the rounding-off cool limit with the maximum firing rate of 100 Hz to one and lower firing rates to zero: (*1, 0*) ≡ (*big, small*). When these neuronal firing rates broadcast among themselves, they form the *Hippocampus* [AM] at the synaptic gap junctions denoted by the weight matrix $W_{i,j}$. For an example, when a small child is first introduced to his/her Aunt and Uncle, in fact the image of Uncle gets compared with Aunt employing the 5 senses. Further, fusion of information from all senses is conducted beneath the cortical layer through the Claustrum[13]. The child can distinguish Uncle by multi-sensing and noticing that he has a normal sized mouth (0), a bigger (1) nose as compare to Aunt, and normal sized eyes (0). These features can be expressed as firing rates $f_{old} \equiv (n_1, n_2, n_3) \equiv$ (*eye, nose, mouth*) ≡ (*0, 1, 0*) which turns out to be the coordinate $\hat{y}$ axis of the family feature space. Likewise, the perception of an Aunt with big (1) eyes, smaller (0) nose, and smaller (0) mouth (1,0,0) forms another coordinate axis $\hat{x}$. Mathematically k/N=0.3 selection of sparse saliency features satisfies the orthogonality criterion for ANN classifier. This ANN sparse classifier not only satisfies the nearest neighbor classifier principle, but also the Fisher's Mini-Max classifier criterion for intra-class minimum spread and inter-class maximum separation [9]. Alternatively, when Uncle smiles, the child generates a new input feature set $f_{new} \equiv (n_1, n_2, n_3) \equiv$ (*eye, nose, mouth*) ≡ (*0, 1, 1*) through the same neural pathway. Then the response arrive at the hippocampus where the AM system recognizes and/or corrects the new input back to the most likely match, the big-nose Uncle state (*0, 1, 0*) with a fault tolerance of direction cos(45°). We write 'data' to the AM by an outer-product operation between the Uncle's feature vector in both column and row forms and overwrite Aunt's data to the same 2-D storage without cross talk confusion. This MPD happens among hundred thousand neurons in a local unit. The child reads Uncle's smile as a new input. The AM matrix vector inner product represents three feature neurons (0,1,1) that are sent at 100 Hz firing rates through the AM architecture of Fig. 1c. Further, the output (0,1,0) is obtained after applying a sigmoid σ_o threshold to each neuron which confirms that he remains to be the big nose Uncle.

4.2 Write

Write by the vector outer product repeatedly over-written onto the identical storage space forming associative matrix memory [A M]. Orthogonal features are necessarily for soft failure indicated in a 3-dimensional feature subspace of N-D.

$$[AM]_{big\ nose\ uncle} = \overrightarrow{output} \otimes \overrightarrow{input} = \begin{bmatrix}0\\1\\0\end{bmatrix} \begin{matrix}[0 & 1 & 0]\\ \begin{bmatrix}0 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}\end{matrix}$$

$$[AM]_{big\ eye\ aunt} = \overrightarrow{output} \otimes \overrightarrow{input} = \begin{bmatrix}1\\0\\0\end{bmatrix} \begin{matrix}[1 & 0 & 0]\\ \begin{bmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}\end{matrix}$$

MPD over-writing storage:

$$[AM]_{big\ nose\ uncle} + [AM]_{big\ eye\ aunt} = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$$

4.3 Read

Read by the vector inner product recalling from the sparse memory template and employing the nearest neighbor to correct input data via the vector inner product:

$$Recall\ Vector = [AM][error\ transmitted] \cong$$

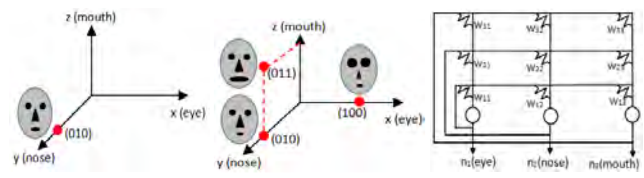
$$\sigma_o\left(\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}0\\1\\1\end{bmatrix}\right) = \begin{bmatrix}0\\1\\0\end{bmatrix}\ smiling\ uncle\ remains\ to\ be\ uncle$$

4.4 Fault Tolerant

AM erases the one-bit error (the lowest bit) recovering the original state which is equivalent to a semantic generalization: a big nosed smiling uncle is still the same big nose uncle. Thus for storage purpose, the orthogonality can produce either fault tolerance or a generalization as two sides of the same coin according to the orthogonal or independent feature vectors. In other words, despite his smile, the AM self corrected the *soft failure degree about the degrees of sparseness* $30\% \cong \frac{k}{N} = 0.3$, or *generalized* the original uncle feature set depending on Claustrum fusion layer [13] for supervision. We demonstrate the necessary and sufficient conditions of admissible AM states that are sampled by the ***selective attention*** called the *novelty detection* defined by significant changes forming an ***orthogonal*** subspace. Further, the measure of significance is defined as degree of orthogonal within the subspace or not.

(i) We take a binary threshold of all these orthogonal novelty change vectors as the picture index vectors [12].
(ii) We take a sequential pair of picture index vectors forming a vector outer-product in the 2-D AM fashion.
(iii) Moreover, we take the outer product between the picture index vector and its original high resolution image vector in a hetero-associative memory (HAM) for instantaneous image recovery.

Thus, these 2-D AM & HAM matrix memory will be the MPD storage spaces where all orthogonal pair products are over-written and overlaid without the need of search and the confusion of orthogonal PI retrieval. Consequently, AM enjoys the generalization by discovering a new component of the degree of freedom, cf. Section 4.



(a) (b) (c)

Figur 4a: Feature organized sparseness (FOS) may serve as the fault tolerance attribute of a distributive associative memory matrix. When a child meets his/her Aunt and Uncle for the first time, the child pays attention to extract three features neurons which fire at 100 Hz or less, represented by 1 or 0 respectively. Figure 4b: Even if uncle smiled at the child (indicated by (0,1,1) in the first quadrant), at the next point in time, the child can still recognize him by the vector inner product read procedure of the [AM] matrix and the new input vector (0, 1, 1). A smiling uncle is still the uncle as it should be. Mathematically speaking, the brain's Hippocampus storage has generalized the feature vector (0, 1, 0) to (0, 1, 1) for a smiling big nose uncle, at the [AM] matrix. However, if the feature vector is over-learned by another person (0, 0, 1), the degree of freedom is no longer sparse and is saturated. In this case, one can no longer have the NI capability of the innate generalization within the subspace. Fig.4c: This broadcasting communication circuitry network is called the artificial neural network (ANN) indicating adjustable or learnable weight values {W_{ij}; i,j=1,2,3} of the neuronal synaptic gaps among three neurons indicated by nine adjustable resistances where both uncle and aunt features memory are concurrently stored.

5. Spatiotemporal Compressive Sensing

The software can take over tracking the local center of gravity (CG) changes of the chips- 1) seeded with the supervised associative memory of pairs of image foreground chip (automatically cut by a built-in system on chip (SOC)), and 2) its role play (by users in the beginning of videotaping). The vector CG changes frame by frame are accumulated to form a net vector of CG change. The tail of a current change vector is added to the head of the previous change vector until the net change vector becomes orthogonal to the previously stored net CG vector. Then, the code will update the new net CG change vector with the previous one in the outer product hetero-associative memory (HAM), known as Motion Organized Sparseness (MOS), or Feature-role Organized Sparseness (FOS). Then, an optical expert system (Szu, Caulfield, 1987) can be employed to fuse the interaction library (IL) matrix [HAM] (IL-HAM) in a massive parallel distributive (MPD) computing fashion. Building the time order [AM] of each FOS, MOS, and [HAM] of IL, we wish to condense by ICA unsupervised learning a composite picture of a simple storyline, e.g. YouTube/BBC on eagle hunting a rabbit.

We have defined [12] a significant event involving a local Center of Gravity (CG) movement such as tiger jumping out off luffing around bushes (Fig.5). The processing window size may have a variable resolution with learnable window sizes in order to determine the optimum LCG movement. This may be estimated by a windowed Median filters (not Mean filter) used to select majority gray-value delegating-pixel locations and image weight values according to the local grey value histogram (64x64, 32x32, 16x16, etc.). Then we draw the optical flow vector from one local delegator pixel location to the next pointing from one to the next. The length of the vector is proportional to the delta change of gray values as the delegator weights. Similarly, we apply this Median Filter over all windows for two frames. We can sequentially update multiple frames employing optical flow vectors for testing the net summation. In this process, one vector tails-to-another vector head is plotted to cover a significant movement over half of the window size. Then the net is above threshold with the value of 'one' representing the whole window population to build a Picture Index (PI) (indicating a tiger might be jumping out with significant net CG movement); otherwise, the net CG will be threshold at zero (as the wind is blowing tree branches or bushes in a cyclic motion without a net CG motion). We could choose the largest jump CG among f frames.

Toward digital automation, we extracted the foreground from background by computing the local histogram based optical flow without tracking, in terms of a simplified medium filter finding a local center of gravity (CG). Furthermore, we generated the picture index (PI-AM) and the image-index (Image-HAM) MPD AM correlations [12]. We conjecture (TBD) another [HAM] of an interaction library (IL-AM) for fusion of storyline subroutine (Szu & Caulfield "Optical Expert System," App Opt. 1988) sketched as follows: AI pointer relational database, e.g. Lisp 1-D array (attribute (color), object (apple), values (red, green)) are represented by the vector outer products as 2-D AM maps. These maps are added with map frequency and restore a missing partial 2-D pattern as a new hypothesis. This type of interaction library can discover significant

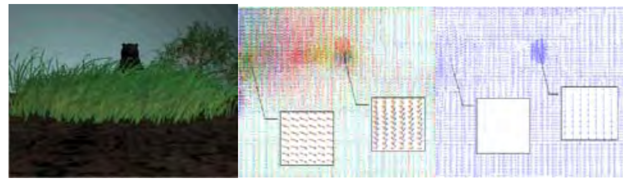


Figure 5: Net C.G. Automation [12]: (5a) Video images of a tiger is jumping out of wind-blowing bushes (Augmented Reality); (5b) The net Center of Gravity (CG) optical flow vectors accumulating f=5 frames reveals the orthogonal property to the previous net CG, capturing a Tiger jumping out region. This net CG motion of moving object (tiger) is different than wind-blowing region (bushes) in cyclic fluctuations; (5c) indicate an associated Picture Index sparse representation (dark blue dot consists of net CG vectors represented by ones, among short lines by zeros).

roles from selected foreground frames by generalizing AM. Further, this IL AM will follow the constructed storyline to compose these significant roles into a Video Cliff Note for tourist picture diary. For an example, a predator-prey video of about 4.5 minutes long was BBC copyrighted. Following the steps listed above, we have developed **compressive sampling (CSp)** video based on AM in terms of motion organized sparseness (MOS) as the picture index forming [AM] and its image as Hetero-AM [12]. Moreover, we have extended the concept with major changes shown as an automatic ***Video Image Cliff Notes.***

The lesson learned from the predator-prey BCC video is summarized in the Cliff Note, Fig.6, where a rapid change & stay-put as the keys for survivor(optical expert system, video Cliff Notes, SPIE DSS/ICA conf. Baltimore April, 2012).

6. Spatial-spectral Compressive Sensing Theory

We sketch a design of a new Smartphone camera that can take either daytime or nighttime picture with a single HVS focal plane array (FPA). Each pixel has a 2x2 Byres filter, which splits 1μ~1eV in quarter sizes and corrects different wavelength differences. Thus the filter trades off the spatial size resolution for increasing the spectral resolution. The camera adopts MPD [AM] & [HAM] storage in SSD medium. Such a handheld device may eventually become a personal secretary that can self-learn owner's habits, follow the itinerary with GPS during travel, and keep diary and send significant events. We can relax the 'purely random' condition of the sparseness sampling matrix $[\Phi]$ with feature organized sparseness $[\Phi_s]$, where 1s indicate the locations of potential discovery of features.

Theorem: Feature Organized Sparseness (FOS)

We shall derive a theorem to design ICA Unsupervised Learning methodology can help design FOS CS sampling matrix $[\Phi_s]$

$$[\Phi_s][\Psi] \equiv [ICA]; \quad [\Phi_s] = [ICA][\Psi]^{-1} \tag{9}$$

where $\{\psi_n\}$ is the Hubel-Wiesel wavelet modeled by the

Figure 6: A predator-prey video of about 4.5 minutes long was taken from YouTube/BBC for education & research purposes. It emulated unmanned vehicle UXV (X=A,G,M) useful Intelligence Surveillance Reconnaissance. An eagle *cruising* gathered the intelligence by a few glimpse of a moving prey during the *surveillance* in the sky; after identifying it as a jumping rabbit, the eagle made a *chase engagement*, closed its wings and dropped at the rabbit. Rabbit was detected via moving shadow, stayed motionless avoiding motion detection. The rabbit jumped away from the ground zero whereas the eagle lost its ability to maneuver due to semi-closed wings at terminal velocity and suffered a heavy fall. (2.11 AM Expert System is adopted to compile the image chiplets).

digital sub-band wavelet bases successfully applied to JPEG 2000 image compression and $\vec{s}$ is a column vector of feature sources $[\Phi_s]$ by solving ICA unsupervised learning. Feature Organized Sparseness (FOS) Compressive Sensing works not only for video motion features, but also for color spectral features if we treat the spectral index as time index.

Proof: We can readily verify the result by comparing the CS linear algebra with the ICA unsupervised learning algebra side-by-side as follows:

$$R^N \ \vec{X} = \sum_{n=1}^{N} s_n \psi_n = \sum_{n_k=1}^{k} s_{n_k} \psi_{n_k} = [\Psi], \qquad (10)$$

where k non-zero wavelets are denoted n_k=1,2,…,k <<N.

$$R^m\!: \ \vec{Y} = \sum_{i=1}^{m} x_i \phi_i^T = [\Phi_s]\vec{X} \ ; \qquad (11)$$

Substituting Eq(7) into Eq(8), the linear matrix relationship yields a desired exemplar image $\vec{Y}$ which has the unknown mixing matrix [ICA] and the unknown feature sources $\vec{s}$

$\vec{Y} = [\Phi_s][\Psi]\vec{s} \equiv [ICA]\vec{s}.$ ***Q.E.D.***

Adaptive Compressive Sensing:

We can exploit the full machinery of unsupervised learning ANN community about how to solve the Blind Sources Separation (BSS). We can either follow the Lagrange Constraint Neural Network based on minimizing the thermodynamic physics Helmholtz free energy by maximizing the a-priori source entropy [9] or the engineering filtering concept of maximizing the posterior de-mixed entropy of the output components [10]. For the edifice of the CS community that BSS is indeed possible, we have recapitulated the simplest possible linear algebra methodology with a simple proof as follow.

(i) *Symmetric Wiener Whitening in* ensemble average *matrix* $[W_z]^T = [W_z] =< \left[\vec{Y}\vec{Y}^T\right] >^{-\frac{1}{2}}$.

By definition $\vec{Y}' \equiv [W_z]\vec{Y}$ satisfying

$$< \vec{Y}'\vec{Y}'^T > \equiv [W_z] < \vec{Y}\vec{Y}^T > [W_z]^T = [I]$$

$$\because [W_z] < \vec{Y}\vec{Y}^T > [W_z]^T [W_z] = [I][W_z] = [W_z];$$

$\therefore [W_z]^T [W_z] =< \left[\vec{Y}\vec{Y}^T\right] >^{-1}; [W_z] =< \left[\vec{Y}\vec{Y}^T\right] >^{-\frac{1}{2}}$ Q.E.D.

(ii) Orthogonal Transform*:* $[W]^T = [W]^{-1}$

By definition

$[W]\vec{Y}' = [W][W_z]\vec{Y} = [W][W_z][\Phi_s][\Psi]\vec{s} \equiv [W][W_z][ICA]\vec{s} = \vec{s}$

$\because [W] < \vec{Y}'\vec{Y}'^T > [W]^T \equiv [W][I][W]^T =< \vec{s}\vec{s}^T > \cong [I];$

$\therefore [W]^T = [W]^{-1}$ Q.E.D.

The Step (ii) can reduce ICA de-mixing to orthogonal rotation. We can compute from these desired exemplar images from their corresponding sources employing simple geometrical solutions called the killing vector. This vector is orthogonal to all row vectors except for one, cf. Fig. 1. Further the rotation procedure generates a corresponding independent source along the specific gradient direction.

Since we have applied (i) Wiener whitening in image domain, and (ii) orthogonal matching pursuit to derive the feature sources, we can estimate by a pair the desired exemplar images with so-constructed feature sources by the rank-1 AM approximation of ICA mixing matrix [ICA]:

$$[ICA] = \sum \vec{y}\vec{s}^T = [\vec{y}_1, \vec{y}_2, \ldots][\vec{s}_1, \vec{s}_2, ..]^T. \qquad (12)$$

The correct CS linear programming could be used to compute a l_1-norm sparse constrained source representation $\vec{s}$ of the input $\vec{Y}$ in the LMS error sense. Our experience indicates a desirable orthogonality post-processing. Given all independent sources, we construct the orthogonal 1s (by the Gram-Schmidt procedure) $< \vec{s}\vec{s}^T > \cong [I]$.

$$[\Phi_s] \cong [\vec{y}_1, \vec{y}_2, \ldots][\vec{s}_1, \vec{s}_2, ..]^T [\Psi]^{-1}. \qquad (13)$$

Furthermore, we prefer the orthogonal feature extraction $[\Phi_s]$ such that $< [\Phi_s] \ [\Phi_s]^T > \cong [I]$. In doing so, we can increase the efficiency of multi/hyper-spectral compressive sensing methodology helping "finding a needle in a haystack" by sampling only the image correlated to the needle sources $\{\vec{s}_1, \vec{s}_2, .\}$ without unnecessarily creating a haystack of data cube blindly. (cf. Balvinder Kaur, et al., 2012 SPIE DSS/ICA Comp Sampling etc Conf. Baltimore)

7. Handheld Day-Night Smartphone Camera

Our goal is making a new handheld smartphone camera which can take both daytime and nighttime pictures with a single photon detector array. It can automatically keep and send only those significant frames capable of discovering motions and features. Our design logic is simple: never imaging daytime pictures with nighttime spectral, and vice versa, in a photon poor lighting or in the night do not take daytime color spectral picture. Of course, a simple clock time will do the job; but a smarter approach is through the correlation between exemplar images and desired features. We wish to design the camera with over-written 2-D storage in a MPD fashion, in terms of a FOS following the AM FTP rinciple. We can avoid the cross-talk confusion and unnecessary random access memory (RAM) search-delay, based on the traditional 1-D sequential optical CD technology storage concept: a pigeon-a-hole. This is a natural application of our Feature Organized Sparseness. We can build a full EOIR spectrum fovea camera applying a generalized Bayer filters using spectral-blind Photon Detectors (PD) emulating cones and rods per pixels. We mention that a current camera technology applied the Bayer color image filters (for RGB colors). We trade the spatial resolution with spectrum resolution. We take the spectral blind photon detector array of N pixels to measure N/4 color pixels. We modify the Bayer filter to be 4x4 per pixel and the extra 4th one is for extra night vision at near infrared 1 micron spectrum. We further correct optical path difference at new Bayer filter media in order to focus all spectrum on the same FPA, without the need of expensive achromatic correction in a compound lens.

Our mathematical basis is derived by combining both CS and ICA formulism, Eqs(6,7,8), and applying ICA unsupervised learning steps (i) & (ii) to design a FOS sampling matrix $[\Phi_s]$. Finding all the independent sources vectors from input day or night images $\vec{y}'s$ we collect expected sources $\vec{s}'s$ into a ICA mixing matrix $[ICA]=\sum \vec{y}\vec{s}^T$, then substituting its equivalence to CS sampling we can design FOS sampling matrix as $[\Phi_s]=[ICA][\Psi]^{-1}$ where $[\Psi]$ is usual image wavelet basis. The hardware is mapping the sparse feature sampling matrix onto 2x2 Bayer Filters per pixel that can afford to trade the spatial resolution with the spectral resolution in close up shots.

In this paper, we have further extended Motion Organized Sparseness [12] with Feature Organized Sparseness compositing two main players as a prey and a predator, namely a rabbit and an eagle. Their interaction is discovered by their chase after each other optical flows as shown in Fig. 6 as an automatic ***Video Image Cliff Notes***. Instead the purely random sparseness, we have generalized CS sampling matrix $[\Phi]$ with FOS sampling matrix$[\Phi_s]$.

In closing, we could estimate the complexity effect of replacing purely random sparseness $[\Phi]$ with FOS $[\Phi_s]$ upon the CRT&D RIP theorem. We could apply the complexity analysis tool called **Permutation Entropy** [14]. PE computes computed in a moving window of the size L=2,3, etc., counting the up-down shape feature of ones over the zeros: $H(L) \equiv -p(\pi)\sum p(\pi)$ of the k-organized sparseness to set a bound the sampling effect from purely random ones. For example, an organized sampling mask $[\Phi_s]_{m,N}$ had a row of {0,1,1,0,0,0…} which yielded a moving window of size L=2: in 4 cases[{01} up, {11}flat, {1,0}down; {0,0}flat, etc.]; size L=3 yields 3 cases[{0,1,1}up, {1,1,0}down; {1,0,0}down, {0,0,0} flat, etc.]. They had shown H(L) to be bounded from organized structure with ones locations (degree of complexity) to purely randomness (zero complexity) as $0 \le H(L) \le log\, L! \cong L\, logL - L$, by Sterling formula where $L \ll k \ll N$. $0 \cong PE([\Phi_s]_{m,N}) \ll PE([\Phi]_{m,N}) \cong 0(L)$ Therefore, instead of intractable l_0-constraint, we could equally use l_1-constrained LMS to both $[\Phi_s]_{m,N}$ and $[\Phi]_{m,N}$, if we were not already choosing HAM MPD for real time image recover [12].

Our teaching of the fittest survival may be necessary for early behaviors. The true survival of human species has to be co-evolved with other species and the environment we live in. This natural intelligence should be open and fair to all who are not so blindly focused by a narrowly defined discipline and ego. This imbalance leads to unnecessary greediness, affecting every aspect of our life. I publish this not for my need to survive; but to pay back the Communities who have taught me so much. The reader may carry on the unsupervised learning running on the fault tolerant and subspace-generalize-able connectionist architectures. Incidentally, Tai Chi practitioners by the walking meditation consider the Lao Tze's advocated mindless state, which is a balance between a fast thinking (Yin, light gravity weight) and a slow analyzing (Yang, heavy gravity weight). The transcendental meditation could achieve a low frequency brain wave (EEG Delta type), having the long wavelength reaching both sides of the hemispheres as the lateralization. If we were just relaxing your conscious-mind controlling muscles, and let the gravity potential takeover, the internal fluids that circulates freely insider our internal organs known as 'the Chi' can which can enhance the wellbeing about our physiology metabolism.

Acknowledgement

The author wishes to thank Dr. Charles Hsu and Army NVESD Mr. Jeff Jenkins, Ms. Lein Ma, Ms. Balvinder Kaur for their technical supports. Dr. Soo-Young Lee for his patience and encouragement. The author acknowledges US AFOSR grant support of CUA R/D as facilitated by Dean Prof. Charles C. Nguyen.

References

[1] E. J. Candes, J. Romberg, and T. Tao "Robust Uncertainty Principle: Exact Signal Reconstruction from Highly Incomplete Frequency Information," IEEE Trans IT, 52(2), pp.489-509, Feb.2006

[2] E. J. Candes, and T. Tao," Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies," IEEE Trans IT 52(12), pp.5406-5425, 2006.

[3] D. Donoho," Compressive Sensing," IEEE Trans IT, 52(4), pp. 1289-1306,2006

[4] W. A. Hagins (NIH), R. D. Penn, and S. Yoshikami; "Dark

current and photocurrent in retinal rods," pp. 380-412, Biophys. J. VOL. 10, 1970; F. Rieke (U. Wash), D.A. Baylor (Stanford), "Single-photon detection by rod cells of the retina," Rev. Mod. Phys., pp. 1027-1036, Vol. 70, No. 3, July 1998.
[5] Hubel, D. H., and Wiesel, T. N. (1962). "Receptive fields and functional architecture in the cat's visual cortex." J. Neurosci, 160, 106–154.
[6] J. Jay Todd and Rene Marols, "Capacity limit of visual short-term memory in human posterior parietal cortex," Nature 428, pp.751-754, 15 Apr. 2004.
[7] P. Pazo-Alvarez, RE. Amenedo, and F. Cadaveira, "Automatic detection of motion direction changes in the human brain, " E. J. Neurosci. **19**, pp.1978-1986, 2004.
[8] Martin A. Giese and Tomaso Poggio, "Neural mechanisms for the recognition of biological movements," Nature Reviews Neuroscience 4, 179-192 (March 2003)
[9] H. Szu, IN: Handbook of Computational Intelligence IEEE Ch.16; H. Szu, L. Miao, H. Qi, "Thermodynamics free-energy Minimization for Unsupervised Fusion of Dual-color Infrared Breast Images," SPIE Proc. ICA etc. V. 6247, 2006; H. Szu, I. Kopriva, "Comparison of LCNN with Traditional ICA Methods," WCCI/IJCNN 2002; H. Szu, P. Chanyagorn, I. Kopriva: "Sparse coding blind source separation through Powerline," Neurocomputing 48(1-4): 1015-1020 (2002); H. Szu: Thermodynamics Energy for both Supervised and Unsupervised Learning Neural Nets at a Constant Temperature. Int. J. Neural Syst. 9(3): 175-186 (1999); H. Szu and C. Hsu, "Landsat spectral Unmixing à la superresolution of blind matrix inversion by constraint MaxEnt neural nets, *Proc. SPIE 3078,* pp.147-160, 1997; H. Szu and C. Hsu, "Blind de-mixing with unknown sources," *Proc. Int. Conf. Neural Networks,* vol. 4, pp. 2518-2523, Houston, June, 1997. (Szu et al. Single Pixel BSS Camera, US PTO Patent 7, 355,182 Apr 8, 2008; US PTO Patent 7,366,564, Apr 29, 2008).
[10] A. J. Bell and T.J. Sejnowski, "A new learning algorithm for blind signal separation," adv. In Inf. Proc. Sys. **8**, MIT Press pp. 7547-763, 1996; A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," neu. Comp. 9, pp 1483-1492, July 1997; S. Amari, "Information Geometry," in Geo. and Nature Contemporary Math (ed. Nencka and Bourguignon) v.203, pp. 81-95, 1997.
[11] S. Y. Lee, et al. Colleagues, "Blind Sources Separation of speeches," and "NIO ffice Mate", cf. KAIST http://bsrc.kaist.ac.kr/new/future.htm.
[12] H. Szu, C. Hsu, J. Jenkins, J. Willey, J. Landa, "Capturing Significant Events with Neural Networks," J. Neural Networks (2012); M.K. Hsu, T.N. Lee and Harold Szu, "Re-establish the time order across sensors of different modalities," Opt Eng. V. 50(4), pp.047002-1~047002-15.
[13] F.C. Crick, C. Koch. 2005. "What is the function of the Claustrum?" Phil. Trans. of Royal Society B -Biological Sciences 360:1271-9.
[14] C. Brandt & B. Pompe, " Permutation Entropy-a natural complexity measure of time series,"PRL 2002; "Early detection of Alzheimer's onset with Permutation Entropy analysis of EEG," G. Morabito, A. Bramanti, D. Labate, F. La Foresta, F.C. Morabito, *Nat. Int*. (INNS) V.1, pp.30~32, Oct 2011.

Regular Paper

Evolving, Probabilistic Spiking Neural Networks and Neurogenetic Systems for Spatio- and Spectro-Temporal Data Modelling and Pattern Recognition

Nikola Kasabov, FIEEE, FRSNZ

Knowledge Engineering and Discovery Research Institute - KEDRI, Auckland University of Technology,
and Institute for Neuroinformatics - INI, ETH and University of Zurich
nkasabov@aut.ac.nz; www.kedri.info, ncs.ethz.ch/projects/evospike

Abstract

Spatio- and spectro-temporal data (SSTD) are the most common types of data collected in many domain areas, including engineering, bioinformatics, neuroinformatics, ecology, environment, medicine, economics, etc. However, there is lack of methods for the efficient analysis of such data and for spatio-temporal pattern recognition (STPR). The brain functions as a spatio-temporal information processing machine and deals extremely well with spatio-temporal data. Its organisation and functions have been the inspiration for the development of new methods for SSTD analysis and STPR. The brain-inspired spiking neural networks (SNN) are considered the third generation of neural networks and are a promising paradigm for the creation of new intelligent ICT for SSTD. This new generation of computational models and systems are potentially capable of modelling complex information processes due to their ability to represent and integrate different information dimensions, such as time, space, frequency, and phase, and to deal with large volumes of data in an adaptive and self-organising manner. The paper reviews methods and systems of SNN for SSTD analysis and STPR, including single neuronal models, evolving spiking neural networks (eSNN) and computational neuro-genetic models (CNGM). Software and hardware implementations and some pilot applications for audio-visual pattern recognition, EEG data analysis, cognitive robotic systems, BCI, neurodegenerative diseases, and others are discussed.

Keywords: Spatio-temporal data, spectro-temporal data, pattern recognition, spiking neural networks, gene regulatory networks, computational neuro-genetic modeling, probabilistic modeling, personalized modelling; EEG data.

1. Spatio- and Spectro-Temporal Data Modeling and Pattern Recognition

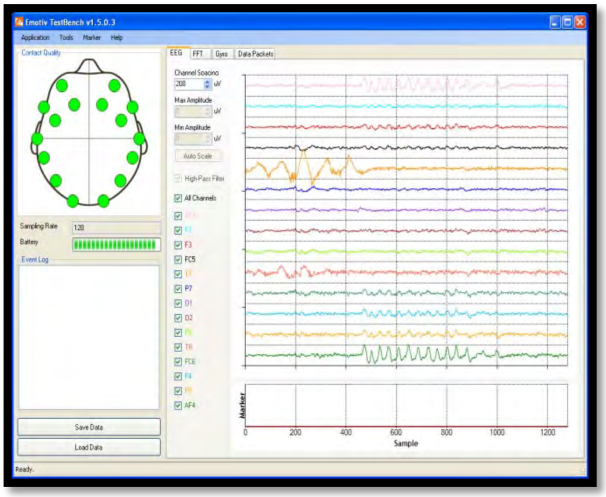
Most problems in nature require spatio- or/and spectro-temporal data (SSTD) that include measuring spatial or/and spectral variables over time. SSTD is described by a triplet (**X**,**Y**,**F**), where **X** is a set of independent variables measured over consecutive discrete time moments *t*; **Y** is the set of dependent output variables, and **F** is the association function between whole segments ('chunks') of the input data, each sampled in a time window d_t, and the output variables belonging to Y:

$$\mathbf{F}: \mathbf{X}(d_t)\rightarrow \mathbf{Y}, \quad \mathbf{X}(t)=(\mathbf{x}_1(t),\mathbf{x}_2(t),\ldots,\mathbf{x}_n(t)),\ t=1,2,\ldots,n \quad (1)$$

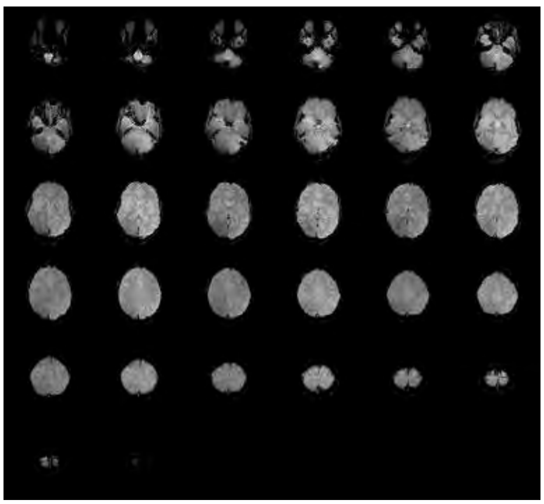
It is important for a computational model to capture and learn *whole* spatio- and spectro-temporal patterns from data streams in order to predict most accurately future events for new input data. Examples of problems involving SSTD are: brain cognitive state evaluation based on spatially distributed EEG electrodes [70, 26, 51, 21, 99, 100] (Fig.1(a)); fMRI data [102] (Fig.1(b)); moving object recognition from video data [23, 60, 25] (Fig.15); spoken word recognition based on spectro-temporal audio data [93, 107]; evaluating risk of disease, e.g. heart attack [20]; evaluating response of a disease to treatment based on clinical and environmental variables, e.g. stroke [6]; prognosis of outcome of cancer [62]; modelling the progression of a neuro-degenerative disease, such as Alzheimer's Disease [94, 64]; modelling and prognosis of the establishment of invasive species in ecology [19, 97]. The prediction of events in geology, astronomy, economics and many other areas also depend on accurate SSTD modeling.

The commonly used models for dealing with temporal information based on Hidden Markov Models (HMM) [88] and traditional artificial neural networks (ANN) [57] have limited capacity to achieve the integration of complex and long temporal spatial/spectral components because they usually either ignore the temporal dimension or over-simplify its representation. A new trend in machine learning is currently emerging and is known as *deep machine learning* [9, 2-4, 112]. Most of the proposed models still learn SSTD by entering single time point frames rather than learning whole SSTD patterns. They are also limited in addressing adequately the interaction between temporal and spatial components in SSTD.

The human brain has the amazing capacity to learn and recall patterns from SSTD at different time scales, ranging from milliseconds to years and possibly to millions of years (e.g. genetic information, accumulated through evolution). Thus the brain is the ultimate inspiration for the development of new machine learning techniques for SSTD



(a)



(b)

Fig.1(a) EEG SSTD recorded with the use of Emotive EEG equipment (from McFarland, Anderson, MÜller, Schlögl, Krusienski, 2006); (b) fMRI data (from http://www.fmrib.ox.ac.uk)

modelling. Indeed, brain-inspired Spiking Neural Networks (SNN) [32, 33, 68] have the potential to learn SSTD by using trains of spikes (binary temporal events) transmitted among spatially located synapses and neurons. Both spatial and temporal information can be encoded in an SNN as locations of synapses and neurons and time of their spiking activity respectively. Spiking neurons send spikes via connections that have a complex dynamic behaviour, collectively forming an SSTD memory. Some SNN employ specific learning rules such as Spike-Time-Dependent-Plasticity (STDP) [103] or Spike Driven Synaptic Plasticity (SDSP) [30]. According to the STDP a connection weight between two neurons increases when the pre-synaptic neuron spikes before the postsynaptic one. Otherwise, the weight decreases.

Models of single neurons as well as computational SNN models, along with their respective applications, have been already developed [33, 68, 73, 7, 8, 12], including evolving connectionist systems and evolving spiking neural networks (eSNN) in particular, where an SNN learns data incrementally by one-pass propagation of the data via creating and merging spiking neurons [61, 115]. In [115] an eSNN is designed to capture features and to aggregate them into audio and visual perceptions for the purpose of person authentification. It is based on four levels of feed-forward connected layers of spiking neuronal maps, similarly to the way the *cortex* works when learning and recognising images or complex input stimuli [92]. It is an SNN realization of some computational models of vision, such as the 5-level HMAX model inspired by the information processes in the cortex [92].

However, these models are designed for (static) object recognition (e.g. *a picture of a cat*), but not for moving object recognition (e.g. *a cat jumping to catch a mouse*). If these models are to be used for SSTD, they will still process SSTD as a sequence of static feature vectors extracted in single time frames. Although an eSNN accumulates incoming information carried in each consecutive frame from a pronounced word or a video, through the increase of the membrane potential of output spike neurons, they do not learn complex spatio/spectro-temporal associations from the data. Most of these models are deterministic and do not allow to model complex stochastic SSTD.

In [63, 10] a computational neuro-genetic model (CNGM) of a single neuron and SNN are presented that utilize information about how some proteins and genes affect the spiking activities of a neuron, such as fast excitation, fast inhibition, slow excitation, and slow inhibition. An important part of a CNGM is a dynamic gene regulatory network (GRN) model of genes/proteins and their interaction over time that affect the spiking activity of the neurons in the SNN. Depending on the task, the genes in a GRN can represent either biological genes and proteins (for biological applications) or some system parameters including probability parameters (for engineering applications).

Recently some new techniques have been developed that allow the creation of new types of computational models, e.g.: probabilistic spiking neuron models [66, 71]; probabilistic optimization of features and parameters of eSNN [97, 96]; reservoir computing [73, 108]; personalized modelling frameworks [58, 59]. This paper reviews methods and systems for SSTD that utilize the above and some other contemporary SNN techniques along with their applications.

2. Single Spiking Neuron Models

2.1 A biological neuron

A single biological neuron and the associated synapses is a complex information processing machine, that involves short term information processing, long term information storage, and evolutionary information stored as genes in the nucleus of the neuron (Fig.2).

2.2 Single neuron models

Some of the-state-of-the-art models of a spiking neuron include: early models by Hodgkin and Huxley [41] 1952;

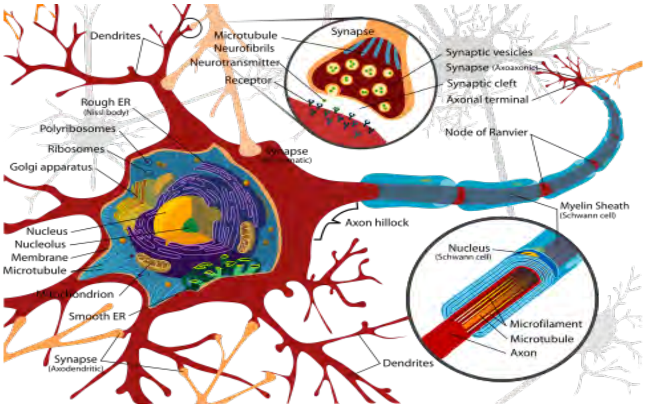


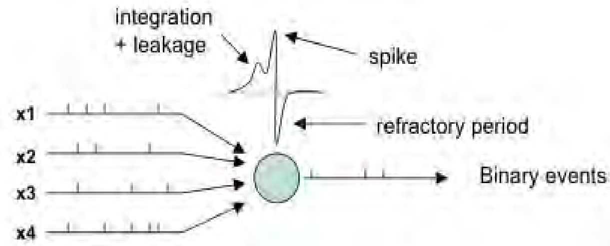
Fig.2. A single biological neuron with the associated synapses is a complex information processing machine (from Wikipedia)

more recent models by Maas, Gerstner, Kistler, Izhikevich and others, e.g.: Spike Response Models (SRM) [33, 68]; Integrate-and-Fire Model (IFM) [33, 68]; Izhikevich models [52-55], adaptive IFM, and others.

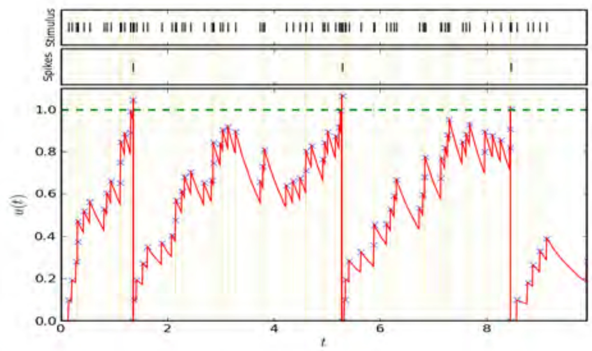
The most popular for both biological modeling and engineering applications is the IFM. The IFM has been realised on software-hardware platforms for the exploration of patterns of activities in large scale SNN under different conditions and for different applications. Several large scale architectures of SNN using IFM have been developed for modeling brain cognitive functions and engineering applications. Fig. 3(a) and (b) illustrate the structure and the functionality of the Leaky IFM (LIFM) respectively. The neuronal post synaptic potential (PSP), also called membrane potential u(t), increases with every input spike at a time *t* multiplied to the synaptic efficacy (strength) until it reaches a threshold. After that, an output spike is emitted and the membrane potential is reset to an initial state (e.g. 0). Between spikes, the membrane potential leaks, which is defined by a parameter.

An important part of a model of a neuron is the model of the synapses. Most of the neuronal models assume scalar synaptic efficacy parameters that are subject to learning, either on-line or off-line (batch mode). There are models of dynamics synapses (e.g. [67, 71, 72]), where the synaptic efficacy depends on synaptic parameters that change over time, representing both long term memory (the final efficacy after learning) and short term memory – the changes of the synaptic efficacy over a shorter time period not only during learning, but during recall as well.

One generalization of the LIFM and the dynamic synaptic models is the probabilistic model of a neuron [66] as shown in fig.4a, which is also a biologically plausible model [45, 68, 71]. The state of a spiking neuron n_i is described by the sum $PSP_i(t)$ of the inputs received from all m synapses. When the $PSP_i(t)$ reaches a firing threshold $\vartheta_i(t)$, neuron n_i fires, i.e. it emits a spike. Connection weights ($w_{j,i}$, j=1,2,...,m) associated with the synapses are determined during the learning phase using a learning rule. In addition to the connection weights $w_{j,i}(t)$, the probabilistic spiking neuron model has the following three probabilistic parameters:



(a)



(b)

Fig.3. (a) The structure of the LIFM. (b) functionality of the LIFM

- A probability $p_{cj,i}(t)$ that a spike emitted by neuron n_j will reach neuron n_i at a time moment *t* through the connection between n_j and n_i. If $p_{cj,i}(t)=0$, no connection and no spike propagation exist between neurons n_j and n_i. If $p_{cj,i}(t) = 1$ the probability for propagation of spikes is 100%.
- A probability $p_{sj,i}(t)$ for the synapse $s_{j,i}$ to contribute to the PSPi(t) after it has received a spike from neuron n_j.
- A probability $p_i(t)$ for the neuron n_i to emit an output spike at time *t* once the total PSP_i (t) has reached a value above the PSP threshold (a noisy threshold).

The total $PSP_i(t)$ of the probabilistic spiking neuron n_i is now calculated using the following formula [66]:

$$PSP_i(t) = \sum_{p=t_0,.,t} \left(\sum_{j=1,..,m} e_j f_1(p_{cj,i}(t-p)) f_2(p_{sj,i}(t-p)) w_{j,i}(t) + \eta(t-t_0) \right) \quad (2)$$

where e_j is 1, if a spike has been emitted from neuron n_j, and 0 otherwise; $f_1(p_{cj,i}(t))$ is 1 with a probability $p_{cji}(t)$, and 0 otherwise; $f_2(p_{sj,i}(t))$ is 1 with a probability $p_{sj,i}(t)$, and 0 otherwise; t_0 is the time of the last spike emitted by n_i; $\eta(t-t_0)$ is an additional term representing decay in the PSP_i. As a special case, when all or some of the probability parameters are fixed to "1", the above probabilistic model will be simplified and will resemble the well known IFM. A similar formula will be used when a leaky IFM is used as a fundamental model, where a time decay parameter is introduced.

It has been demonstrated that SNN that utilises the probabilistic neuronal model can learn better SSTD than traditional SNN with simple IFM, especially in a nosy environment [98, 83]. The effect of each of the above three probabilistic parameters on the ability of a SNN to process

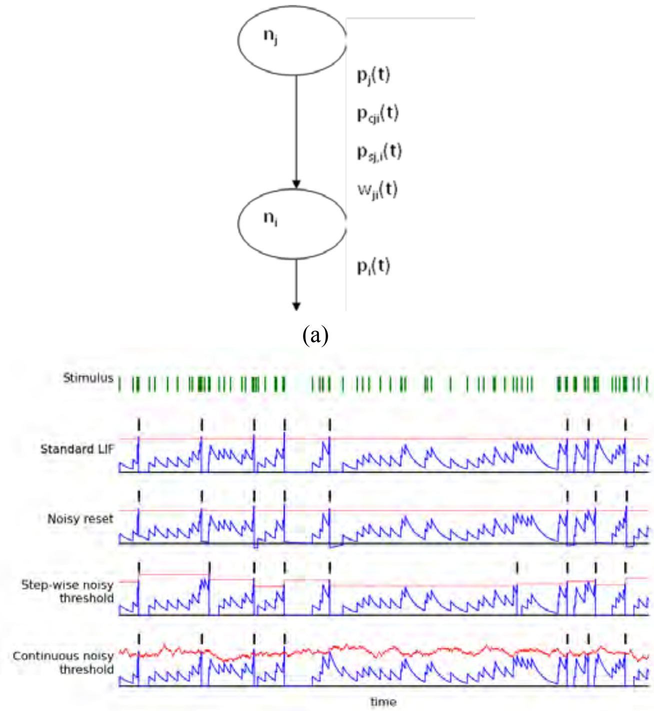


Fig.4 (a) A simple probabilistic spiking neuron model (from [66]); (b) Different types of noisy thresholds have different effects on the output spikes (from [99, 98]).

noisy and stochastic information was studied in [98]. Fig. 4(b) presents the effect of different types of nosy thresholds on the neuronal spiking activity.

2.3 A neurogenetic model of a neuron

A neurogenetic model of a neuron is proposed in [63] and studied in [10]. It utilises information about how some proteins and genes affect the spiking activities of a neuron such as *fast excitation, fast inhibition, slow excitation, and slow inhibition*. Table 1 shows some of the proteins in a neuron and their relation to different spiking activities. For a real case application, a part from the GABAB receptor some other metabotropic and other receptors could be also included. This information is used to calculate the contribution of each of the different synapses, connected to a neuron n_i, to its post synaptic potential PSPi(t):

$$\varepsilon_{ij}^{synapse}(s) = A^{synapse}\left(\exp\left(-\frac{s}{\tau_{decay}^{synapse}}\right) - \exp\left(-\frac{s}{\tau_{rise}^{synapse}}\right)\right) \quad (3)$$

where $\tau_{decay/rise}^{synapse}$ are time constants representing the rise and fall of an individual synaptic PSP; A is the PSP's amplitude; $\varepsilon_{ij}^{synapse}$ represents the type of activity of the synapse between neuron j and neuron i that can be measured and modelled separately for a fast excitation, fast inhibition, slow excitation, and slow inhibition (it is affected by different genes/proteins). External inputs can also be added to model background noise, background oscillations or environmental information.

An important part of the model is a dynamic gene/protein regulatory network (GRN) model of the dynamic interactions between genes/proteins over time that affect the spiking activity of the neuron. Although biologically plausible, a GRN model is only a highly simplified general model that does not necessarily take into account the exact chemical and molecular interactions. A GRN model is defined by:

(a) a set of genes/proteins, G= (g_1,g_2,…, g_k);
(b) an initial state of the level of expression of the genes/proteins G(t=0);
(c) an initial state of a connection matrix L = (L_{11},…, L_{kk}), where each element L_{ij} defines the known level of interaction (if any) between genes/proteins g_j and g_i;
(d) activation functions f_i for each gene/protein g_i from G. This function defines the gene/protein expression value at time (t+1) depending on the current values G (t), L (t) and some external information E(t):

$$g_i(t+1)= f_i\,(G(t), L(t), E(t)) \quad (4)$$

3. Learning and Memory in a Spiking Neuron

3.1 General classification

A learning process has an effect on the synaptic efficacy of the synapses connected to a spiking neuron and on the information that is memorized. Memory can be:

- Short-term, represented as a changing PSP and temporarily changing synaptic efficacy;
- Long-term, represented as a stable establishment of the synaptic efficacy;
- Genetic (evolutionary), represented as a change in the genetic code and the gene/ protein expression level as a result of the above short-term and long term memory changes and evolutionary processes.

Learning in SNN can be:

- Unsupervised - there is no desired output signal provided;
- Supervised – a desired output signal is provided;
- Semi-supervised.

Different tasks can be learned by a neuron, e.g:

- Classification;
- Input-output spike pattern association.

Several biologically plausible learning rules have been introduced so far, depending on the type of the information presentation: (3)

- Rate-order learning, that is based on the average spiking activity of a neuron over time [18, 34, 43];
- Temporal learning, that is based on precise spike times [44, 104, 106, 13, 42];
- Rank-order learning, that takes into account the order of spikes across all synapses connected to a neuron [105, 106].

Rate-order information representation is typical for cognitive information processing [18].

Table 1. Neuronal action potential parameters and related proteins and ion channels in the computational neuro-genetic model of a spiking neuron: AMPAR - (amino- methylisoxazole- propionic acid) AMPA receptor; NMDR - (N-methyl-D-aspartate acid) NMDA receptor; GABAAR - (gamma-aminobutyric acid) $GABA_A$ receptor, $GABA_BR$ - $GABA_B$ receptor; SCN - sodium voltage-gated channel, KCN - kalium (potassium) voltage-gated channel; CLC - chloride channel (from Benuskova and Kasabov, 2007)

Different types of action potential of a spiking neuron used as parameters for its computational model	Related neurotransmitters and ion channels
Fast excitation PSP	AMPAR
Slow excitation PSP	NMDAR
Fast inhibition PSP	$GABA_AR$
Slow inhibition PSP	$GABA_BR$
Modulation of PSP	mGluR
Firing threshold	Ion channels SCN, KCN, CLC

Temporal spike learning is observed in the auditory [93], the visual [11] and the motor control information processing of the brain [13, 90]. Its use in neuro-prosthetics is essential, along with applications for a fast, real-time recognition and control of sequence of related processes [14].

Temporal coding accounts for the precise time of spikes and has been utilised in several learning rules, most popular being Spike-Time Dependent Plasticity (STDP) [103, 69] and SDSP [30, 14]. Temporal coding of information in SNN makes use of the exact time of spikes (e.g. in milliseconds). Every spike matters and its time matters too.

3.2 The STDP learning rule

The STDP learning rule uses Hebbian plasticity [39] in the form of long-term potentiation (LTP) and depression (LTD) [103, 69]. Efficacy of synapses is strengthened or weakened based on the timing of post-synaptic action potential in relation to the pre-synaptic spike (example is given in Fig.5(a)). If the difference in the spike time between the pre-synaptic and post-synaptic neurons is negative (pre-synaptic neuron spikes first) than the connection weight between the two neurons increases, otherwise it decreases. Through STDP, connected neurons learn consecutive temporal associations from data. Pre-synaptic activity that precedes post-synaptic firing can induce long-term potentiation (LTP), reversing this temporal order causes long-term depression (LTD).

3.3 Spike Driven Synaptic Plasticity (SDSP)

The SDSP is an unsupervised learning method [30, 14], a modification of the STDP, that directs the change of the synaptic plasticity V_{w0} of a synapse w_0 depending on the time of spiking of the pre-synaptic neuron and the post-synaptic neuron. V_{w0} increases or decreases, depending on the relative timing of the pre- and post-synaptic spikes.

If a pre-synaptic spike arrives at the synaptic terminal before a postsynaptic spike within a critical time window, the synaptic efficacy is increased (potentiation). If the post-synaptic spike is emitted just before the pre-synaptic spike, synaptic efficacy is decreased (depression). This change in synaptic efficacy can be expressed as:

$$\Delta V_{w0} = \frac{I_{pot}(t_{post})}{C_p}\Delta t_{spk} \quad \text{if } t_{pre} < t_{post} \tag{5}$$

$$\Delta V_{w0} = -\frac{I_{dep}(t_{post})}{C_d}\Delta t_{spk} \quad \text{if } t_{post} < t_{pre} \tag{6}$$

where Δt_{spk} is the pre- and post-synaptic spike time window.

The SDSP rule can be used to implement a supervised learning algorithm, when a teacher signal, that copies the desired output spiking sequence, is entered along with the training spike pattern, but without any change of the weights of the teacher input.

The SDSP model is implemented as an VLSI analogue chip [49]. The silicon synapses comprise bistability circuits for driving a synaptic weight to one of two possible analogue values (either potentiated or depressed). These circuits drive the synaptic-weight voltage with a current that is superimposed on that generated by the STDP and which can be either positive or negative. If, on short time scales, the synaptic weight is increased above a set threshold by the network activity via the STDP learning mechanism, the bi-stability circuits generate a constant weak positive current. In the absence of activity (and hence learning) this current will drive the weight toward its potentiated state. If the STDP decreases the synaptic weight below the threshold, the bi-stability circuits will generate a negative current that, in the absence of spiking activity, will actively drive the weight toward the analogue value, encoding its depressed state. The STDP and bi-stability circuits facilitate the implementation of both long-term and short term memory.

3.4 Rank-order learning

The rank-order learning rule uses important information from the input spike trains – the rank of the first incoming

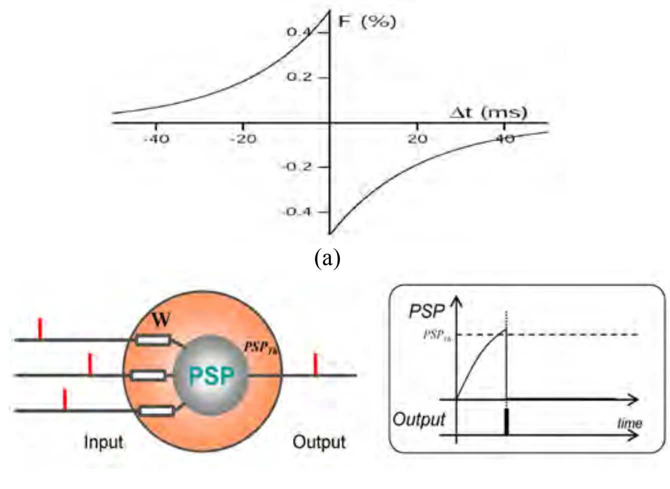


Fig.5. (a) An example of synaptic change in a STDP learning neuron [103]; (b) Rank-order learning neuron. .

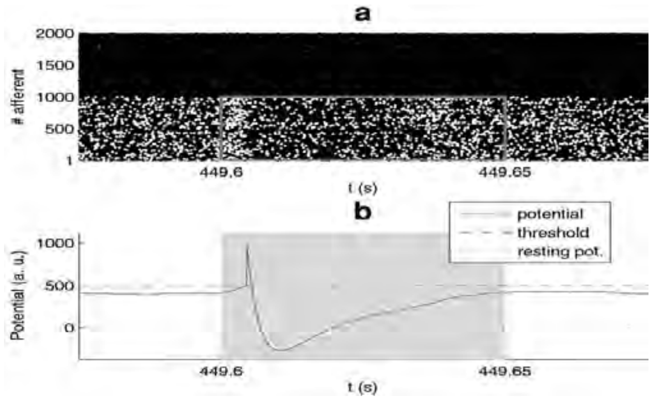


Fig.6. A single LIF neuron with simple synapses can be trained with the STDP unsupervised learning rule to discriminate a repeating pattern of synchronised spikes on certain synapses from noise (from : T. Masquelier, R. Guyonneau and S. Thorpe, PlosONE, Jan2008))

spike on each synapse (Fig.5(b)). It establishes a priority of inputs (synapses) based on the order of the spike arrival on these synapses for a particular pattern, which is a phenomenon observed in biological systems as well as an important information processing concept for some STPR problems, such as computer vision and control [105, 106]. This learning makes use of the extra information of spike (event) order. It has several advantages when used in SNN, mainly: fast learning (as it uses the extra information of the order of the incoming spikes) and asynchronous data entry (synaptic inputs are accumulated into the neuronal membrane potential in an asynchronous way). The learning is most appropriate for AER input data streams [23] as the events and their addresses are entered into the SNN 'one by one', in the order of their happening.

The postsynaptic potential of a neuron i at a time t is calculated as:

$$PSP(i,t) = \sum_{j} \mathrm{mod}^{order(j)} w_{j,i} \qquad (7)$$

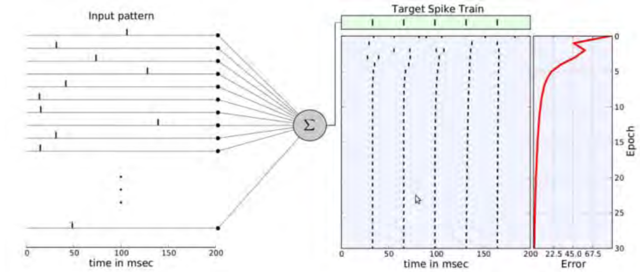
where *mod* is a modulation factor; j is the index for the incoming spike at synapse j,i and $w_{j,i}$ is the corresponding synaptic weight; $order(j)$ represents the order (the rank) of the spike at the synapse j,i among all spikes arriving from all m synapses to the neuron i. The $order(j)$ has a value 0 for the first spike and increases according to the input spike order. An output spike is generated by neuron i if the PSP (i,t) becomes higher than a threshold PSPTh (i).

During the training process, for each training input pattern (sample, example) the connection weights are calculated based on the order of the incoming spikes [105]:

$$\Delta w_{j,i}(t) = mod^{\ order\ (j,i\ (t))} \qquad (8)$$

3.5 Combined rank-order and temporal learning

In [25] a method for a combined rank-order and temporal (e.g. SDSP) learning is proposed and tested on benchmark data. The initial value of a synaptic weight is set according to the rank-order learning based on the first incoming spike on this synapse. The weight is further modified to



A single output neuron is trained to respond with a temporally precise output spike train to a specific spatio-temporal input.

(a)

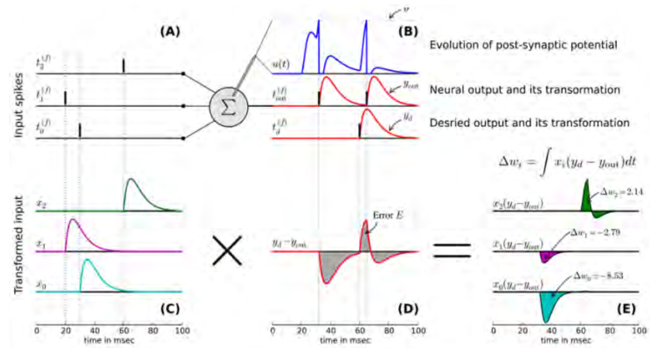
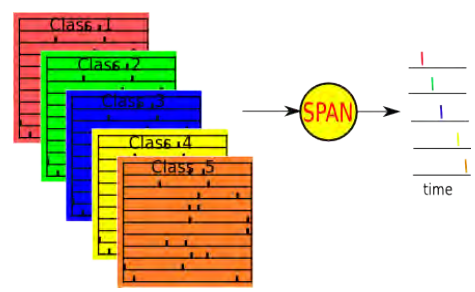
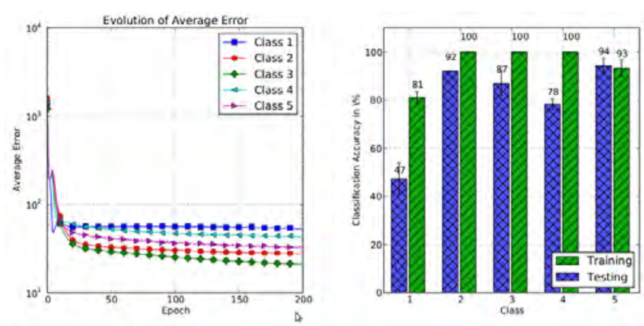


Illustration of the proposed training algorithm.

(b)



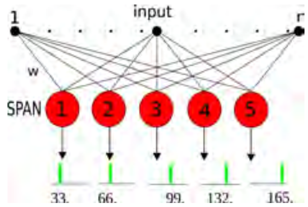
(c)



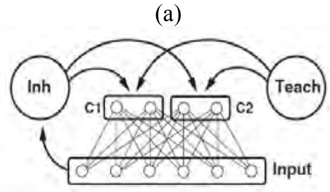
Evolution of the average erros obtained in 30 independent trails for each class of the training samples, and the average accuracies obtained in the training and testing phase.

(d)

Fig.7 (a) The SPAN model [77]. (b) The Widrow-Hoff Delta learning rule applied to learn to associate an output spike sequence to an input STP [77, 30]. (c) The use of a single SPAN neuron for the classification of 5 STP belonging to 5 different classes [77]. (d) The accuracy of classification is rightly lower for the class 1 – spike at the very beginning of the input pattern as there is no sufficient input data).



(a)



(b)

Fig.8: (a) Multiple SPAN neurons [76]. (b) Multiple SDSP trained neurons [14]

accommodate following spikes on this synapse with the use of a temporal learning rule – SDSP.

4. STPR in a Single Neuron

In contrast to the distributed representation theory and to the widely popular view that a single neuron cannot do much, some recent results showed that a single neuronal model can be used for complex STPR.

A single LIF neuron, for example, with simple synapses can be trained with the STDP unsupervised learning rule to discriminate a repeating pattern of synchronised spikes on certain synapses from noise (from: T. Masquelier, R. Guyonneau and S. Thorpe, PlosONE, Jan2008) – see Fig. 6.

Single neuron models have been introduced for STPR, such as: Temportron [38]; Chronotron [28]; ReSuMe [87]; SPAN [76, 77]. Each of them can learn to emit a spike or a spike pattern (spike sequence) when a certain STP is recognised. Some of them can be used to recognise multiple STP per class and multiple classes [87, 77, 76].

'Fig.7(a)-(d) show a SPAN neuron and its use for classification of 5 STP belonging to 5 different classes [77]. The accuracy of classification is rightly lower for the class 1 (the neuron emits a spike at the very beginning of the input pattern) as there is no sufficient input data – Fig.7(d).) [77].

5. Evolving Spiking Neural Networks

Despite the ability of a single neuron to conduct STPR, a single neuron has a limited power and complex STPR tasks will require multiple spiking neurons.

One approach is proposed in the evolving spiking neural networks (eSNN) framework [61, 111]. eSNN evolve their structure and functionality in an on-line manner, from incoming information. For every new input pattern, a new neuron is dynamically allocated and connected to the input neurons (feature neurons). The neurons connections are established for the neuron to recognise this pattern (or a similar one) as a positive example. The neurons represent centres of clusters in the space of the synaptic weights. In some implementations similar neurons are merged [61, 115]. That makes it possible to achieve a very fast learning in an eSNN (only one pass may be necessary), both in a supervised and in an unsupervised mode.

In [76] multiple SPAN neurons are evolved to achieve a better accuracy of spike pattern generation than a single SPAN – Fig.8(a).

In [14] the SDSP model from [30] has been successfully used to train and test a SNN for 293 character recognition (classes). Each character (a static image) is represented as 2000 bit feature vector, and each bit is transferred into spike rates, with 50Hz spike burst to represent 1 and 0 Hz to represent 0. For each class, 20 different training patterns are used and 20 neurons are allocated, one for each pattern (altogether 5860) (Fig.8(b)) and trained for several hundreds of iterations.

A general framework of eSNN for STPR is shown in Fig.9. It consists of the following blocks:

- Input data encoding block;
- Machine learning block (consisting of several sub-blocks);
- Output block.

In the input block continuous value input variables are transformed into spikes. Different approaches can be used:

- population rank coding [13] – Fig.10(a);
- thresholding the input value, so that a spike is generated if the input value (e.g. pixel intensity) is above a threshold;
- Address Event Representation (AER) - thresholding the difference between two consecutive values of the

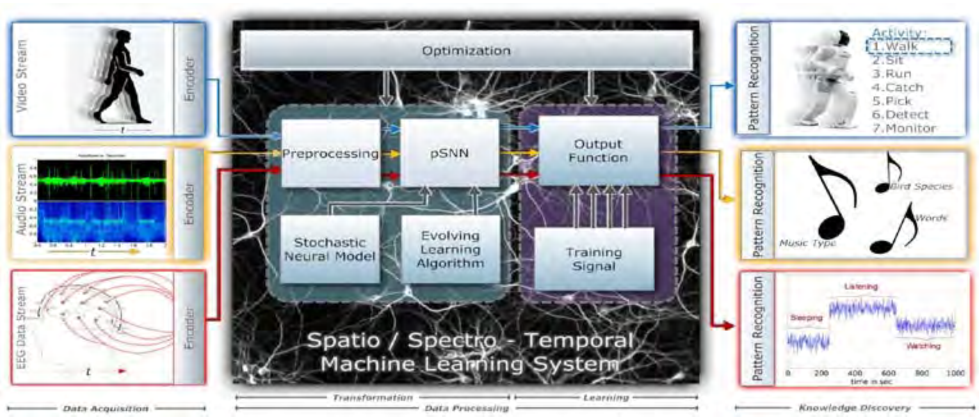
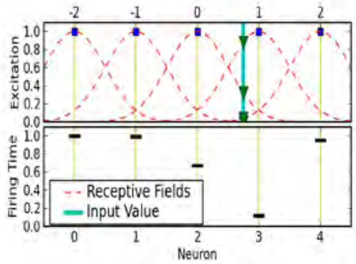
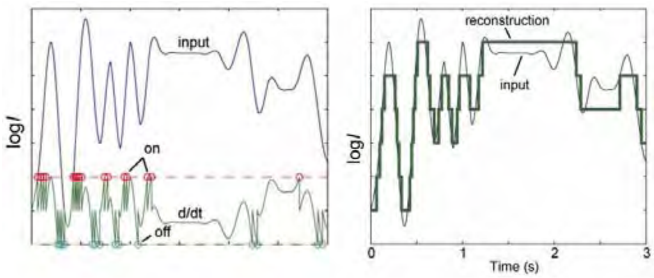


Fig.9. The eSNN framework for STPR (from: http://ncs.ethz.ch/projects/evospike)



(a)



(b)

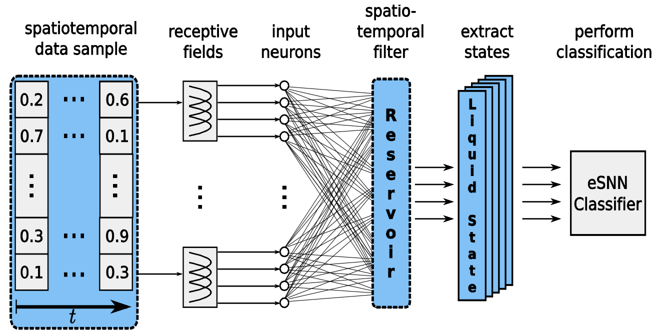
Fig.10. (a) Population rank order coding of input information; (b) Address Event Representations (AER) of the input information [23].

same variable over time as it is in the artificial cochlea [107] and *a*rtificial retina devices [23] – Fig.10(b).

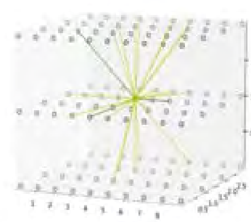
The input information is entered either on-line (for on-line, real time applications) or as a batch data. The *time* of the input data is in principal different from the internal SNN *time* of information processing.

Long and complex SSTD cannot be learned in simple one-layer neuronal structures as the examples in Fig.8(a) and (b). They require neuronal 'buffers' as shown in Fig.11(a). In [82] a 3D buffer was used to store spatio-temporal 'chunks' of input data before the data is classified. In this case the size of the chunk (both in space and time) is fixed by the size of the reservoir. There are no connections between the layers in the buffer. Still, the system outperforms traditional classification techniques as it is demonstrated on sign language recognition, where eSNN classifier was applied [61, 115].

Reservoir computing [73, 108] has already become a popular approach for SSTD modelling and pattern recognition. In the classical view a 'reservoir' is a homogeneous, passive 3D structure of probabilistically connected and fixed neurons that in principle has no learning and memory, neither it has an interpretable structure – fig.11b. A reservoir, such as a Liquid State Machine (LSM) [73, 37], usually uses *small world recurrent connections* that do not facilitate capturing explicit spatial and temporal components from the SSTD in their relationship, which is the main goal of learning SSTD. Despite difficulties with the LSM reservoirs, it was shown on several SSTD problems that they produce better results than using a simple classifier [95, 73, 99, 60]. Some publications demonstrated that probabilistic neurons are suitable for reservoir computing especially in a noisy environment [98, 83].



(a)



(b)

Fig.11. (a) An eSNN architecture for STPR using a reservoir; (b) The structure and connectivity of a reservoir

In [81] an improved accuracy of LSM reservoir structure on pattern classification of hypothetical tasks is achieved when STDP learning was introduced into the reservoir. The learning is based on comparing the liquid states for different classes and adjusting the connection weights so that same class inputs have closer connection weights. The method is illustrated on the phone recognition task of the TIMIT data base phonemes – spectro-temporal problem. 13 MSCC are turned into trains of spikes. The metric of separation between liquid states representing different classes is similar to the Fisher's *t*-test [27].

After a presentation of input data example (or a 'chink' of data) the state of the SNN reservoir $S(t)$ is evaluated in an output module and used for classification purposes (both during training and recall phase). Different methods can be applied to capture this state:

- Spike rate activity of *all* neurons at a certain time window: The state of the reservoir is represented as a vector of n elements (n is the number of neurons in the reservoir), each element representing the spiking probability of the neuron within a time window. Consecutive vectors are passed to train/recall an output classifier.
- Spike rate activity of spatio-temporal clusters C_1, C_2, … C_k of close (both in space and time) neurons: The state $S_{Ci}(t)$ of each cluster C_i is represented by a single number, reflecting on the spiking activity of the neurons in the cluster in a defined time window (this is the internal SNN time, usually measured in '*msec*'). This is interpreted as the current spiking probability of the cluster. The states of all clusters define the current reservoir state $S(t)$. In the output function, the cluster states $S_{Ci}(t)$ are used differently for different tasks.
- Continuous function representation of spike trains: In contrast to the above two methods that use spike rates to evaluate the spiking activity of a neuron or a neuronal

cluster, here the train of spikes from each neuron within a time window, or a neuronal cluster, is transferred into a continuous value temporal function using a kernel (e.g. α-kernel). These functions can be compared and a continuous value error measured.

In [95] a comparative analysis of the three methods above is presented on a case study of Brazilian sign language gesture recognition (see Fig.18) using a LSM as a reservoir.

Different adaptive classifiers can be explored for the classification of the reservoir state into one of the output classes, including: statistical techniques, e.g. regression techniques; MLP; eSNN; nearest-neighbour techniques; incremental LDA [85]. State vector transformation, before classification can be done with the use of adaptive incremental transformation functions, such as incremental PCA [84].

6. Computational Neurogenetic Models (CNGM)

Here, the neurogenetic model of a neuron [63, 10] is utilized. A CNGM framework is shown in Fig.12 [64].

The CNGM framework comprises a set of methods and algorithms that support the development of computational models, each of them characterized by:

- Two-tire, consisting of an eSNN at the higher level and a gene regulatory network (GRN) at the lower level, each functioning at a different time-scale and continuously interacting between each other;
- Optional use of probabilistic spiking neurons, thus forming an epSNN;
- Parameters in the epSNN model are defined by genes/proteins from the GRN;
- Can capture in its internal representation both spatial and temporal characteristics from SSTD streams;
- The structure and the functionality of the model evolve in time from incoming data;
- Both unsupervised and supervised learning algorithms can be applied in an on-line or in a batch mode.
- A concrete model would have a specific structure and a set of algorithms depending on the problem and the application conditions, e.g.: classification of SSTD; modelling of brain data.

The framework from Fig.12 supports the creation of a multi-modular integrated system, where different modules, consisting of different neuronal types and genetic parameters, represent different functions (e.g.: vision; sensory information processing; sound recognition; motor-control) and the whole system works in an integrated mode.

The neurogenetic model from Fig.12 uses as a main principle the analogy with biological facts about the relationship between spiking activity and gene/protein dynamics in order to control the learning and spiking parameters in a SNN when SSTD is learned. Biological support of this can be found in numerous publications (e.g. [10, 40, 117, 118]).

The Allen Human Brain Atlas (www.brain-map.org) of the Allen Institute for Brain Science (www.alleninstitute.org) has shown that at least 82% of the human genes are expressed in the brain. For 1000 anatomical sites of the brains of two individuals 100 mln data points are collected that indicate gene expressions of each of the genes and underlies the biochemistry of the sites.

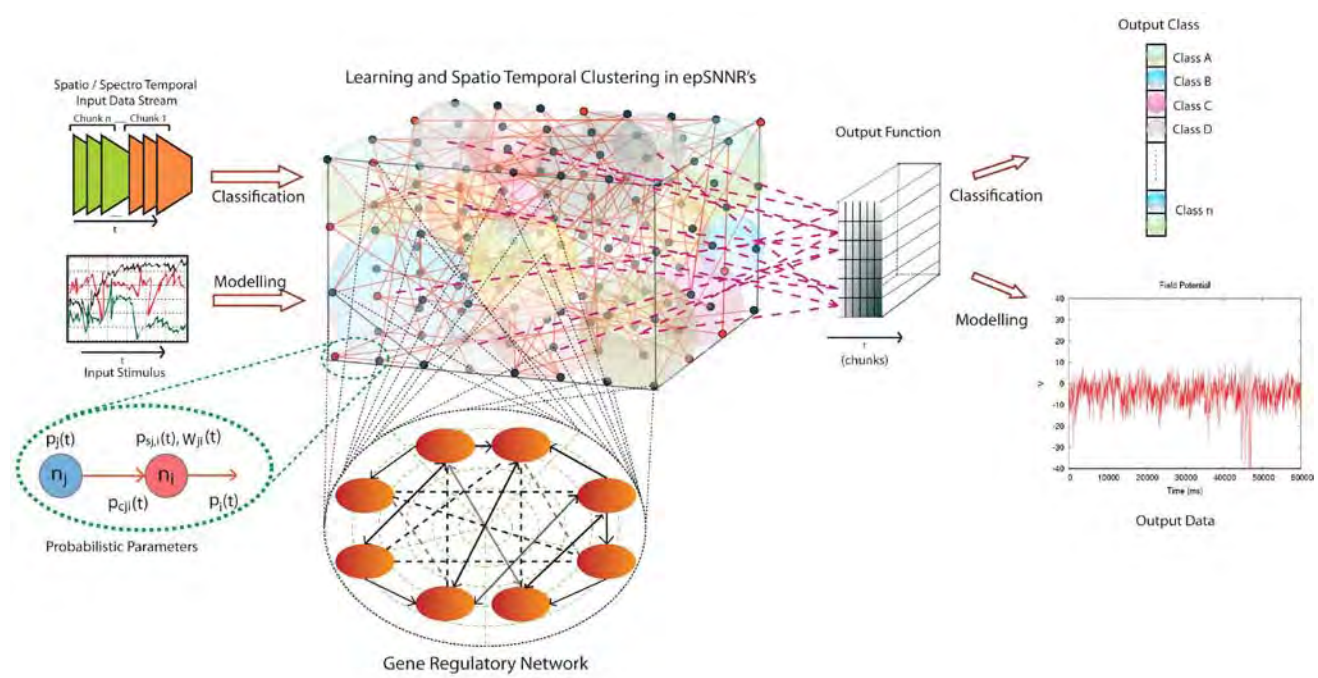


Fig.12. A schematic diagram of a CNGM framework, consisting of: input encoding module; a SNN reservoir output function for SNN state evaluation; output classifier; GRN (optional module). The framework can be used to create concrete models for STPR or data modelling (from [64]).

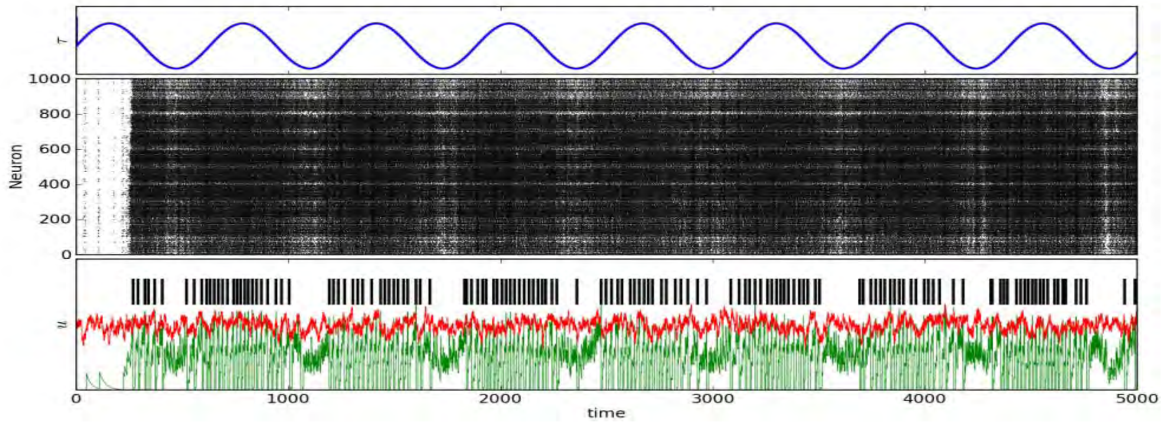


Fig.13. A GRN interacting with a SNN reservoir of 1000 neurons. The GRN controls a single parameter, i.e. the τ parameter of all 1000 LIF neurons, over a period of five seconds. The top diagram shows the evolution of τ. The response of the SNN is shown as a raster plot of spike activity. A black point in this diagram indicates a spike of a specific neuron at a specific time in the simulation. The bottom diagram presents the evolution of the membrane potential of a single neuron from the network (green curve) along with its firing threshold ϑ (red curve). Output spikes of the neuron are indicated as black vertical lines in the same diagram (from [65]).

In [18] it is suggested that both the firing rate (rate coding) and spike timing as spatiotemporal patterns (rank order and spatial pattern coding) play a role in fast and slow, dynamic and adaptive sensorimotor responses, controlled by the cerebellar nuclei. Spatio-temporal patterns of population of Purkinji cells are shaped by activities in the molecular layer of interneurons. In [40] it is demonstrated that the temporal spiking dynamics depend on the spatial structure of the neural system (e.g. different for the hippocampus and the cerebellum). In the hippocampus the connections are scale free, e.g. there are hub neurons, while in the cerebellum the connections are regular. The spatial structure depends on genetic pre-determination and on the gene dynamics. Functional connectivity develops in parallel with structural connectivity during brain maturation. A growth-elimination process (synapses are created and eliminated) depend on gene expression [40], e.g. glutamatergic neurons issued from the same progenitors tend to wire together and form ensembles, also for the cortical GABAergic interneuron population. Connections between early developed neurons (mature networks) are more stable and reliable when transferring spikes than the connections between newly created neurons (thus the probability of spike transfer). Postsynaptic AMPA-type glutamate receptors (AMPARs) mediate most fast excitatory synaptic transmissions and are crucial for many aspects of brain function, including learning, memory and cognition [10, 31].

It was shown the dramatic effect of a change of single gene, that regulates the τ parameter of the neurons, on the spiking activity of the whole SNN of 1000 neurons – see Fig.13 [65].

The spiking activity of a neuron may affect as a feedback the expressions of genes [5]. As pointed in [118] on a longer time scales of minutes and hours the function of neurons may cause the changes of the expression of hundreds of genes transcribed into mRNAs and also in microRNAs, which makes the short-term, the long-term and the genetic memories of a neuron linked together in a global memory of the neuron and further - of the whole neural system.

A major problem with the CNGM from fig.12 is how to optimize the numerous parameters of the model. One solution could be using evolutionary computation, such as PSO [75, 83] and the recently proposed quantum inspired evolutionary computation techniques [22, 97, 96]. The latter can deal with a very large dimensional space as each quantum-bit chromosome represents the whole space, each point to certain probability. Such algorithms are faster and lead to a close solution to the global optimum in a very short time.

In one approach it may be reasonable to use same parameter values (same GRN) for all neurons in the SNN or for each of different types of neurons (cells) that will results in a significant reduction of the parameters to be optimized. This can be interpreted as 'average' parameter value for the neurons of the same type. This approach corresponds to the biological notion to use one value (average) of a gene/protein expression for millions of cells in bioinformatics.

Another approach to define the parameters of the probabilistic spiking neurons, especially when used in biological studies, is to use prior knowledge about the association of spiking parameters with relevant genes/proteins (neuro-transmitter, neuro-receptor, ion channel, neuro-modulator) as described in [64]. Combination of the two approaches above is also possible.

7. SNN Software and hardware implementations to support STPR

Software and hardware realisations of SNN are already available to support various applications of SNN for STPR. Among the most popular software/hardware systems are [24, 16, 29]:

- jAER (http://jaer.wiki.sourceforge.net) [23];

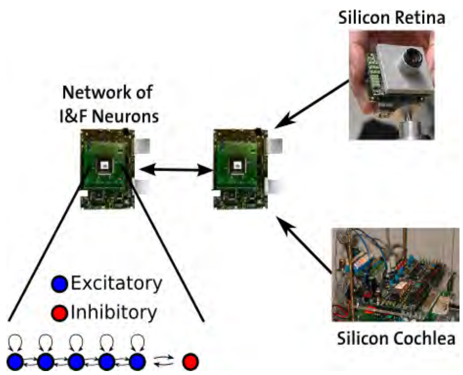


Fig.14. A hypothetical neuromorphic SNN application system (from http://ncs.ethz.ch)

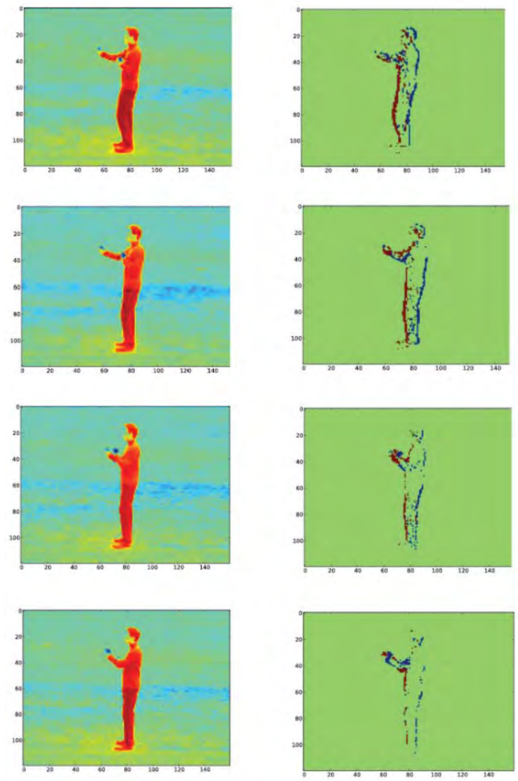
- Software simulators, such as Brian [16], Nestor, NeMo [79],etc;
- Silicon retina camera [23];
- Silicon cochlea [107];
- SNN hardware realisation of LIFM and SDSP [47-50];
- The SpiNNaker hardware/software environment [89, 116];
- FPGA implementations of SNN [56];
- The IBM LIF SNN chip, recently announced.

Fig.14 shows a hypothetical engineering system using some of the above tools (from [47, 25]).

8. Current and Future Applications of eSNN and CNGM for STPR

Numerous are the applications of eSNN for STPR. Here only few of them are listed:

- Moving object recognition (fig. 15) [23, 60];
- EEG data modelling and pattern recognition [70, 1, 51, 21, 26, 99, 35, 36] directed to practical applications, such as: BCI [51], classification of epilepsy [35, 36, 109] - (fig.16);
- Robot control through EEG signals [86] (fig.17) and robot navigation [80];
- Sign language gesture recognition (e.g. the Brazilian sign language – fig.18) [95];
- Risk of event evaluation, e.g. prognosis of establishment of invasive species [97] – fig.19; stroke occurrence [6], etc.
- Cognitive and emotional robotics [8, 64];
- Neuro-rehabilitation robots [110];
- Modelling finite automata [17, 78];
- Knowledge discovery from SSTD [101];
- Neuro-genetic robotics [74];
- Modelling the progression or the response to treatment of neurodegenerative diseases, such as Alzheimer's Disease [94, 64] – fig.20. The analysis of the obtained GRN model in this case could enable the discovery of unknown interactions between genes/proteins related to a brain disease progression and how these interactions can be modified to achieve a desirable effect.



(a) (b)

Fig.15.Moving object recognition with the use of AER [23]. (a) Disparity map of a video sample; (b) Address event representation (AER) of the above video sample.

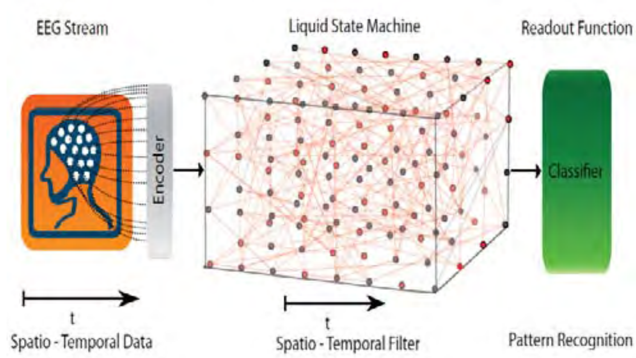


Fig.16. EEG based BCI.

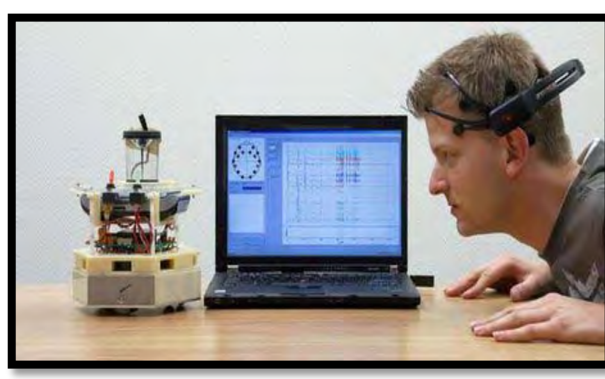


Fig.17. Robot control and navigation

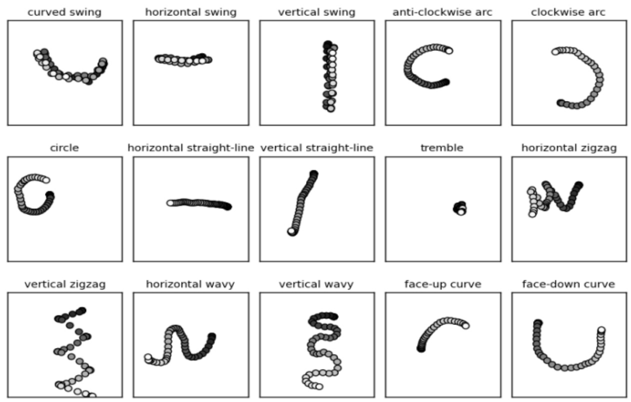


Fig.18. A single sample for each of the 15 classes of the LIngua BRAsileira de Sinais (LIBRAS) - the official Brazilian sign language is shown. The colour indicates the time frame of a given data point (black/white corresponds to earlier/later time points) [95].

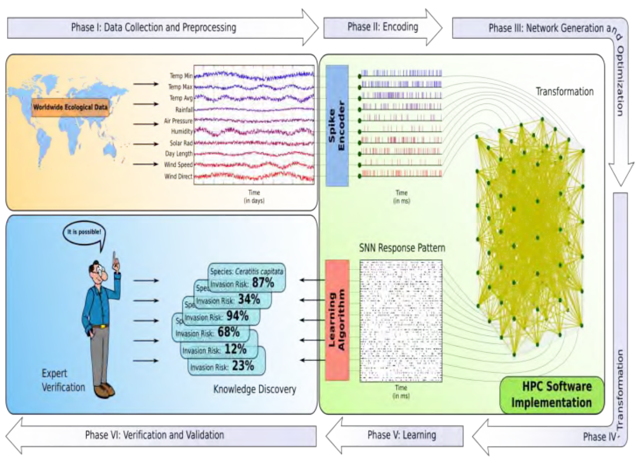


Fig 19. Prognosis of the establishment of invasive species [97]

- Modelling financial and economic problems (neuro-economics) where at a 'lower' level the GRN represents the dynamic interaction between time series variables (e.g. stock index values, exchange rates, unemployment, GDP, prize of oil), while the 'higher' level epSNN states represents the state of the economy or the system under study. The states can be further classified into pre-define classes (e.g. buy, hold, sell, invest, likely bankruptcy) [113];
- Personalized modelling, which is concerned with the creation of a single model for an individual input data [58, 59, 62]. Here as an individual data a whole SSTD pattern is taken rather than a single vector.

Acknowledgement

I acknowledge the discussions with G. Indivery and also with: A.Mohemmed, T.Delbruck, S-C.Liu, N.Nuntalid, K.Dhoble, S.Schliebs, R.Hu, R.Schliebs, H.Kojima, F.Stefanini. The work on this paper is sponsored by the Knowledge Engineering and Discovery Research Institute, KEDRI (www.kedri.info) and the EU FP7 Marie Curie International Incoming Fellowship project PIIF-GA-2010-272006 *EvoSpike,* hosted by the Institute for Neuroinformatics – the Neuromorphic Cognitive Systems Group, at the University of Zurich and ETH Zurich (http://ncs.ethz.ch/projects/evospike). Diana Kassabova helped with the proofreading.

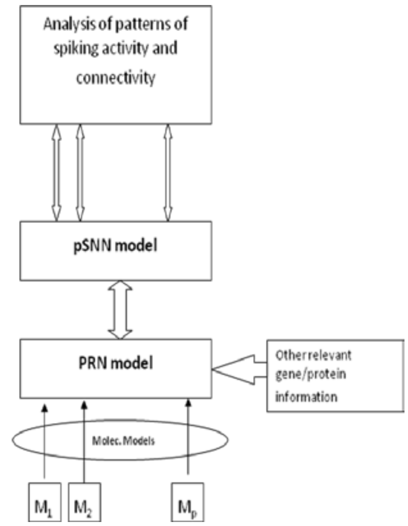


Fig.20.Hierarchical CNGM [64]

References

1. Acharya, R., Chua, E.C.P., Chua, K.C., Min, L.C., and Tamura, T. (2010), "Analysis and Automatic Identification of Sleep Stages using Higher Order Spectra," *Int.Journal of Neural Systems*, 20:6, pp. 509-521.
2. Arel, I., D. C. Rose, T. P. Karnowski, Deep Machine Learning: A New Frontier in Artificial Intelligence Research, Computational Intelligence Magazine, IEEE, vol.5, no.4, pp.13-18, 2010.
3. Arel, I., D.Rose, and B.Coop (2008) DeSTIN: A deep learning architecture with application to high-dimensional robust pattern recognition, in: Proc. 2008 AAAI Workshop Biologically Inspired Cognitive Architectures (BICA)
4. Arel, I., D.Rose, T.Karnovski (2010) Deep Machine Learning – A New Frontier in Artificial Intelligence Research, IEEE CI Magazine, Nov.2010, 13-18
5. Barbado, M., Fablet, K., Ronjat, M. And De Waard, M. (2009) Gene regulation by voltage-dependent calcium channels, Biochimica et Biophysica Acta, 1793, 1096-1104.
6. Barker-Collo, S., Feigin, V. L., Parag, V., Lawes, C. M. M., & Senior, H. (2010). Auckland Stroke Outcomes Study. Neurology, 75(18), 1608-1616.
7. Belatreche, A., Maguire, L. P., and McGinnity, M. Advances in Design and Application of Spiking Neural Networks. Soft Comput. 11, 3, 239-248, 2006
8. Bellas, F., R. J. Duro, A. Faiña, D. Souto, MDB: Artificial Evolution in a Cognitive Architecture for Real Robots, IEEE Transactions on Autonomous Mental Development, vol. 2, pp. 340-354, 2010
9. Bengio, Y. (2009) Learning Deep Architectures for AI, Found. Trends. Mach. Learning, vol.2, No.1, 1-127.
10. Benuskova, L, and N.Kasabov, Computational neuro-genetic modelling, Springer, New York, 2007, 290 pages

11. Berry, M.J., D. K. Warland, and M. Meister. The structure and precision of r etinal s piketrains. P NAS, 94(10): 5411–5416, May 1997.
12. Bohte, S., J. Kok, J. LaPoutre, Applications of spiking neural networks, Information Processing Letters, vol. 95, no. 6, 519-520, 2005.
13. Bohte, S.M., The evidence for neural information processing with pre cise s pike-times: A sur vey. NAT URAL COMPUTING, 3:2004, 2004.
14. Brader, J., W. Senn, and S. Fusi, Learning real-world stimuli in a n eural n etwork w ith s pike-driven s ynaptic d ynamics, Neural computation, vol. 19, no. 11, pp. 2881–2912, 2007.
15. Brader, J.M., Walter Senn, Stefano Fusi, Learning Real-World Stimuli i n a N eural N etwork with S pike-Driven S ynaptic Dynamics, Neural Computation 2007 19(11): 2881-2912, 2007.
16. Brette R., Rudolph M., Carnevale T., Hines M., Beeman D., Bower J. M., Diesmann M., Morrison A., Goodman P. H., Harris F. C., Zirpe M., Natschläger T., Pecevski D., Ermentrout B., Djurfeldt M., Lansner A., Rochel O., Vieville T., Muller E., Davison A. P., Boustani S. E., Destexhe A. (2007). Simulation of networks of spiking neurons: a review of tools and strategies. J. Comput. Neurosci. 23, 349–398.
17. Buonomano, D., W. Maass, State-dependent c omputations: Spatio-temporal proc essing in c ortical n etworks, N ature Reviews, Neuroscience, vol. 10, Feb. 2009, 113-125.
18. Chris I. De Zeeuw, Freek E. Hoebeek, Laurens W. J. Bosman, Martijn Schonewille, Spatiotemporal fi ring p atterns i n the cerebellum, Nature Reviews Neuroscience 12, 327-344 (June 2011) | doi:10.1038/nrn3011
19. Chris R. Shortall, Alison Moore, Emma Smith, Mike J. Hall, Ian P. Woiwod, and Richard Harrington. Long-term changes in the a bundance of fl ying i nsects. Ins ect Cons ervation a nd Diversity, 2(4):251–260, 2009
20. Cowburn, P. J., Cleland, J. G. F., Coats, A. J. S., & Komajda, M. (1996). Ri sk s tratification in c hronic he art fa ilure. E ur Heart J, 19, 696-710.
21. Craig, D.A., H.T. Nguyen, Adaptive EEG Thought Pattern Classifier for Advanced Wheelchair Control, Engineering in Medicine and Biology Society- EMBS'07, pp.2544-2547, 2007.
22. Defoin-Platel, M., S.Schliebs, N.Kasabov, Quantum-inspired Evolutionary Algorithm: A multi-model EDA, IEEE Trans. Evolutionary Computation, vol.13, No.6, Dec.2009, 1218-1232
23. Delbruck, T., jAER open source project, 2007, http://jaer.wiki.sourceforge.net.
24. Douglas, R. and Mahowald, M. (1995) Silicon Neurons, in: The Handbook of Brain Theory and Neural Networks, pp. 282-289, M. Arbib (Ed.), MIT Press.
25. Dhoble, K., N.Nuntalid, G.Indivery and N.Kasabov, Online Spatio-Temporal Pattern Recognition with Evolving Spiking Neural Networks utilising Address Event Representation, Rank Order, and Temporal Spike Learning, Proc. IJCNN 2012, Brisbane, June 2012, IEEE Press
26. Ferreira A., C. Almeida, P. Georgieva, A. Tomé, F. Silva (2010): Advances in EEG-based Biometry, International Conference on Image Analysis and Recognition (ICIAR), June 21-27, 2010, Povoa de Varzim, Portugal, to appear in Springer LNCS series.
27. Fisher, R.A., The use of multiple measurements in taxonomic problems, Annals of Eugenics, 7 (1936) 179-188)
28. Florian, R.V., The chronotron: a neuron that learns to fire temporally-precise spike patterns.
29. Furber, S., Temple, S., Neural systems engineering, Interface, J.of the Royal Society, vol. 4, 193-206, 2007
30. Fusi, S., M. Annunziato, D. Badoni, A. Salamon, and D. Amit, Spike-driven synaptic plasticity: theory, simulation, VLSI implementation, Neural Computation, vol. 12, no. 10, pp. 2227–2258, 2000.
31. Gene and Disease (2005), NCBI, http:/www.ncbi.nlm.nih.gov
32. Gerstner, W. (1995) Time structure of the activity of neural network models, Phys. Rev 51: 738-758.
33. Gerstner, W. (2001) What's different with spiking neurons? Plausible Neural Networks for Biological Modelling, in: H.Mastebroek and H. Vos (Eds.), Kluwer Academic Publishers, pp. 23- 48.
34. Gerstner, W., A. K. Kreiter, H. Markram, and A. V. M. Herz. Neural codes: firing rates and beyond. Proc. Natl. Acad. Sci. USA, 94(24):12740–12741, 1997.
35. Ghosh-Dastidar S. and Adeli, H. (2009), "A New Supervised Learning Algorithm for Multiple Spiking Neural Networks with Application in Epilepsy and Seizure Detection," *Neural Networks*, 22:10, 2009, pp. 1419-1431.
36. Ghosh-Dastidar, S. and Adeli, H. (2007), Improved Spiking Neural Networks for EEG Classification and Epilepsy and Seizure Detection, *Integrated Computer-Aided Engineering*, Vol. 14, No. 3, pp. 187-212
37. Goodman, E. and D. Ventura. Spatiotemporal pattern recognition via liquid state machines. In Neural Networks, 2006. IJCNN '06. International Joint Conference on, pages 3848–3853, Vancouver, BC, 2006)
38. Gutig, R., and H. Sompolinsky. The tempotron: a neuron that learns spike timing-based decisions. Nat Neurosci, 9(3):420–428, Mar. 2006.
39. Hebb, D. (1949). The Organization of Behavior. New York, John Wiley and Sons.
40. Henley, J.M., E.A.Barker and O.O.Glebov, Routes, destinations and delays: recent advances in AMPA receptor trafficking, Trends in Neuroscience, May 2011, vol.34, No.5, 258-268
41. Hodgkin, A. L. and A. F. Huxley (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. Journal of Physiology, 117: 500-544.
42. Hopfield, J., Pattern recognition computation using action potential timing for stimulus representation. Nature, 376:33–36, 1995.
43. Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. PNAS USA, vol.79, 2554-2558.
44. Hugo, G.E. and S. Ines. Time and category information in pattern-based codes. Frontiers in Computational Neuroscience, 4(0), 2010.
45. Huguenard, J. R. (2000) Reliability of axonal propagation: The spike doesn't stop here, PNAS 97(17): 9349-50.
46. Iglesias, J. and Villa, A.E.P.(2008), Emergence of Preferred Firing Sequences in Large Spiking Neural Networks During Simulated Neuronal Development, *Int. Journal of Neural Systems*, 18(4), pp. 267-277.
47. Indiveri, G., B. Linares-Barranco, T. Hamilton, A. Van Schaik, R. Etienne-Cummings, T. Delbruck, S. Liu, P. Dudek, P. H¨afliger, S. Renaud et al., "Neuromorphic silicon neuron circuits," Frontiers in Neuroscience, 5, 2011.
48. Indiveri, G ; Chicca, E ; Douglas, R J (2009). Artificial cognitive systems: From VLSI networks of spiking neurons to neuromorphic cognition. Cognitive Computation, 1(2):119-127.
49. Indiveri, G ; Stefanini, F ; Chicca, E (2010). Spike-based learning with a generalized integrate and fire silicon neuron. In: 2010 IEEE Int. Symp. Circuits and Syst. (ISCAS 2010), Paris, 30 May 2010 - 02 June 2010, 1951-1954.
50. Indiviery G., and T.Horiuchi (2011) Frontiers in Neuromorphic Engineering, Frontiers in Neuroscience, 5:118.

51. Isa, T., E. E. Fetz, K. Muller, Recent advances in brain-machine interfaces, Neural Networks, vol 22, issue 9, Brain-Machine Interface, pp 1201-1202, November 2009.
52. Izhikevich, E (2003) Simple model of spiking neurons, IEEE Trans. on Neural Networks, 14, 6, 1569-1572.
53. Izhikevich, E. M. (2004) Which model to use for cortical spiking neurons? IEEE TNN, 15(5): 1063-1070.
54. Izhikevich, E.M., G.M.Edelman (2008) Large-Scale Model of Mammalian Thalamocortical Systems, PNAS, 105: 3593-3598.
55. Izhikevich, E .(2006) Polychronization: Computation with Spikes, Neural Computation, 18, 245-282.
56. Johnston, S .P., Prasad, G .,Maguire, L., McGinnity, T .M. (2010), FPGA Hardware/software co-design methodology - towards evolvable spiking networks for robotics application, *Int. J. Neural Systems*, 20:6, 447-461.
57. Kasabov, N. Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering. Cambridge, Massachussets, MIT Press (1996) 550p
58. Kasabov, N., and Y. Hu (2010) Integrated optimisation method for personalised modelling and case study applications, Int. Journal of Functional Informatics and Personalised Medicine, vol.3,No.3,236-256.
59. Kasabov, N., Data Analysis and Predictive Systems and Related Methodologies – Personalised Trait Modelling System, PCT/NZ2009/000222, NZ Patent.
60. Kasabov, N., Dhoble, K., Nuntalid, N., & Mohemmed, A. (2011). Evolving probabilistic spiking neural networks for spatio-temporal pattern recognition: A preliminary study on moving object recognition. In 18th International Conference on Neural Information Processing, ICONIP 2011, Shanghai, China, Springer LNCS .
61. Kasabov, N., Evolving connectionist systems: The knowledge engineering approach, Springer, 2007(2003).
62. Kasabov, N., Global, local and personalised modelling and profile discovery in Bioinformatics: An integrated approach, Pattern Recognition Letters, Vol. 28, Issue 6 , April 2007, 673-685
63. Kasabov, N., L .Benuskova, S .Wysoski, A Computational Neurogenetic Model of a Spiking Neuron, IJCNN 2005 Conf. Proc., IEEE Press, 2005, Vol. 1, 446-451
64. Kasabov, N., R.Schliebs, H .Kojima (2011) Probabilistic Computational Neurogenetic Framework: From Modelling Cognitive Systems to Alzheimer's Disease. IEEE Trans. Autonomous Mental Development, vol.3, No.4,2011, 1-12.
65. Kasabov, N., S.Schliebs and A.Mohemmed (2011) Modelling the Effect of Genes on the Dynamics of Probabilistic Spiking Neural Networks for Computational Neurogenetic Modelling, Proc. 6th meeting on Computational Intelligence for Bioinformatics and Biostatistics, CIBB 2011, 30 June – 2 July, Gargangio, Italy, Springer LNBI
66. Kasabov, N., To spike or not to spike: A probabilistic spiking neuron model, Neural Netw., 23(1), 16–19, 2010.
67. Kilpatrick, Z .P., Bresloff, P .C. (2010) Effect of synaptic depression and adaptation on spatio-temporal dynamics of an excitatory neural networks, Physica D, 239, 547-560.
68. Kistler, G., and W. Gerstner, Spiking Neuron Models - Single Neurons, Populations, Plasticity, Cambridge Univ. Press, 2002.
69. Legenstein, R., C. Naeger, W. Maass, What Can a Neuron Learn with Spike-Timing-Dependent Plasticity? Neural Computation, 17:11, 2337-2382, 2005.
70. Lotte, F., M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain–computer interfaces, J. Neural Eng 4(2):R1-R15, 2007.
71. Maass, W. and H. Markram, Synapses as dynamic memory buffers, Neural Network, 15(2):155–161, 2002.
72. Maass, W., and A. M. Zador, Computing and learning with dynamic synapses, In: Pulsed Neural Networks, pages 321–336. MIT Press, 1999.
73. Maass, W., T. Natschlaeger, H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Computation, 14(11), 2531–2560, 2002.
74. Meng, Y., Y. Jin, J. Yin, and M. Conforth (2010) Human activity detection using spiking neural networks regulated by a gene regulatory network. Proc. Int. Joint Conf. on Neural Networks (IJCNN), IEEE Press, pp. 2232-2237, Barcelona, July 2010.
75. Mohemmed, A., Matsuda, S., Schliebs, S., Dhoble, K., & Kasabov, N. (2011). Optimization of Spiking Neural Networks with Dynamic Synapses for Spike Sequence Generation using PSO. In Proc.Int. Joint Conf. Neural Networks, pp. 2969-2974, California, USA, IEEE Press.
76. Mohemmed, A., Schliebs, S., Matsuda, S., Kasabov, N., Evolving Spike Pattern Association Neurons and Neural Networks, Neurocomputing, Elsevier, in print.
77. Mohemmed, A., Schliebs, S., Matsuda, S., Kasabov, N., SPAN: Spike Pattern Association Neuron for Learning Spatio-Temporal Sequences, International Journal of Neural Systems, in print, 2012.
78. Natschläger, T., W. Maass, Spiking neurons and the induction of finite state machines, Theoretical Computer Science - Natural Computing, Vol. 287, Issue 1, pp.251-265, 2002.
79. NeMo spiking neural network simulator, http://www.doc.ic.ac.uk/~akf/nemo/index.html
80. Nichols, E., McDaid, L .J., and Siddique, N .H. (2010), Case Study on Self-organizing Spiking Neural Networks for Robot Navigation, *International Journal of Neural Systems*, 20:6, pp. 501-508.
81. Norton, D. and Dan Ventura, Improving liquid state machines through iterative refinement of the reservoir, Neurocomputing, 73 (2010) 2893-2904
82. Nuzlu, H., Kasabov, N., Shamsuddin, S., Widiputra, H., & Dhoble. (2011) . An Extended Evolving Spiking Neural Network Model for Spatio-Temporal Pattern Classification. In Proceedings of International Joint Conference on Neural Networks (pp. 2653-2656). California, USA, IEEE Press.
83. Nuzly, H. , N. Kasabov, S .Shamsuddin (2010) Probabilistic Evolving Spiking Neural Network Optimization Using Dynamic Quantum Inspired Particle Swarm Optimization, Proc. ICONIP 2010, Part I, LNCS, vol.6443.
84. Ozawa, S., S. Pang and N Kasabov, Incremental Learning of Chunk Data for On-line Pattern Classification Systems, IEEE Transactions of Neural Networks, vol .19, no .6, June 2008 , 1061-1074,
85. Pang, S., S. Ozawa and N. Kasabov, Incremental Linear Discriminant Analysis for Classification of Data Streams, IEEE Trans. SMC-B, vol. 35, No. 5, 2005, 905 – 914
86. Pfurtscheller, G., R. Leeb, C. Keinrath, D. Friedman, C. Neuper, C. Guger, M. Slater, Walking from thought, Brain Research 1071(1): 145-152, February 2006.
87. Ponulak, F., and A. Kasi´nski. Supervised learning in spiking neural networks with ReSuMe: sequence learning, classification, and spike shifting. Neural Computation, 22(2):467–510, Feb. 2010. PMID:19842989.
88. Rabiner, L .R., A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE, vol. 77, no. 2, pp. 257 - 285, 1989.
89. Rast, A .D., Xin Jin, Francesco Galluppi, Luis A . Plana, Cameron Patterson, Steve Furber, Scalable Event-Driven Native Parallel Processing: The SpiNNaker Neuromimetic System, Proc. of the ACM International Conference on

Computing Frontiers, pp. 21-29, May 17-19, 2010, Bertinoro, Italy, ISBN 978-1-4503-0044-5

90. Reinagel, P. and R. C. Reid. Precise firing events are conserved across neurons. Journal of Neuroscience, 22(16):6837–6841, 2002.
91. Reinagel, R. and R. C. Reid. Temporal coding of visual information in the thalamus. Journal of Neuroscience, 20(14):5392–5400, 2000.
92. Riesenhuber, M. and T.Poggio (1999) Hierarchical Model of Object Recognition in Cortex, Nature Neuroscience, 2, 1019-1025.
93. Rokem, A., S. Watzl, T. Gollisch, M. Stemmler, A. V. Herz, and I. Samengo. Spike-timing precision underlies the coding efficiency of auditory receptor neurons. J Neurophysiol, 2005.
94. Schliebs, R. (2005). Basal forebrain cholinergic dysfunction in Alzheimer´s disease – interrelationship with β-amyloid, inflammation and neurotrophin signaling. Neurochemical Research, 30, 895-908.
95. Schliebs, S., Hamed, H. N. A., & Kasabov, N. (2011). A reservoir-based evolving spiking neural network for on-line spatio-temporal pattern learning and recognition. In: 18th International Conference on Neural Information Processing, ICONIP 2011, Shanghai, Springer LNCS.
96. Schliebs, S., Kasabov, N., and Defoin-Platel, M. (2010), "On the Probabilistic Optimization of Spiking Neural Networks," *International Journal of Neural Systems*, 20:6, pp. 481-500.
97. Schliebs,S., M .Defoin-Platel, S .Worner, N .Kasabov, Integrated Feature and Parameter Optimization for Evolving Spiking Neural Netw.: Exploring Heterogeneous Probabilistic Models, Neural Netw.,22,623-632, 2009.
98. Schliebs, S., Mohemmed, A., & Kasabov, N. (2011). Are Probabilistic Spiking Neural Networks Suitable for Reservoir Computing? In: Int. Joint Conf. Neural Networks IJCNN, pp. 3156-3163, San Jose, IEEE Press.
99. Schliebs,S., Nuntalid, N., & Kasabov, N. (2010). Towards spatio-temporal pattern recognition using evolving spiking neural networks. Proc. ICONIP 2010, Part I, Lect. Notes in Computer Science (LNCS), 6443, 163-170.
100. Schrauwen, B., and J. Van Campenhout, "BSA, a fast and accurate spike train encoding scheme, in Neural Networks, 2003. Proceedings of the International Joint Conference on, vol. 4. IEEE, 2003, pp. 2825–2830.
101. Soltic and S. Kasabov, N. (2010), "Knowledge extraction from evolving spiking neural networks with rank order population coding," *International Journal of Neural Systems*, 20:6, pp. 437-445.
102. Sona, D., H .Veeramachaneni, E .Olivetti, P .Avesani, Inferring cognition from fMRI brain images, Proc. of IJCNN, 2011, IEEE Press.
103. Song, S., K. Miller, L. Abbott et al., Competitive hebbian learning through spike-timing-dependent synaptic plasticity, Nature Neuroscience, vol. 3, pp. 919–926, 2000.
104. Theunissen, F. and J. P. Miller. Temporal encoding in nervous systems: a rigorous definition. Journal of Computational Neuroscience, 2(2):149–162, 1995.
105. Thorpe, S., and J. Gautrais, Rank order coding, Computational neuroscience: Trends in research, vol. 13, pp. 113–119, 1998.
106. Thorpe,S., Delorme, A., et al. (2001). Spike-based strategies for rapid processing. Neural Netw.,14(6-7),715-25.
107. van Schaik, A., L. Shih-Chii Liu, AER EAR: a matched silicon cochlea pair with address event representation interface, in: Proc. of ISCAS - IEEE Int. Symp. Circuits and Systems, pp. 4213- 4216, vol. 5, 23-26 May 2005.
108. Verstraeten, D., B. Schrauwen, M. D'Haene, and D. Stroobandt, An experimental unification of reservoir computing methods, Neural Networks, 20(3):391 – 403, 2007.
109. Villa, A .E.P., et al., (2005) Cross-channel coupling of neuronal activity in parvalbumin-deficient mice susceptible to epileptic seizures. Epilepsia, 46(Suppl. 6) p. 359.
110. Wang, X., Hou, Z.G., Zou, A., Tan, M., and Cheng, L., A behavior controller for mobile robot based on spiking neural networks, Neurocomputing (Elsevier), 2008, vol. 71, nos. 4-6, pp. 655-666.
111. Watts, M. (2009) A Decade of Kasabov's Evolving Connectionist Systems: A Review, IEEE Trans. Systems, Man and Cybernetics- Part C: Applications and Reviews, vol.39, no.3, 253-269.
112. Weston, I., F.Ratle and R.Collobert (2008) Deep learning via semi-supervised embedding, in: Proc. 25th Int. Conf. Machine Learning, 2008, 1168-1175
113. Widiputra, H., Pears, R., & Kasabov, N. (2011). Multiple time-series prediction through multiple time-series relationships profiling and clustered recurring trends. Springer LNAI 6635 (PART 2), 161-172.
114. Widrow, B. and M. Lehr. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. Proceedings of the IEEE, 78(9):1415 –1442, sep 1990.
115. Wysoski, S., L. Benuskova, N. Kasabov, Evolving spiking neural networks for audiovisual information processing, Neural Networks, vol 23, 7, pp 819-835, 2010.
116. Xin Jin, Mikel Lujan, Luis A. Plana, Sergio Davies, Steve Temple and Steve Furber, Modelling Spiking Neural Networks on SpiNNaker, Computing in Science & Engineering, Vol: 12 Iss:5, pp 91 - 97, Sept.-Oct. 2010, ISSN 1521-961
117. Yu, Y .C. et al (2009) Specific synapses develop preferentially among sister excitatory neurons in the neocortex, Nature 458, 501-504.
118. Zhdanov, V .P. (2011) Kinetic models of gene expression including non-coding RNAs. Phys. Reports, 500, 1-42.

Regular Paper

Biologically Motivated Selective Attention Model

Minho Lee

Kyungpook National University, Korea
**corresponding author*: mholee@knu.ac.kr

Abstract

Several selective attention models partly inspired by biological visual attention mechanism are introduced. The developed models consider not only binocular stereopsis to identify a final attention area so that the system focuses on the closer area as in human binocular vision, but also both the static and dynamic features of an input scene. In the models, I show the effectiveness of considering the symmetry feature determined by a neural network and an independent component analysis (ICA) filter, which are helpful to construct an object preferable attention model. Also, I explain an affective saliency map (SM) model including an affective computing process that skips an unwanted area and pays attention to a desired area, which reflects the human preference and refusal in subsequent visual search processes. In addition, I also consider a psychological distance as well as the pop-out property based on relative spatial distribution of the primitive visual features. And, a task specific top-down attention model to locate a target object based on its form and color representation along with a bottom-up attention based on relativity of primitive visual features and some memory modules. The object form and color representation and memory modules have an incremental learning mechanism together with a proper object feature representation scheme. The proposed model includes a Growing Fuzzy Topology Adaptive Resonance Theory (GFTART) network which plays two important roles in object color and form biased attention. Experiments show that the proposed models can generate plausible scan paths and selective attention for natural input scenes.

Keywords: Selective attention, bottom-up attention, GFTART

1. Introduction

The visual selective attention can allow humans to pay attention to an interesting area or an object in natural or cluttered scenes as well as to properly respond for various visual stimuli in complex environments. Such a selective attention visual mechanism allows the human vision system to effectively process high complexity visual scene.

Itti, Koch, and Niebur (1998) introduced a brain-like model in order to generate the bottom-up saliency map (SM). Koike and Saiki (2002) proposed that a stochastic winner take all (WTA) enables the saliency-based search model to change search efficiency by varying the relative saliency, due to stochastic attention shifts. Kadir and Brady (2001) proposed an attention model integrating saliency, scale selection and a content description, thus contrasting with many other approaches. Ramström and Christensen (2002) calculated saliency with respect to a given task by using a multi-scale pyramid and multiple cues. Their saliency computations were based on game theory concepts. Rajan et al. (2009) presented a robust selective attention model based on the spatial distribution of color components and local and global behavior of different orientations in images. Wrede et al. (2010) proposed a random center surround bottom up visual attention model by utilizing the stimulus bias techniques such as the similarity and biasing function. Ciu et al. (2009), Hou and Zhang (2007) developed a new type of bottom-up attention model by utilizing the Fourier phase spectrum. Based on the psychological understanding, Wang and Li (2008) presented a saliency detection model by combining localization of visual pop-outs using the spectrum residual model (Hou and Zhang, 2007) and coherence propagation strategy based on Gestalt principles. Frintrop, Rome, Nüchter, and Surmann (2005) proposed a bimodal laser-based attention system that considers both static features including color and depth for generating proper attention. Fernández-Caballero, López, and Saiz-Valverde (2008) developed a dynamic stereoscopic selective visual attention model that integrates motion and depth in order to choose the attention area. Maki, Nordlund, and Eklundh (2000) proposed an attention model integrating image flow, stereo disparity and motion for attentional scene segmentation. Ouerhani and Hügli (2000) proposed a visual attention model that considers depth as well as static features. Belardinelli and Pirri (2006) developed a biologically plausible robot attention model, which also considers depth for attention.

Carmi and Itti (2006) proposed an attention model that considers seven dynamic features in MTV-style video clips, and also proposed an integrated attention scheme to detect an object by combining bottom-up SM with top-down attention based on the signal-to-noise ratio (Navalpakkam & Itti, 2006). As well, Walther et al. (2005) proposed an object preferred attention scheme that considers the bottom-up SM results as biased weights for top-down object-perception. Li and Itti (2011) represented a visual attention model to solve the target detection problem in satellite images by combining biologically-inspired features such as saliency and gist features. Guo and Zhang (2010) extended their previous approach (Hou and Zhang, 2007) called spectral residual (SR) to calculate the spatiotemporal saliency map of an image by its quaternion representation.

This paper presents several types of selective attention models that generate an attention area by considering psychological distance of visual stimuli. The previous stereo SM models proposed by Lee et al. (2008) were modified and enhanced through learning the characteristics of familiar and unfamiliar objects and generating a preference and refusal bias signals reflecting the psychological distances to change the scan path obtained by conventional stereo SM models (Jeong, Ban, and Lee, 2008). Also, a human scan path was measured to verify whether the proposed SM model successfully reflects human's psychological distance for familiar and unfamiliar objects according to spatial distance from human subjects to attention candidates (Ban et al., 2011).

To effectively process and understand complex visual scenes, a top-down selective attention model with efficient biasing mechanism to localize a candidate target object area is essential. In order to develop such a top-down object biased attention model, we combine a bottom-up saliency map (SM) model that utilizes biologically motivated primitive visual features with a top-down attention model that can efficiently memorize the object form and color characteristics, and generates a bias signal corresponding to the candidate local area containing desired object characteristics. In addition, the proposed system comprises of an incremental object representation and memorization model with a growing fuzzy topology adaptive resonant theory (GFTART) network for building the object color and form feature clusters, in addition to the primitive knowledge building and inference (Kim et al., 2011). The GFTART network is a hybrid model by combining an ART based network and a growing cell structure (GCS) (Grossberg, 1987, Fritzke, 1994, Kim et al., 2011). In the GFTART network, each node of the F2 layer in the conventional fuzzy ART network is replaced with a GCS unit, thereby increasing the stability, while maintaining plasticity, and preserving the topology structures. The inclusion of the GCS unit allows the model to dynamically handle the incremental input features considering topological information (Kim et al., 2011). I propose a new integrated visual selective attention model that can generate an attention area by considering the psychological distance as well as top-down bias signals in the course of GFTART networks for continuous input scenes. It also includes both the human's affective computing process and stereo vision capability in selective attention.

In Section 2, we present the biological background on visual information processing. Sections 3 & 4, describes the proposed integrated selective attention model in detail. The experimental results of the proposed integrated selective attention model are presented in Section 5. Discussion and conclusions follow in Section 6.

2. Biological Background on Visual Information Processing

Fig. 1 shows the visual pathway from the retina to the V1 in brain. When the rods and cones cells in retina are excited, signals are transmitted through successive neurons in the retina itself and finally into the optic nerve fibers and cerebral cortex (Guyton 1991, Goldstein 1995, Kuffler 1984, Majani 1984, Bear 2001). The various visual stimuli on the visual receptors (or retina) are transmitted to the visual cortex through ganglion cells and the LGN. As shown in Fig. 1, there are three different types of retinal ganglion cells, W, X, and Y cells and each of these serves a different function (Guyton 1991, Majani 1984). Those preprocessed signal transmitted to the LGN through the ganglion cell, and the on-set and off-surround mechanism of the LGN and the visual cortex intensifies the phenomena of opponency (Guyton 1991, Majani 1984). The LGN has a six-layered structure, which serves as a relay station for conveying visual information from the retina to the visual cortex by way of the geniculocalcarine tract (Guyton 1991). This relay function is very accurate, so much so that there is an exact point-to-point transmission with a high degree of spatial

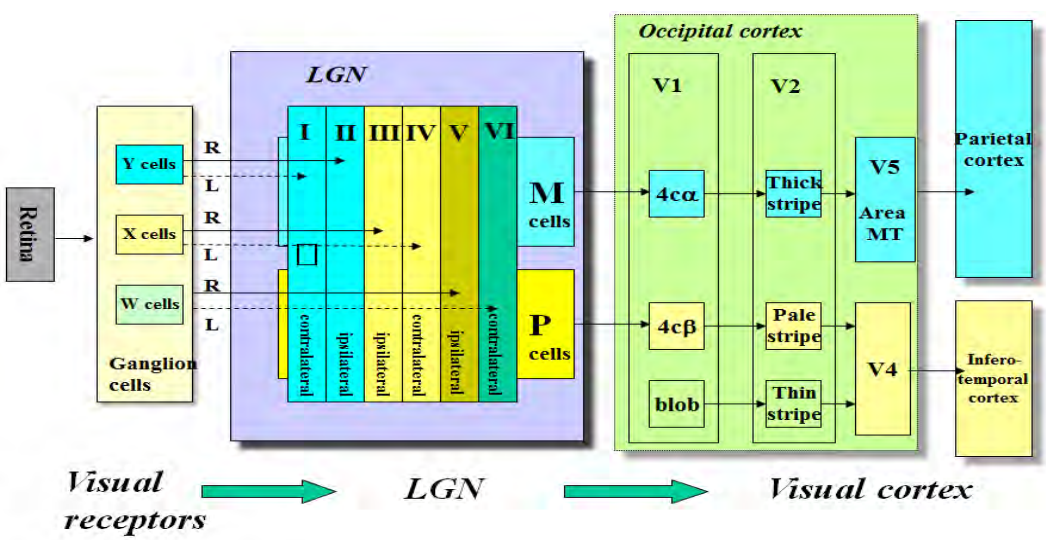


Figure 1. Overall visual pathways from the retina to the secondary visual cortex.

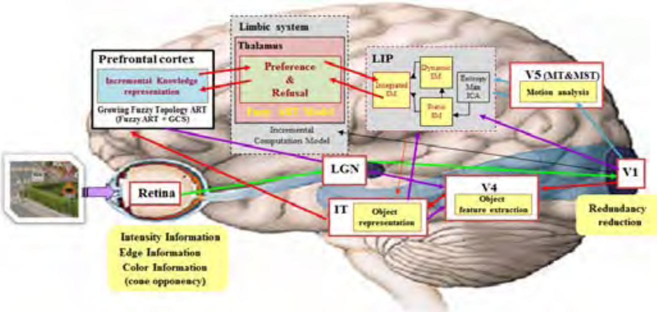


Figure 2. The visual pathway for visual environment perception.

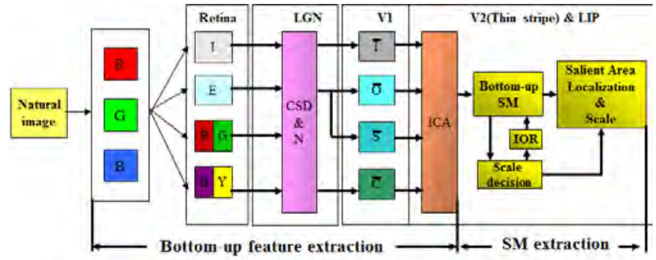


Figure 3. Static bottom-up saliency map model (I: intensity image, E:edge image, RG: red-green opponent coding image, BY: blue-yellow opponent coding image, CSD & N: center-surround difference and normalization, $\overline{I}$: intensity FM, $\overline{O}$: orientation FM, $\overline{S}$: symmetry FM, $\overline{C}$: color FM, ICA: independent component analysis, SM: saliency map).

fidelity all the way from the retina to the visual cortex. Finally, some detail features sensed by X-cells in the retina are slowly transferred to a higher level brain area from the LGN-parvo cells to the V4 and the inferior-temporal area (IT) through the V1-4cβ, which is related to the ventral pathway (Guyton, 1991). In contrast, some rapidly changing visual information sensed by Y-cells in the retina is rapidly transferred from LGN-magno cells to the middle temporal area (MT) and medial superior temporal area (MST) through the V1-4cα, which is related to the visual dorsal pathway (Bear et al., 2001). In order to develop a plausible visual selective attention model, we consider both object related information such as color and shape in the ventral pathway and motion information used in the dorsal pathway.

Fig. 2 shows the relation between visual pathway and proposed computational model that is related with visual environment perception. According to understanding the roles of brain organs that are related with visual environment perception, the visual pathway from the retina to the V1 mainly works for bottom-up visual processing, and the V4 and the IT area focus on top-down visual processing such as object perception. The hippocampus and the hypothalamus in the limbic system provide novelty detection and reflect top-down preference, respectively. The lateral-intra parietal cortex (LIP) works as an attention controller or center. Actually, the prefrontal cortex (PFC) plays a very important function in high-level perception such as knowledge representation, reasoning and planning. Each part will be described in detail in following sections.

3. Bottom-up visual attention model

3.1 Static bottom-up saliency map model

The human visual system can focus on more informative areas in an input scene via visual stimuli. From a bottom-up processing point of view, more informative areas in an input scene can be considered as 'pop-out' areas. The "pop-out" areas are places where relative saliency, compared with its surrounding area, is more based upon primitive input features such as brightness, odd color, and etc.

Fig. 3 shows the bottom-up processing for selective attention reflecting simple biological visual pathway of human brain to decide salient areas. Based on the Treisman's feature integration theory (Treisman & Gelde, 1980), Itti and Koch used three basis feature maps: intensity, orientation and color information (Itti et al., 1998). Extending Itti and Koch's SM model, I previously proposed SM models which include a symmetry feature map based on the generalized symmetry transformation (GST) algorithm and an independent component analysis (ICA) filter to integrate the feature information (Park et al., 2002). Through intensive computer experiments, I investigate the importance of the proposed symmetry feature map and the ICA filter in constructing an object preferable attention model. I also incorporate the neural network approach of Fukushima (Fukushima, 2005) to construct the symmetry feature map, which is more biologically plausible and takes less computation than the conventional GST algorithm (Park et al., 2002). Symmetrical information is also an important feature to determine the salient object, which is related with the function of LGN and primary visual cortex (Li, 2001). Symmetry information is very important in the context free search problems (Reisfeld et al., 1995). In order to implement an object preferable attention model, we emphasize using a symmetry feature map because an object with arbitrary shape contains symmetry information, and our visual pathway also includes a specific function to detect a shape in an object (Fukushima, 2005). In order to consider symmetry information in our SM model, I modified Fukushima's neural network to describe a symmetry axis (Fukushima, 2005). Fig. 3 shows the static bottom-up SM model. In the course of computing the orientation feature map, we use 6 different scale images (a Gaussian pyramid) and implement the on-center and off-surround functions using the center surround and difference with normalization (CSD & N) (Itti et al., 1998; Park et al., 2002).

As shown in Fig. 4, the orientation information in three successive scale images is used for obtaining the symmetry axis from Fukushima's neural network (Fukushima, 2005). By applying the center surround difference and normalization (CSD&N) to the symmetry axes extracted in four different scales, we can obtain a symmetry feature map. This procedure mimics the higher-order analysis mechanism of complex cells and hyper-complex cells in the posterior visual cortex area, beyond the orientation-selective simple cells in the V1. Using CSD & N in Gaussian pyramid images (Itti et al. 1998), we can construct intensity ($\overline{I}$), color ($\overline{C}$), and orientation ($\overline{O}$) feature maps as well as the symmetry feature map ($\overline{S}$) (Fukushima. 2005, Jeong et al., 2008).

Based on the Barlow's hypothesis that human visual cortical feature detectors might be the end result of a redundancy reduction process (Barlow & Tolhurst, 1992),

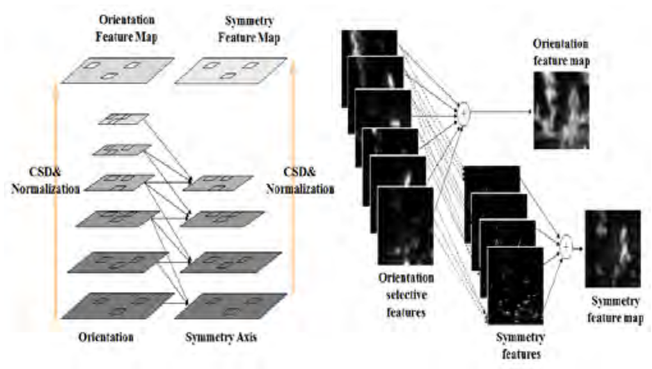


Figure 4. Symmetry feature map generation process.

and Sejnowski's results showing that independent component analysis (ICA) is the best alternative to reduce redundancy (Bell & Sejnowski, 1997), the four constructed feature maps ($\bar{I}$, $\bar{C}$, $\bar{O}$, and $\bar{S}$) are then integrated by an ICA algorithm based on maximization of entropy (Bell & Sejnowski, 1997).

Fig. 5 shows the procedure for computing the SM. In Fig. 5, *S(x,y)* is obtained by the summation of the convolution between the *r*-th channel of input image(I_r) and the *i*-th filters (IC_{sri}) obtained by the ICA learning (Bell & Sejnowski, 1997). A static SM is obtained by Eq. (1).

$$S(x,y) = \sum I_r * IC_{Sri} \qquad \text{for all } I \tag{1}$$

Since we obtained the independent filters by ICA learning, the convolution result shown in Eq. (1) can be regarded as a measure for the relative amount of visual information. The LIP plays a role in providing a retinotopic spatio-feature map that is used to control the spatial focus of attention and fixation, which is able to integrate feature information in its spatial map (Lanyon & Denham, 2004). As an integrator of spatial and feature information, the LIP provides the inhibition of return (IOR) mechanism required here to prevent the scan path from returning to previously inspected sites (Lanyon & Denham, 2004).

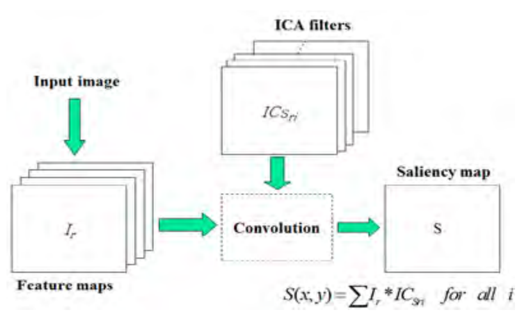


Figure 5. Saliency map generation process using ICA filter.

3.2 Scale saliency

The localized salient area, which is obtained from the bottom-up saliency map model, has suitable scale/size by considering an entropy maximization approach. The size of salient area was adapted to select a proper scale of the salient area by using the saliency map. The scale selection algorithm is based upon Kadir's approach (Kadir & Brady, 2001). Fig. 2 shows that proposed model selectively localizes the salient area in the input scene and in addition, it can decide the proper scale of the salient area. For each salient location, the proposed model chooses those scales at which the entropy is at its maximum, or has peaked, and then the entropy value is weighted by some measure of self-dissimilarity in the scale-space of the saliency map. The most appropriate scale for each salient area, centered at location x, is obtained by Eq. (2):

$$scale(\mathbf{x}) = \arg\max_{s} \{H_D(s,\mathbf{x}) \times W_D(s,\mathbf{x})\} \tag{2}$$

where D is the set of all descriptor values, $H_D(s,\mathbf{x})$ is entropy as defined by Eq. (3) and $W_D(s,\mathbf{x})$ is the inter-scale measure as defined by Eq. (4):

$$H_D(s,\mathbf{x}) = -\sum_{d \in D} p_{d,s,\mathbf{x}} \log_2 p_{d,s,\mathbf{x}} \tag{3}$$

$$W_D(s,\mathbf{x}) = \frac{s^2}{2s-1} \sum_{d \in D} \left| p_{d,s,\mathbf{x}} - p_{d,s-1,\mathbf{x}} \right| \tag{4}$$

where $p_{d,s,\mathbf{x}}$ is the probability mass function for scale *s*, position x, and the descriptor value *d* that takes on values in *D*. The probability mass function $p_{d,s,\mathbf{x}}$ is obtained from the histogram of the pixel values of the salient area centered at the location x with size *s* in the saliency map. As shown in Fig. 6, the proposed scale decision model can select suitable scale for the face (Kadir & Brady, 2001, Park et al., 2002).

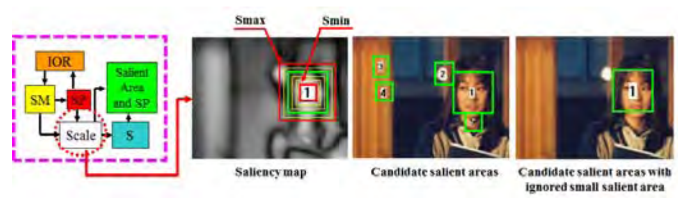


Figure 6. Scale decision in saliency map.

3.3 Dynamic bottom-up saliency map model

The human visual system sequentially interprets dynamic input scenes as well as still input images. A conventional bottom-up SM model, however, considers only static visual features in single frame. Most of selective attention models, including our previous model (Park, An, and Lee, 2002), consider only static scenes. Humans, however, can decide the constituents an interesting area within a dynamic scene, as well as static images. The dynamic SM model is based upon the analysis of successive static SMs. The entropy maximization is considered to analyze the dynamics of the successive static SMs, which is an extension of Kadir's approach (Kadir & Brady, 2001), since the dynamic SM model considers time-varying properties as well as spatial features. The selective attention model is the first such a model to handle dynamic input scenes. Fig. 7 shows the procedure adopted to acquire a final SM by integrating both of the static SM and the dynamic SM from natural input images. (Jeong et al., 2008, Fernández-Caballero et al., 2008).

The entropy value at each pixel represents a fluctuation of visual information according to time, through which a dynamic SM is generated. Finally, the attention model decides the salient areas based upon the dynamic bottom-up SM model as shown in Fig. 7, which is generated by the integration of the static SM and the dynamic SM. Therefore,

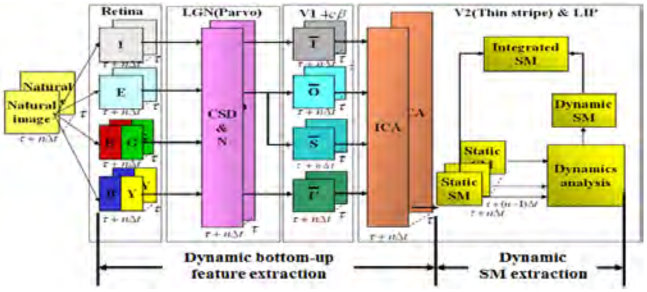


Figure 7. The proposed dynamic bottom-up saliency map model.

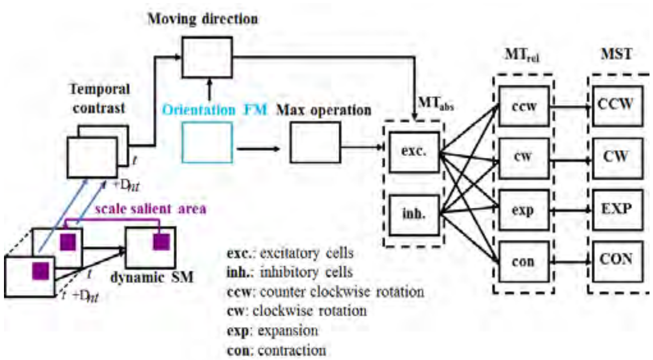


Figure 8. Motion analysis model based on dynamic saliency map

the proposed dynamic bottom-up attention model can selectively decide an attention area by considering not only static saliency, but also the feature information of dynamics, which are obtained from consecutive input scenes.

3.4 Motion analysis based on the dynamic saliency map model

Fig. 8 shows a proposed motion analysis model which integrates the dynamic SM with the motion analysis model as proposed by Fukushima (2008). The model is partly inspired by the roles of the visual pathway in the brain, from the retina to the MT and the MST through the LGN, by means of the V1 and the V2, including the lateral intra-parietal cortex (LIP).

As shown in Fig. 8, motion analysis networks are related to rotation, expansion, contraction and planar motion for the selected area obtained from the dynamic and static SM models. The model analyzes the motion within a salient area obtained by the SM model. In the Fukushima's neural network, MT cells extract the absolute and relative velocities (MT_{abs}-cells and MT_{rel}-cells), and MST cells extract optic flow in a large visual field. The proposed model can automatically select the size of a receptive field at each cell using the factors, taken from Fukushima's neural network (Fukushima, 2008). The relative velocity is then extracted by using orientation and local velocity information as proposed by Fukushima (2008). MT_{abs}-cells extract absolute-velocity stimuli. The MT_{abs}-cells consist of two sub-layers, namely excitation and inhibition cells. Only the receptive-field size of an inhibition cell is larger than that of an excitation cell. MT_{rel}-cells extract relative velocity of the stimuli by receiving antagonistic signals from excitation and inhibition cells of MT_{abs}-cells. MT_{abs}-cells integrate responses of many MT_{rel}-cells by summation and then extract the counter-clockwise rotation, clockwise rotation, expansion and contraction of optic flow (Jeong et al., 2008, Fukushima. 2008).

3.5 Stereo saliency map model

Based on the single eye alignment hypothesis (Thorn et al., 1994), Lee et al. developed an active vision system that can control two cameras by partly mimicking a vergence mechanism to focus two eyes at the same area in the human vision system. This stereo vision system used the static selective attention model to implement an active vision system for vergence control (Choi et al., 2006). I use depth information from disparities of most salient regions in left and right cameras to construct the stereo SM, which can then support pop-outs for closer objects. In the model, selective attention regions in each camera are obtained from static and dynamic saliency and are then used for selecting a dominant landmark. Comparing the maximum salient values within selective attention regions in the two camera images, we can adaptively decide the camera with larger salient value as master eye. After successfully localizing the corresponding landmarks on both left and right images with master and slave eyes, we are able to get depth information by simple triangulation. Fig. 9 shows a stereo saliency map model including the bottom-up SM process and depth perception (Jeong et al., 2008).

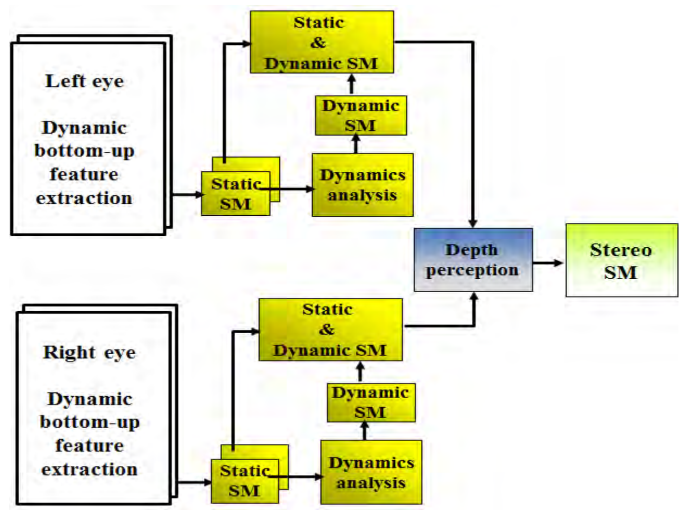


Figure 9. Stereo saliency map model including the bottom-up SM process and depth perception.

The stereo SM uses the depth information specifically, in which the distance between the camera and a focused region is used as a characteristic feature in deciding saliency weights. The stereo SM is obtained by Eq. (5):

$$S_c(v) = S_p(v)\cdot(1+\exp^{-z/\tau})\cdot L(sp, v, \sigma) \tag{5}$$

where v denotes a pixel in the salient area, and $s_c(v)$ and $s_p(v)$ are the current and previous SMs, respectively. z represents the distance between the camera and a focused region, and τ determines the rate at which distance effects decay. $L(\cdot)$ is a Laplacian function as shown in Eq. (6):

$$L(sp,\nu,\sigma)=\mathrm{C}\cdot\left(\frac{|\nu-sp|^2-2\sigma^2}{\sigma^4}\right)\exp^{-|\nu-sp|^2/2\sigma^2} \tag{6}$$

where sp represents the center location of the salient area, ν denotes a pixel in the salient area, σ is the width of the salient area and C is a constant. The Laplacian in Eq. (6) reflects brain-cell activity characteristics such as on-center (excitatory) and off-surround (inhibitory) signals within the attention region. The stereo SM is constructed using not only depth information, but also spatial information within a salient area.

4. Top-down visual attention

4.1 Affective saliency map model

To enhance the previously described bottom-up SM models, we need to consider affective factors that reflect human preference and refusal. As an affective computing process, the proposed model considers such a simple process that can reflect human's preference and refusal for visual features by inhibiting an uninterested area and reinforcing an interested area respectively, which are decided by human. Top-down modulation of visual inputs to the salience network may facilitate a visual search (Mazer & Gallant, 2003). We avoid focusing on a new area having similar characteristics to a previously learned uninteresting area by generating a top-down bias signal obtained through a training process. Conversely, humans can focus on an interesting area even if it does not have salient primitive features, or is less salient relative to another area. In Lee's trainable selective attention scheme (Choi et al., 2006), fuzzy adaptive resonance theory (ART) networks learn the characteristics of unwanted and interesting areas (Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992), but the training process was not considered to generate suitable top-down bias weight values. They only considered fixed weight values for biasing four different features according to inhibition or reinforcement control signals. I introduce affective SM model using a top-down bias signal that is obtained by means of the Hebbian learning process, which generates adaptive weight values intensified according to co-occurrence of the similar feature between input and memorized characteristics in every feature (Haykin, 1999).

Fig. 10 shows the selective attention model with affective factors. The top-down weight values are trained by the Hebbian learning method the activation values of the SM and those of each feature map (FM) as shown in preference Eq. (7) and refusal Eq. (8), where sp is the center location of the salient area and ν denotes a pixel in a salient area. The weight values, $w(x,y)$, in each feature map increase or decrease according to the training results.

$$\Delta w(\nu)_{k+}=\Sigma\left(\alpha(n,c,N)k+\exp(|\nu-sp|^2/\sigma)+FM(sp,N,\nu)\right) \tag{7}$$

$$\Delta w(\nu)_{k-}=\Sigma\left(\alpha(n,c,N)k-\exp(|\nu-sp|^2/\sigma)+FM(sp,N,\nu)\right) \tag{8}$$

The training weights for refusal and preference of human $\alpha_{(n,c,N)k-}$ and $\alpha_{(n,c,N)k+}$ are calculated by Eqs. (9) and (10).

$$\alpha_{(n+1,c,N)k-}=(1+\beta)\alpha_{(n,c,N)}-\eta\,\overline{FM}_{(n+1,sp,N)}\,\overline{SM}_{(n+1,sp)} \tag{9}$$

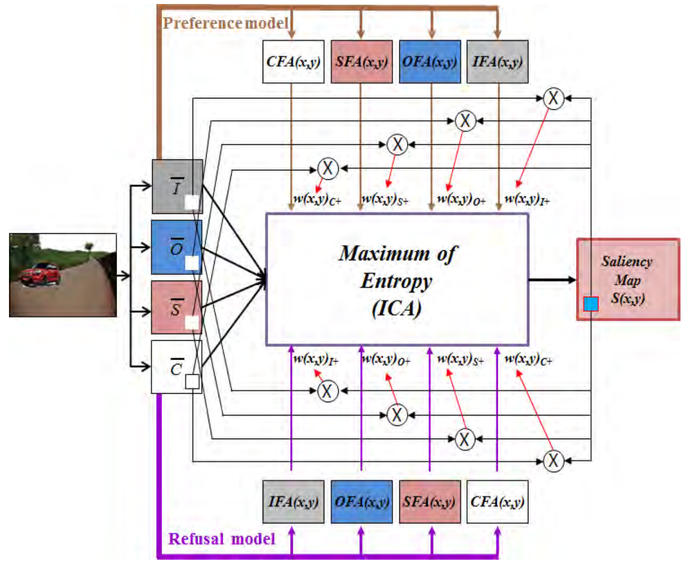


Figure 10. Selective attention model considering affective factors

$$\alpha_{(n+1,c,N)k+}=(1-\beta)\alpha_{(n,c,N)}+\eta\,\overline{FM}_{(n+1,sp,N)}\,\overline{SM}_{(n+1,sp)} \tag{10}$$

Eqs. (9) and (10) are obtained by the Hebbian learning method based on coincidence of two activities, which are the activity of the SM and FM. n represents training times, N (=1,…,4) is a FM index, and c represents a node in an F2 layer of the fuzzy ART, of which each node reflects a training pattern class. $\overline{FM}$ and $\overline{SM}$ represent local activities of the FM and the SM, respectively, and are obtained by Eqs. (11) and (12), where η is a training rate and β represents the influence of the previous α on the current value.

$$\overline{FM}_{(n,sp,N)}=\sum_{|\nu|<SA_{(n,sp)}}{}^{n}FM(sp,N,\nu) \tag{11}$$

$$\overline{SM}_{(n,sp)}=\sum_{|\nu|<SA_{(n,sp)}}{}^{n}SM(sp,\nu)\ ,\ SM(sp,\nu)=\sum_{N=1}^{4}FM(sp, \tag{12}$$

${}^{n}FM(sp,N,\nu)$ and ${}^{n}SM(sp,\nu)$ represent the FM and the SM at n^{th} training time, respectively.

4.2. Object oriented attention based on top-down bias

When humans pay attention to a target object, the prefrontal cortex gives a competitive bias signal, related with the target object, to the IT and the V4 area. Then, the IT and the V4 area generates target object dependent information, and this is transmitted to the low-level processing part in order to make a competition between the target object dependent information and features in whole area in order to filter the areas that satisfy the target object dependent features.

Fig. 11 shows the overview of the proposed model. The lower part in Fig. 11 generates a bottom-up SM based on primitive input features such as intensity, edge and color opponency. In training mode, each salient object decided by the bottom-up SM is learned by a GFTART. For each object area, the log-polar transformed features of RG and BY color opponency features represents color features of an object.

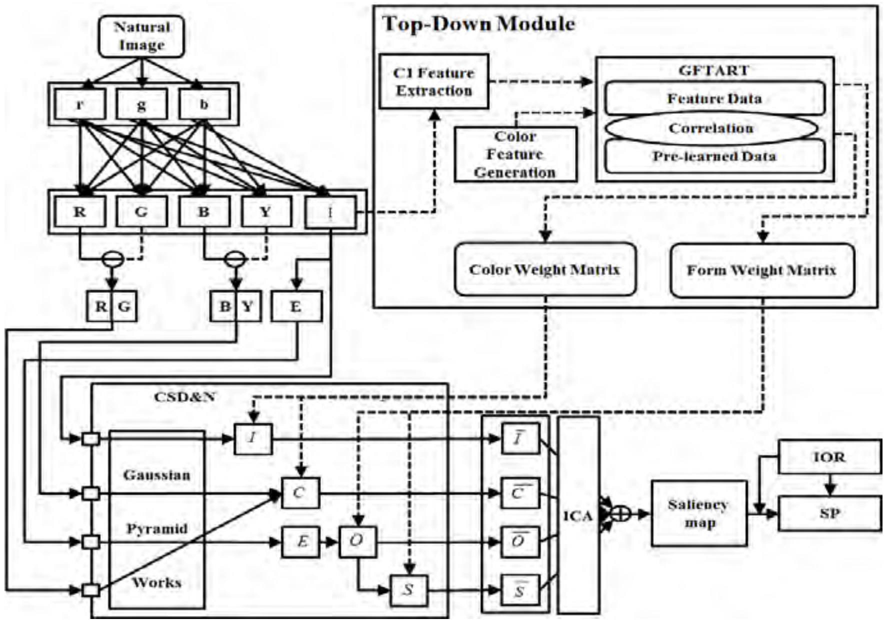


Figure 11. Top-down object biased attention using GFTART.

Orientation histogram and Harris corner based C1 features in the hierarchical MAX model proposed by Riesenhuber and Poggio are used as form features. Those extracted color and form features are used as the inputs of the GFTART. In top-down object biased attention, the GFTART activates one of memorized color and form features according to a task to find a specific object. The activated color and form features related with a target object are involved in competition with the color and form features extracted from each bottom-up salient object area in an input scene. By such a competition mechanism, as shown in Fig. 11, the proposed model can generate a top-down signal that can bias the target object area in the input scene.

Finally the top-down object biased attention model can generate a top-down object biased SM, in which the target object area is mostly popped out.

4.2.1 Top-down biasing using GFTART

Fig. 12 shows the architecture of GFTART network. The inputs of the GFTART consist of the color and form features. Those features are normalized and then represented as a one-dimensional array X that is composed of every pixel value a_i of the three feature maps and each complement a_i^c is calculated by $1- a_i$, the values of which are used as an input pattern in the F1 layer of the GFTART model. Next, the GFTART finds the winning growing cell structure (GCS) unit from all GCS units in the F2 layer, by calculating the Euclidean distance between the bottom-up weight vector W_i, connected with every GCS unit in the F2 layer, and X is inputted. After selecting the winner GCS unit, the growing fuzzy TART checks the similarity of input pattern X and all weight vectors W_i of the winner GCS unit. This similarity is compared with the vigilance parameter ρ, if the similarity is larger than the vigilance value, a new GCS unit is added to the F2 layer. In such situation, resonance has occurred, but if the similarity is less than the vigilance, the GCS algorithm is applied. The detailed GCS algorithm is described in (Marsland, Shapiro and Nehmzow, 2002).

Our approach hopefully enhances the dilemma regarding the stability of fuzzy ART and the plasticity of GCS (Marsland, et al. 2002, Carpenter, et al. 1992). The advantages of this integrated mechanism are that the stability in the convention fuzzy ART is enhanced by adding the topology preserving mechanism in incrementally-changing dynamics by the GCS, while plasticity is maintained by the fuzzy ART architecture (Kim et al., 2010). Also, adding GCS to fuzzy ART is good not only for preserving the topology of the representation of an input distribution, but it also self adaptively creates increments according to the characteristics of the input features.

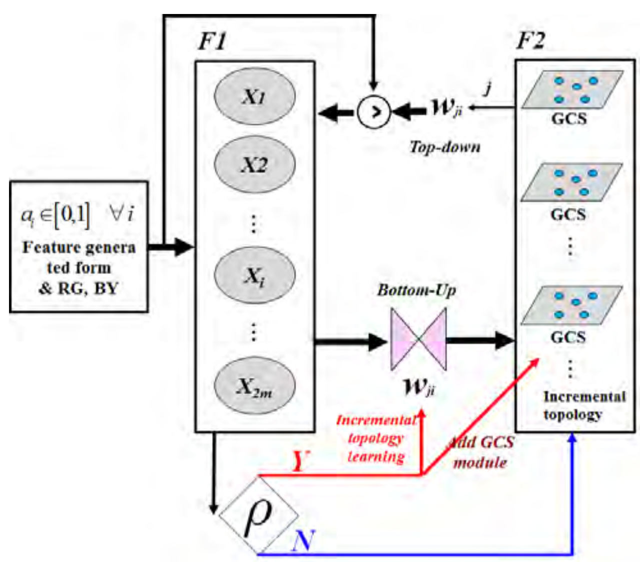


Figure 12. The architecture of GFTART network.

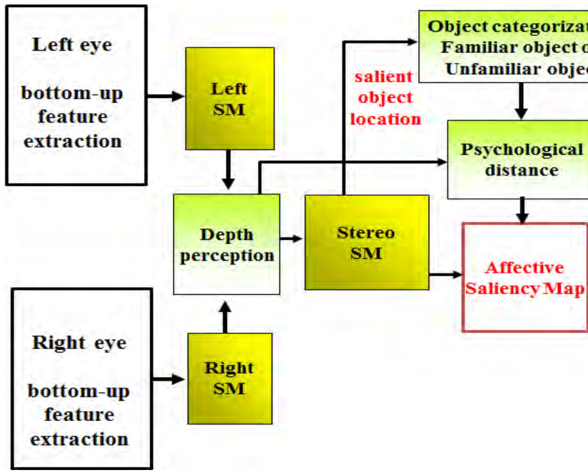


Figure 13. Visual selective attention model considering a psychological distance as well as primitive visual features.

4.3 Selective attention reflecting psychological distance

Fig. 13 illustrates the proposed visual attention model, which is partly inspired by biological visual pathway from the retina to the visual cortex through the LGN for bottom-up processing, which is extended to the IT and PFC for top-down processing. In order to implement a visual selective attention function, three processes are combined to generate an affective SM (Ban et al. 2011). One generates a stereo SM from binocular visions. Second considers object perception for categorizing and memorizing social proximal objects and social distal objects. Finally, an affective SM is constructed by considering the psychological distance that reflects the relationship between social distance and spatial distance for an attended object. Social proximity or distance of an attended object is perceived by an object categorization module. And the spatial distance to an attended object from an observer is obtained from a depth perception module using the stereo SM.

In order to develop more human like visual selective attention, we need to consider a stereo SM model for binocular visions. The stereo visual affective SM model is constructed by two mono SM models, which can give spatial distance information precisely. Then, the affective SM model considering the psychological distance can be plausibly developed by reflecting more accurate relation based on spatial distance information.

Affective stereo SM is generated by reflecting a psychological distance based on both social distance and depth information of a salient area, in which the final stereo SM, $StereoSM_c(v)$, is obtained by Eq. (13):

$$StereoSM_c(v) = S_c(v)\,(Psycho\_distance\ v) \qquad (13)$$

where the quantitative value of psychological distance is obtained by Eq. (14). The psychological distance, $Psycho\_distance(v)$ as shown in Eq. (14), is obtained by the ratio between response time for congruent condition and that for incongruent condition obtained from experiments. In congruent condition, the congruent object area becomes more highly salient, which induces faster selection area in a visual scan path. On the other hand, in the case of incongruent condition, the incongruent object area becomes less salient, which induces slower selection area in a visual scan path.

$$\begin{aligned}
&if\ congruent_condition\\
&\quad then\ Psycho_distance(v) = \frac{incongruency_mean_response_time}{congruency_mean_response_time}\\
&else\\
&\quad Psycho_distance(v) = \frac{congruency_mean_response_time}{incongruency_mean_response_time}
\end{aligned} \qquad (14)$$

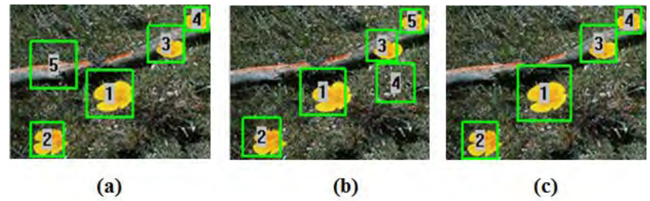


Figure 14. Comparison of salient areas selected by various static SMs.

5. Experimental Results

Fig. 14 shows an experiment example in which the proposed static bottom-up SM model generates a better attention path by using symmetry information as an additional input and ICA for feature integration. The numbers in Fig. 14 show the attention priority according to the degree of saliency using different SM models. Fig. 14 (a) shows the experimental result of the SM model considering intensity, color and orientation as features. Fig. 14 (b) shows the scan path result of the SM model considering intensity, color, orientation and symmetry as features. Fig. 14 (c) shows the experimental result of the SM model considering intensity, color, orientation and symmetry features together with ICA method for feature integration. The symmetry feature with ICA method successfully reduces redundant information in FMs so that the final scan path focus on the flower as shown in Fig. 14 (c).

Table 1 compares the performance for preferred object attention of three different SM models using hundreds of test images. As shown in table 1, we achieve suitable scan path by considering both the symmetry feature and ICA method, which means that object regions are mostly selected as attention areas by the SM model considering intensity, color, orientation, symmetry and ICA method.

Fig. 15 shows the comparison of SMs such as static SM, dynamic SM, static and dynamic SM, and integrated SM that combines the static and dynamic SM with depth information. We use five successive image frames for extracting dynamic feature. As shown in Fig. 15 the static and dynamic SM generates a maximum salient value for the attention area in each camera image. Comparing the maximum salient values in two camera images, we can adaptively decide the camera with a large salient value as the master eye. And we generated the integrated SM that combines the static and dynamic SM with depth information from the master eye (Choi et al., 2006).

TABLE 1. COMPARISON OF THREE DIFFERENT BOTTOM-UP SM MODELS FOR OBJECT PREFERRED ATTENTION.

salient area	$\bar{I}+\bar{C}+\bar{O}$	$\bar{I}+\bar{C}+\bar{O}+\bar{S}$	$\bar{I}+\bar{C}+\bar{O}+\bar{S}+ICA$
1st salient area	143	150	165
2nd salient area	103	104	106
3rd salient area	64	77	68
4th salient area	47	57	66
5th salient area	28	32	46
# of total	385	420	451
Detection rate	77 %	84 %	90 %

TABLE 2. COMPARISON OF THE DEGREES OF SALIENCY IN STATIC, DYNAMIC AND INTEGRATED SM MODELS.

Salient objects	Static SM	Dynamic SM	Static & Dynamic SM
Right human	183 (1st)	139 (2nd)	161 (1st)
Center kettle	143(2nd)	113 (3rd)	128 (3rd)
Left human	135 (3rd)	171 (1st)	154 (2nd)

TABLE 3. COMPARISON OF THE DEGREES OF SALIENCY ACCORDING TO DEPTH INFORMATION IN THE INTEGRATED SM MODEL.

Salient objects	Depth (m) in each salient area from static SM	Degree of saliency in selected area		
		Static SM	Static & Dynamic SM with depth information	
			τ_2 =0.5	τ_3 = 1.5
Right human	0.86	183 (1st)	190 (1st)	252 (1st)
Center kettle	**2.4**	**143(2nd)**	**129 (3rd)**	**154 (3rd)**
Left human	**1.3**	**135 (3rd)**	**164 (2nd)**	**218 (2nd)**

Table 2 shows the degrees of saliency of the static and the dynamic SMs, which is calculated by the average of saliency values in the salient areas. The degree of saliency changes while the integrated SM is generated as in Fig. 15, through which the plausibility of salient area choices can be verified even if the key biological mechanism for integrating static and dynamic features is not reflected since it is not known well.

Fig. 16 shows the selective motion analysis results generated by the neural network model for motion analysis in conjunction with the integrated SM model. As Fig. 16 (a) shows, the proposed model only analyzes attention areas that are selected by the integrated SM model, and Figs. 16(b), (c), (d) and (e) represent the relative degree of motion information for counter-clockwise, clockwise, expansion and contraction, respectively. The area 'c' in Fig. 16 (a) is moving to camera and rightward direction, the area 'a' in Fig. 16(a) is moving away from camera and leftward direction. Fig. 16(d) shows that the proposed model properly describes motion characteristic of area 'a' with contraction movement away from camera. Also, our model successfully responds to the rightward motion in area 'c' by producing an increased amount of motion information and a little response for motions of static object in area 'b' as shown in Figs. 16(b), (c), (d) and (e).

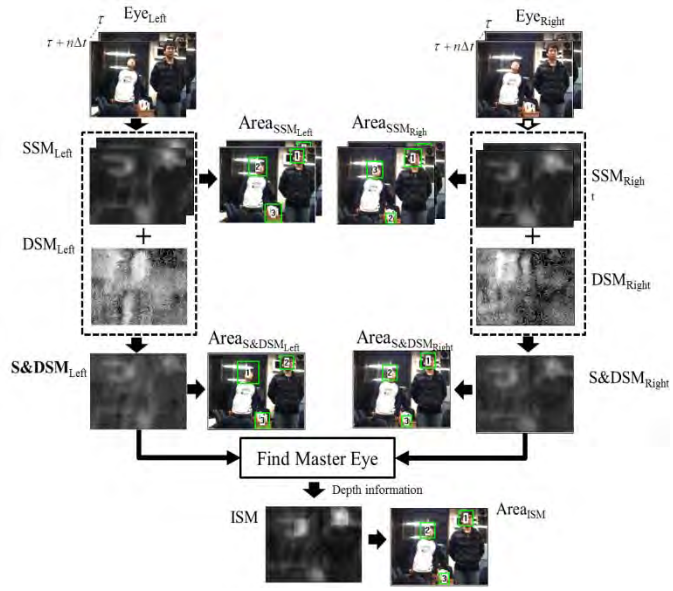


Figure 15. Comparison of SMs among static SM (SSM), dynamic SM (DSM), the static and dynamic SM (S&DSM), and the integrated SM (ISM)

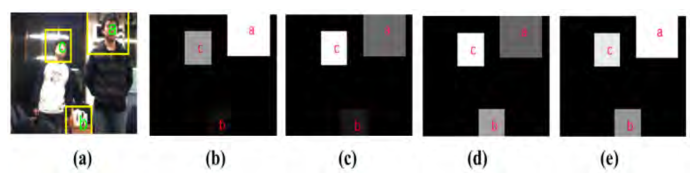


Figure 16. Experimental results of motion analysis using integraed SM.

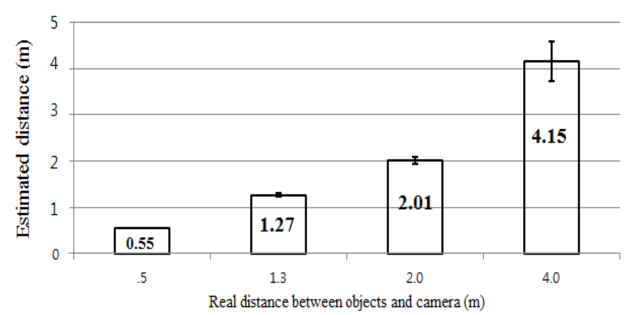


Figure 17. Comparison of real depth with estimated depth obtained by the stereo SM model.

Fig. 17 compares real depth with mean estimated one by the stereo SM model, in which we use the tens of data for estimating for each depth error. As Fig. 17 shows, the stereo SM model properly estimates depth. Although humans can perceive relative depth well, they may not estimate real depth correctly, whereas the proposed stereo SM model successfully estimates real depth within the range between 0.5m and 4m.

Table 3 shows how the average of saliency values in each salient area can be changed by using different τ values in Eq. (5) and constant value C that is fixed as 1 in Eq. (6). It is hard to decisively fix the τ value based on known biological mechanism. However, the data in Table 4 shows the importance of depth information in generating attention.

TABLE 4. PERFORMANCE OF THE AFFECTIVE SM MODEL AFTER TRAINING LIP AREAS AS A PREFERENTIAL REGION.

	96 Training Images	90 Test images	
		Bottom-up SM	Affective SM
# of with lip	96	35	88
# of without lip	0	55	2
Correct rate	100 (%)	39 (%)	98 (%)

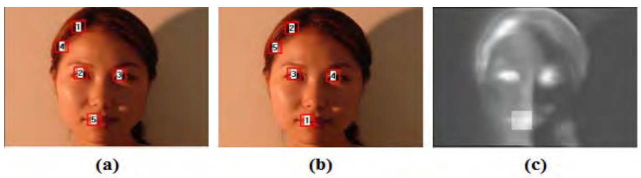


Figure 18. Affective saliency experiments

Fig. 18 shows experimental results using the affective saliency of our model. Fig. 18 (a) shows the results of the bottom-up SM model. Fig. 18 (b) shows the results of preferable one according to human affective factors after being trained by the affective SM model for preferential processing. Fig. 18 (b) shows modified scan paths after preference processing for the lips: the lip area in Fig. 18 (a) became the most salient area in Fig. 18 (b). Fig. 18 (c) sows the SMs generated by the affective SM model with lip preference.

Table 4 compares the performance of the bottom-up SM model with that of the affective SM model for focusing on the lip area in face images. POSTECH Face Database 2001(FD01) (Kim, Sung, Je, Kim, Kim, Jun, & Bang, 2002) was used for the experiments. As shown in Table 4, the bottom-up SM model can pay attention to the lip area with an accuracy of 39%. However, the affective SM model shows 98 % accuracy to focus on the lip area in the test images.

To verify the performance, the proposed GFTART was tested on two practical categorization problems. The first problem is to categorize pedestrians and cars on real traffic roads obtained from KNU and MIT CBCL databases. Fig. 19 shows some example images from 12 data sets consisting images of cars and pedestrians from the KNU database. Three among the 12 data sets were different class data sets (pedestrian images) with different characteristics from the other 9 data sets (car images). The 3rd, 8th, and 11th data sets were pedestrian images, while the other data sets were all car images.

We have shown that the relation between psychological distance and spatial distance can affect to visual selective attention process by the previous experiments. Fig. 20 shows the experimental results of the stereo affective SM model. Fig. 20 (a) shows a process for generating a stereo SM from two input images by left and right cameras. Left and right SMs are generated from corresponding left and right input image by bottom-up feature extraction and integration process. Then those two SMs are integrated as one stereo SM by reflecting depth information obtained from a depth

Figure 19. Sample images for cars and pedestrians in real traffic roads.

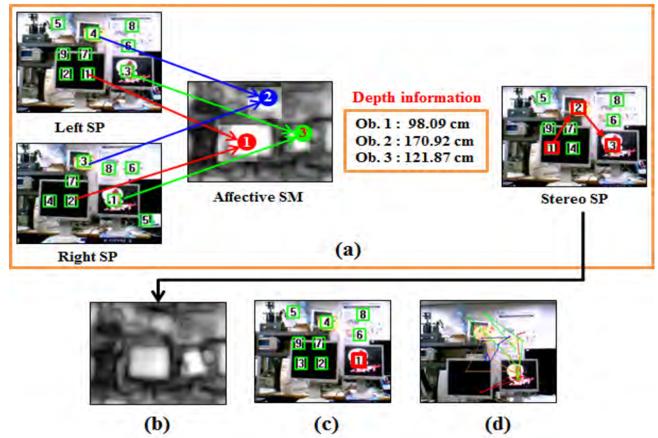


Figure 20. Comparison of visual scan path generated by the proposed affective SM model considering a psychological distance with that by the stereo SM without considering a psychological distance and human real visual scan path

perception module. Fig. 20 (a) shows a visual scan path generated by the visual attention model without considering the psychological distance. An affective stereo SM, as shown in Fig. 20 (b), is constructed by considering the psychological distance to construct the final stereo SM results. Finally a modified visual scan path by the affective SM is generated as shown in Fig. 20 (c), in which social distal visual stimuli given at a distant monitor (right monitor among two monitors) becomes more salient by considering psychological distance since congruent condition is occurred. In order to verify plausibility of the finally obtained visual scan path by the affective SM, we measured real human visual scan path for the same visual stimulus as shown in Fig. 20 (d). Human visual scan path in Fig. 20 (d) shows more similar scan path with the visual scan path in Fig. 20 (c) generated by the affective SM model than the visual scan path without considering psychological distance as shown in Fig. 20 (a).

Fig. 21 shows that the proposed visual attention model successfully reflects the psychological distance in both congruent conditions and incongruent conditions. As shown in Fig. 21, congruent visual stimuli become most salient through intensifying the degree of saliency in the course of reflecting psychological distance concept. On the other hand, in incongruent conditions shown in Fig. 21, incongruent visual stimuli become less salient through diminishing the degree of saliency by the proposed visual attention model. For all trials as shown in Fig. 21, the visual scan path generated by the proposed visual attention model shows higher similarity with real human visual scan path than the visual scan path generated by without considering a psychological distance.

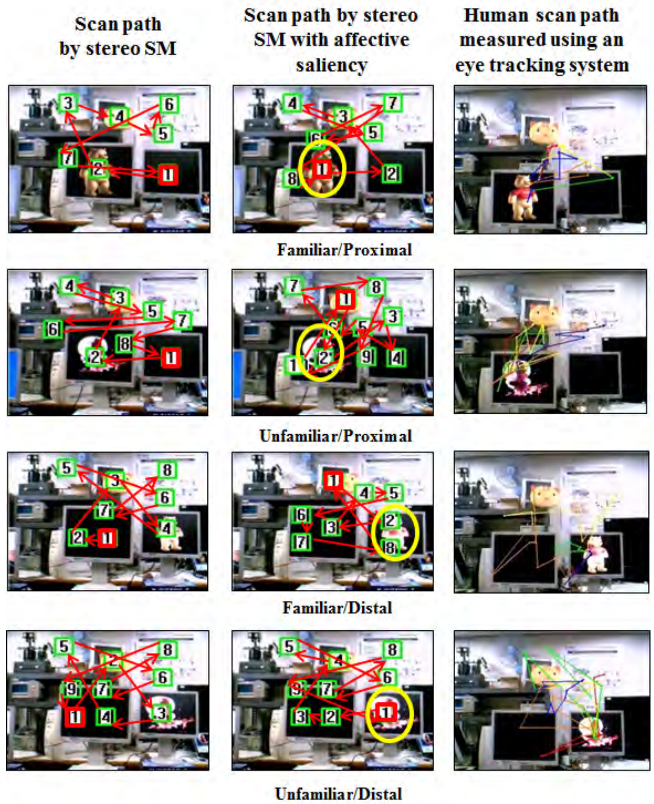


Figure 21. Comparison of visual scan paths by the stereo SM without considering a psychological distance and the affective SM considering a psychological distance with real human visual scan paths for congruent trials and incongruent trials.

6. Conclusion

I present several kinds of biologically motivated SM that is partly inspired by human visual selective attention mechanisms. Our experiments also illustrate the importance of including a symmetry FM and ICA filtering, which provide enhanced performance in generating preferred object attention. The presented selective attention model can also generate a SM that integrates the static and dynamic features as well as affective factors and depth information in natural input scenes. In particular, we added a Hebbian learning process to generate a top-down bias signal based on human affective factors, which enhances the performance of the previous Lee's trainable selection attention scheme (Choi et al., 2006). Another proposed selective attention model, which is motivated from Bar-Anan's psychological distance experiments, is a novel approach that considers psychological distance related with familiarity and preference as well as spatial distance in a stereo saliency map.

Moreover, an incremental neural network was introduced, which was based on combining the conventional fuzzy ART model and the GCS model. It plays important role for generating bias signals for the proposed object-oriented top-down attention model. Experimental results verified that the proposed model is able to utilize the advantages of each model, while alleviating their respective disadvantages. Nonetheless, a more appropriate vigilance measure is still needed for the GFTART model to enable a proper comparison of the topology of each object class represented by a GCS unit in the F2 layer.

The attention mechanism is so complex that we need to find more biological mechanisms related to generating attention or indirectly get insights from known biological mechanism in further work.

Acknowledgment

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (2011K000659)

References

Ban, S.-W., Lee, I., and Lee, M., 2008, Dynamic visual selective attention. Neurocomputing, 71(4-6), pp. 853-856.

Ban, S.-W., Jang Y.-M., and Lee, M., 2011, Affective saliency map considering psychological distance. Neurocomputing, 74, pp. 1916–1925

Barlow, H. B., and Tolhurst, D. J., 1992, Why do you have edge detectors? Optical society of America: Technical Digest, 23, 172.

Bar-Anan, Y., Trope, Y., Liberman, N., Algom, D., 2007, Automatic Processing of Psychological Distance: Evidence From a Stroop Task. J. Experimental Psychology: General 136(4), pp. 610–622.

Bear, M. F., Connors, B. W., and Paradiso M. A., 2001, Neuroscience Exploring The Brain, Lippincott Williams & Wilkins Co.

Belardinelli, A., and Pirri, F., 2006, A biologically plausible robot attention model, based on space and time. Cognitive Processing, 7(Suppl. 1), pp. 11-14.

Bell, A. J., & Sejnowski, T. J. 1997. The 'independent components' of natural scenes are edge filters. Vision Research, 37(23), 3327-3338.

Carmi, R., and Itti, L., 2006, Visual causes versus correlates of attentional selection in dynamic scenes. Vision Research, 46(26), pp. 4333-4345.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B., 1992, Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Transactions on Neural Networks, 3(5), pp. 698-713.

Choi, S.-B., Jung, B.-S., Ban, S.-W., Niitsuma, H., and Lee, M., 2006, Biologically motivated vergence control system using human-like selective attention model. Neurocomputing, 69(4-6), pp. 537-558.

Cui, X., Liu, Q., and Metaxas, D., 2009, Temporal spectral residual: fast motion saliency detection. Proceedings of the ACM international Conference on Multimedia.

Fernández-Caballero, A., López, M. T., and Saiz-Valverde, S., 2008, Dynamic stereoscopic selective attention (DSSVA): Integrating motion and shape with depth in video segmentation. Expert Systems with Applications, 34(2), pp. 1394-1402.

Frintrop, S., Rome, E., Nüchter, A., and Surmann, H., 2005, A bimodal laser-based attention system. Computer Vision and Image Understanding, 100(1-2), pp. 124-151.

Fritzke, B., 1994, Growing cell tuructures-A self-organizing network for unsupervised and supervised learning. Neural Networks, 7(9), pp. 1441-1460.

Fukushima, K., 2005, Use of non-uniform spatial blur for image comparison: symmetry axis extraction. Neural Networks, 18(1), pp. 23-32.

Fukushima, K., 2008, Extraction of visual motion and optic flow. Neural Networks, 21(5), pp. 774-785.

Gopalakrishnan, V., Hu, Y., and Rajan, D., 2009, Salient Region Detection by Modeling Distributions of Color and Orientation. IEEE Trans. Multimedia, 11(5), pp. 892-905.

Grossberg, S., 1987, Competitive learning: from interactive activation to adaptive resonance. Cognitive Science, 11, pp. 23-63.

Guyton, A. C., 1991, Textbook of medical physiology (8th ed.). USA: W.B. Saunders Company.

Guo, C. and Zhang, L., 2010, A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression. IEEE Trans. Image Processing, 19(1), pp. 185-198.

Guyton A. C., 1991, Textbook of medical physiology, 8th ed., W.B. Saunders Company, USA.

Goldstein E. B., 1995, Sensation and Perception. 4th edn. An International Thomson Publishing Company, USA.

Hou, X. and Zhang, L., 2007, Saliency Detection: A Spectral Residual Approach. Proceedings of Int. Conf. on Computer Vision and Pattern Recognition. pp. 1-8.

Itti, L., Koch, C., and Niebur, E., 1998, A model of saliency-based visual attention for rapid scene analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence, 20(11), pp. 1254-1259.

Jeong, S., Ban, S.-W., and Lee, M., 2008, Stereo saliency map considering affective factors and selective motion analysis in a dynamic environment. Neural networks, 21, 1420–1430.

Kadir, T., and Brady, M., 1996, Scale, saliency and image description. International Journal of Computer Vision, 45(2), pp. 83-105.

Kim, B., Ban, S. W., and Lee, M., 2011, Growing fuzzy topology adaptive resonance theory models with a push-pull learning algorithm. Neurocomputing, 74, pp. 646-655.

Kim, H.-C., Sung, J.-W., Je, H.-M., Kim, S.-K., Kim, D., Jun, B.-J., and Bang, S.-Y., 2002, Asian face image database PF01. Technical Report, Intelligent Multimedia Lab, Department of CSE, POSTECH.

Koike, T., and Saiki, J., 2002, Stochastic guided search model for search asymmetries in visual search tasks. Lecture Notes in Computer Science, 2525, pp. 408-417.

Kuffler, S. W., Nicholls, J. G., and Martin, J. G., 1984, (Ed.), From neuron to brain, Sunderland U.K. Sinauer Associates.

Li, Z., 2001, Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex, Neural Computation, 13(8), pp. 1749-1780.

Li, Z. and Itti, L., 2011, Saliency and Gist Features for Target Detection in Satellite Images. IEEE Trans. Image Processing, 20(7), pp. 2017-2029.

Maki, A., Nordlund, P., and Eklundh, J. O., 2000, Attentional scene segmentation: Integrating depth and motion. Computer Vision and Image Understanding, 78(3), pp. 351-373.

Majani, E., Erlanson, R., and Abu-Mostafa, Y., 1984, (Ed.), The eye, New York: Academic.

Marsland, S., Shapiro, J., Nehmzow, U., 2002, A self-organising network that grows when required. Neural Networks, Special Issue 15(8-9), pp. 1041–1058.

Mazer, J. A., Gallant, J. L., 2003, Goal-related activity in V4 during free viewing visual search: evidence for a ventral stream visual saliency map, Neuron 40, pp. 1241-1250.

Navalpakkam, V., and Itti, L., 2006, An integrated model of top-down and bottom-up attention for optimal object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2049-2056.

Ouerhani, N., and Hügli, H., 2000, Computing visual attention from scene depth. Proceedings of the International Conference on Pattern Recognition, pp. 375-378.

Park, S.-J., An, K.-H., and Lee, M., 2002, Saliency map model with adaptive masking based on independent component analysis. Neurocomputing, 49(1), pp. 417-422.

Ramström, O., and Christensen, H. I., 2002, Visual attention using game theory. Lecture Notes in Computer Science, 2525, pp. 462-471.

Reisfeld, D., Wolfson, H., & Yeshurun, Y., 1995, Context-free attentional operators: The generalized symmetry transform. International Journal of Computer Vision, 14, pp. 119-130.

Thorn, F., Gwiazda, J., Cruz, A. A. V., Bauer, J. A., and Held, R., 1994, The development of eye alignment, convergence, and sensory binocularity in young infants. Investigative Ophthalmology and Visual Science, 35, pp. 544-553.

Treisman, A. M., and Gelde, G., 1980, A feature-integration theory of attention. Cognitive Psychology, 12(1), pp. 97-136.

Vikram, T. N., Tscherepanow, M., and Wrede, B., 2010, A Random Center Surround Bottom up Visual Attention Model useful for Salient Region Detection. Proceedings of IEEE Workshop on Applications of Computer Vision (WACV), pp. 166-173.

Walther, D., Rutishauser, U., Koch, C., and Perona, P., 2005, Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. Computer Vision and Image Understanding, 100(1-2), pp. 41-63.

Wang, Z. and Li, B., 2008, A TWO-STAGE APPROACH TO SALIENCY DETECTION IN IMAGES. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 965-968

Regular Paper

CASA: Biologically Inspired Approaches for Auditory Scene Analysis

Azam Rabiee[1*], Saeed Setayeshi[2], and Soo-Young Lee[3]

[1]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2]Department of Medical Radiation, Amirkabir University of Technology, Tehran, Iran
[3]Brain Science Research Center, Korea Advanced Institute of Science & Technology, Korea
**corresponding author*: azrabiee@gmail.com

Abstract

This review presents an overview of computational auditory scene analysis (CASA), as biologically inspired approaches for machine sound separation. In this review, we address human auditory system containing early auditory stage, binaural combining, cortical stage, and top-down attention. We compared the models employed for CASA, especially for early auditory and cortical stages. We emphasized on how the existing models are similar to human auditory mechanism for sound separation. Finally, we discussed current issues and future of this task.

Keywords: Auditory model, CASA, auditory scene analysis

1. Introduction

In a natural environment, speech usually occurs simultaneously with acoustic interference. The acoustic interference, as a noise, reduces the performance of automatic speech recognition (ASR) systems. The most challenging issue is when the interference is another speech signal. Hence, many researchers are interested in speech signal separation task.

Some researchers have tried to separate signals explicitly using conventional signal processing approaches, such as blind signal separation (BSS) methods [1-6]. In this set of methods, microphone arrays are usually required to prepare input mixtures of signals. Independency of the sources is also an essential requirement of the methods. Other studies have tried to model human auditory system to overcome the problem implicitly [7-13].

Physiologically, with no more than two ears, the human auditory system shows a remarkable capacity for scene analysis. This is what Cherry called it *the cocktail party effect* for the first time [14][15]. Cherry in [15] wrote: "One of our most important faculties is our ability to listen to, and follow, one speaker in the presence of others. This is such a common experience that we may take it for granted; we may call it 'the cocktail party problem'. No machine has been constructed to do just this, to filter out one conversation from a number jumbled together."

According to Bregman [16], the auditory system separates the acoustic signal into streams, corresponding to different sources, based on auditory scene analysis (ASA) principles. Research in ASA has inspired considerable work to build computational auditory scene analysis (CASA) systems for sound separation. Physiological models of ASA may in turn lead to useful engineering systems for sound separation and speech enhancement. Although there is no requirement that a practical system should be based on a physiological account, ASA approaches based on neural models are attractive because they are comprised of simple, parallel and distributed components and are therefore well suited to hardware implementations [17].

Generally, in the conventional CASA systems, the input is supposed to be a mixture of target speech and the interferences. Hence, the interferences are removed by a binary (or gray) mask in a time-frequency representation, using two main stages: segmentation (analysis) and grouping (synthesis) [13]. In segmentation, the acoustic input is decomposed into continuous time-frequency units or segments. Each of these segments originates from a single speaker. In grouping, those segments that likely come from the same source are grouped together.

Although, many CASA approaches have presented in the last two decades, the current models of the human auditory system for this task still need to be improved. In the rest of the current review, we mention the physiology of hearing, and then investigate different human auditory models in CASA, especially for early auditory and cortical stages.

2. Human Auditory System

Generally, human auditory system contains the ears and the central auditory system [18]. As the early auditory stage, ear receives the sound waves and generates corresponding neural signals for the central auditory system. Figure 1(a) illustrates the physiology of the ear containing the outer, middle and inner ears. The outer ear includes the pinna, the ear canal, and the very most superficial layer of the eardrum. The outer ear acts as a sound collector and enhances the sound vibrations best at the human audible frequency range. Moreover, it serves sound amplification and localization.

The middle ear, including most of the eardrum and three bones, converts the acoustic energy of the sound into the mechanical vibration. The mechanical vibration of the eardrum in the middle ear helps the last bone (stapes) to push the fluid in and out on the cochlea. The most complicated part of the ear, the inner ear includes the cochlea and the vestibular system. Since the vestibular system is not related to our topic, we avoid its description. The cochlea is a system of coiled tubes consisting of two liquid-filled tubes coiled side by side as shown in the cross section in Figure 1 (a). The two main tubes are separated from each other by the basilar membrane. Along the coil of the cochlea, the basilar membrane is approximately 35 mm in length and its stiffness varies by a factor of 100 along its length. The physical characteristics of the basilar membrane make it act like a frequency analyzer, which responds tonotopically to the sound.

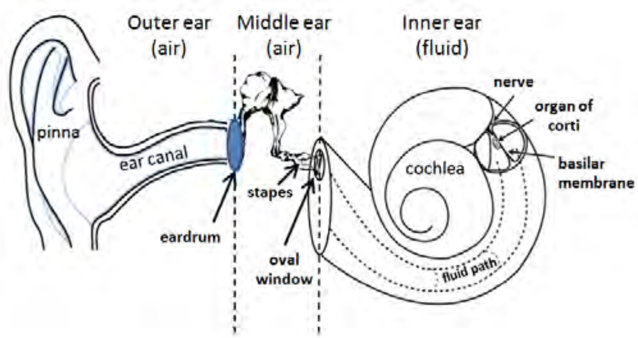
On the surface of the basilar membrane lies the organ of Corti, which contains a series of electromechanically sensitive cells, called the hair cells. The inner hair cells convert the vibrating fluid into neural signals, i.e. spikes. In fact, the output of each inner hair cell represents acoustic input signal with specific frequency filtering and nonlinear characteristics through the spikes. Furthermore, the outer hair cells are believed to conduct the function of automatic gain controls.

The neural spikes generated in the early auditory stage go to the next stage, central auditory system, for further processing. Figure 1(b) demonstrates a simplified schematic diagram of the human auditory pathway in which the early auditory stage is shown by left and right cochleas. In the central auditory system, the sound information travels through intermediate stations such as the cochlear nucleus and superior olivary complex (SOC) of the brainstem and the inferior colliculus (IC) of the midbrain. As shown in the figure, the information eventually reaches the medial geniculate nucleus (MGN) in thalamus, and from there it is relayed to the primary auditory cortex, which is located in the temporal lobe of the brain.

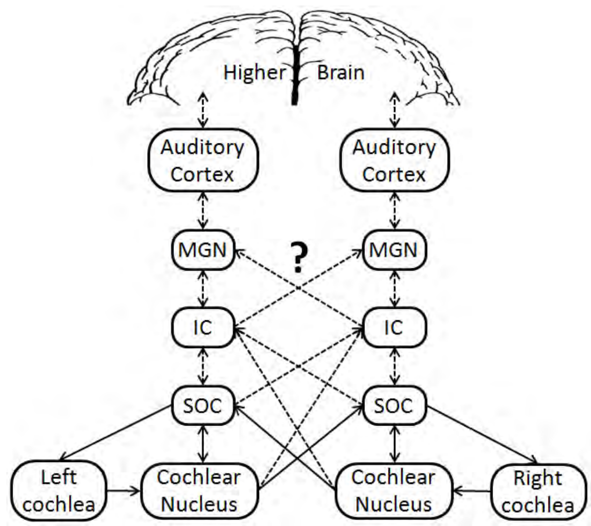
Signals from both left and right ears merge at SOCs. Physiologically, the medial superior olive (MSO) in SOC is a specialized nucleus that is believed to measure the interaural time difference (ITD). The ITD is a major cue for determining the azimuth of low-frequency sounds, i. e., localizing them on the azimuthal plane, their degree to the left or the right. On the other hand, the lateral superior olive (LSO) is believed to be involved in measuring the interaural level difference (ILD). The ILD is a second major cue in determining the azimuth of high-frequency sounds.

Major ascending auditory pathway converges in IC before sending to the thalamus and cortex. IC appears as an integrative station and switchboard. It is involved in the integration and routing of multi-modal sensory perception. It is also responsive to specific amplitude modulation frequencies, and this might be responsible for detection of pitch. In addition, spatial localization by binaural hearing is a related function of IC, specifically regarding the information from the SOC.

Moreover, MGN represents the thalamic relay between the IC and the auditory cortex. It is thought that the MGN influences the direction and maintenance of attention. MGN



(a)



(b)

Fig. 1. (a) The ear including outer, middle and inner ears converts the acoustic signal into neural patterns for the central auditory system. (b) A simplified schematic diagram of the human auditory pathway. Signals from both left and right ears merge at SOCs, and go to auditory cortexes through ICs and MGNs. Also, there exist backward paths from the higher brain through auditory cortex to the cochlea. The signal processing mechanism between SOC and auditory cortex is less understood, and represented as dotted lines.

is primarily responsible for relaying frequency, intensity and binaural information to the cortex. In addition, some of the neurons in MGN respond to other stimuli often from somatosensory. The behavior of these cells is complicated by the fact that sensory stimulation from other modalities modifies the responsiveness of many of them. Moreover, it is not clear whether there truly is one, none, or many tonotopic organizations maps present in the MGN.

Eventually, the auditory cortex receives the information from the thalamus. Functionally, the cortical stage estimates the spectral and temporal modulation content of the early stage output. The current understanding of cortical processing reveals that cortical units exhibit a wide variety of receptive field profiles. These response fields, also called spectrotemporal receptive fields (STRFs), represent a time-frequency transfer function of each neuron and summarize the way each cell responds to the stimulus, hence capturing the specific sound features that selectively drive the cell best.

Speech recognition and language understanding take place at the higher brain via the interaction with other regions of the brain. Furthermore, there exist backward paths from the higher brain through auditory cortex to the cochlea. Although the earlier auditory signal processing mechanisms at cochlea and possibly up to SOC are relatively well understood, the signal processing mechanism between SOC and auditory cortex is less understood [19], and represented as dotted lines in Figure 1(b).

Several biologically inspired models of the human auditory system are reported in the literatures. According to [19], the developed mathematical models of the human auditory pathway include three components: (1) the nonlinear feature extraction model from the cochlea to the auditory cortex, (2) the binaural processing model in brainstem and midbrain, and (3) the top-down attention model from higher brain to the cochlea. In the following, we mostly focus on the feature extraction models utilized in CASA in two parts: early auditory and cortical processing.

3. Models for the Early Auditory Stage

Mainly, the early auditory models mimic the function of basilar membrane in the cochlea by either a transmission line, i.e. a cascade of filter section, or a filter bank, in which each filter models the frequency response associated with a particular point on the basilar membrane. Then, the outputs of the basilar membrane are further processed to derive a simulation of auditory nerve activity using a representation of firing rate or spike-based representation by a half-wave rectification of the filterbank output followed by a nonlinear function. A more sophisticated approach may model the automatic gain control of outer hair cells and midbrain integration, as well.

Proposed in the last two decades, many CASA systems have investigated the role of the early auditory stage in performing frequency analysis and transforming the waveform signal into a 2D time-frequency representation. Conventional CASA systems utilize ASA cues to decompose this 2D representation into sensory segments in segmentation stage, as well as to assign those segments to corresponding speakers in grouping stage [8]. Among several models for early auditory stage, we explain two well-known models to generate auditory spectrogram and cochleagram representations in the following.

3.1. Auditory Spectrogram

A self-normalized and noise-robust auditory spectrogram for early auditory representation is introduced by Wang and Shamma in 1994 [20]. In brief, the early auditory stage consists of cochlear filter bank, hair cell transduction, lateral inhibitory network (LIN), and midbrain integration. The schematic diagram of the model for the auditory spectrogram is illustrated in Figure 2(a).

In the first stage, the cochlear filter bank contains a bank of 128 overlapping band pass filters with center frequencies uniformly distributed along a logarithmic frequency axis (x), over 5.3 oct (24 filters/octave). Let $f(t;x)$ be the impulse response of each filter. Then, given $s(t)$, the input signal in time domain, the cochlear filter output is calculated by

$$y_{coch}(n,x) = s(t) *_t f(t;x) \tag{1}$$

where $*_t$ is convolution in time domain.

These cochlear filter outputs are transduced into auditory-nerve patterns $y_{AN}(t,x)$ by a hair cell stage consisting of a high-pass filter, a nonlinear compression $g(.)$, and a membrane leakage low-pass filter $\omega(t)$ accounting for decrease of phase-locking on the auditory nerve beyond 2 kHz, as follows:

$$y_{AN}(t,x) = g\big(\partial_t y_{coch}(t,x)\big) *_t \omega(t). \tag{2}$$

The next transformation simulates the action of laterally inhibition. The LIN is simply approximated by a first-order derivative with respect to the tonotopic axis and followed by a half-wave rectifier, as follows:

$$y_{LIN}(t,x) = max(\partial_x y_{AN}(t,x), 0). \tag{3}$$

The final output of this step, the auditory spectrogram $p(t,x)$, is obtained by integrating $y_{LIN}(t,x)$ over a short window, $\mu(t,\tau) = e^{-t/\tau}u(t)$, with time constant $\tau = 8$ ms mimicking the further loss of phase locking observed in the midbrain, as

$$p(t,x) = y_{LIN}(t,x) *_t \mu(t,\tau). \tag{4}$$

3.2. Cochleagram

The well-known cochleagram introduced by Wang and Brown in [12] is another model for the early auditory stage which is utilized in many CASA systems [8-11]. The stages of generating the cochleagram are shown in Figure 2 (b). In the cochleagram, the basilar membrane is modeled by gammatone filters. The gammatone is a bandpass filter, whose impulse response, $g_{f_c}(t)$, is the product of a gamma function and a tone:

$$g_{f_c}(t) = t^{N-1} e^{-2\pi t b(f_c)} \cos(2\pi f_c t + \phi)\, u(t). \tag{5}$$

Here, N is the filter order, f_c is the filter center frequency in Hz, ϕ is the phase, and $u(t)$ is the unit step function. The function $b(f_c)$ determines the bandwidth for a given center frequency. The bandwidth of the gammatone filter is usually set according to measurements of the equivalent rectangular bandwidth (ERB), which is a good match to human data, given by

$$ERB(f) = 24.7 + 0.108f. \tag{6}$$

The center frequencies are linear in ERB domain and usually from 50 Hz to 8 kHz.

The gammatone filterbank is often paired with the model of hair cell transduction proposed by Meddis [21]. Physiologically, in the inner hair cells, movements of the stereocilia, hairs attach to the hair-cell, cause a depolarisation of the inner hair cell, which in turn results in a receptor potential. The receptor potential in the hair-cell causes the release of neurotransmitter into the auditory nerve i.e., the synaptic cleft. The change in neurotransmitter concentration generates a spike. After such a spike it takes a while to prepare for the next spike. This no-spike period is called the absolute refractory period and lasts approximately 1 ms.

Specifically, in the meddis model, the rate of change of the amount of neurotransmitter in the synaptic cleft is calculated by

$$\frac{dc(t)}{dt} = k(t)q(t) - lc(t) - rc(t), \tag{7}$$

where $k(t)$ is the permeability, $q(t)$ is the transmitter level, $c(t)$ is the amount of transmitter in the synaptic cleft, l is a loss factor, and r is a return factor. Thus the term $k(t)q(t)$ is the amount of transmitter received from the hair-cell, $lc(t)$ is the amount of transmitter lost from the cleft and $rc(t)$ is the amount of transmitter returned to the hair-cell.

Eventually, from the assumption that the spike probability is proportional to the amount of transmitter in the synaptic cleft, the probability of spike generation is calculated as follows,

$$P = hc(t)dt, \tag{8}$$

where h is the proportionality factor. The probability is computed for every output of the gammatone filterbank, independently.

3.3. Discussion

An example of the auditory spectrogram and the cochleagram is demonstrated in Figure 3. By comparing the figures (a) and (b), the auditory spectrogram and the

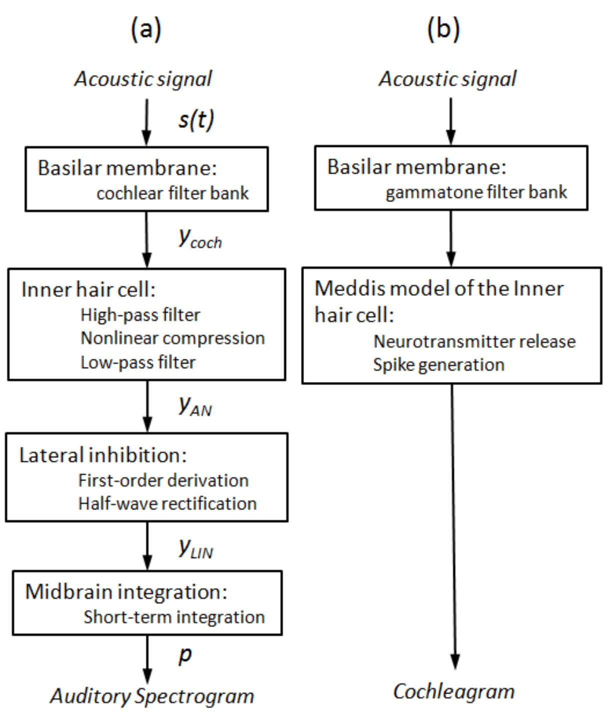
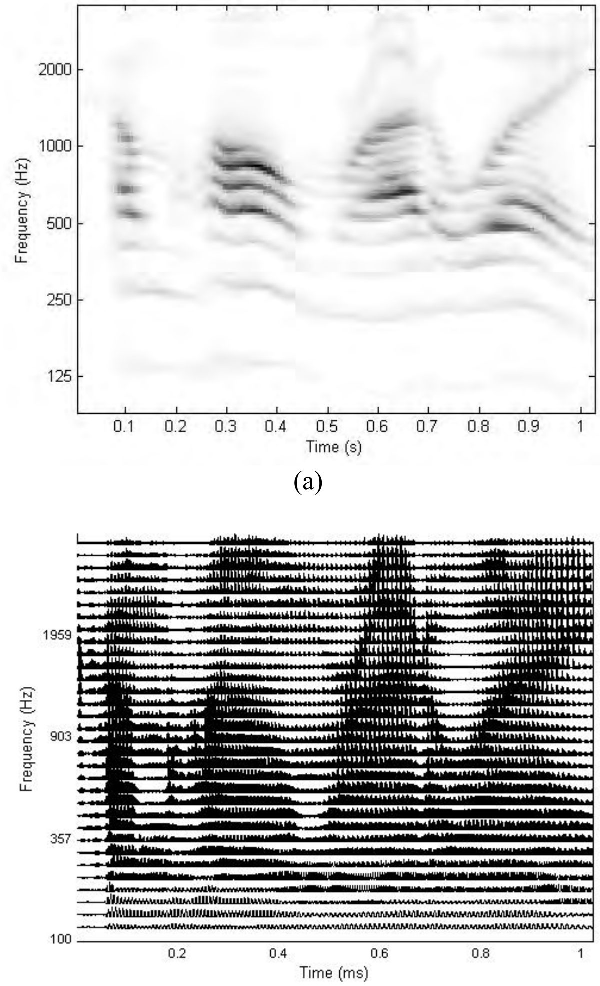


Fig. 2. The stages of two early auditory models: (a) Auditory spectrogram, and (b) Cochleagram.



(b)

Fig. 3. Different time-frequency representations of the sentence "come home right away" in TIMIT corpus uttered by a male speaker. (a) Auditory spectrogram of Wang and Shamma [20]. (b) Cochleagram introduced by Wang and Brown [12].

cochlearagm for the same sentence "come home right away" show different time-frequency representation of similar characteristics.

It was demonstrated that the auditory spectrogram has a significant advantage over conventional representations in noise robustness when employed as a front-end for ASR systems and for source separation [7][20]. The spectrogram is also self-normalized which means it has relative stability with respect to an overall scaling. Furthermore, the representation is suitable for music processing, because of its 1/12-octave spacing of center frequencies, which matched to the note spacing. It is also appropriate for harmonic analysis, because of its sharpness in frequency axis as a result of the laterally inhibition process, as it is clear from Figure 3(a).

On the other hand, the Meddis model in the cochleagram represents a good compromise between accuracy and computational efficiency. The model replicates many of the characteristics of auditory nerve responses, including rectification, compression, spontaneous firing, saturation effects and adaptation.

Although some CASA research have established their methods on cochleagram domain [10][22][23], correlogram extracted from the cochleagram also shows a robust time-frequency representation, especially for pitch estimation in multiple simultaneous sources in several research [9][24][25]. The correlogram is usually computed in the time domain by autocorrelating the simulated auditory nerve firing activity at the output of each cochlear filter channel, resulting in a 3D time-frequency-lag representation of the acoustic signal.

4. Models for the Cortical Stage

As described in the previous section, an early stage captures the process from the cochlea to the midbrain. It transforms the acoustic stimulus to an auditory time-frequency spectrogram-like representation. Although, many CASA systems have employed human early auditory modeling, a few papers have explored the role of cortical mechanisms in organizing complex auditory scenes. In fact, auditory cortex in or near Heschl's gyrus, as well as in the planum temporale are involved in sound segregation [26].

Generally, the role of the cortical stage is to analyze the spectrotemporal content of the spectrogram. In the following, we mention two cortical models known as multiresolution spectrotemporal analysis and localized spectrotemporal analysis. The utilization of the models for CASA is considered as well.

4.1. Multireseolution Spectrotemporal Analysis

Chi et. al. in [27] have described a computational model of auditory analysis that is strongly inspired by psychoacoustical and neurophysiological findings over the past two decades in both early and central stages of the auditory system. The model for the early auditory stage is the auditory spectrogram described in Section 3.1.

The central stage, specifically, models the process in primary auditory cortex. It does so computationally via a bank of filters that are selective to different spectrotemporal modulation parameters that range from *slow* to *fast* rates temporally, and from *narrow* to *broad* scales spectrally. Various temporal and spectral characteristics of cells are revealed in their STRFs. In the model, they assumed a bank of directional selective STRF's (downward [−] and upward [+]) that are real functions formed by combining two complex functions of time and frequency as follows,

$$STRF_{+} = \Re\{H_{rate}(t;\omega,\theta).H_{scale}(f;\Omega,\phi)\},$$
$$STRF_{-} = \Re\{H^{*}_{rate}(t;\omega,\theta).H_{scale}(f;\Omega,\phi)\}, \tag{9}$$

where $\Re$ denotes the real part, * is the complex conjugate, ω and Ω are rate and scale parameters of STRF respectively, and θ and ϕ are characteristic phases that determine the degree of asymmetry along time and frequency, respectively. Functions H_{rate} and H_{scale} are analytic signals obtained from h_{rate} and h_{scale}:

$$H_{rate}(t;\omega,\theta) = h_{rate}(t;\omega,\theta) + j\hat{h}_{rate}(t;\omega,\theta),$$

$$H_{scale}(f;\Omega,\phi) = h_{scale}(f;\Omega,\phi) + j\hat{h}_{scale}(f;\Omega,\phi), \tag{10}$$

where $\hat{.}$ denotes Hilbert transform. h_{rate} and h_{scale} are temporal and spectral impulse responses defined by sinusoidally interpolating between symmetric seed functions $h_r(.)$ and $h_s(.)$, and their asymmetric Hilbert transforms:

$$h_{rate}(t;\omega,\theta) = h_r(t;\omega)\cos\theta + \hat{h}_r(t;\omega)\sin\theta,$$
$$h_{scale}(f;\Omega,\phi) = h_s(f;\Omega)\cos\phi + \hat{h}_s(f;\Omega)\sin\phi. \tag{11}$$

Eventually, the impulse responses for different scales and rates are given by dilation

$$h_r(t;\omega) = \omega h_r(\omega t),$$
$$h_s(f;\Omega) = \Omega h_s(\Omega f), \tag{12}$$

in which

$$\mathrm{h_r(t) = t^2 e^{-3.5t}\sin 2\pi t},$$
$$\mathrm{h_s(f) = (1-s^2)e^{-s^2/2}}\ . \tag{13}$$

As shown in the Figure 4(a), the convolution between STRFs and the spectrogram gives an estimate of the time-varying firing rate of the neurons, generating a multidimensional representation of the waveform signal with time, frequency, scale, and rate axis.

Numerous studies have utilized the computational model for feature extraction [28][29] and speech enhancement [7][30]. Among them, Elhilali and Shamma in [7] have utilized it for a sound separation task consists of a multidimensional feature representation stage followed by an integrative clustering stage in which the segregation was performed based on an unsupervised clustering and the statistical theory of Kalman prediction.

In the model, they broke down the cortical analysis into a spectral mapping and a temporal analysis. The spectral shape analysis was considered to be part of the feature analysis stage of the model, as it further maps the sound patterns into a spectral shape axis organized from narrow to broad spectral features. On the other hand, the slow temporal dynamics (<30Hz) came into play in the next integrative and clustering stage.

In the study, they have demonstrated that the model can successfully tackle aspects of the "cocktail party problem" and provide an account of the perceptual process during stream formation in auditory scene analysis. The model directly tested the premise that cortical mechanisms play a fundamental role in organizing sound features into auditory objects by (1) mapping the acoustic scene into a multidimensional feature space and (2) using the spectral and temporal context to direct sensory information into corresponding perceptual streams.

One of the drawbacks of the multiresolution spectrotemporal analysis [27] is that the spectrotemporal responses are organized into a very large multi-dimensional representation, which is very hard to visualize and interpret. In contrast, the following localized spectrotemporal analysis presented in [31] is less computationally complex.

4.2. Localized Spectrotemporal Analysis

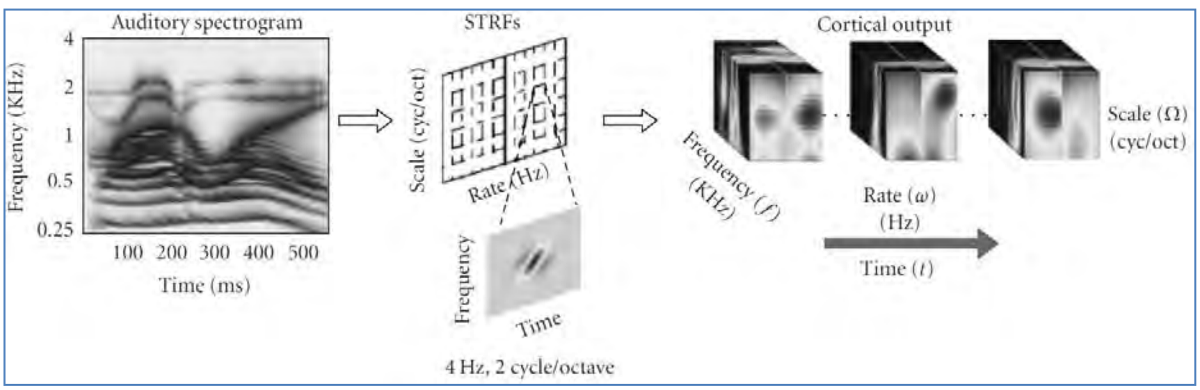
A simplified model of the auditory cortex is presented in [31] for analyzing the spectrotemporal content of the early auditory output. Wang and Quatieri in [31] have proposed the spectrotemporal analysis via a 2D Gabor filterbank over localized time-frequency patches of a narrowband spectrogram.

To generate the patches, a moving window, usually of size 50ms by 700Hz, sweeps whole the narrowband spectrogram with a rational jump in time and frequency, usually 5ms and 100Hz, respectively. As an example a simplified harmonic patch of the conventional short time Fourier transform (STFT) spectrogram is illustrated in the Figure 4 (b) left, in which the parallel lines show the harmonics. The localized spectrotemporal analysis can be done by a simple 2D Fourier transform. Certainly, a 2D Hamming or Gaussian window before the transform prevents the aliasing effect.

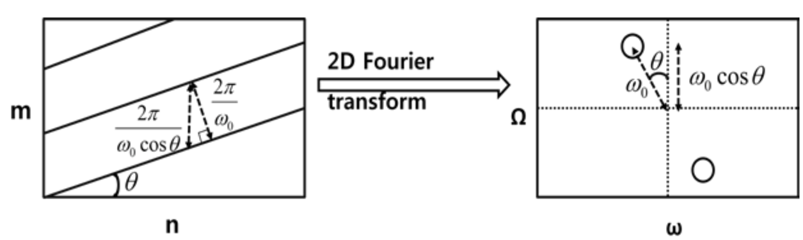
As shown in Figure 4(b), the transform of the parallel harmonic lines in the patch leads to two compressed points shown in the right figure, in which their vertical distance is related to the pitch value. Indeed, the 2D transform analyzes the temporal and spectral dynamics of the spectrogram. Therefore, the localized spectrotemporal analysis is spiritually very similar to the previous model [27], except its localized processing and consequently, its lower computational complexity.

Wang and Quatieri have evaluated their model for pitch processing, since pitch is an essential cue for speech and music perception and separation. Moreover, psychoacoustical and neurophysiological experiments show pitch processing mainly appears in the cortical stage of several species of mammals [32]. They illustrated the usability of their model in 1) multi-pitch estimation, especially in the case of two close pitch values and crossing trajectories [33], and 2) speech separation using a-priori pitch estimates of individual speakers [34]. Furthermore, we in [35] indicated how harmonic magnitude suppression can be integrated with the localized spectrotemporal processing to separate voiced speech signals.

Although the model is nicely fit for multi-pitch extraction via a spectrotemporal process mimicking the primary auditory cortex function, the processing of pitch in the auditory cortex of mammals is much sophisticated, and requires higher-order cortical areas and interactions with the frontal cortex [32]. In fact, a fixed pitch seems to active the Heschl's gyrus and the planum temporale. Moreover, when the pitch is varied the activation is found in the regions beyond Heschl's gyrus and planum temporale, specifically in the superior temporal gyrus and planum polare [32]. Hence, the model, lonely, is too simplified to be used for CASA.



(a)



(b)

Fig. 4. Different models of spectrotemporal analysis, mimicking the auditory cortical stage. (a) The model of the cortical stage presented by Chi et. al. in [27] (adapted from [30]). To analyze its spectrotemporal content, the auditory spectrogram is convolved with the STRFs of the cortical cells, generating the time-frequency representation in different scales and rates (b) The localized spectrotemporal analysis presented by Wang and Quatieri in [31]. Left figure is a patch of the conventional STFT spectrogram of a harmonic signal, in which the parallel lines are the harmonics. The 2D Fourier transform of the parallel harmonic lines in the patch leads to two compressed points shown in the right figure, in which their vertical distance is related to the pitch value.

4.3. Future of the Cortical Models

The two previous models of cortical stage are utilized for sound separation in some experiments [7][35]. There should be other cortical models that are not necessarily utilized for this task yet. For example, since auditory and visual cortices are structurally similar, in addition to the mentioned studies, it is worth it to evaluate if a visual cortex model such as Neocognitron [36][37] or HMAX [38][39] can be customized for audio processing. Spiritually similar, Neocognitron and HMAX are cortex-like mechanisms for visual object recognition.

Moxham et. al. in [40] incorporated the Neocognitron for pitch estimation and voice detection. They emphasized the reliability of the Neocognitron-based method via some experiments and comparison with the existing methods. Yamauchi et.al. in [41] also employed the Neocognitron as the recognition module of a speed invariant speech recognizer, which benefits from velocity-controlled delay lines. Both mentioned research may encourage customizing visual cortex models for more complicated auditory tasks like source separation in the future.

In some sense, models for visual cortex are more matured than the auditory cortex. Nevertheless, it is not easy to develop a comprehensive model for auditory cortex, especially because of the stochastic nature and temporal dynamics of sound. Moreover, physiologically, there are not many findings about how individual cortical areas compute, the nature of the output from those areas to cognitive and higher brain, as well as the interaction of the auditory cortex with other sensory areas, which is called *the binding problem* in neuroscience [17]. The problem talks about how information is encoded in different areas of the brain bound together into a coherent whole. Hence, we need to consider a richer model of sound processing and auditory object recognition, especially in brain.

5. Discussions and Future

5.1. Binaural vs. Monaural models

Human auditory system as a reference model is strongly capable of separating sound sources employing either one or two ears. Hence, there are two groups of CASA approaches depending on the number of available input mixtures. Specifically, many speech separation and recognition approaches competed in the monaural challenge in Interspeech 2006 [11], and consequently, binaural solutions in Pascal CHiME challenge in Interspeech 2011 [42].

Nowadays, equipping two microphones in sound separation platforms, even cell phone devices, is not expensive or inaccessible demand. Hence, binaural methods can be feasibility utilized in practice. Indeed, the extraction of the spatial properties of the sources is an informative cue for separation in mammalian auditory system and is performed in SOC utilizing the interaural difference information received from left and right cochlear nucleuses [43], which can be easily simulated in binaural methods.

Despite various mechanisms are suggested by binaural CASA researchers for extraction of the ITD, ILD, and interaural phase difference (IPD) [44], we would not ignore the remarkable monaural CASA approaches. Certainly, monaural source separation is the extreme case of the separation with respect to the number of available mixtures and is a challenging task, especially when the number of sources is more than two. Harmonicity, onset/offset, amplitude modulation, frequency modulation, timbre and so on are the dominant cues usually considered in monaural approaches [11]. Undoubtedly, integration of the above mentioned monaural cues and binaural cues (ITD, ILD and IPD) improves the performance of separation [45].

5.2. Integration of Bottom-Up and Top-Down Models

The bottom-up models use information from the sound to group components and understand an auditory scene. Except for information such as temporal change of pitch or onsets, there is little high-level knowledge to guide the scene analysis process. Instead, our brains seem to abstract sounds, and solve the auditory scene analysis problem using high-level representations of each auditory object. In fact, the cortical feedback in the auditory system exerts its effect all the way down to the outer hair cells in the cochlea via the midbrain structure to reinforce the signal stream of interest and maximize expectation through feedback [46]. Although the physiological findings of the top-down process are not matured, in order to improve the performance, the CASA approaches go toward the combination of low-level and high-level models. Generally, according to the literatures, the best example of a top-down auditory understanding system is a probabilistic model such as hidden Markov model- (HMM-) based speech recognition system; but nobody has evaluated the suitability of modeling human language perception with a HMM [47]. Srinivasan and Wang in [23] combined bottom-up and top-down cues in order to simultaneously improve both mask estimation and recognition accuracy. They incorporated the top-down information in a probabilistic model for missing data recognizer. Similarly, Barker et. al. introduced a speech fragment decoding system integrating data-driven techniques and missing data techniques for simultaneous speaker identification and speech recognition in presence of a competing speaker [24].

Shao et. al. also employed an uncertainty decoding technique as a top-down model for missing data recognition in the back end of a two-stage segmentation and grouping CASA system [25]. On the other hand, in [10] the top-down integration is carried out by training Gaussian mixture models (GMMs) and vector quantizers (VQs) of the isolated clean data for each speaker, and incorporating in the bottom-up separation process that is based in speaker identification.

5.3. Recognition or Synthesizing

Human auditory system isolates separate representations of each sound object and never turns it back into sound.

Instead, it seems more likely that the sound understanding and sound separation occur in concert, and the brain only understands the concepts [47]. Human sound separation work should not strive to generate acoustic waveforms of the separated signals. In [48], Stark et. al. investigated both strategies, designing and applying a binary mask on the spectrogram of the mixture and synthesizing from the estimated speech features. In similar conditions, the target to masker ratio (TMR) results of the mask-based separation significantly outperformed the synthesized-based one. In fact, the separated signals in time domain carry an additional noise following the synthesis process. Hence, in the last stage of CASA, recognition without synthesizing the separated signals not only is biologically plausible, but also shows dominant performance.

5.4. CASA vs. BSS

Using a standard corpus of voiced speech mixed with interfering sounds, Kouwe et. al. in [49] have reported a comparison between CASA and BSS techniques, which have been developed independently. Eventually, they concluded that if the requirements, such as enough number of available mixtures and independency of the sources are met, BSS is a powerful technique and performs precisely; but the requirements may not be equitable with a natural environment.

On the other hand, in the natural environment, CASA brings the flexibility of the physiological systems which they model to bear on a variety of signal mixtures, so that they can achieve a reasonable level of separation in the absence of many of the requirements of BSS; but they are still weak in noisy conditions.

The different performance profiles of the CASA and BSS techniques suggest that there would be merit in combining the two approaches. More specifically, scene analysis heuristics that are employed by CASA systems (such as continuity of F0 and spatial location) could be exploited by BSS algorithms in order to improve their performance on real-world acoustic mixtures. Conversely, blind separation techniques could help CASA in decomposing mixtures that overlap substantially in the time-frequency plane [49]. Moreover, CASA solutions come to help BSS in underdetermined condition when the number of sources exceeds the number of mixtures. The research in [48][50] are examples of the bridge between underdetermined BSS and CASA.

5.5. Phase Information

According to the literatures, magnitude spectrum plays a dominant role for sound processing. Although, about 150 years ago, Ohm observed that the human auditory system is phase-deaf, recent studies show the importance of both phase and magnitude of the spectrogram [51]. However, traditionally, phase showed noise-like behavior, perhaps because of the low computational resolution. Nowadays, by increasing the speed and precision of the processors, a growing group of sound processing research is trying to investigate more features by the phase information [52-56]. Utilizing phase information is investigated in complex matrix factorization (CMF) for source separation in [57][58]; but there is not enough research in CASA methods in this case. The relationship of the phase of the harmonics introduced in [56] is a useful cue for multi-pitch extraction and may help separation of the harmonic signals that have overlap in some harmonics.

6. Conclusion

Physiological models of auditory scene analysis are still in their infancy. An efficient cortical model together with the higher brain incorporations is still an interesting research topic in this field. Moreover, the integration of top-down and bottom-up processing in ASA is an issue for future work, as is the role of attention. In addition, most computer models of ASA assume that the listener and sound sources are static. In natural environments, sound sources move over time, and the listener is active; as a result, factors such as head movement and dynamic tracking of spatial location need to be accounted for in more sophisticated models. Eventually, current CASA models cannot deal with more-than-two-source mixtures, efficiently.

Acknowledgment

S.Y. Lee was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (2009-0092812 and 2010-0028722).

References

[1] S. Makino, T. W. Lee, and H. Sawada, *Blind Speech Separation*. Springer, 2007.

[2] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, "Blind Source Separation Exploiting Higher-Order Frequency Dependencies," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 1, pp. 70-79, 2007.

[3] H. M. Park, S. H. Oh, and S. Y. Lee, "A modified infomax algorithm for blind signal separation," *Neurocomputing*, vol. 70, pp. 229-240, 2006.

[4] C. S. Dhir and S. Y. Lee, "Discriminant Independent Component Analysis," *IEEE Trans. on Neural Networks*, vol. 22, no. 6, June, pp. 845-857, 2011.

[5] H. M. Park, C. S. Dhir, S. H. Oh, and S. Y. Lee, "A filter bank approach to independent component analysis for convolved mixtures," *Neurocomputing*, vol. 69, pp. 2065-2077, 2006.

[6] C. S. Dhir, H. M. Park, and S. Y. Lee, "Directionally Constrained Filterbank ICA," *IEEE Signal Processing Letters,* vol. 14, no. 8, pp. 541-544, Aug., 2007.

[7] M. Elhilalia and S. A. Shamma, "A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation," *J. Acoust Soc. Am.*, vol. 124, no. 6, pp. 3751-71, 2008.

[8] G. Hu and D. L. Wang "Monaural speech segregation based on pitch tracking and amplitude modulation," *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1135-1150, 2004.

[9] N. Ma, P. Green, J. Barker, and A. Coy, "Exploiting correlogram structure for robust speech recognition with multiple speech sources," *Speech Communication*, pp. 874-891, 2007.

[10] P. Li, Y. Guan, S. Wang, B. Xu, and W. Liu, "Monaural speech separation based on MAXVQ and CASA for robust speech recognition," *Computer Speech & Language*, pp. 30-44, 2010.

[11] M. Cooke, J. R. Hershey, and S. J. Rennie, "Monaural speech separation and recognition challenge," *Computer Speech & Language*, pp. 1-15, 2010.

[12] D.L. Wang and G. J. Brown, "Fundamentals of computational auditory scene analysis," in *Computational Auditory Scene Analysis:*

Principles, Algorithms, and Applications, D. L. Wang and G.J. Brown, (Chapter 1, pp. 1-44), Eds., Wiley/IEEE Press, 2006.

[13] D. L. Wang, "Feature-based speech segregation," in *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, D. L. Wang and G. J. Brown, Eds., Wiley/IEEE Press, Hoboken NJ, pp. 81-114, 2006.

[14] E. Colin Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *J. Acoust. Soc. Am.* vol. 25, no. 5, pp. 975-979, 1953.

[15] E. C. Cherry, *On Human Communication: A Review, a Survey and a Criticism*. MIT Press, John Wiley & Sons, Inc., Chapman & Hall Limited, 1957.

[16] A. S. Bregman, *Auditory scene analysis*, Cambridge MA: MIT Press, 1990.

[17] G. J. Brown, "Physiological models of auditory scene analysis," in *Computational Models of the Auditory System,* R. Meddis, E .A. Lopez-Poveda, Eds., AN Popper and RR Fay, Springer Verlag, pp. 203-236, 2010.

[18] J. E. Hall and A. C. Guyton, *Textbook of medical physiology*, 12th ed., Saunders/Elsevier, 2011.

[19] S. Y. Lee, "Modeling Human Auditory Perception for Noise-Robust Speech Recognition," *Int. Conf. Neural Networks and Brain*, 2005.

[20] K. Wang and S. A. Shamma, "Self-normalization and noise-robustness in early auditory representations," *IEEE Trans. Audio Speech Lang. Process.*, vol. 2, no. 3, pp. 421–435, July 1994.

[21] R. Meddis and E. A. Lopez-Poveda, "Auditory Periphery: From Pinna to Auditory Nerve," in *Computational Models of the Auditory System*, R. Meddis and E. A. Lopez-Poveda, Eds., AN Popper and RR Fay, Springer Verlag, pp. 7-38, 2010.

[22] G. Hu and D. Wang, "Auditory Segmentation Based on Onset and Offset Analysis," *IEEE Trans. Audio Speech Lang. Process.*, pp. 396-405, 2007.

[23] S. Srinivasan and D. L. Wang, "Robust speech recognition by integrating speech separation and hypothesis testing," Speech Communication, pp. 72-81, 2010.

[24] J. Barker, N. Ma, A. Coy, and M. Cooke, "Speech fragment decoding techniques for simultaneous speaker identification and speech recognition," *Computer Speech & Language*, pp. 94-111, 2010.

[25] Y. Shao, S. Srinivasan, Z. Jin, and D. Wang, "A computational auditory scene analysis system for speech segregation and robust speech recognition," *Computer Speech & Language*, pp. 77-93, 2010.

[26] C. Alain, K. Reinke, K. L. McDonald, W. Chau, F. Tam, A. Pacurar, and S. Graham, "Left thalamo-cortical network implicated in successful speech separation and identification," *NeuroImage*, vol. 26, no. 2, pp. 592-599, 2005.

[27] T. Chi, P. Ru, and S. A. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," *J. Acoust Soc. Am.*, vol. 118, no. 2, pp. 887-906, Aug. 2005.

[28] N. Mesgarani, M. Slaney, and S. Shamma, "Discrimination of speech from non-speech based on multiscale spectro-temporal modulations," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 3, pp. 920–930, 2006.

[29] Q. Wu, L-Q Zhang, and G-C Shi, "Robust Feature Extraction for Speaker Recognition Based on Constrained Nonnegative Tensor Factorization," *J. Comput. Sci. Tech.*, vol. 25, no. 4, pp. 783–792, July 2010.

[30] N. Mesgarani and S. A. Shamma, "Denoising in the Domain of Spectrotemporal Modulations," *EURASIP J. Audio, Speech and Music Processing*, 2007.

[31] T. T. Wang and T. F. Quatieri, "2-d processing of speech for multi-pitch analysis," in *Proc. INTERSPEECH*, pp. 2827-2830, 2009.

[32] J. J. Eggermont, "The auditory cortex: The final frontier," in *Computational Models of the Auditory System*, R Meddis, EA Lopez-Poveda, AN Popper and RR Fay, Springer-Verlag, pp. 97-127, 2010.

[33] T. T. Wang and T. F. Quatieri, "Multi-Pitch Estimation by a Joint 2-D Representation of Pitch and Pitch Dynamics," in *Proc. INTERSPEECH*, pp. 645-648, 2010.

[34] T. T. Wang and T. F. Quatieri, "Towards Co-Channel Speaker Separation by 2-D Demodulation of Spectrograms," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 65-68, 2009.

[35] A. Rabiee, S. Setayeshi, and S. Y. Lee, "Monaural Speech Separation Based on a 2D Processing and Harmonic Analysis," in *Proc. INTERSPEECH*, Florence, Italy, 2011.

[36] K. Fukushima: "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193-202, April, 1980.

[37] K. Fukushima, "Neocognitron trained with winner-kill-loser rule," *Neural Networks*, pp. 926-938, 2010.

[38] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nat. Neurosci.* vol. 2, pp. 1019-1025, 1999.

[39] T. Serre, L. Wolf, S. M. Bileschi, M. Riesenhuber, and T. Poggio, "Robust Object Recognition with Cortex-Like Mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 411-426, 2007.

[40] J. R. E. Moxham, P. A. Jones, H. J. McDermott, and G. M. Clark, "a new algorithm for voicing detection and voice pitch estimation based on the Neocognitron," in *Proc. 1992 IEEE-SP Workshop Neural Networks Signal Proc.*, pp. 204-213, 1992.

[41] K. Yamauchi, M. Fukuda, and K. Fukushima, "Speed invariant speech recognition using variable velocity delay lines," *Neural Networks*, pp. 167-177, 1995.

[42] "The PASCAL 'CHiME' Speech Separation and Recognition Challenge," http://spandh.dcs.shef.ac.uk/projects/chime/challenge.html.

[43] T. R. Jennings and H. S. Colburn, "Models of the Superior Olivary Complex," in *Computational Models of the Auditory System*, R. Meddis and E. Lopez-Poveda, Eds., New York: Springer, 2010.

[44] N. Roman and D. L. Wang, "Binaural Speech Segregation," in *Speech and Audio Processing in Adverse Environments*, Hänsler E. et.al., Eds., Springer. pp. 525-549, 2008.

[45] J. Woodruff, R. Prabhavalkar, E. Fosler-Lussier, and D. Wang, "Combining monaural and binaural evidence for reverberant speech segregation," in *Proc. INTERSPEECH*, 2010, pp. 406-409.

[46] S. Haykin and Z. Chen, "The Cocktail Party Problem," *Neural Computation*, pp. 1875-1902, 2005.

[47] P. Divenyi, *Speech Separation by Human and Machine*, Kluwer Academic Publishers, 2005.

[48] M. Stark, M. Wohlmayr, and F. Pernkopf, "Source-Filter-Based Single-Channel Speech Separation Using Pitch Information," *IEEE Trans. Audio Speech Lang. Process.*, pp. 242-255, 2011.

[49] A. J. W. van der Kouwe, D. L. Wang and G. J. Brown, "A comparison of auditory and blind separation techniques for speech segregation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 9, pp. 189-195, 2001.

[50] M. H. Radfar, R. M. Dansereau, and A. Sayadiyan, "A Maximum Likelihood Estimation of Vocal-Tract-Related Filter Characteristics for Single Channel Speech Separation," *EURASIP J. Audio, Speech and Music Proc*., 2007.

[51] K. K. Paliwal and L. D. Alsteris, "Usefulness of phase spectrum in human speech perception," in *Proc. INTERSPEECH,* 2003.

[52] L. Wang, S. Ohtsuka, and S. Nakagawa, "High improvement of speaker identification and verification by combining MFCC and phase information," in Proc. ICASSP, 2009, pp. 4529-4532.

[53] L. Wang, K. Minami, K. Yamamoto, and S. Nakagawa, "Speaker identification by combining MFCC and phase information in noisy environments," in *Proc. ICASSP*, 2010, pp. 4502-4505.

[54] I. Saratxaga, I. Hernáez, D. Erro, E. Navas, and J. Sánchez, "Simple representation of signal phase for harmonic speech models," *Electronics Letters*, 45, pp. 381-383, 2009.

[55] I. Saratxaga, I. Hernáez, I. Odriozola, E. Navas, I. Luengo, and D. Erro, "Using harmonic phase information to improve ASR rate," in *Proc. INTERSPEECH*, 2010, pp. 1185-1188.

[56] I. Hernáez, I. Saratxaga, J. Sánchez, E. Navas, and I. Luengo, "Use of the Harmonic Phase in Speaker Recognition," in *Proc. INTERSPEECH,* 2011, pp. 2757-2760.

[57] R. M. Parry and I. Essa, "Incorporating phase information for source separation via spectrogram factorization." in *Proc. ICASSP*, 2007.

[58] B. King and L. E. Atlas, "Single-channel source separation using simplified-training complex matrix factorization," in *Proc. ICASSP*, 2010, pp. 4206-4209.

2012 INNS Awards

By Leonid Perlovsky, Ph.D.
Chair of the Awards Committee of the INNS

As the chair of the Awards Committee of the INNS, I am pleased and proud to announce the recipients of the 2012 INNS Awards:

2012 **Hebb Award** goes to: Moshe Bar

2012 **Helmholtz Award** goes to: Kunihiko Fukushima

2012 **Gabor Award** goes to: Nikola Kasabov

2012 **INNS Young Investigator Awards** go to: Sebastien Helie and Roman Ilin

These awards were decided after careful deliberations by the Awards Committee and the Board of Governors.

Moshe Bar, the Hebb Award recipient, is recognized for his long-standing contribution and achievements in biological and computational learning.

Kunihiko Fukushima, the Helmholtz Award recipient, is recognized for his many years of contribution and achievements in understanding sensation/perception.

Nikola Kasabov, the Gabor Award recipient, is recognized for his achievements in engineering/ application of neural networks.

Sebastien Helie and Roman Ilin, the Young Investigator Award recipients, are recognized for significant contributions in the field of Neural Networks by a young person (with no more than five years postdoctoral experience and who are under forty years of age).

These awards will be presented at IJCNN 2012 in Brisbane. ■

LET'S CONGRATULATE THE AWARDEES!

Moshe Bar
Recipient of 2012 INNS Hebb Award

Dr. Bar graduated Ben-Gurion University in Israel in 1988 with a Bachelor of Science in Biomedical Engineering. After graduating, Dr. Bar spent the next six years as a member of the Israeli Air Force, during which time he began his Masters work in Computer Science at the Weizmann Institute of Science. After completing his Masters education in 1994, he entered a PhD program in Cognitive Neuroscience at the University of Southern California, where he was awarded the Psychology department's 'Outstanding Doctoral Thesis Award'. He completed his post-doctoral fellowship at Harvard University in Cambridge, Massachusetts in 2000. Since this time he has been the recipient of many distinguished awards and research grants including, McDonnel-Pew Award in Cognitive Neuroscience, the prestigious Mc Donnell Foundation's 21st Century Science Initiative Award, and several Research Awards from the National Institutes of Health and the National Science Foundation. He has been an Associate Professor in both Psychiatry and Radiology at the Harvard Medical School and at the Martinos Center for Biomedical Imaging at Massachusetts General Hospital in Boston, Massachusetts.

His research interests encompass a wide-range of domains from episodic memory and spatial cognition, to the cognitive neuroscience of major depression. Some of the recent questions his lab has addressed include: cognitive and cortical processes that underlie visual awareness, the flow of information in the cortex during visual recognition, including the mechanism and representations mediating "vague-to-crisp" processes, contextual associative processing of scene information, predictions in the brain, the processes and mechanisms mediating the formation of first impressions, the visual elements that determine our preference, and the implications of his research to clinical disorders.

Kunihiko Fukushima
Recipient of 2012 INNS Helmholtz Award

Kunihiko Fukushima received a B.Eng. degree in electronics in 1958 and a PhD degree in electrical engineering in 1966 from Kyoto University, Japan. He was a professor at Osaka University from 1989 to 1999, at the University of Electro-Communications from 1999 to 2001, at Tokyo University of Technology from 2001 to 2006, and a visiting professor at Kansai University from 2006 to 2010. Prior to his professorship, he was a Senior

Research Scientist at the NHK Science and Technical Research Laboratories. He has now part-time positions at several laboratories: Senior Research Scientist, Fuzzy Logic Systems Institute; Research Consultant, Laboratory for Neuroinformatics, RIKEN Brain Science Institute; and Research Scientist, Kansai University.

He received the Achievement Award and Excellent Paper Awards from IEICE, the Neural Networks Pioneer Award from IEEE, APNNA Outstanding Achievement Award, Excellent Paper Award from JNNS, and so on. He was the founding President of JNNS (the Japanese Neural Network Society) and was a founding member on the Board of Governors of INNS. He is a former President of APNNA (the Asia-Pacific Neural Network Assembly).

He is one of the pioneers in the field of neural networks and has been engaged in modeling neural networks of the brain since 1965. His special interests lie in modeling neural networks of the higher brain functions, especially the mechanism of the visual system. In 1979, he invented an artificial neural network, "Neocognitron", which has a hierarchical multilayered architecture and acquires the ability to recognize visual patterns through learning. The extension of the neocognitron is still continuing. By the introduction of top-down connections and new learning methods, various kinds of neural networks have been developed. When two or more patterns are presented simultaneously, the "Selective Attention Model" can segment and recognize individual patterns in tern by switching its attention. Even if a pattern is partially occluded by other objects, we human beings can often recognize the occluded pattern. An extended neocognitron can now have such human-like ability and can, not only recognize occluded patterns, but also restore them by completing occluded contours. He also developed neural network models for extracting visual motion and optic flow, for extracting symmetry axis, and many others. He is recently interested in new learning rules for neural networks. He proposed "winner-kill-loser" rule for competitive learning. By the use of this rule, a high recognition rate can be obtained with a smaller scale of the network. He also proposed and is trying to extend "interpolating vectors", which is a learning rule suited for training the highest stage of the neocognitron.

Nikola Kasabov
Recipient of 2012 INNS Gabor Award

Nikola Kasabov, FIEEE, FRSNZ is the Director of the Knowledge Engineering and Discovery Research Institute (KEDRI), Auckland. He holds a Chair of Knowledge Engineering at the School of Computing and Mathematical Sciences at Auckland University of Technology. Currently he is an EU FP7 Marie Curie Visiting Professor at the Institute of Neuroinformatics, ETH and University of Zurich.

Kasabov is a Past President of the International Neural Network Society (INNS) and also of the Asia Pacific Neural Network Assembly (APNNA). He is a member of several technical committees of IEEE Computational Intelligence Society and a Distinguished Lecturer of the IEEE CIS. He has served as Associate Editor of Neural Networks, IEEE TrNN, IEEE TrFS, Information Science, J. Theoretical and Computational Nanosciences, Applied Soft Computing and other journals.

Kasabov holds MSc and PhD from the Technical University of Sofia, Bulgaria. His main research interests are in the areas of neural networks, intelligent information systems, soft computing, bioinformatics, neuroinformatics. He has published more than 450 publications that include 15 books, 130 journal papers, 60 book chapters, 28 patents and numerous conference papers. He has extensive academic experience at various academic and research organisations in Europe and Asia. Prof. Kasabov has received the AUT VC Individual Research Excellence Award (2010), Bayer Science Innovation Award (2007), the APNNA Excellent Service Award (2005), RSNZ Science and Technology Medal (2001), and others. He is an Invited Guest Professor at the Shanghai Jiao Tong University (2010-2012). More information of Prof. Kasabov can be found on the KEDRI web site: http://www.kedri.info.

Sebastien Helie
Recipient of 2012 INNS Young Investigator Award

Sebastien Helie is a Researcher in the Department of Psychological & Brain Sciences at the University of California, Santa Barbara. Prior to filling this position, he was a postdoctoral fellow and adjunct professor in the Cognitive Science Department at the Rensselaer Polytechnic Institute (2006-2008). Dr. Helie completed a Ph.D. in cognitive computer science at the Universite du Quebec A Montreal, and graduate (M.Sc.) and undergraduate (B.Sc.) degrees in psychology at the Universite de Montreal.

His research interests are related to neuroscience and psychological modeling in general and more precisely to computational cognitive neuroscience, cognitive neuroscience, categorization, automaticity, rule learning, sequence learning, skill acquisition, and creative problem solving. Dr. Helie has published 15 articles in peer-reviewed journals, 17 articles in peer-reviewed conference proceedings, and 2 book chapters. He regularly serves on the program committee of the Annual Conference of the Cognitive Science Society and the International Joint Conference in Neural Networks. Dr. Helie has also chaired many tutorials on the CLARION cognitive architecture presented at various international conferences.

Roman Ilin
Recipient of 2012 INNS Young Investigator Award

Roman Ilin is a research scientist at the Air Force Research Laboratory, Wright Patterson Air Force Base, OH. He received his doctorate degree in Computer Science from the University of Memphis, Memphis, TN, in 2008. His graduate research focused on several areas including computational neurodynamics, where he investigated population level neural models, approximate dynamic programming, and text document clustering.

Before joining AFRL, he worked as an NRC research associate at AFRL, Hanscom Air Force Base, MA. His postdoctoral work focused on developing cognitive dynamic logic based algorithms for target detection, tracking, and situation learning. He authored 19 publications. His current research interests include cognitive algorithms for automatic situation assessment, target tracking and characterization, multi-sensor data fusion, optimal control, reinforcement learning, and neural networks.

INNS AWARD ACCEPTANCE STATEMENT:
Moshe Bar - INNS Hebb Awardee

It is an honor to receive this year's INNS Hebb award. The Hebb Award recognizes achievement in biological learning. Over the past several decades, there has been increasing recognition that the brain is not simply a passive information-processing device that operates on incoming stimuli and generates output. Rather, the brain is proactive, constantly generating predictions about what to expect which guide perception and other cognitive processes. The ability of the brain to generate these predictions is dependant on experience; we encode statistical regularities in the world over time. Thus, learning engenders predictions, and predictions are in turn a fundamental operating principle in the brain. Work in my lab has focused on the kinds of predictions that the brain makes during visual perception and how these processes are instantiated in the brain. A large portion of my research has so far focused on two specific predictive systems:

I. *A cortical context network supports the production of contextual predictions*

Contextual learning is a rich source of information for the predictive brain. When entering an office, for example, many previous encounters with offices (or images of offices) have taught us to expect computers, desks, and fax machines. In the 1970s, separate studies by Ivring Biederman and Steven Palmer showed that subjects identify target objects faster and more accurately when the target is located in an appropriate context rather than in an inappropriate one. These results suggest that contextual information may sensitize the representations of associated objects, facilitating their recognition.

The brain areas underlying this facilitation, however, have only recently been revealed. Work from my lab has shown that images of objects strongly associated with a particular context (e.g. golf cart, roulette wheel) elicit greater activity in a network of cortical regions than do objects weakly associated with any particular context (e.g., a pen). These regions are: parahippocampal cortex (PHC), retrosplenial complex (RSC), and ventromedial prefrontal cortex (vmPFC). Furthermore, temporally sensitive magneto-encephalography techniques have shown that the regions within this network begin to synchronize as early as 150 ms post-stimulus. This early synchronization suggests that contextual information is indeed activated early enough to facilitate recognition. Futher studies will clarify how these regions encode contextual regularities over time.

II. *Global information based on low special frequencies facilitates predictions*

Even with the aid of contextual learning, vision is a remarkable feat. Consider a driver who turns a corner to face a deer standing in the road. Within seconds, the driver's brain transforms the electrical signals leaving the retina into a 3D representation of the 'object' in the middle of the road, matches this percept with a representation in memory identifying this object as a 'deer', and computes the necessary motor movements necessary to avoid collision.

I have proposed that the efficiency of vision is due, in part, to a cortical mechanism that makes use of learned global object properties to facilitate recognition. In brief, a low special frequency (LSF) representation of an image (essentially a blurred version of the input) is projected rapidly from early visual cortex to orbitofrontal cortex (OFC) via the dorsal magnocellular pathway. This coarse representation, despite lacking fine detail, conveys sufficient gross detail to activate a number of candidate objects based on learning. For example, from experience it is clear that a thin cylinder might be a pen or a laser pointer, but not a computer mouse. These predictions may then constrain the slower, more detail oriented processes taking place in the ventral visual stream. Studies in my lab have accrued significant support for this model, including but not limited to the findings that 1) LSF information preferentially activates OFC 2) LSF stimuli elicit synchrony across early visual cortex, OFC, and ventral areas, and 3) orbitofrontal activity elicited by object stimuli designed to excite magnocellular cells predicts faster recognition.

I believe that the strength and efficiency of these two anticipatory systems suggests that predictions may be a universal principal in the operation of the brain. Future research exploring predictive processes outside the realm of vision holds great promise.

INNS AWARD ACCEPTANCE STATEMENT: Kunihiko Fukushima - INNS Helmholtz Awardee

It is a great pleasure and honor to receive the prestigious Helmholtz Award, which recognizes achievement in sensation/perception.

I have been working modeling neural networks since around 1965. At that time, I was working for NHK (Japan Broadcasting Corporation), and I joined the Broadcasting Science Research Laboratories, which was newly established in NHK. In the laboratory, there were groups of engineers, neurophysiologists and psychologist, working together to discover the mechanism of the visual and auditory systems of the brain. I was fascinated by the neurophysiological findings on the visual systems, such as the ones by Hubel and Wiesel, and started constructing neural network models of the visual system.

Since then, I have been working on modeling neural networks for higher brain functions. In 1975, I proposed a multi-layered network, "cognitron". The cognitron has a function of self-organization and acquires an ability to recognize patterns through learning. In 1979, the cognitron was extended to have a function of recognizing shifted and deformed visual patterns robustly. The new model was named "neocognitron".

After that, the idea of the neocognitron has been extended in various directions. By introducing top-down signal paths, I proposed a model that has a function of selective attention. The model focuses its attention to one of the objects in the visual field, and recognizes it by segmenting it from other objects. As applications of the model, various systems have been developed: such as, a network extracting a face and its parts from a complicated visual scene, a system recognizing connected characters in English words, a model for the mechanism of binding form and motion, and many others.

I also proposed neural networks extracting symmetry axis, extracting optic flow, recognizing and restoring partly occluded patterns, extracting binocular parallax, associative memory for spatio-temporal patterns, and so on.

One of my recent interests resides in developing new learning algorithms for multi-layered neural networks. Using new learning methods, I am trying to improve the recognition rate of the neocognitron, and simplifying the designing process of the network.

Many scientists and engineers are now working for modeling neural networks. The ability of neural network models is increasing rapidly but is still far from that of the human brain. It is my dream that neural networks for higher brain functions be modeled from various aspects, and that systems much more like human brain be developed.

INNS AWARD ACCEPTANCE STATEMENT: Nikola Kasabov - INNS Gabor Awardee

It is my great honor to receive the top INNS Gabor Award for Engineering Applications of Neural Networks. I consider my contribution mainly in two directions: (1) the development of both generic and applied methods and systems that lead to a better quality of information processing and knowledge discovery across application areas; (2) dissemination of knowledge.

The distinctive feature of my research is the integration of principles of information processing inspired by nature. In the late 1970s I introduced methods for the design of novel parallel computational architectures utilizing algebraic theory of transformational groups and semi-groups. Later I introduced a hybrid connections production rule-based model and developed connectionist-based expert systems.

However, my major contribution began when I integrated connectionist and fuzzy logic principles into efficient neuro-fuzzy knowledge based models. The monograph book "Foundations of neural networks, fuzzy systems and knowledge engineering", MIT Press 1996, proposed tools and techniques along with their applications going beyond usual neuro-fuzzy models of that time.

In the late 1990s I developed and published several neuro-fuzzy self-adapting, evolving models, such as Evolving Fuzzy Neural Network (EFuNN, 2001) and DENFIS (2002). These methods provide a remarkable additional value in effectiveness and efficiency. The main methods of evolving connectionist systems (ECOS) and their applications were published in a monograph book "Evolving connectionist systems", Springer 2003 (second edition, 2007).

I also developed a series of novel methods for transductive reasoning for personalised modelling that created new opportunities for the application of computational intelligence to personalized medicine.
Recently I proposed novel evolving spiking neural networks (eSNN) with applications for spatio- and spectro temporal pattern recognition, multimodal audiovisual information processing, taste recognition (Proc. ICONIP 2007-2011; IEEE WCCI, 2010; Neural Networks 2010). The proposed computational neuro-genetic models for brain data modeling and engineering applications integrated for the first time principles from genetics and neuronal activities into more efficient spiking neural models (another monograph book by Springer, 2007 and a paper in IEEE TAMD 2011). A method to integrate a brain gene ontology system with evolving connectionist systems to enhance the brain knowledge discovery was also proposed.

Integrative connectionist-, genetic- and quantum-inspired methods of computational intelligence is also a topic of my interest and research. In 2009 a quantum inspired evolutionary computation method is proposed and proved that it belongs to the class of probability estimation of distribution algorithms (IEEE Transactions of Evolutionary Computation, 2009). This early stage work suggests a way for the integration of neuronal- and quantum principles with the probability theory for the development of principally new algorithms and machines, exponentially faster and more accurate than the traditional ones.

I have developed some practical engineering applications with the use of the introduced generic methods, to mention only some of them: neuro-fuzzy methods and systems for speech and image analysis; integrated methods for time series prediction; personalised medical decision support systems; connectionist-based models for bioinformatics gene and protein data analysis; methods for cancer drug target discovery; ecological data modeling; neuroinformatics and brain data analysis.

INNS AWARD ACCEPTANCE STATEMENT: Roman Ilin - INNS Young Investigator Awardee

I would like to thank the INNS Awards Committee for this award. I wanted to use this opportunity and to thank Dr. Robert Kozma, who served as my Ph.D. adviser at the University of Memphis and Dr. Leonid Perlovsky, who was my postdoctoral adviser at the Air Force Research Laboratory and is my current research collaborator, for all the valuable guidance and support that I received from them over the years of my graduate and postgraduate studies. I would also like to thank my wife, Victoria for giving me inspiration, support and encouragement over the past 13 years.

My recent and current research interests can be divided in three categories.

(1) *Computational Neurodynamics*

I have been investigating, through computational modeling, the properties of K-sets named after Dr. Aharon Katchalsky, and utilized in the chaotic brain theory advocated by Dr. Walter J. Freeman. K sets is a hierarchical family of models describing a population of about 10,000 cortical neurons at the lowest level and the whole brain at the top level. On the higher hierarchical level these models perform pattern classification tasks and multi-sensory processing and thus can be applied to adaptive control problem. The K sets are described by nonlinearly coupled second order differential equations. The system contains hundreds of independent parameters which affect its dynamic properties. I conducted analytical and numerical studies to identify structural stability regions of the K sets and various bifurcation types occurring in the borders between stability regions. Such studies contribute to the understanding of mechanisms used to generate the aperiodic background activity of the brain. The results have been used to design a simple K model with chaotic switching between several attractor regions.

(2) *Approximate Dynamic Programming*

This study concerned the Model-Action-Critic networks advocated by Dr. Paul Werbos, and used for control of autonomous agents. The Critic network is the most challenging part of the control as it has to approximate the long term utility function, which is the solution of the Bellman equation. The Bellman equation gives the exact solution to the problem however its computational complexity grows exponentially with the number of possible states of the agent/environment system, and it is known that the brain does not solve equations. The brain approximates solutions of the Bellman equation and this is what an artificial neural network is challenged to do. Studies have been done to solve a generalized 2D maze navigation problem using a new type of neural network, called Cellular Simultaneous Recurrent Network (CSRN). This task could not be solved by a feed forward network in the previous studies. Due to the size of the CSRN network and its recurrent nature the standard back propagation learning is inefficient. The more efficient Extended Kalman Filtering (EKF) has been applied to the network to speed up the learning by the order of amplitude. The results show that this network is capable of learning the long term utility function for the case of 2D navigation.

(3) *Dynamic Logic*

Dynamic Logic is a cognitively inspired mathematics framework based on current understanding of how the bran processes information in efficient way. Its main feature is the process of transition from vague to crisp data associations and parameter values. This algorithm has been successfully applied in the context of multiple targets tracking with radar sensors. I conducted successful studies to extend this method to tracking with optical sensors, and to the task of situation recognition. My current research continues along the lines of investigating the use of Dynamic Logic and other computational intelligence techniques for solving challenging real world problems in the areas of tracking and data fusion.

INNS AWARD ACCEPTANCE STATEMENT: Sebastien Helie - INNS Young Investigator Awardee

It is with great pride and honor that I accept the 2012 INNS Young Investigator Award. I would like to thank Prof. Denis Cousineau for introducing me to fundamental research, Prof. Robert Proulx for introducing me to neural network research, Prof. Ron Sun for introducing me to the INNS, and Prof. F. Gregory Ashby for furthering my interest in computational cognitive neuroscience.

My research over the last five years has mainly focused on the interaction between explicit (e.g., rule-based) and implicit (e.g., procedural) processing in psychological tasks using artificial neural networks. This research led to contributions in three different research areas: (1) perceptual categorization, (2) sequence learning, and (3) creative problem solving.

(1) *Categorization*

My research on the cognitive neuroscience of categorization involves both empirical research (e.g., behavioral, fMRI, genetics) and neural network modeling. My modeling research has focused on the role of dopamine in early and late perceptual categorization performance as a function of category structure. This research led to the development of neural network models of positive affect, Parkinson's disease, and rule-based automaticity.

(2) *Sequence learning*

My research on sequence learning is focused on the development of explicit knowledge during procedural learning, as well as the development of automaticity in sequence learning. This research has produced a neural network model that has been used to model the emergence of explicit knowledge from procedural processing and

2012 INNS New Members at the Board of Governors

By Ron Sun
President of INNS

It is my pleasure to announce the result of the recent INNS Board of Governors election.

Those elected to the **BoG for the 2012-2014 term** are:

Soo-Young Lee
Asim Roy
Jacek Zurada
Kumar Vengayamoorthy
Peter Erdi
Hava Siegelmann

Also, **Danil Prokhorov** has been confirmed as **President-elect for 2012**.

Please join me in congratulating them. Thank you all for participating in the vote.

My thanks go to the nomination committee, especially Carlo Francesco Morabito. ■

another, more biologically-detailed, neurocomputational model of automatic motor sequence processing.

(3) *Creative problem solving*

My research on creative problem solving has focused on the development of an integrative framework called the Explicit-Implicit Interaction (EII) theory of creative problem solving. Implicit processing, referred to as 'incubation', plays a key role in the theory. EII is one of the first psychological theory of creative problem solving to be formulated with sufficient precision to allow for a neural network implementation based on the CLARION cognitive architecture.

Photo at the Board of Governors meeting on Nov. 7, 2011

Call for Papers

IJCNN2012

2012 International Joint Conference on Neural Networks

June 10-15, 2012, Brisbane, Australia

Call for Abstracts: Neuroscience & Neurocognition

Following the successful experience of IJCNN11, abstract submissions are invited for a special **Neuroscience and Neurocognition Track** at IJCNN 2012. Abstracts must focus on areas broadly related to neurobiology, cognitive science and systems biology, including - but not limited to - the following:

- Theory & models of biological neural networks.
- Computational neuroscience.
- Computational models of perception, cognition and behavior.
- Models of learning and memory in the brain.
- Brain-machine interfaces and neural prostheses.
- Brain-inspired cognitive models.
- Neuroinformatics.
- Neuroevolution and development.
- Models of neurological diseases and treatments.
- Systems and computational biology

Recognizing that some of the most exciting current research in neural networks is being done by researchers in neuroscience, psychology, cognitive science, and systems biology, the abstracts program seeks participation from the broader community of scientists in these areas by offering an accessible forum for the interdisciplinary exchange of ideas. It will also provide researchers - especially doctoral students and postdocs - with an opportunity to showcase ongoing research in advance of its publication in journals.

Abstracts must be no longer than **500 words** plus as many as 4 bibliographic citations. No figures or tables can be included. Abstracts should be submitted through the IJCNN 2012 online submission system**.**

Unlike full papers, abstracts will receive only limited review to ensure their appropriateness for IJCNN and consistency with the focus areas of the abstracts program. Authors of accepted abstracts will be guaranteed a poster presentation at the conference after regular registration. Abstracts will not be included in the conference proceedings, but will be published in the IJCNN 2012 program (including the printed conference book) and on-line at the IJCNN 2012 web site along with abstracts of all presentations.

Important Due Dates

Abstract Submission :	**March 15, 2012**
Decision Notification :	**March 20, 2012**
Final Submission :	**April 2, 2012**

For IJCNN inquiries please contact Conference Chair:
Cesare Alippi at cesare.alippi@polimi.it

WIRN 2012

22nd Italian Workshop on Neural Networks

May 17-19, Vietri sul Mare, Salerno, Italy

The Italian Workshop on Neural Networks (WIRN) is the annual conference of the Italian Society of Neural Networks (SIREN). The conference is organized, since 1989, in co-operation with the International Institute for Advanced Scientific Studies (IIASS) of Vietri S/M (Italy), and is a traditional event devoted to the discussion of novelties and innovations related to field of Artificial Neural Networks. In recent years, it also became a multidisciplinary forum on psychological and cognitive theories for modelling human behaviors. The 22nd Edition of the Italian Workshop on Neural Networks (*WIRN 2012*) will be held at the IIASS of Vietri sul Mare, near Salerno, Italy.

Call for Papers and Special Session Proposals

Prospective authors are invited to contribute high quality papers in the topic areas listed below and proposals for special sessions. Special sessions aim to bring together researchers in special focused topics. Each special session should include at least 3 contributing papers. A proposal for a special session should include a summary statement (1 page long) describing the motivation and relevance of the proposed special session, together with the article titles and author names of the papers that will be included in the track. Contributions should be high quality, original and not published elsewhere or submitted for publication during the review period. Please visit the web site for further details of the required paper format. Papers will be reviewed by the Program Committee, and may be accepted for oral or poster presentation. All contributions will be published in a proceeding volume by IOS Press. Authors will be limited to one paper per registration. The submission of the manuscripts should be done through the following website (page limit: 8 pages):
https://www.easychair.org/conferences/?conf=wirn2012

Topic Areas
Suggested topics for the conference include, but are not limited to, the following research and application areas:

General Topics of Interest about Computational Intelligence: Neural Networks, Fuzzy Systems, Evolutionary Computation and Swarm Intelligence, Support Vector Machines, Complex Networks, Bayesian and Kernel Networks, Consciousness and Models of Emotion Cognitive and Psychological Models of Human Behavior

Algorithms & Architectures: ICA and BSS, Opportunist Networks, Metabolic Networks, Bio-inspired Neural Networks, Wavelet Neural Networks, Intelligent Algorithms for Signal (Speech, Faces, Gestures, Gaze, etc) Processing and Recognition, and others

Implementations: Hardware Implementations and Embedded Systems, Neuromorphic Circuits and Hardware, Spike-based VLSI NNs, Intelligent Interactive Dialogue Systems, Embodied Conversational Agents, and others

Applications: Finance and Economics, Neuroinformatics and Bioinformatics, Brain-Computer Interface and Systems, Data Fusion, Time Series Modelling and Prediction, Intelligent Infrastructure and Transportation Systems, Sensors and Network of Sensors, Process Monitoring and Diagnosis, Intelligent and Adaptive Systems for Human-Machine Interaction, and others.

CALL FOR PROF. EDUARDO R. CAIANIELLO Ph.D. THESIS PRIZE

During the Workshop the "*Premio E.R. Caianiello*" will be assigned to the best Italian Ph.D. thesis in the area of Neural Networks and related fields. The prize consists in a diploma and a 800,00 € check. Interested applicants must send their CV and thesis in *pdf* format to "Premio Caianiello-WIRN 2012" c/o IIASS before April 20, 2012 to the addresses (wirn2012@associazionesiren.org, iiass.segreteria@tin.it). To participate, the Ph.D. degree has to be obtained after January 1, 2009 and before March 31, 2012. A candidate can submit his/her Ph.D. thesis to the prize at most twice. Only SIREN members are admitted (subscription forms can be downloaded from the SIREN web site). For more information, contact the conference Secretary at I.I.A.S.S. "E. R. Caianiello", Via G. Pellegrino, 19, 84019 Vietri Sul Mare (SA), ITALY.

Important Dates

Special Session/Workshop proposals: February 19, 2012
Paper Submission: March 25, 2012
Notification of acceptance: April 29, 2012
Camera-ready copy: on site, May 17, 2012

Chair:
Francesco Carlo Morabito

Co-Chair:
Simone Bassis

STEERING COMMITTEE SIREN
Bruno Apolloni, *Universita di Milano*
Simone Bassis, *Universita di Milano*
Anna Esposito, *Seconda Universita di Napoli*
Francesco Masulli, *Universita di Genova*
Francesco Carlo Morabito, *Universita Mediterranea di Reggio Calabria*
Francesco Palmieri, *Seconda Universita di Napoli*
Eros Pasero, *Politecnico di Torino*
Stefano Squartini, *Universita Politecnica delle Marche*
Roberto Tagliaferri, *Universita di Salerno*
Aurelio Uncini, *Universita di Roma "La Sapienza"*
Salvatore Vitabile, *Universita di Palermo*

INTERNATIONAL ADVISORY COMMITTEE SIREN
Metin Akay, *Arizona State University*
Pierre Baldi, *University of California in Irvine*
Piero P. Bonissone, *Computing and Decision Sciences*
Leon O. Chua, *University of California at Berkeley*
Jaime Gil-Lafuente, *University of Barcelona, Spain*
Giacomo Indiveri, *Institute of Neuroinformatics, Zurich*
Nikola Kasabov, *Auckland University of Technology, New Zealand*
Vera Kurkova, *Academy of Sciences, Czech Republic*
Shoji Makino, *NTT Communication Science Laboratories*
Dominic Palmer-Brown, *London Metropolitan University*
Witold Pedrycz, *University of Alberta, Canada*
Harold H. Szu, *Army Night Vision Electronic Sensing Directory*
Jose Principe, *University of Florida at Gainesville, USA*
John G. Taylor, *King's College London*
Alessandro Villa, *Universita Joseph Fourier, Grenoble 1, France*
Fredric M. Ham, *Florida Institute of Technology*

Papers submitted could be also sent by electronic mail to the address: wirn2012@associazionesiren.org