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
Editor-in-Chief
Soo-Young Lee, Korea
sylee@kaist.ac.kr

Associate Editors
Wlodek Duch, Poland
wduch@is.umk.pl
Marco Gori, Italy
marco@di.unisi.it
Nik Kasabov, New Zealand
nkasabov@aut.ac.nz
Irwin King, China
irwinking@gmail.com
Robert Kozma, US
rkozma@memphis.edu
Minho Lee, Korea
mholee@gmail.com
Francesco Carlo Morabito, Italy
morabito@unirc.it
Klaus Obermayer, Germany
oby@cs.tu-berlin.de
Ron Sun, US
dr.ron.sun@gmail.com
Regular Papers

5  NI Can Do Better Compressive Sensing
   by Harold Szu

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

38  Biologically Motivated Selective Attention Model
    by Minho Lee

50  CASA: Biologically Inspired approaches for Auditory Scene Analysis
    by Azam Rabiee, Saeed Setayeshi, and Soo-Young Lee

Columns

4  Message from the Vice President for Membership

News

59  2012 INNS Awards

64  2012 INNS New Members at the Board of Governors

Call for Papers

65  Call for Abstracts on Neuroscience & Neurocognition, IJCNN2012, 2012 International Joint Conference on Neural Networks, Brisbane, Australia

65  WIRN 2012, 22nd Italian Workshop on Neural Networks, Salerno, Italy
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 addition 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 a new exciting 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
suzuharoldh@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 a new enhanced nighttime 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 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 other people {software: generating video Cliff Notes; hardware: designing day-night spectral camera}. Each high resolution image on a smartphone has mega pixels on target (pot). Each face has a smaller pot denoted as mega pixels on target (pot). Each face has a smaller pot denoted by \( N \approx 10^4 \). Since the FD-SOC can cut equal-size facial pictures \( \{x_t \mid t = 1,2,3,4\} \) at different poses, likewise, the other person \( \{y_t \mid t = 1,2,3,4\} \), etc., if so wish, forms a database \( [A]_{N,m} = [\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{y}_1, \ldots, \tilde{z}_1, \ldots,]_{N,m} \) with \( m = 3x4 \) faces. The app CS algorithm matches an incoming face \( \tilde{y}_N \) 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 \( \tilde{x}_m \) of \( \tilde{y}_N \). The \( \tilde{x}_m \) has few ones (=yes) matched, among many mismatched zeros (=no).

\[
\tilde{y}_N = [A]_{N,m} \tilde{x}_m
\]  

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

Medical Imaging App: Emmanuel Candés 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 f 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 matrix \([\Phi]_{m,N}\) consisting of \( m \equiv k \) number of one’s (passing the radiation) among 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:

\[
\tilde{y}_m = [B]_{m,m} \tilde{x}_m,
\]  

where \( \tilde{y}_m \equiv [\Phi]_{m,N} \tilde{y}_N \) and \( [B]_{m,m} \equiv [\Phi]_{m,N} [A]_{N,m} \). Remarkably, given a set of sparse orthogonal measurements \( \tilde{y}_m \), they reproduced the original resolution medical image \( \tilde{x}_N \). CRT&D used an iterative hard

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 implementations. The experience of the International Neural Networks Society (INNS) working on traditional computational systems in the last decades developing the unique capabilities of INNS working on traditional computational systems in the experience of the International Neural Network Society: 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 \( N \approx 10^4 \). Since the FD-SOC can cut equal-size facial pictures \( \{x_t \mid t = 1,2,3,4\} \) at different poses, likewise, the other person \( \{y_t \mid t = 1,2,3,4\} \), etc., if so wish, forms a database \( [A]_{N,m} = [\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{y}_1, \ldots, \tilde{z}_1, \ldots,]_{N,m} \) with \( m = 3x4 \) faces. The app CS algorithm matches an incoming face \( \tilde{y}_N \) 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 \( \tilde{x}_m \) of \( \tilde{y}_N \). The \( \tilde{x}_m \) has few ones (=yes) matched, among many mismatched zeros (=no).

\[
\tilde{y}_N = [A]_{N,m} \tilde{x}_m
\]  

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

Medical Imaging App: Emmanuel Candés 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 f 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 matrix \([\Phi]_{m,N}\) consisting of \( m \equiv k \) number of one’s (passing the radiation) among 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:

\[
\tilde{y}_m = [B]_{m,m} \tilde{x}_m,
\]  

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

ANN supervises learning adopts the same LMS errors energy between the desired outputs $v' = \hat{y}_m & the actual weighted output $v_i = \sigma(u_i) = \sigma(\sum_{i'=1}^{N} [W_{i,i'}] u_{i'})$. Neurodynamics $\rho/i_s$ given $d\Delta u/dt = -\Delta E/d\hat{v}_i$; Lyapunov's convergence theorem $dE/dt \leq 0$ is proved for the m onotonic sigmoid intelligence $\Delta u/d\hat{v}_i \geq 0$ 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 unsophisticated utilities.

Mathematically speaking, the true sparseness measure is not the $l_1$-norm but the $l_0$-norm: $|\hat{x}|_0 = (\sum_{n=1}^{N} |x_n|^0)^{1/0} = k$, counting the number of non-zero elements because only non-zero entries raise zero power to equal 1. Nevertheless, a practical lesson from CRT&D is that the min. $|\hat{x}|_1$ subject to min. $|\hat{x} - \hat{v}|_2^2$ is sufficient to avoid the computationally intractable min. $|\hat{x}|_0$. In fact, without the constraint of minimization $l_1$ norm, LMS is blind to all possible direction co-sines within the hyper-sphere giving rise to the Penrose’s inverse $[A]^{-1} \equiv [A]^T([A][A]^T)^{-1}$ or $[A]^{-1} \equiv ([A]^T[A])^{-1}[A]^T$ (simplified by $A = QR$ decomposition). To be sure, CRT&D proved a Restricted Isometry Property (RIP) Th eorem, stating that a l bounded on a purely random sparse sampling: $\|{\hat{x}}\|_m/n \|{\hat{x}}\|_2 \approx O(1+\epsilon_0)$, $m \geq 3.1k < N$. A result, the min. $l_1$-norm is equivalent to the min. $l_0$-norm at the same random sparseness. Such an equal variant 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 C.S. 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 generalization (SG).

**Hebb Learning Rule**: If blinking traffic lights at all street intersections have b uilt-in d ata storage f or m a h e transceivers, then traffic lights function like neurons with synaptic junctions. They send a n d 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 $s$ ynapse m atrix $[W_{i,j}]$ is adj usted in proportion to b e outputs of $u_i$ through $i$-th s treet we igh ted b y t he output change, $\Delta u_i$ at the $j$-th street as the vector product code:

\[
\text{Do 10:} \quad \Delta W_{i,j} \approx \Delta u_i u_j, \quad (3a)
\]

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

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 synergetic 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 b rakes down, t he dis tributed associative memory (AM) can be retrieved. This is the FT as the nearest neighbor classifier in a finite solid angle around e ach orthogonal a xis of the subspace; t hen, t he 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 systems reports accordingly. Thus, the constraint of unsupervised learning requires a judging 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 a algorithm is t hat his n atural s top ping c riterion for the given set of input vectors $\{\mathbf{X}\} \subseteq R^N$. The BSAO projection pursuit algorithm is merely a rotation within a hyper-sphere. It stops when t he weight vector $[W_{i,j}] = [\mathbf{w}_i, \mathbf{w}_j, \ldots] \Rightarrow$ becomes parallel in time to the input vectors of any magnitude $\mathbf{w}_i/\|\mathbf{w}_i\|$. The following stopping criterion of a n unsupervised learning will be discovered thrice in Sect.2

\[
[\delta_{a,\beta} - x_a x_\beta] \mathbf{x}_a \mathbf{x}_\beta = 0, \quad (3c)
\]

\[
\Delta \mathbf{w} \equiv \mathbf{w} - \mathbf{v} = \epsilon \delta_{a,\beta} - \mathbf{w}^\top \mathbf{w} \mathbf{x}_a \frac{dK(u_i)}{du_i}, \quad (3c)
\]

where $K$ is a reasonable co-nattraction function for the source separation of $t$ weighted input $u_i = [W_{i,j} \mathbf{x}_j]$. The contrast f u nction c ould be (i) the m aximum a-posteriori entropy (filtered output entropy) used in Bell & S enowski algorithm in 1996; (ii) the fixed point a logarithm of 4-th order c umulant Kurtosis (Fast ICA) adopted in Hyvarinen & Oja in 1997; (iii) the isothermal equilibrium of minimum thermodynamic Helmholtz free energy (Brain $T_{\text{Brain}} = 37^\circ$ C) known as a s L a grange C onstraint N eural Net. In this terms of $\min. H = E - T_{\text{Brain}}$ (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 t he 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 f o r c reactive emotional s ide e-Brain i s 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 $[\Phi]_{ls,m,n}$. 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 $t$ he e nd of $S c t. 7$. In intuitively speaking, we do not change the number of ones within $[\Phi]_{ls,m,n}$, on ly $we$ e ndow a feature learning to $t$ he ones’ locations, beyond the original mappingless ill-pass filtering. In other words, we found that the admissible ANN storage demands the orthogonal sparseness of moments of spotting dramatic orthogonal changes of salient features. Consequently, we will not alter the ev allue of unk nown image v ector $\mathbf{X}$ more than $\delta_k$. In fact, for or al-time video, we have bypassed random C S coding a nd image recovery algorithms, and chose instantaneous retrieval by y MPD Retro-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-disciplinary governors. Notably, Stephen Grossberg, Teuvo Kohonen, Shun-ichi Amari serve as editors in chief of INNS, which as published quarterly by Pergamon Press a nd subsequently, monthly by Elsevier Publisher.

In the last two decades, both I NNS a nd Co mputational Intelligence S ociety of I EEE have av e omplished a lot of recorded in the IJCNN proceedings. Being the f ounding Secretary and Treasurer, a f former President and d owon Secretary 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 approaches using the lookup table, the curse is only a t the ‘static’ or c losed loop table. A iso, ‘this limitation of f supervised learning is not due to the connectionist concept, rather, due t o t he de eply entrenched “near e quilibrium” concept’; Norbert Wiener developed ne ar equilibrium Cybernetics in 1948 & 1961. ‘What’s missing is the ability to c reate a new class far a way from equilibrium.’ ‘INNS took the out of the box, interdisciplinary approach to learn from t he Ne urosiences h ow t o de velop unsupervised learning paradigm f rom Neurobiology.’

This is an important leg of NI tripod. The other two legs are the fault tolerance and d the usp space g eneralization. I us ed ed 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 a 45° solid a ngle 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 $m-D$, any within the turf family is recognized as the original axis. This is the reason that the read o ut is f ault tolerant. Th us, $[\Phi]$ m atrix storage can enjoy a s a soft failure in a graceful degradation fashion, if and only if (iff) all storage state vectors are mutually orthogonal within t he subspace; and goi ng completely out side the subspace in a new orthogonal direction t o a ll i s the subspace generalization (SG).

Subspace Generalization (SG): We introduce t he i nner product $B r a C k e t n o t a t i o n < B r a | K e t > = c$, in the d ual spaces $f < B r a [ a]$, and $[ket]$, while t he o ut product matrix is conveniently in the reverse order $\{w|\}$ i ntroduced 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. De fined, the Ortho-Normal (ON) basis $i s < n|n' >= \delta_{n,n'}; n,n' = 1,..m$. Then, SG is t he $T$ rac e- less ON $[\Phi]_{m,m} = \sum_{n=1}^{m}|n| < n| - Tr[AM]_{m,m}$. Trace operator $Tr$: summing all diagonal elements is the projection operator defined $Tr^2 = Tr$.

SG Theorem: Without s upervision, a t raceless matrix storage of ON sub-space can self-determine admitting $|x > = m + 1 >, i f f m < |m| n > = \delta_{m+1,n}$ satisfying the fixed point of cycle 2 rule: $[\Phi]_{m,m} = \sum_{n=1}^{m}|n| < n| - Tr[AM]_{m+1,m+1}$. Q.E.D.

AM is MPD computing, more than t he nearest ne ighbor Fisher c lassifier. These FT & SG a re trademarks of connectionist, which w ill b e o r b asi s a f CS a n d p roach. Unsupervised learning is a dynamic trademark of NI. New learning c apability o f m ean R-Brain combination. This FT & SG are trademarks of connectionist, which will be or basis of f CS a pproach. Unsupervised learning is a dynamic trademark of NI. New learning capability of mean R-Brain combination.
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:

\[ \min E = \langle (u(m) - y)^2 \rangle \]

where the scalar input \( u(m) = \overline{w m^T} \overline{x_m} \equiv \overline{x_m} \) has weighted average of the past \( m \) data vector

\[ \overline{x_m} = (x_0, x_{-1}, x_{-2}, \ldots, x_{-(m-1)})^T \]

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

A simple near equilibrium a nalysis is derived s oltion i s as follows

\[ \frac{\partial E}{\partial w} = 2 \langle (\overline{w m^T} \overline{x_m} - x_{m+1}) \overline{x_m} \rangle = 0, \quad \text{e.g.,} \quad m = 3 \]

\[ \begin{bmatrix} c_0 & c_1 & c_2 \\ c_1 & c_0 & c_2 \\ c_2 & c_1 & c_0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \]

\[ c_{t-t'} = \langle x_t x_{t'} \rangle; \quad c_1 = \langle x_{t-1} x_{t-1} \rangle; \quad c_2 = \langle x_t x_{t-2} \rangle; \ldots \]

Solving the Toeplitz 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, \( \tilde{X} = [A] \tilde{S} + \tilde{N} \) where additive noisy images become a vector \( \tilde{X} \) represented by a lexicographic row-by-row order over 2-D space \( \tilde{X} \). Wiener images filter is derived using AR fixed point (f.p.) algorithm in the Fourier transform domain:

\[ \tilde{X}(\tilde{k}) = \frac{1}{2\pi} \frac{1}{\tilde{S}^* + \tilde{N}} \exp(\tilde{j} \tilde{k} \cdot \tilde{x}) \tilde{X}(\tilde{x}) \tilde{j} = \sqrt{\frac{1}{\pi}} \]

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

\[ \tilde{X} = \tilde{A} \tilde{S} + \tilde{N} \]

Wiener sought \( \tilde{S} = \tilde{W} \tilde{X} \) to minimize the LMS errors

\[ E = \langle (\tilde{W} \tilde{X} - \tilde{S}^*)^2 \rangle \]

\[ \therefore f.p. \quad \frac{\partial E}{\partial \tilde{W}^*} = 2 \langle \tilde{X}^* (\tilde{W} \tilde{X} - \tilde{S}) \rangle = 0; \]

Interesting is the termination: \( \frac{\text{data}}{\text{error}} \rightarrow 0; \tilde{S}^* \rightarrow \tilde{S} \)

\[ \tilde{W} = \langle \tilde{X}^* \tilde{S}^* \rangle^* \]

where noise to signal ratio is \( \varepsilon \equiv \langle \tilde{N}^* \tilde{N} \rangle / \langle \tilde{A}^2 \tilde{S} \tilde{S}^* \rangle \).

Wiener filtering is the inverse filtering \( \tilde{W} = \tilde{A}^{-1} \) at strong signals, and becomes Van der Lught filtering \( \tilde{W} = \tilde{A}^{-1} \tilde{S}^* \langle \tilde{N}^2 \rangle \) for weak signals. A mini-max filtering is given by Szu (Appl. Opt. V. 24, pp.1426-1431, 1985).

Such a near equilibrium a djustment i nfluenced 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 s hort 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

McCulloch & 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 \( \frac{dv_i}{dt} = v_i(1 - v_i) = 0 \), at ‘no or yes’ limits \( v_i = 0; v_i = 1 \). Exact solution is:

\[ v_i = \sigma(u_i) \equiv [1 + \exp(-u_i)]^{-1} = \exp(u_i^2)[\exp(u_i^2) + \exp(-u_i^2)]^{-1}; \quad (4a) \]

Three interdisciplinary interpretations are given:

**Thermodynamics**, this is a two state equilibrium solution expressed in firing or not, the cnical ensemble of the brain at the equilibrium temperature \( T \), and the Boltzmann’s constant \( K_B \), as well as an arbitrary threshold \( \theta \):

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

**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 \((1,0,0,0,0,0,\ldots)\) as a sparse network representation.

**Computer Science**, an overall cooling limit, \( K_B T \rightarrow 0 \), the sigmoid logic is reduced to the binary logic used by John von Neumann for the digital computer: \( 1 \equiv \sigma_0(x \geq \theta) \equiv 0 \).

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 \( W_{ij} \propto v_i u_j \) measuring the weight matrix of inter-neuron synaptic gap junction. A weight summation of \( \tilde{x_i} \) given by **Compressive Sensing** to a potential sparse input \( \tilde{u_i} = \left| W_{i0} \right| x_0 \) Eq(4b).

2.3.3 Non-stationary dynamics is insured by Laponov control theory:

\[ \frac{d\tilde{u_i}}{dt} = -\frac{\partial E}{\partial \tilde{u_i}} \quad \text{Eq}(4c) \]

2.3.4 Non-supervised learning is based on nonconvex energy landscape: \( E \equiv H(\text{open/no exemplars}) \quad \text{Eq}(4d) \)

2.4 Energy Landscape Supervised Algorithm

Physicist John Hopfield broadened the near-equilibrium Wiener notion a nd introduced a non-convex energy landscape \( E(v_i) \) at the output \( v_i \) space to accommodate the (neurophysiologic) associative memory storage.
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:

\[
\frac{dE}{dt} = \sum_v \left( \frac{\partial E}{\partial v_i} \right) \frac{dv_i}{dt} = -\sigma'(\frac{\partial E}{\partial v_i})^2 , \text{ Q.E.D.}
\]

independent of energy landscapes, as long as a real monotonic positive logic \( dv_i/dt \equiv \sigma' \geq 0 \), in terms of (in, out) = \((u_i, v_i)\) defined by

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

Physicist E. R. Caianiello is considered a t hinking machine by and W iener’s AR. H e u sed c ausality physics principles t o generalize the instantaneous McCulloch & 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 hebbian learning rule 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 a self-learning neural network. 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 t he c onnectionist viewpoint, Shun-ichi Amari indicated in 1980 that t he binary logic a proach m ight suffer a premature locking in the corners of f hyper-cubes topology.

### 2.5. Backprop Algorithm

It took a team of scientists known as the Cambridge PDP group (Neuropsychologists D avid R umelhart, J ames McClelland, G eoffrey H inston, R. J. W illiams, Michael Jordan, T erreceph S ejnowski, Francis Cr ick, a nd g raduate students) to determine t he backprop algorithm. They improved Wiener’s LMS error \( E = \sum \left[ v_i - v_i^* \right]^2 \) with a parallel distributed processing (PDP) double decker architecture, consisting of 2 layers o f f eedback (uplink \( w_{kj} \) & d ownlink \( w_{ji}' \)) s andwiched between in 3 layers o f buns made o f neurons. The sigmoid logic: \( \sigma' \equiv \partial \sigma/\partial u_i < \infty \) is analytic, they unlocked the bipolar ‘bang-hang’ 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 turn, to the bottom layer which has exemplars inputs.

\[
\frac{\partial w_{ji}}{\partial t} \equiv \frac{\Delta w_{ji}}{\Delta t} = -\frac{\partial E}{\partial w_{ji}} \tag{5a}
\]

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

\[
\Delta w_{k,j} = -\frac{\partial E}{\partial w_{k,j}} \delta_k \equiv -\sum_n \frac{\partial E}{\partial u_n} \frac{\partial u_k}{\partial w_{k,j}} \Delta t
\]

\[
= \sum_n \delta_n \delta_{k,v'_j} v'_j \Delta t = \delta_k v'_j \Delta t, \tag{5b}
\]

Kronecker \( \delta_{n,k} \equiv \partial u_n/\partial u_k \) selects t op layer post-synaptic \( \delta_j \) (error energy slope) and hidden layer pre-synaptic \( v'_j \):

\[
\delta_j \equiv -\frac{\partial E}{\partial u_k} = -\frac{\partial E}{\partial v_k} \frac{\partial v_j}{\partial u_k} \equiv -(v_k - v'_k)\sigma'(\delta_k) . \tag{5c}
\]

Each layer’s 1/O firing rates are denoted in the a plhabetic order as (input, output) = \((u,v)\) respectively; the top, hidden, and bottom layers are labeled accordingly:

\[
\begin{align*}
(v_k, u_k) & \leftarrow (v'_j, u'_j) \leftarrow (v''_{ji}, u'')
\end{align*}
\]

where \( v_k = \sigma(u_k) \equiv \sigma(\sum_j w_{kj}v'_j); v'_j = \sigma(u'_j) \equiv \sigma(\sum_i w_{ji}'v''_{ji}).\) Hebbian rule turns out to be self similar at every layer, e.g.,

\[
\sum_k \delta'_j - \frac{\partial E}{\partial u'_j} = \sum_k \delta'_j w_{k,j} \sigma'(\delta'_k), \text{ etc. Such an self-similar chain relationship is known as backprop.}
\]

### 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 loop in real-world financial-like applications (IEEE Handbook Learning & Approx. Dyn. Prog., 2004). As if this were a ‘carrot and stick’ model controlling a system, t o be effective, these feedback controls must be applied at different parts of the system. 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 S., George L empties, Kumpati Narendra, et al. produced heuristic dynamic programming, stochastic, chaotic, multi-plants, multi-scales, etc., bio-control theories.

### 2.7 Self-Organization Map (SOM)

Teuvo Kohonen computed the batched centroid update rule sequentially:

\[
< \tilde{x}_{N+1} = < \tilde{x}_N + \frac{N+1}{N+1} \tilde{x}_{N+1} = \frac{N+1}{N+1} \tilde{x}_{N+1} + \rho(\tilde{x}_{N+1} - < \tilde{x}_N >), \tag{6}
\]

replacing the uniform update weight with adaptive learning \( \rho = 1/N+1 < 1 \). SOM has significantly contributed to database
applications with annual world-wide events, 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 TSP) by non-deterministic polynomial (NP), e.g. the Travelling Salesman Problem, Job scheduling, etc.) The possible tours are combinatorial express 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 equal distance. T-H solved this by using the powerful MPD computing capability of A NN (Cybernetics, & Sci. Am. Mag).

\[
E = \sum_{i=1}^{N} \sum_{j=1}^{N} v_{i} [W_{c,f}]_{ij} + \text{Constraints}.
\]

Their contribution is similar to DNA computing for cryptography RSA de-coding. Unfortunately, the T-H constraints of the permutation matrix \([W_{c,f}]\) are not readily translatable to all the other classes of the NP complete problems:

\[
[W_{c,f}] = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix},
\]

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

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

\[
\min |\hat{A}|_2^2 = \min |\hat{B}|_2^2 + \min |\hat{C}|_2^2, \text{ if } f \hat{B} \cdot \hat{C} = 0. \tag{7}
\]

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 M exican standoff dual-like in a Hollywood western movies.

Improved TSP with D&C Algorithm: A necessary and sufficient constraint turns out to be sparse orthogonal sampling Matrix \([Φ_{s}]\) which is equivalent to a permutation mix up of the identity matrix. If each row and column is added up to one, similar to T-H 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 an unlimited number of rank-identical black or white stones or two competing groups roaming over the same non-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 a point in time, including any of the same color stones putting down early by foresights. T hus, 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 \(\hat{X}\).

Without the need of a boundary search among cities, adding a ghost city \(\hat{X}\) finds two neighborhood cities \(\hat{Y}\) and \(\hat{Z}\) with two vector distances: \(\vec{B} = \vec{Y} - \vec{X}, \vec{C} = \vec{X} - \vec{Z}\). If \(\vec{B} \cdot \vec{C} = 0\) satisfies the D&C theorem, we accept \(\hat{X}\). 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 \(N/2=9\); on one continues the procedure without solving the remaining half until one can carry out TSP in smaller units and 4, 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 proposed a convergence theorem that short and long term traces \(Z_{ij}\) 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. Be sides the PDP 3 layers bums and 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 \(\rho = (\rho_{t+1}, <\vec{w}_{t}> = \cos(\leq \pi/4) \geq 0.7\). 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 C enter to f L earning Excellence. Notably, M. Cader, et. al. at the World Bank
implemented ART expert systems for typing seeds choice for s aving t he c ostly di agnosis ne eds by m eas of PCR amplification in or der to b ui ld up enough DNA samples (pico grams); a decision prediction system based on the past Federal R eserve open f orum r eports (Neural Network Financial Expert S ystems, G. J. Deboeck a nd M. C ader, 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 shape function is not the probability measure which must be sigmoid logic is crucial for John Hopfield’s convergence proof. If the neuron had a piecewise negative response in ANN Modeling of Fuzzy Memberships.

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 shape adjustable by the \( \lambda \)-knot (by M. Feigenbaum for the tuning of period doubling bifurcation cascade). If we represent each pixel by the sick neuron model \( v_i = \sigma_N(u_i) \), then recursively we produce the n onlinear Baker t ransform of im age m ixing. Such a chaotic N N i s 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 L ett. and I EEE 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 \( T_c = \frac{T_0}{1 + t'} \) for any initial temperature \( T_0 \) that guarantees the r eaching of the equilibrium ground state at the minimum energy. Given a sufficient low temperature \( T_0 \) Geman and Geman proved in 1984 the c onverging to the minimum energy ground state by a n inversely logarithmic time s tep: \( T_0 = \frac{T_0}{1 + log(1 + \epsilon)} \).

Sejnowski & Hinton used the Gaussian random walks in the Boltzmann’s m achine for a s imulated a mealing l earning algorithm e mutating a b aby l earning t he t alk, c alled Net-talk, or Boltzmann Machine.

Cauchy Machine: Y. Takefuji & Szu designed a n electronic implementation of s tochastic Langevin e quations. S tochastic ne urons a re c oupled through t he s ynapse AM l earning rule a nd recursively driven by Levy flights and Brown motions governed by the Cauchy pdf. The set of Cauchy-Langevin dynamics enjoys the f ast i nversely linear cooling s chedule to r each the equilibrium state. Do 10

\[
t' = t' + 1; \quad T_c(t') = \frac{T(t)}{1 + t'}
\]

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

\[
x(t') = x(t) + \Delta x; \quad E(t') = \sum_{i=1}^{n} \frac{1}{2}k(x(t') - x_i)^2;
\]

\[
\Delta E = E(t') - E(t' - 1);
\]

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

compute Metropolitan \( \exp(\Delta E/K_B T_c(t')) > \epsilon_0 \);

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 \( \theta \) and Joseph Landa. The Cauchy noise is optically generated and driven by Levy flights and Brown motions governed by the Cauchy pdf. The set of Cauchy-Langevin dynamics enjoys the fast inversely linear cooling schedule to reach the equilibrium state.

\[
\rho_c(\Delta x) = \frac{1}{\sqrt{\pi}} \exp(\frac{-\Delta x^2}{T});
\]

\[
\rho_c(\Delta x) = \frac{1}{\sqrt{\pi}} (1 + \frac{\Delta x^2}{T})^{-1} = \frac{1}{\sqrt{\pi}} (1 + \frac{\Delta x^2}{T} + \cdots).
\]

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

\[
\pi = \int d\theta = \int \frac{dx}{d\theta} = \int \frac{dx}{1 + tan \theta^2} = \frac{\pi}{2} \int d(\frac{x}{1+\frac{x^2}{T}}).
\]

Global Levy Flights \( < \Delta x^2 > _{\rho_c} = \infty \)
Local Brownian motion \( < \Delta x^2 > _{\rho_c} \equiv T(t) \)

2.12 NI Expert System

Szu and John Caulfield published an optical expert system in 1987, generalized the AI Lisp programming the pointer linkage m ap from 1-D vector arrays o f f = (A, O, V)^T to \( \tilde{f} = (A', O', V')^T \), etc. The co lor, “A attribute,” o f apple, “O o bject,” i s r ed, “V value”. We represent the Lisp m ap w ith the MPD \( [HAM] = \sum_{i=1}^{n} f_i^{T-r} \) storage which ha s demonstrated both t he F TA nd t he Generalization capabilities. This FT & SG of AM NI Expert System is a key d riving e ngine f or V ideo i mage Cliff Notes.

2.13 Unsupervised Learning of l-Brain

In order to make sure nothing but the desired independent sources c oming out of the filter, C. J utten a nd 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 . C. ardoso has systematically investigated t he blind de - convolution of f unknown impulse re sponse function; he called a matrix f orm as B lined Sources Separation (BSS) by non - Gaussian higher order st atistics (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. S ejnowski ( BS), e t a l. [10] have systematically formulated an unsupervised learning of ANN algorithm, solving both the unknown mixing w eight matrix and the unknown sources. Their solutions are subject to the constraints of m aximum f ilter e ntry H (y_i) of the output y_i = [W_{ia}]x_a , where x_j = [A_{ja}]^2, and the repeated G reek i ndices r e present t he s ummat i on. A NN model uses a robust saturation of linear filtering in terms of a n online sigmoid out put y_i = σ(χ_i) = (1 + exp(−[W_{ia}]x_a))−1. S ince 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 H abb’s synaptic weight updates. Again, a bona fide unsupervised learning does not need to prescribe the desirable outputs for exemplars input. For ICA, B S chose t o m aximize t he Shannon output e ntry H(y) i ndicating t hat he i nverse f iltering has b lindly de - convoluted and found the unknown i ndependent sources w ithout knowing the impulse response function or mixing m atrix. Thus, the filter weight adjustment is defined in the following and the BS result is derived as follows:

$$\frac{\partial H(y)}{\partial \delta t} = \frac{\partial \log f(y)}{\partial y} dy = \delta w \approx |w|^{-1} \int f(y) log f(y) dy $$

Derivation: From the normalized probability definitions:

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

$$H(y) = - \log f(y) > 0$$

we express the ou tpu t pdf in terms of the input pdf with changing J acobian variables. We exchange the o ders of operation of t he ensemble a verage brackets a nd t he derivat ives to compute:

$$\frac{\partial H(y)}{\partial w} = \frac{\partial \log f(y)}{\partial y} dy \approx |w|^{-1} a \frac{dy}{\partial w};$$

Ricati sigmoid:

$$y = [1 + \exp(-w x)]^{-1};$$

$$\frac{dy}{\partial w} = \sigma(1-y) \frac{dy}{dx} = \sigma(1-y) \frac{dy}{dx} = \sigma y(1-y).$$

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 a stable weight increasing N nodes. This non-Hebbian learning en tering through t he loga rithmic derivative of J acobian gives ving $\frac{dy}{dx}$ of T. If improve the computing speed, S. Amari et a. l. assumed identity $[\delta_{ij}] = [W_{ij}] [W_{jk}]^{-1}$ and multiplied it to the BS algorithm

$$\frac{dH}{aw_{ij}} [\delta_{ij}] = ([\delta_{ij}] - (2y - 1)y^2) [W_{ij}]^{-1},$$

which use is made of $y_i = [W_{ia}]x_a$ to change the input $x_i$ to the synaptic gap by its weighted output $y_i$. In information geometry, A m ari et al. derived the natural gradient ascent BSA algorithm:

$$\frac{dH}{aw_{ij}} [W_{ij}] = ([\delta_{ij}] - (2y - 1)y^2),$$

which is not in the direction of original $\frac{dH}{aw_{ij}}$ 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.

\[ <\vec{x}_k >^T \theta = \lambda \theta; \]

\[ w' - w = \lambda \theta (\vec{x}_k^T \theta) \approx <\vec{x}_k >^T \theta; \]

\[ \frac{d\vec{w}}{dt} = <\vec{x}_k >^T \theta \equiv \sigma(\vec{x}_k^T \theta) \equiv \frac{dK(u_i)\overline{u_i}}{du_i} \equiv k(\vec{x}_k^T \theta)\theta; \]

where Oja changed the unary logic to bipolar hyperbolic tangent logic $u_i = \sigma(u_i) = u_i - \frac{1}{2} u_i^3 \equiv \frac{dK(u_i)}{du_i}$, $u_i = w_{ia}x_a$.

It becomes similar to a Kurtosis s lop e, which suggested to Oja a new contrast f unction K. The following i s t he geometric b asis of f a s t t opping c riterion of f unsupervised learning. Taylor expansion of the normalization, Eq(8c) and set ||w||^2 = 1:

$$|\vec{w}|^{-1} = ([\vec{w} + \varepsilon \vec{k} (\vec{u}_k^T \theta)]^T [\vec{w} + \varepsilon \vec{k} (\vec{u}_k^T \theta)])^{-1}$$

$$= 1 - \varepsilon \frac{1}{2} (\vec{k} (\vec{w}_k^T \theta)^T (\vec{k} (\vec{w}_k^T \theta))^T + O(\varepsilon^2).$$

$$\vec{w}'' = \overline{\vec{w}} / |\vec{w}|^{-1}$$

$$= \left(\vec{w} + \varepsilon \vec{k} (\vec{u}_k^T \theta) \right) \left(1 - \varepsilon \frac{1}{2} (\vec{k} (\vec{w}_k^T \theta)^T (\vec{k} (\vec{w}_k^T \theta))^T + O(\varepsilon^2)\right)$$

$$\Delta \vec{w}'' = \vec{w}'' - \vec{w}'' = \epsilon \delta_{ab} - w'' a w'' b \delta_{a} \delta_{b} \frac{dK(u_i)}{du_i}$$

This kind of derivation is therefore referred to as a 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 r out oot: $\frac{dK(u_i)}{du_i} = 0$, of a specific contrast function named Kurtosis. Rather than max imizing an arbitrary contrast function, o r the BS filtered output entropy, they considered the 4th order cumulant 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, a nd $K < 0$ for sub-Gaussian, e.g. a speech amplitude L aplacian histogram that is narrower than Gaussian. Every f acts a nd o ve c uses h a ve d i fferent f ixed value o f Kurtosis t o s et t hem a part. At t he bottom of a fixed p oint, t hey s et t hes lope of Kurtosis t o zero a nd efficiently and analytically solved its cubic r oots. T his i s
called (Fast) ICA, a s c oined by P eter C omo (S ig. P roc., circa '90).

\[ \hat{x}_i = \tilde{a}(\theta_i)\alpha \hat{\omega}_i = \cos \theta_i s_1 + \sin \theta_i s_2; \quad i = 1, 2 \]

\[ \bar{u}_j = k^T(\varphi_j) \tilde{\omega}_a = \cos \varphi_1 x_1 + \sin \varphi_1 x_2; \quad j = 1, 2 \]

\[ (u_1, u_2) = \begin{bmatrix} \cos \varphi_1 - \sin \varphi_1 \cos \theta_1 \sin \theta_2 \sin \theta_1 \\ \sin \varphi_1 \cos \varphi_2 - \cos \theta_2 \sin \theta_1 \cos \theta_1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \]

\[ = \begin{bmatrix} \cos (\varphi_1 - \theta_1) & \cos (\varphi_1 - \theta_2) \\ \sin (\varphi_1 - \theta_1) & \sin (\varphi_1 - \theta_2) \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}. \]

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 rule: Iff \( \varphi_j = \theta_i \pm \frac{\pi}{2} \); then \( K(u_1) = K(s_2); K(u_2) = K(s_1) \), verifying Fast ICA \( \frac{\partial K}{\partial s_1} = 0 \). Given arbitrary unknown \( \theta_1 \), not necessarily orthogonal to each other, the killing weight \( \hat{K}(\varphi_j) \) can eliminate a mixing vector \( \tilde{a}_i(\theta_i) \).

### 2.14 Sparse ICA

New application is applying a sparse constraint of non-negative matrix factorization (NMF), which is useful for image learning of parts: eyes, noses, mouths, (D. D. Lee and H. S. Seung, Nature 401(6755):788–791, 1999); following a sparse neural code for natural images (B. A. Olshausen and D. J. Field, Nature, 381:607–609, 1996). P. Hoyer provided Matlab code to run sparse NMF \([X] = [A][S] \) (2004). \( \min_{[A], [S]} \text{min} \|[A]\|_1; \text{min} \|[S]\|_1 \) subject to \( E = \|[X] - [A][S]\|_F^2 \). The projection operator is derived from the Grand-Schmidt decomposition \( \hat{B} = \hat{B}_\perp + \hat{B}_1 \); where \( \hat{B}_\perp \equiv (\hat{B} \ast \hat{A})/\|\hat{A}\|_F^2 \), and \( \hat{B}_1 \equiv \hat{B} - \hat{B}_1 \). Alternative gradient descent solutions between unknown matrices \([A, [S]\) : new \( [Z'] = [Z] - \frac{\partial^2 E([X]= [A][S])}{\partial [Z]} = 0; \min_{[Z]} \|[Z]\|_1 \); where alternatively substituting \([Z] \) with \([A] \) or \([S]\) \). (Q. Du, I. Kopriva & H. Szu, “ICA hyperspherical emotes s ensing,” O p t E ng.V45, 2006; “Constrained Matrix Factorization for Hyperspectral,” IEEE IGARS 2005). Recently, T-W. Lee and Soo-Young Lee, et al. at KAIST have solved the source permutation challenge of ICA speech sources in the Fourrier domain by de-mixing for Office mate automation. They grouped similar Fourrier 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].

Figure 1: (a) De-mixing by killing vector (Yamakawa & Szu); (b) a sources image (Szu) (not shown Yamakawa) 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) 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, Izuka, Japan, Oct. 1, 1998).
2.14 Effortless Learning Equilibrium Algorithm

An effortless thought process that emulates how the e-Brain intuitive idea works. Such a ne 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 Bo Itzmann constant, \( T_o = 273 + 37°C = 310^o \) Kelvin). Therefore, the thermodynamic physics gives an inverse solution that must satisfy the minimum Helmholtz free energy: \[ \text{min. } E(s_i) = E - T_o S \]. 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.

Derivation: Convert discrete Boltzmann-Shannon entropy 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:

\[
\text{Min. } H = E - T_o S = 0;
\]

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:

\[
E = \text{intercept} + \text{slope } s_1 = E_o + \frac{dE}{ds_1} (s_1 - 0)
\]

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 N nonlinear L CNN based on Kuhn-Tucker augmented Lagrange methodology. (IEEE IP V.16, pp1008-1021, 2007).

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

\[
H = E - T_o S
\]

\[
S = 1 - \exp(-\frac{E_o}{K_B T_o})
\]

Figure 2: Exact LCNN pixel by pixel Solution

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 mixing). 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 a ware of the interdisciplinary contributions made by mathematicians (e.g. V. Cherkassky, Jose Principe, Lei Xu; Asim R 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, Tacee Zurada, Cheng-Yuan Liou of Taiwan U, You-Shu Wu of Tsing Hwa U, Huiseng Chi of Peking U, Toshio Fukuda, Hideo Aiso of 5th Gen Computing, et al.). The author apologized that he could not overt heirs and o thers younger contributors in this short survey.

Combining ICA a nd C S line ar a nd ic ege pr oduced the Feature Org anized Spar seness (FOS) t heorem i n 7 . Contrary to p urely r andom sparseness, the bi o-inspired t urns out to have the additional meaning, i.e., the locations indicating dramatic moment of f c changes a t s alient s patial pixel features. The measure of significance is quantified by the o rthogon al d egree among ad missible s tates of f AM s torage [12].

3.1 Human Visual Systems (HVS)

Physiologically speaking, the human visual system has a unique long visual pathway of 4 million R G c olor v i~s ion cones for high resolution spot size. 2 millions B cones are distributed in the peripheral outside the t h e c entral f ovea i n order t o r eceive t he h igh bl ue sky a nd t he low bl ue l ake water at 0.5μ wavelengths. This is understood by a simple geometrical ray inversion of our orbital lens w hen t he 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, grass, 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 lighter fructose content at a longer wavelength of about 0.7μ. These primates could feed m ore o ff springs, a nd m ore off springs inherited the same trait, who then had m ore intelligence m ay be e ndowed by (God). Ow ing to t he 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 sp arse pattern am ong housekeeping g lial (Muller) c e lls a nd B c ones i n t he peripheral of the fovea. What is the biological mechanism for organizing sparse representation? If too many of them directly a nd i nstantly s end t heir responses t o t he 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, and fired in order to be retrieved instantly by those orthogonal salient features stored in the Hippocampus.

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 (codenamed AM of e-Brain). Thus, we developed a paranoid bias toward unknown events. For example, miracles must have messages, 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, a e m ust b e given w henever rounding of f de cimals. A double-blind protocol (to t he analyst and volunteer participants) with a (negative) control is often demanded, in order to suppress the bias of False Positive Rate. “In God we trust, all the rest show data.” NIH Motto. We now know e ven 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 co uld have been a advocating a n intuitive thought about the dual therapy of hypertension drugs; that both ACE inhibitor upon a certain hormone and ARB acting in a differential 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 index in a different way on the same hormones should be thought about the dual therapy of hypertension drugs; that multiple stacks of disks, converts the hydrocarbon chain of Rhodopsin pigment molecules from the cis configuration to trans configuration. As a result, the alternatizing single and double carbon bonds of trans carbon chains are switching continuously in a domino effect until it reaches a nd 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 some semiconductor Carbon Nano-Tube (CNT) IR sensor, Xi Ning & H. Szu). To detect a single moonlight photon, we must combat against thermal fluctuations at $300^\circ K \approx \frac{1}{40} eV$. How the thermal noise is celled without cooling operated at physiology temperatures. T his is a accomplished by s y naptic ga p junctions of a single ganglion integrating 100 rods. These 100 Rod bundles can sense a single moonlight photon ($1\mu$~$1 eV$) because there exist a ‘dark light’ photon current where h e s i o n a l i g n a t e d, a s discovered b y yagi in [4]. Our eye s supplies t he eлектric current energy necessary to generate the ‘dark currents.’ It is an ion c urrent m ade of Potassium i nside t he R od a nd Sodium outside the Rod, circulating around each Rods. (i) Nature s eparates t he signal pr ocessing e nergy from t he f unctional s ignal information energy, because (ii) a single night vision photon does not have enough e nergy t o d rive t he s ignal current to the back of brain; but may be enough (iii) to depolarize the membrane potential to switch off the 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 detect the ‘Hagins dark c urrents’ a way f rom t he R od. C onsequently, it changes t he ganglion pre-synaptic junction membrane potential. A s a r esult, t he icoming photon c hanges t he membrane potential and the ganglion fires at 100Hz using different s ervoir e nergy bu dget f or reporting the information [ 4]. A s single ganglion s y naptic j union ga p integrating over these 100 rods bundle provides a larger size of the bundle t oo vercomes ( v) t he s patial uncertainty principle of a single photon wave mechanics. These (i-v) are lesson learned f rom bi osensors. Another bi osensor 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. T hereby, the peripheral night vision can achieve a graceful degradation of imaging object size. This fan-in architecture allows t he HVS to achieve scale invariance mathematically, as follows. These 1.4 millions night vision ganglion a xon f iber s a re s queezed uniformly through the fovea channel, which closely packs them uniformly toward the LGN and visual cortex in the back of the head. The densities of f Rods’ and B-cones increase f irst and d rop gently along the radial direction, in an exponential increase and decrease fashion:

\[
\text{Input locations} = \exp(\pm \text{Output uniform location}),
\]

which can therefore achieve a graceful degradation of the size variances b y m eans 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 \(= \log(\text{Input}) \equiv \text{Output}^\prime\), when Input \(= 2 \times 1\text{ input}^\prime\) because \(\log(2)\) is negligible. T his s i ze invariance allows our ancestor to run in the moonlight while chasing after a significant other to integrate the i ntensity rapidly a nd continuously over r the time w ithout computational slow down. For photon-rich day vision, the high de nsity fovea g anglion s equire 100 H z f iring rate, which m ight require a sharing o f t he c ommon p ool of resources, b efore replenishing because the molecular kinetics produces a n atural s upply d elay. A s a result, the ganglions who use up t heir e nergy n eed in hibit neighborhood g angli ons t o incr ease firing r a t e, producing the lateral inhibition o f n-center-off-surround, the s o-called H ubel a nd Wiesel oriented edge wavelet feature map \(\psi_n\). [5]

### 3.5 Division of Labor

It’s natural t o d i vide o ur l arge b rai n i nto l eft a nd r ight hemispheres c orresponding t o o ur s ymmetric bo dy l imbs reversely. Ne urophysiology s peaking, we shall divide o ur ‘learning/MPD s toring/thinking’ p rocess i nto a balanced slow and fast process. In fact, Nobel Laureate Prof. Daniel Kahneman w rote a bout the decision making by s low and fast thinking in his recent book published in 2011. W e 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 t he c omputing t ask i s complete. We w ish 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 ourselves as a self-correction mechanism. By the lateralization seems to be natural balance to build in 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. C. Rick & C. Koc in 2005, the consciousness layer is a wide & thin layer, called Claustrum, located underneath the enter brain and a bove the l outer part lizard brain. The Claustrum acts like a music conductor of brain sensory orchestra, tuning at a certain C note for all sensory i instruments (by t be winner-take-all masking effect). The existence of conscious toning remains t o be experimentally confirmed (e.g. studying an anesthesia awakening might be good i dea). I t c ould be a bove t he normal EEG brain waves types known as alpha, beta, theta, etc., and underneath the decision making neuron firing rate waves at 1 00 Hz. This par of hippocampus requires the connection mediated by t he Claustrum known as t he Lateralization. According to the equilibrium minimum of thermodynamic He lnoltz the energy, t he s ensory processing indeed happens effortlessly at the balance between m inimum e nergy and a nd m aximum entropy, we a re operating at.

The s parse o rthonal is necessary for HVS, but also natural for brain euron r epresentation. W e hav e 1 0 neurons and 100 billion synapses with some replenishment and regeneration, t he s ynapses e ould t ast over 125 years. An other reason is the parse o rthonal representation is not loaded up with a ll t he degree of freedoms and no longer has a free will for generalization. In ot her w ords, unlike a robot having a limited memory capacity and computing capability, we pri te to keep our brain degrees of freedom a s a ppossible, a bout 10 ~15% level (so-called t he enter d eveloped place o n t he Earth) about 1 0% x 10^20 2 encyclopedia Britannica. T he odd and M aros in Na ture 2004 [6] summarized the e apacity limit o f visual s hort-term memory i n human P osterior Parietal Cortex (PPC) where s parsely a range n euron population called grandmother neurons fires intensely for l second without disturbing ot hers, supporting o ur independence concept yielding our orthogonality attribute. The ‘grandmother neuron’ may be activated by other stimulus and nd memories, but is the sole representation of ‘grandmother’ for t he i di vidual. 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 f brain i maging methodologies, and confirmed the biological base of feature organized sparseness (FOS) to be based on a utomatic comparison–selection of changes. “How m any views or r frames doe s a monkey need i n order t o e ll a good zookeeper from a bad one?” Monkeys select 3 distinctive views, which w e refer to as m frames: frontal, side and a 45° view [8]. Interestingly, humans need only m = 2 views when constructing a 3-D building from architectural blueprints, or for visualizing a human head. This 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 he is aniconductor storaget echnology has become inexpensive or ‘silicon dirt cheap,’ we can apparently afford wasteful 2-D MPD A M storage for 1-D vectors. Here, we illustrate how MPD A M can replace a current digital disk drive s storage, a pigeon-hole, with storage offering recall confusion and search delays. The necessary and sufficient condition of such AM admissible states requires that rank-1 vector o uter product is orthogonal a s depicted i n Fig.4. Thus, we recapitulate the essential attributes, sparseness and orthogonality as follows.

4.1 Connectionist Storage

Given facial images $\mathbf{X}_{nt}$, 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 hippocampus as a monkey need to tell a good one? Monkeys select 3 distinctive views, which we refer to as m frames: frontal, side and a 45° view [8]. Interestingly, humans need only m = 2 views when constructing a 3-D building from architectural blueprints, or for visualizing a human head. This kind of questions, posed by Tom Poggio et al in 2003 [8], can be related to an important medical imaging application.
the big-nose Uncle state \((0, 1, 0)\) with a fault tolerance of direction \(\cos(45^\circ)\). 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 crosstalk 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 \(\sigma_o\) 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 \([\text{AM}]\). Orthogonal features are necessarily for or soft failure indicated in a 3-dimensional feature subspace of N-D.

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

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

MPD over-writing storage:

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

### 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:

\[
\sigma_o \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = 1 \quad \text{smiling uncle remains to be uncle}
\]

### 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 itself 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 the gravity \((\text{CG})\) changes of the child’s eyes - 1s seeded with the supervised as associative memory of pairs of image.

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 IR retrieval. Consequently, AM enjoys the generalization by discovering a new component of the degree of freedom, cf. Section 4.

### 5. Spatiotemporal Compressive Sensing
The software can take over tracking the local enter of the gravity \((\text{CG})\) changes of the child’s eyes - 1s seeded with the supervised as 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 change vector becomes orthogonal to the previously stored net CG vector. Then, the code will update the new CG change vector with the previous one n t h e outer pr oduct hetero-associative memory (HAM), known as a Motion Organized Sparserness (MOS), or Feature-role Organized Sparserness (FOS). Then, an optical c xpert system (Szu, Caulfield, 1987) can b e employed to fuse the interaction library (IL) matrix [HAM] (IL-HAM) in a massive p arallel d istributive (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 C enter of Gravity (CG) m ovement such as a tiger jumping out of frame luffing a branch and 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 filter (not mean filter) where select majority gray-value de legating-pixel locations a n d image w eight values according to the local gray 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. T he length of the vector is proportional to the delta change of gray values as the wind is blowing tree branches or bushes in a cyclic motion without a net CG motion. 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 the window size of a significant movement over half of the window. Then the net is above the 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 another [HAM] of an interaction library (IL-AM) for fusion of storyline subroutine (Szu & Caulfield “Optical Expert System,” App Opt. 1988) sketched as follows: A 1 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 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 B BC c opyrighted. 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] in the motion organized sparseness (MOS) as the picture index forming [AM] an d image as Hetero-AM [12]. Moreover, we have extended the concept with more 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 & s taget at s the e k eys for sur vivor(optical c xpert s ystem, 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 2-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 MD [AM] & [HAM] s torage i n 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 Φ with feature organized sparseness [Φs], where Φ is the locations of potential discovery of features.

Theorem: Feature Organized Sparseness (FOS)

We shall derive a theorem to design ICA U nsupervised Learning methodology can help to design FOS C S sampling matrix [Φs]

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

(9)

where \{ψ_n\} is the Hubel-Wiesel wavelet modeled by the
digital sub-band wavelet bases successfully applied to JPEG 2000 image compression and is a common 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 side-by-side as follows:

\[
R^n \bar{x} = \sum_{n=1}^{N} s_n \psi_n = \sum_{n=1}^{N} s_n \psi_n = [\Psi].
\]

where \( k \) non-zero wavelets are denoted \( k_{1,2,\ldots,k} \ll N \).

\[
R^m: \bar{y} = \sum_{m=1}^{M} x_m \phi_m^T = [\Phi_s] \bar{x} ;
\]

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

\[
\bar{y} = [\Phi_s][\Psi] \bar{s} \equiv [ICA] \bar{s}.
\]

**Q.E.D.**

**Adaptive Compressive Sensing:**

We can exploit the full machinery of unsupervised learning ANN community about how to solve the Beit Sours Separation (BSS). We can either follow the L agrangian Constraint Neural Net workba sed on minimizing the thermodynamic physics Helmholtz free energy by maximizing the a priori source entropy \( [9] \) or the engineering filtering concept of f m amzing 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 \( W \) hitenig i n ensemble av erage matrix \( [W_s]^T = [W_s] \equiv < [\bar{y}\bar{y}^T] >^{-\frac{1}{2}} \).

By definition \( \bar{y} \equiv [W_s] \bar{y} \) satisfying

\[
< \bar{y}\bar{y}^T > \equiv [W_s] < \bar{y}\bar{y}^T > [W_s]^T = [I]
\]

\[
\therefore [W_s] < \bar{y}\bar{y}^T > [W_s]^T = [I][W_s] = [W_s];
\]

\[
\therefore [W_s]^T [W_s] = < [\bar{y}\bar{y}^T] >^{-1}; [W_s] = < [\bar{y}\bar{y}^T] >^{-\frac{1}{2}} \text{ Q.E.D.}
\]

(ii) Orthogonal Transform: \( [W] = [W]^{-1} \)

By definition

\[
[W]\bar{y} = [W][W_s]\bar{y} = [W][W_s][\Phi_s][\Psi] \bar{s} \equiv [W][W_s][ICA] \bar{s} = \bar{s}
\]

\[
\therefore [W] < \bar{y}\bar{y}^T > [W]^T = [W][I][W]^T = < \bar{s}\bar{s}^T > \equiv [I];
\]

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

The step (i) can reduce ICA d e-mixing to o rthogonal rotation. We can compute from the desired 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, c.f. 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 independent source according to the rank-1 AM approximation of ICA mixing matrix \( [ICA] \):

\[
[ICA] = \sum s^2 \bar{s} \bar{s}^T = [\bar{s}_1, \bar{s}_2, \ldots][\bar{s}_1, \bar{s}_2, \ldots]^T. \tag{12}
\]

The correct CS linear pr ogramming could be us ed to compute a \( \ell_1 \)-norm sparse constrained source representation \( \bar{s} \) of the input \( \bar{y} \) in the LMS error sense. Our experience indicates a de sirable or thogonality post-processing. Given all independent sources, we construct the orthogonal \( 1 \)-s (by the Gram-Schmidt procedure) \( < \bar{s}\bar{s}^T > \equiv [I] \).

\[
[\Phi_s] \equiv [\bar{s}_1, \bar{s}_2, \ldots][\bar{s}_1, \bar{s}_2, \ldots]^T[\Psi]^{-1}. \tag{13}
\]

Furthermore, we refer to the orthogonal feature extraction \( [\Phi_s] \) such that \( < [\Phi_s][\Psi]^{-1} > \equiv [I] \). In doing so, we can i ncrease t he e ficiency of the hiper-spectral compressive sensing methodology he pping “finding a needle in a haystack” by sampling only the image correlated to the n edle s sources \( [\bar{s}_1, \bar{s}_2, \ldots] \) without u ncessarily creating a haystack of data cu be 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 cam era which can take both daytime and nighttime pictures with a single photon detector array. It can automatically keep and send on thy tho significant frames cap able o discovering motions and features. Our design logic is s s imple: never imaging daytime pictures with nighttime spectral, and v s versa, in a photon poor lighting or in the night do not take daytime color spectral picture. Of course, a s simple clock time will do the job; but a smarter approach is through the correlation between example images and desired features. We wish to design the camera with a row of N pixels to measure N/4 resolution with spectrum resolution. We take the spectral color image filters (for RGB colors). We trade the spatial pixel and the extra 4th one is for extra night vision at near infrared 1 micron spectrum. We further correct optical path difference at the wave Baye r filter media in order to focus a 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 formalism, Eq (6,7,8), a nd a pplying ICA unsupervised learning steps (i) & (ii) to de sign a FOS sampling matrix \( \Phi \). Finding all the independent sources vectors from input data y or migt images \( y_s \) we e collect expected s sources \( s_s \) into a \( ICA \) m ixing matrix \( \mathcal{A} \) = \( \sum y_s^2 \), then substituting its equivalence to CS sampling we can design FOS sampling matrix a s \( \Phi = \mathcal{A}^\dagger \Phi \) where \( \Phi \) is usual image w avelet basis. The hardware is mapping the sparse feature sampling matrix onto 2x2 Bayer Filters p er pixel t hat c an a fiord t o t ake t he s patial resolution with the spectral resolution in close up shots.

In his paper, we have further c xtended M otion Organized S parseness [12] with F eature Organized Sparseness compositing two main p layers as a prey and a predator, namely a rabbit and an eagle. Their interaction is discovered by their chase a fier eac h other optical flows as shown in F ig. 6 as an automatic Video Image Cliff Notes.

Instead the purely random sparseness, we have generalized CS sampling matrix \( \Phi \) with FOS sampling matrix[11]. In cl osing, we c oULD es timate t he c omplexity ef fect o f r eplacing purely r andom s parseness \( \Phi \) with F OS \( \Phi \) upon t he C RT&D R IP t heorem. We c ould a pply the complexity analysis tool called Permutation Entropy [14]. PE c omputes c omputes i n a moving window of t he s ize L=2,3, etc., c ounting t he up-down shape feature of one s over t he z eros: \( H(L) = -p(\pi) \sum p(\pi) \) of the k-organized sparseness to set a bound on the sampling effect from purely random ones. For e xample, a n organized sampling mask \( \Phi_{m,N} \) 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, \{0,0,0\} flat, etc.}. They had shown H(L) to be bounded from organized structure with one s locations (degree of c omplexity) t o purely randomness (zero complexity) as \( 0 \leq H(L) \leq \log L - L \log L - L \), by Sterling f ormula where \( L \ll N \). 0 \( \leq PE(\Phi_{m,N}) \leq PE(\Phi_{m,N}) \leq 0(L) \). Therefore, instead of intractable \( l_0 \)-constraint, w e could equally use \( l_1 \)-constrained LMS to both \( \Phi_{m,N} \) and \( \Phi_{m,N} \), if we were not a la reedy choosing H AM 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 de fined discipline a nd eg0. T his imbalance leads to unnecessary greediness, affecting every aspect of our life. I publish this not \( f \) or \( m \) y need t o survive; but t o pay back t he Communities who have taught me so much. The reader may carry on t he unsupervised learning running t on the \( f \) ault tolerant and d s uspace-generalize-able c onnectionist architectures. Incidentally, Tai Ch i p ractitioners b y t he walking m editation co nsider t he \( l_0 \) a nd \( l_1 \) as t he laterization. If w e were just relaxing your conscious-mind controlling m uscles, and let the gravity potential takeover, the internal fluids that circulates freely i nsider our internal organs known a s ‘t he Ch’i’ can which c an e nhance t he well-being a bout our p hysiology metabolism.

Acknowledgement

The a uthor w ishes t o t hank D r. C harles Hsu and Army NVESD Mr. J eff Jenkins, Ms. L ein M a, Ms. Ba linder Kaur for t heir technical supports. Dr. S oo-Young L ee for his patience and encouragement. T he author acknowledges US AFOSR grant s upport of CUA R/D a s f acilitated b y Dean Prof. Charles C. Nguyen.

References


Abstract

Spatio- and spectro-temporal data (SSTD) are the most common types of data collected in many domains and research areas, including engineering, bioinformatics, neuroinformatics, ecology, environment, medicine, economics, etc. However, there is a lack of methods for efficient analysis of such data for spatio-temporal pattern recognition (STPR). The brain inspires functions as a spatio-temporal information processing machine and self-organising manner. The brain-inspired spiking neural networks and are a promising paradigm for the creation of new intelligent ICT for SSTD. Thus the brain is the ultimate inspiration for the development of new models of the brain's functioning (e.g., genetic information, accumulated through evolution). 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 cognition state evaluation based on EEG and fMRI data; moving object detection in video data; word recognition based on spectro-temporal audio data; evaluating response of a disease to treatment based on clinical and environmental variables; e.g., stroke or cancer; and evaluating risk of disease, e.g., heart attack; Alzheimer’s Disease; modelling and prognosis of outcome of cancer; modelling the progression of a neuro-degenerative disease; e.g., Alzheimer’s Disease; estimation of invasive species in ecology; and predicting the effects of plants on agriculture.

Keywords: Spatio-temporal data, pattern recognition, spiking neural networks, computational modelling, EEG data.

1. Spatio- and 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; \( Y \) 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 \( dt \), and the output variables belonging to \( Y \):

\[
F: X(d) \rightarrow Y, \quad X(t) = (x_1(t), x_2(t), \ldots, x_n(t)), \quad t = 1, 2, \ldots, n \quad (1)
\]

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 cognition state evaluation based on EEG and fMRI data; moving object detection in video data; word recognition based on spectro-temporal audio data; evaluating response of a disease to treatment based on clinical and environmental variables; e.g., stroke or cancer; and evaluating risk of disease, e.g., heart attack; Alzheimer’s Disease; estimation of invasive species in ecology; and predicting the effects of plants on agriculture.

The commonly used models for data analysis 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 time temporal/spatial/cosmic components because they usually either ignore them or oversimplify 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. Thus the brain is the ultimate inspiration for the development of new machine learning techniques for SSTD.
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 a 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 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 online-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 authentication. It is based on our levels of feed-forward connected layers of spiking neuronal maps, similarly to the way the cortex works when learning and recognizing images of complex input stimuli [92].

However, these models are not designed for moving 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 frame from a pronounced word or a video, through the increase of the membrane potential of output spikes, they do not learn complex spatio-spectro-temporal associations from the data. Most of these models do not allow to model complex stochastic SSTD.

In [63, 10] a computational neuro-genetic model (CNGM) of a single neuron and SNN are represented 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 an 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 biological modeling and engineering applications is the IFM. The IFM has been realized 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.

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 $\{w_{ji}, j=1,2,\ldots,m\}$ are associated with the synapses and are determined during the learning phase using a learning rule. In addition to these connection weights $w_{ji}(t)$, the probabilistic spiking neuron model has the following three probabilistic parameters:

- A probability $p_{cji}(t)$ 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 $p_{cji}(t)=0$, no connection and no spike propagation exist between neurons $n_j$ and $n_i$. If $p_{cji}(t)=1$ the probability for propagation of spikes is 100%.

- A probability $p_{sji}(t)$ for the synapse $s_{ji}$ 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 $PSP_i(t)$ 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]:

$$PSP_i(t) = \sum_{p=t_0}^{t} \sum_{j=1}^{m} (\sum_{p=p_{b0},t}^{t} f_1(p_{cji}(t-p)) f_2(p_{sji}(t-p))w_{ji}(t)+\eta(t-t_0))$$

where $e_j$ is 1, if a spike has been emitted from neuron $n_j$ and 0 otherwise; $f_1(p_{cji}(t))$ is 1 with a probability $p_{cji}(t)$, and 0 otherwise; $f_2(p_{sji}(t))$ is 1 with a probability $p_{sji}(t)$, 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. As a special case, when all or some of the probability parameters are fixed to “1”, $t$ be a bove p olaristic m odel wi ll be simplified and will resemble the well known IFM. A similar formula will be used when a leaky LFM is used as a fundamental model, where a time decay parameter is introduced.

It has been demonstrated that SNN that utilizes the probabilistic neuronal model can learn better SSTD than traditional SNN with simple IFM, especially in a noisy environment [98, 83]. The effect of each of the above three probabilistic parameters on the ability of a SNN to process...
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 application, apart 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 \( \text{PSP}_{i}(t) \):

\[
\text{PSP}_{ij}(s) = A_{\text{synapse}} \left( \exp \left(-\frac{s}{\tau_{\text{synapse, decay}}} \right) - \exp \left(-\frac{s}{\tau_{\text{synapse, rise}}} \right) \right)
\]

where \( \tau_{\text{synapse, decay/ rise}} \) are time constants representing the rise and fall of an individual synaptic PSP; \( A \) is the PSP's amplitude; \( \varepsilon_{ij} \) represents the probability 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 (if 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, \( G = (g_1, g_2, \ldots, g_k) \);
(b) an initial state of the level of expression over time \( G(t=0) \);
(c) an initial state of the connection matrix \( L = (L_{11}, \ldots, L_{kk}) \), where each element \( L_{ij} \) defines the known level of interaction (if any) between genes/proteins \( g_i \) and \( g_j \);
(d) activation functions \( f_i \) for each gene/protein \( g_i \) from \( G \). This function defines the gene/protein expression value at time \( t(t+1) \) depending on the current values \( G(t), L(t), g_i \), and some external information \( E(t) \):

\[
g_i(t+1) = f_i(G(t), L(t), E(t))
\]

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:

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

Rate-order information representation is typical for cognitive information processing [18].
Temporal spike learning is observed in the auditory [93], the visual [11] and motor information processing of the brain [13, 90]. Its use in neuro-prosthetics is essential, a long with applications for fast, real-time recognition and control 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 SNN [30, 14]. Temporal coding of information in a spiking neuron [103] and in the computational neuro-genetic model of a spiking neuron [Benuskova and Kasabov, 2007] is used as parameters for its implementation of both long-term and short term memory.

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 postsynaptic action potentials relative to the pre-synaptic spike (see Table 1). If the difference in the spike time between the pre- and post-synaptic spikes is less than a critical time window, the synaptic weight is increased above a set threshold by the STDP learning mechanism. If the difference is more than the critical time window, the synaptic weight is decreased below the threshold, otherwise it remains the same.

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 of a synapse w0, depending on the time of the postsynaptic spike relative to the presynaptic neuron. The synaptic efficacy is increased (potentiation) if the postsynaptic spike is emitted just before the pre-synaptic spike, synaptic efficacy is decreased (depression). This change in synaptic efficacy can be expressed as:

$$\Delta V_{w0} = \frac{I_{p_{post}}(t_{post}) - I_{p_{pre}}(t_{pre})}{C_p} \Delta t_{spk}$$ if $t_{pre} < t_{post}$

$$\Delta V_{w0} = -\frac{I_{d_{dep}}(t_{post}) - I_{d_{depr}}(t_{pre})}{C_d} \Delta t_{spk}$$ if $t_{post} < t_{pre}$

where $\Delta t_{spk}$ is the difference in spike time of the pre- and postsynaptic spikes.

3.4 Rank-order learning

The rank-order learning rule uses information from the input spike trains – the rank of the first incoming synaptic spike is emitted just before the pre-synaptic spike, synaptic efficacy is decreased (depression). This change in synaptic efficacy can be expressed as:

$$\Delta V_{w0} = \frac{I_{p_{post}}(t_{post}) - I_{p_{pre}}(t_{pre})}{C_p} \Delta t_{spk}$$ if $t_{pre} < t_{post}$

$$\Delta V_{w0} = -\frac{I_{d_{dep}}(t_{post}) - I_{d_{depr}}(t_{pre})}{C_d} \Delta t_{spk}$$ if $t_{post} < t_{pre}$

where $\Delta t_{spk}$ is the difference in spike time of the pre- and postsynaptic spikes.
Fig. 6. A single LIF neuron with simple synapses can be trained with the STDP unsupervised learning rule to discriminate a repeating pattern of synchronized spikes on certain synapses from noise (from: T. Masquelier, R. Guillonneau and S. Thorpe, PlosONE, Jan 2008).

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 S TPR 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 at a time \( t \) is calculated as:

\[
PSP(i,t) = \sum_j \text{mod}^{\text{order}(j)} w_{j,i} \quad (7)
\]

where \( \text{mod} \) 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; \( \text{order}(j) \) 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 \( \text{order}(j) \) 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 \( \text{PSPTh} (i) \).

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]:

\[
\Delta w_{j,i} (t) = \text{mod}^{\text{order}(j,i)(t)} \quad (8)
\]

### 3.5 Combined rank-order and temporal learning

In [25] a method for a combined rank-order and temporal learning (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

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.)
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 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, Jan 2008) – see Fig. 6.

Single neuron models have been introduced for STPR, such as: Temporotron [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 recognised 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).

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 online on-line 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 neurons to recognize the 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 an 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 bursts 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
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 STD 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 an SNN classifier was applied [61, 115].

Reservoir computing [73, 108] has already become a popular approach for STD modelling and pattern recognition. In the classical view a ‘reservoir’ is a homogeneous, passively connected and fixed neurons that in principle has no learning and memory, neither it has an interpretable structure – fig. 11(b). 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 STD, which is the main goal of learning STD. Despite difficulties with the LSM reservoirs, it was shown on several STD problems that they produce better results than using a simple classifier [95, 73, 99, 60]. Some publications demonstrated that a reservoir computing is especially suitable in a noisy environment [98, 83].

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**: 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 $S_{C_i}(t)$ are used differently for different tasks.

- **Continuous function representation of spiker spikes**: 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. $\alpha$-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 a framework 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 neurons 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 million 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]).](image-url)
In [18] it is suggested that both the firing rate (rate coding) and spike timing as spatiotemporal patterns (rank order and spatial coding) play a role in fast and slow dynamic responses, controlled by the cerebellar nuclei. Spatio-temporal patterns of population of Purkinje 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) depends on gene expression [40], e.g., glutamatergic neurons expressing FGF17 from the same progenitors tend to form clusters, a node for the cortical GABAergic interneuron population. 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.

It was shown that a single gene, that regulates the $\tau$ 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 scale 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 system 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 particle swarm optimization (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 solution of the global optimum in a very short time.

In one approach it may be reasonable to use the same parameter values (same GRN) for all neurons in the SNN or for each of different types of neurons (cells) that will result 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 realizations of SNN are 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];
Software simulators, such as Brian [16], Nestor, NeMo [79], etc;
- Silicon retina camera [23];
- Silicon cochlea [107];
- SNN hardware realisation of LIF and S DSP [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 the 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 valuation, e.g. prognosis 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.
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 (neuroeconomics) where at a ‘lower’ level the GRN represents the dynamic interaction between time series variables (e.g. stock index values, exchange rates, unemployment, GDP, price of oil), while at a higher level the pSNN states represents the state of the economy or the system under study. The states can be further classified into predefined 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. Indery and also with: A. Mohamed, 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 of Neuroinformatics – the Neuronomorphic Cognitive Systems Group, at the University of Zurich and ETH Zurich (http://ncs.ethz.ch/projects/evospike). Diana Kasabova helped with the proofreading.

References


Research, 30, 895-908.


Biologically Motivated Selective Attention Model

Minho Lee
Kyungrak National University, Korea
*corresponding author: mholee@knu.ac.kr

Abstract
Several selective attention models partly inspried by biological visual attention mechanism are introduced. The developed models consider not only binocular stereopsis to identify a final attention area s o t h e system foc uses on t h e closer area as in human binocular vision, but also both the static and dynamic features of an input scene. In the model, Is how t he n eye 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 effects in subsequent visual processes. In this model, I also consider a psychological distance a s well as t he pop-out and refusal in subsequent visual processes. In addition, I develop a new type of bottom-up attention model based on the relative spatial distribution of 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 the relative spatial location of visual features and some memory modules. The object form and color representation a nd color memory modules are used in a hierarchical learning mechanism together with the context of each scene. The proposed model includes a Growing Fuzzy Topology Adaptive Resonance Theory (GFTART) network which plays two important roles in object color and form based attention. Experiments show how the proposed model considers an generate plausible scan paths and selective attention for natural input scenes.

Keywords: Selective attention, bottom-up attention, GFTART

1. Introduction
The visual s electorate a attention c an a llow humans t o p ay attention t o a n i nteresting area o r a n cl ustered scene a s well as t o pr operly respond for various visual stimuli in natural environments. Such a s elective attention visual mechanism allows the human vision system to effectively process high complexity visual scenes.

Itti, K och, and N eib ur (1998) introduced a brain-like model in order to generate the bottom-up saliency map (SM). K oike a nd S aiki (2002) proposed a stochastic winner take all (WTA) model that 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 in tegrating saliency, scale selection and a content description, thus contrasting with many other approaches. Anström and C hristersen (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 heuristics for detection of interesting or salient points in images. Wrede et al. (2010) proposed a random center 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 F ourier phase spectrum. Based on the psychological understanding, W ang a nd 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, R ome, Nüchter, a nd S urmann (2005) proposed a bi-modal attention system that considers both static and dynamic features including color and depth for generating proper attention. Fernández-Caballero, L ópez, and Sáiz-Valverde (2008) developed a dynamic stereoscopic selective visual attention model that integrates motion and depth in order to choose the attention area. Maki, N oрудn, a nd E klund (2000) proposed an attention model integrating image flow, stereo disparity a nd motion for a tentional s cene segmentation. Ouerhani a nd H ügli (2000) proposed a visual attention model t hat c onsiders the signal-to-noise ratio ( Navalpakkam & Itti, 2006). As well, W alther et al. (2005) proposed an object preferred attention scheme that considers the bottom-up SM results as biased w eights for bottom-up o bject perception. Li and Itti (2011) represented a visual attention model t o s olve t he t arget detection p roblem in n satellite images by combining biologically-inspired features such as saliency and gist features. Guo and Zhang (2010) extended their previous a pproach ( Hou a nd Z hang, 2007) c alled spectral r esidual (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 a neural attention region by considering psychological distance 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 color characteristics, and generates a bias signal corresponding to the candidate object area. In addition, the proposed system comprises of an incremental object representation and memory model with 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 color characteristics, and generates a bias signal corresponding to the candidate object area.

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 the brain. When rods and cones in the retina are excited, signals are transmitted through successive neurons in the retina 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). These preprocessed signals 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 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
fiducial all the way from the retina to the visual cortex. Finally, some detailed 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 area and the inferior-temporal area (IT) through the V1-4c, which is related to the visual dorsal pathway (Guyton, 1991). In contrast, some rapidly changing visual information sensed by Y-cells in the retina is rapidly transferred from the LGN-magnocellular cells to the middle temporal area (MT) and medial superior temporal area (MST) through the V1-4c, which is related to the visual ventral pathway (Bear et al., 2001). In order to develop a plausible visual selective atention model, we consider the role of the IT area in bottom-up visual processing such as attention and motion information used in the dorsal pathway.

Fig. 2 shows the relation between visual pathways and an observed computational model that is related to the visual environment perception. The visual pathway works for bottom-up visual processing, and the IT area focuses on top-down visual processing such as attention. The lateral-intra parietal cortex (LIP) works as an attention controller. Actually, the prefrontal cortex (PFC) plays an important role in the high-level perception such as brightness, odd color, and etc. The prefrontal cortex (PFC) plays a very important function in the high-level perception such as brightness, odd color, and etc.

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 process of sensing or selective attention reflecting simple bi-logical visual saliency (Fukushima, 2005). Itti and Koch used three basis feature maps: intensity (I), color (C), and orientation (O) feature maps as well as the symmetry transformation (GST) algorithm and 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 of object (Reisfeld et al., 1995). In order to implement an object referable atention 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 express symmetry a xis (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 center-surround and of f-surround function 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 with 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, being the orientation-selective simple cells in the V1. Using C SD & N in an Australian pyramid (Itti et al. 1998), we can construct intensity (I), color (C), and orientation (O) feature maps as well as the symmetry feature map (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),
localizes the salient area in the input scene and in addition, it 2001). Fig. 2 shows that proposed models selectively algorithm is based upon Kadir’s approach (Kadir & Brady, 1997). Scale election salient area was adapted to select a proper scale of the salient area by using the saliency map. The scale selection was considered an entropy maximization approach. The size of the salient area is obtained from the analysis of successive static SMs. The entropy value at each pixel represents a fluctuation of dissimilarity in the scale-space of the saliency map. 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 salient map. The most appropriate scale f or each salient area, centered at location x, is obtained by Eq. (2):  

\[
\text{scale}(x) = \arg \max_s \{H_s(x, s) \times W_d(s, x)\} \tag{2}
\]

where \(D\) is the set of all descriptor values, \(H_s(x, s)\) is entropy as defined by Eq. (3), and \(W_d(s, x)\) is the interscale measure as defined by Eq. (4):

\[
H_s(x, s) = -\sum_{d \in D} p_{d,s,x} \log_2 p_{d,s,x} \tag{3}
\]

\[
W_d(s, x) = \frac{s^2}{2s-1} \sum_{d \in D} |p_{d,s,x} - p_{d,s-1,x}| \tag{4}
\]

where \(p_{d,s,x}\) 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 \(P_{d,s,x}\) 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).

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).

3.2 Scale saliency

The localized salient area, which is obtained from the bottom-up saliency map, is suitable for the 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 of 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). 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 how the procedure adopted to acquire a final SM by integrating both of the static and dynamic SMs. 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 and the bottom-up image. (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 and the bottom-up image. (Jeong et al., 2008, Fernández-Caballero et al., 2008).
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 MST through the LGN, by means of the V1 and the V2, including the lateral intraparietal 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 (MTabs-cells and MTrel-cells), and MST cells extract optical 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). MTabs-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. MTrel-cells extract relative velocity of stimuli. The MTabs-cells integrate responses of many MTrel-cells by summation and extract counter-clockwise rotation, clockwise rotation, expansion and contraction of optical flow (Jeong et al., 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 visual system. Th is st creo vision system is used to detect attention regions in the 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 whether the salient regions are in front or behind the 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).

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):

$$ S_c(ν) = S_p(ν) \cdot (1 + \exp -\frac{z}{\tau} ) \cdot L(s_p, ν, σ) $$

where $ν$ denotes a pixel in the salient area, and $s_p(ν)$ and $s_d(ν)$ 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. $L(·)$ is a Laplacian function as shown in Eq. (6):
where \( sp \) represents the center location of the salient area, \( v \) denotes a pixel in the salient area, \( \sigma \) 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 an attention region. The stereo SM is constructed using not only depth information, but also spatial information within a salient area. The attention region is determined by the salient area and \( d \) denotes a pixel in a salient area. The weight values, \( \alpha \), in each feature map increase or decrease according to the training results. They only considered fixed weight values for biasing four different features according to top-down bias weights. The weight values are \( \alpha \) and are calculated by Eq. (9) and (10), where \( \alpha \) is a training rate and \( \beta \) represents the influence of the previous \( \alpha \) on the current feature.

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 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 salient area facilitates a visual search (Mazer & Gallant, 2003). We avoid focusing on an interesting area having similar characteristics to a previously learned uninteresting area by generating a top-down bias signal obtained through the 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 uninterested objects, 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 top-down bias weights. The weight values are calculated by Eqs. (9) and (10), where \( \alpha \) and \( \beta \) represent the influence of the previous \( \alpha \) on the current feature.

\[
\alpha_{(n+1,c,N)k} = (1-\beta)\alpha_{(n,c,N)k} + \eta \cdot FM_{(n+1,sp,N)} \cdot SM_{(n+1,sp)}
\]

(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, \( N \) denotes the number of nodes in an F2 layer, \( \alpha \) is the fuzzy ART, \( \beta \) is the learning rate, and \( \eta \) represents the influence of the previous \( \alpha \) on the current feature.

\[
FM_{(n,sp,N)} = \sum_{y \in C_{(sp,N)}} SM_{(sp,y)} \cdot FM_{(sp,y)}
\]

(11)

\[
SM_{(n,sp)} = \sum_{y \in C_{(sp,N)}} SM_{(sp,y)} \cdot FM_{(sp,y)}
\]

(12)

\[
SM_{(n,sp)} \quad \text{and} \quad FM_{(n,sp)}
\]

(12)

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 to 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 IT and the V4 area, generating target object information in whole area in order to filter the areas that satisfy the target object dependent features. 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 represent salient features of an object.
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 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 $W'_j$, 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 $\rho$, 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 conventional fuzzy ART is enhanced by adding the topology preserving mechanism in an 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.
4.3 Selective attention reflecting psychological distance

Fig. 13 illustrates the proposed visual attention model, which is partly inspired by biological visual pathways 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object processing, which is extended to the IT and PFC for top-down processing.

In order to develop a more human-like visual selective attention, we need to consider a stereo SM model for binocular vision. 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 considers the psychophysical distance of an attended object from an observer obtained from a depth perception module using the stereo SM.

In order to develop a more human-like visual selective attention, we need to consider a stereo SM model for binocular vision. 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 considers the psychophysical distance of an attended object from an observer obtained from a depth perception module using the stereo SM.


\[
\text{Psycho\_distance}(v) = \frac{\text{incongruency\_mean\_response\_time}}{\text{congruency\_mean\_response\_time}}
\]

if congruent condition

\[
\text{Psycho\_distance}(v) = \frac{\text{congruency\_mean\_response\_time}}{\text{incongruency\_mean\_response\_time}}
\]

else

5. Experimental Results

Fig. 14 shows an example in which the proposed stereo bottom-up SM model generates a better attention path than by using symmetric information. A stereo SM model generates a maximum salient value for the 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.

\[
\text{Psycho\_distance}(v) = \frac{\text{incongruency\_mean\_response\_time}}{\text{congruency\_mean\_response\_time}}
\]

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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object 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 for binocular vision. 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 a non-tended object is perceived by an object categorization module. And the spatial distance to an attended object is perceived by an object processing, which is extended to the IT and PFC for top-down processing.
Table 1. Comparison of three different bottom-up SM models for object preferred attention.

<table>
<thead>
<tr>
<th>Salient area</th>
<th>Static SM</th>
<th>Dynamic SM</th>
<th>Static &amp; Dynamic SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st salient area</td>
<td>143</td>
<td>150</td>
<td>165</td>
</tr>
<tr>
<td>2nd salient area</td>
<td>103</td>
<td>104</td>
<td>106</td>
</tr>
<tr>
<td>3rd salient area</td>
<td>64</td>
<td>77</td>
<td>68</td>
</tr>
<tr>
<td>4th salient area</td>
<td>47</td>
<td>57</td>
<td>66</td>
</tr>
<tr>
<td>5th salient area</td>
<td>28</td>
<td>32</td>
<td>46</td>
</tr>
<tr>
<td># of total</td>
<td>385</td>
<td>420</td>
<td>451</td>
</tr>
<tr>
<td>Detection rate</td>
<td>77 %</td>
<td>84 %</td>
<td>90 %</td>
</tr>
</tbody>
</table>

Table 2. Comparison of the degrees of saliency in static, dynamic and integrated SM models.

<table>
<thead>
<tr>
<th>Salient objects</th>
<th>Static SM</th>
<th>Dynamic SM</th>
<th>Static &amp; Dynamic SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right human</td>
<td>183 (1st)</td>
<td>139 (2nd)</td>
<td>161 (1st)</td>
</tr>
<tr>
<td>Center kettle</td>
<td>143 (2nd)</td>
<td>113 (3rd)</td>
<td>128 (3rd)</td>
</tr>
<tr>
<td>Left human</td>
<td>135 (3rd)</td>
<td>171 (1st)</td>
<td>154 (2nd)</td>
</tr>
</tbody>
</table>

Table 3. Comparison of the degrees of saliency according to depth information in the integrated SM model.

<table>
<thead>
<tr>
<th>Salient objects</th>
<th>Depth (m)</th>
<th>Degree of saliency in selected area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right human</td>
<td>0.86</td>
<td>183 (1st) 190 (1st) 252 (1st)</td>
</tr>
<tr>
<td>Center kettle</td>
<td>2.4</td>
<td>143 (2nd) 129 (3rd) 154 (3rd)</td>
</tr>
<tr>
<td>Left human</td>
<td>1.3</td>
<td>135 (3rd) 164 (2nd) 218 (2nd)</td>
</tr>
</tbody>
</table>

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 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) represent the relative degree of motion information from the Static & Dynamic SM model. In Fig. 16(a) i.e. moving to camera and rightward direction, the area ‘c’ from static SM is moving in and out from camera. Also, our model successfully responds to the rightward motion i.e. a rea ‘c’ by producing a mount 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).
Table 4. Performance of the affective SM model after training lip areas as a preferential region.

<table>
<thead>
<tr>
<th></th>
<th>96 Training Images</th>
<th>90 Test images</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bottom-up SM</td>
<td>Affective SM</td>
</tr>
<tr>
<td># of with lip</td>
<td>96</td>
<td>35</td>
</tr>
<tr>
<td># of without lip</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Correct rate</td>
<td>100 (%)</td>
<td>39 (%)</td>
</tr>
</tbody>
</table>

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 the affective SM model. Fig. 18 (a) and (b) show that the affective SM model has better performance than the bottom-up SM model for focusing on lip areas. Table 4 shows the performance of the affective SM model with that of the bottom-up SM model for focusing on lip areas in face images. The affective SM model has a correct rate of 98%, while the bottom-up SM model has a correct rate of 39%. The affective SM model is more accurate in identifying lip areas in face 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. The second problem is to categorize objects in the real world. The GFTART was able to successfully reflect the psychological distance in both congruent and incongruent conditions. The proposed affective SM model was able to generate a visual scan path that is more salient in the real world than the visual scan path generated by the bottom-up SM model.
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.

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 a depth information in natural input scenes. In particular, we added a Hebbian learning process to generate a top-down bias signal based on human afferent factors, which enhance the performance of the previous Lee’s trainable s election 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 in a stereosaliency 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 a stable and reliable model of each ach model, while it is capable of generating attention or indirectly get insights from known biological mechanisms 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


CASA: Biologically Inspired Approaches for Auditory Scene Analysis

Azam Rabiee1*, Saeed Setayeshi2, and Soo-Young Lee3

1Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
2Department of Medical Radiation, Amirkabir University of Technology, Tehran, Iran
3Brain 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 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 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. Independence 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 has a remarkable capacity for sound analysis. This is what Cherry called 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 call it 'the cocktail party problem'. No machine has ever 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 a uditory scene analysis (ASA) principles. Research in ASA has inspired considerable work to build computational auditory systems for sound separation. Physiological models of ASA may induce ideas 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 mixed, the interference is 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 input is decomposed into continuous time-frequency units or segments. Each of these segments originates from a single speaker. In grouping, 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 of the cochlea. The most complicated part of the ear is the inner ear, which 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, each inner hair cell represents a sound 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, the 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 the 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 receives sounds from the ears, basal ganglia, and other structures. The information eventually reaches the SOC of the auditory nerve, which is responsible for controlling the direction and maintenance of attention. MGN 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 from somatosensory and visual systems. 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 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 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 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 the 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 literature. 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 the midbrain, and (3) the output model from the 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 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 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 early auditory stage in performing frequency analysis on the waveform signal into a 2D time-frequency representation. Conventional CASA systems utilize a SA cascade model to decompose this 2D representation into sensory segments in segmentation stage, as well as to assign those segments to corresponding S peaks in the 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 [19]. 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 is illustrated in Figure 2(a).

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

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

where $*$ is convolution in time domain.

These cochlear filter outputs are passed into auditory-nerve pa to allow the output of 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:

$$y_{AN}(t, x) = g(\partial_t y_{coch}(t, x)) *\omega(t)$$

The next transformation simulates the effect of laterally inhibitory. 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:

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

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

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

#### 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 spaced by a membrane leakage low-pass filter $g(t)$. The cochlear membrane stage consists of a bandpass filter, whose impulse response, $g_{fc}(t)$, is the product of a gamma function and a tone:

$$g_{fc}(t) = t^{-1}e^{-2\pi t^2f_c}cos(2\pi f_c t + \phi)u(t)$$

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

$$ERB(f) = 24.7 + 0.108f$$

The center frequencies are related to the ERB domain and overlap 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, and a tach to the hair cell, cause a depolarization 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

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

(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,

$$P = hc(t)dt,$$

(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 cochleagram 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 $1\over 12$-octave scaling of center frequencies, which matched to the note scaling. It is also appropriate for harmonic analysis, because of its sharpness in frequency axis as a result of the laterally inhibition process, as 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, in cluding
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 several research [9][24][25]. The correlogram is usually computed in the time domain by autocorrelating the simulated auditory time-frequency representation and localized spectrotemporal analysis [7][20]. Among them, Elhilali and Shamma in [7][30] have successfully tackled aspects of the “cocktail party problem” and provided an account of the perceptual processing of the model, as it further maps the sound patterns into a perceptual space organized from narrow to broad spectral features. Our model, on the other hand, targets the slow temporal dynamics (<30Hz) and their corresponding spectrotemporal analysis.

In the model, they broke down the cortical analysis into a spectral mapping and temporal analysis. The s pectrogram and h i lbert transform are used to estimate the time-varying firing rate of the auditory neurons, generating a multidimensional representation of the waveforms 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][30] have utilized it for a sound separation task as an unsupervised clustering and integrative stage in which the segregation was performed based on an unsupervised clustering and a statistical theory of Kalman prediction.

In the model, they broke down the cortical analysis into a spectral mapping and temporal analysis. The spectrogram and Hilbert transform are used to estimate the time-varying firing rate of the auditory neurons, generating a multidimensional representation of the waveforms with time, frequency, scale, and rate axis.

In the model, they broke down the cortical analysis into a spectral mapping and temporal analysis. The spectrogram and Hilbert transform are used to estimate the time-varying firing rate of the auditory neurons, generating a multidimensional representation of the waveforms with time, frequency, scale, and rate axis.

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 signal to a non-linear, 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 pure wavelet 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. Multiresolution 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 processes of 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 the STRFs are revealed in their STRFs. In the model, they assumed a bank of directional selective STRFs (downward [−] and upward [+] ) that are real and equal functions formed by combining two complex functions of time and frequency as follows,

$$
\begin{align*}
    S_{T R F +} &= \Re \{H_{r a t e}(t; \omega, \theta), H_{s c a l e}(f; \Omega, \phi)\}, \\
    S_{T R F -} &= \Re \{H_{r a t e}(t; \omega, \theta), H_{s c a l e}(f; \Omega, \phi)\},
\end{align*}
$$

where \( \Re \) denotes the real part, * is the complex conjugate, \( \omega \) and \( \Omega \) are rate and scale parameters of the STRF, respectively, and \( \theta \) and \( \phi \) are characteristic phases that determine the degree of asymmetry along the time and frequency, respectively. Functions \( H_{r a t e} \) and \( H_{s c a l e} \) are analytic signals obtained from \( h_{r a t e} \) and \( h_{s c a l e} \):

$$
    H_{r a t e}(t; \omega, \theta) = h_{r a t e}(t; \omega, \theta) + j h_{s c a l e}(t; \omega, \theta),
$$

where * denotes Hilbert transform. \( h_{r a t e} \) and \( h_{s c a l e} \) are temporal and spectral impulse responses defined by sinusoidally interpolating between symmetric seed functions \( h_{r}(.), \) and \( h_{s}(.), \) at the output of each cochlear filter channel, resulting in a 3D time-frequency-lag representation of the acoustic signal.

$$
    h_{r}(t; \omega) = \omega h_{r}(\omega t),
$$

$$
    h_{s}(f; \Omega) = \Omega h_{s}(\Omega f),
$$

in which

$$
    h_{r_{r}}(t) = t^{2} e^{-3.5t} \sin 2 \pi t,
$$

$$
    h_{r_{s}}(f) = (1 - s^{2}) e^{-s^{2}/2}.
$$

As shown in the Figure 4(a), the convolution between STRFs and the spectrogram gives an estimate of the time-varying firing rate at the neuron, generating a multidimensional representation of the waveforms. In the model, they broke down the cortical analysis into a spectral mapping and temporal analysis.

In the model, they broke down the cortical analysis into a spectral mapping and temporal analysis. The spectrogram and Hilbert transform are used to estimate the time-varying firing rate of the auditory neurons, generating a multidimensional representation of the waveforms 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][30] have utilized it for a sound separation task as an unsupervised clustering and integrative stage in which the segregation was performed based on an unsupervised clustering and a statistical theory of Kalman prediction.

In the model, they broke down the cortical analysis into a spectral mapping and temporal analysis. The spectrogram and Hilbert transform are used to estimate the time-varying firing rate of the auditory neurons, generating a multidimensional representation of the waveforms with time, frequency, scale, and rate axis.

In the model, they broke down the cortical analysis into a spectral mapping and temporal analysis. The spectrogram and Hilbert transform are used to estimate the time-varying firing rate of the auditory neurons, generating a multidimensional representation of the waveforms with time, frequency, scale, and rate axis.
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. Clearly, 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 activate 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, alone, is too simplified to be used for CASA.

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 to evaluate if a visual cortex model such as Neocognitron [36][37] or H MAX [38][39] can be customized for audio processing. Spiritually similar, Neocognitron and H MAX 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 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.

In some sense, models for visual cortex are more mature than the auditory cortex. Nevertheless, it is not easy to develop a comprehensive model for a visual auditory cortex, especially because of the stochastic nature of temporal dynamics of sound. Moreover, physiologically, there are not many findings about how ow i individual cortical areas compute, the nature of how the brain interprets auditory objects, except for some artificial models. 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 several 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 with sound separation platforms, even cell phones, 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 the cochlear nuclei, which can be easily simulated in binaural methods.

Despite various mechanisms are suggested by binaural CASA re searchers for extraction of interaural differences [44], we would not ignore the remarkable monaural C ASA approaches. Certainly, monaural source separation is a challenging task, especially when the number of sources is more than two. However, the situation is different in 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, t here 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 the way down to the outer hair cells in the cochlea via the midsbrain, which helps these neurons of interest and maximize expectation through feedback.

Although the physiological findings of the top-down process are not matured, in order to improve the performance, the CASA approaches use combination of low-level and high-level models. Generally, according to the literature, the best example of a top-down auditory understanding system is a probabilistic model such as hidden Markov model (HMM) as a 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 recognition. Similarly, Baraker 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 process [25]. On the other hand, in [10] the top-down integration is achieved out by training Gaussian mixture models (GMMs) and vector quantizers (VQs) of each clean data for each speaker, and incorporating in the bottom-up separation process that is based on 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 concept of the separated signals. In [48], Stark et. al. investigated both strategies, designing a model of a binary mask on the spectrogram of the mixture with the estimated speech features. In similar conditions, the target 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 CA SA, recognition without synthesizing the separated signals not only is biologically plausible, but also shows dominant performance.

5.4. CASA vs. BSS

Using standard corpus of voiced speech with interfering sounds, Kouwe et. al. in [49] have reported a comparison between C ASA and B SS 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 equivalent with a natural environment.

On the other hand, in the natural environment, CA SA brings the flexibility of the physiological systems in 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, a model of auditory scene analysis that has played a crucial role in the development of CASA systems (such as continuity of F0 and spatial location) can be exploited by BSS algorithms in order to improve their performance on real-world acaustic mixtures. Conversely, blind separation techniques could help CASA in decomposing mixtures that overlap substantially in time-frequency plane [49]. Moreover, CA SA solutions come to help when BSS is not able to n the absence of many of the requirements of BSS. Eventually, current CASA models cannot deal with more-than-two-source mixtures, efficiently.

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 how the importance of both phase and magnitude of the psectrogram [51]. However, traditionally, phase is deemed unnecessary because of the low computational resolution. Nowadays, by increasing the precision of phase, the researchers are trying to investigate more features by the phase information [52-56].

Utilizing phase information in investigating complex matrix factorization (CMF) or source separation in [57][58]; but here is not enough research in C ASA 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 processes in CASA 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 National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (2009-0092812 and 2010-0028722).

References

Natural Intelligence: the INNS Magazine                 58                             Volume 1, Issue 2, Winter  2012

Pinna to Auditory Nerve, “ in

“Combining monaural and binaural evidence for reverberant speech segregation,” in


Natural Intelligence: the INNS Magazine                 58                             Volume 1, Issue 2, Winter  2012

Pinna to Auditory Nerve, “ in

“Combining monaural and binaural evidence for reverberant speech segregation,” in


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 outstanding contributions to 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 that time he has been the recipient of many distinguished awards and research grants including, M Donnel-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.

Kunihiko Fukushima
Recipient of 2012 INNS Helmholtz Award

Kunihiko Fukushima received a B.Eng. degree in Electronics in 1958 and a Ph.D. 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...
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 at Auckland University of Technology. Currently he is a Chair of 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) a nd al so of t he A sia-P acific Neural Network Assembly (APNNA). He is a member of several t echnical c ommittees o f t he IEEE Computational Intelligence Society and a Distinguished Lecturer of t he IEEE CIS. He has served as a Associate Editor of Neural Networks, IEEE TRNN, IEEE TRFS, Information Science, J. Theoretical and Computational Nanosciences, Applied Soft Computing and other journals.

Kasabov holds MS c and P hD f rom t he T echnical 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 Achievement Award and Excellent Paper Awards from IEEE, the Neural Networks Pioneer Award, the APNNA Excellent Service Award (2005), R SNZ Science and Technology Medal (2001), and others. He is an invited Professor at the Shanghai Jiao Tong University (2010-2012). More information of Prof. Kasabov can be found on the KEDRI website: http://www.kedri.info.

Sebastien Helie
Recipient of 2012 INNS Young Investigator Award

Sebastien Helie is a Researcher in the Department of Psychology & Brain Sciences at the University of California, Santa Barbara. Prior to filling this position, he was a postdoctoral fellow at the Cognitive Science Department at the Rensselaer Polytechnic Institute (2006-2008). Dr. Helie completed a Ph.D. in cognitive computer science at the University of Quebec A Montreal, an M.Sc. degree at the University of Montreal.

His research interests are related to neuroscience and psychological modeling in general and more precisely to computational cognitive neuroscience, categorization, automaticity, active learning, sequence learning, skill acquisition, and creative problem solving. Dr. Helie ha s 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 Journal of Cognitive Science and the EU NERIAL Networks, D. R. Elie has served as al so chair of the Annual Conference of the ARION 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 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 algorithms for target detection, tracking, and situational understanding. He authored 9 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 an information-processing device that operates on incoming stimuli and generates output. Rather, the brain is proactive, constantly generating predictions about what we see, which guide perception and other cognitive processes. The ability of the brain to generate predictions is dependant on experience; we encode statistical regularities in the world over time. Thus, learning engages predictions, and predictions are representations of the fundamental operating principle in the brain. Work in my lab has shown that images of objects strongly associated with a particular context are encoded in the brain in a network of cortical regions rather than as objects weakly associated with any particular context (e.g., a pen). These regions include parahippocampal cortex (PHC), retrosplenial cortex, ventromedial prefrontal cortex (vmPFC), and temporal parietal cortex (RSC). This coarse (LSF) representation of an image is projected rapidly from early visual cortex to orbitofrontal cortex (OFC) via the dorsal magnocellular pathway. This coarse representation, in turn, activates OFC and other regions in the ventral visual stream. Studies in my lab have shown that the OFC regions 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. Further 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 omissions necessary for a void collision.

I have proposed that the efficiency of vision is due in part to a cortical mechanism that makes us see objects in the world as whole objects, not a collection of features. This model, known as the global information model, has been supported by numerous studies showing that low frequency information is indeed activated early enough to facilitate recognition. 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 the sense of perception.

I have been working on 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 1965. In the laboratory, there were groups of engineers, neurophysiologists, and psychologists, working together to discover the mechanisms of the visual system, such as the ones by Hubel and Wiesel, a nd simplifying the auditory system 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, called the "cognitron." The network has a function of self-organization, and it is able to recognize patterns through learning. In 1979, the cognitron was extended to have a function of recognizing shifted and deformed visual patterns, extracting binocular parallax, and so on. 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 it recognizes individuals by segmenting and a network. As a result of the top-down model, various systems have been developed: such as a network extracting a face and its parts from a complicated 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 a nd modeling 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 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 systems that lead to a better quality of information processing and knowledge discovery are now being developed. They 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; and rule-based systems for personalized medicine.

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 a novel parallel computational architecture utilizing algebraic theory of transformational groups and semantic 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 methods with the probability theory for the development of novel parallel computational architectures utilising algebraic theory of transformational groups and semantic groups. Later I introduced a hybrid connections production rule-based model and developed connectionist-based expert systems.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

I also developed a series of novel methods for transductive reasoning for personalized medicine that created new opportunities for the application of computational intelligence to personalized medicine.

Recent work on evolving spiking neural networks (eSNN) with applications for spatio-temporal pattern recognition, ultramodal audiovisual information processing, and taste recognition (Proc. ICONIP 2007-2011; IEEE WCCI, 2010; Neural Networks 2010). The proposed computational neuro-genetic models for brain knowledge discovery was also proposed.

Integrative connectionist, genetic, and quantum-inspired methods are now being developed and applied to various problems, including personalized medicine. The brain knowledge discovery was also proposed.

I am greatly honored to receive the top INNS Gabor Award for 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 a novel parallel computational architecture utilizing algebraic theory of transformational groups and semantic 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 methods with the probability theory for the development of novel parallel computational architectures utilising algebraic theory of transformational groups and semantic groups. Later I introduced a hybrid connections production rule-based model and developed connectionist-based expert systems.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

I also developed a series of novel methods for transductive reasoning for personalized medicine that created new opportunities for the application of computational intelligence to personalized medicine.

Recent work on evolving spiking neural networks (eSNN) with applications for spatio-temporal pattern recognition, ultramodal audiovisual information processing, and taste recognition (Proc. ICONIP 2007-2011; IEEE WCCI, 2010; Neural Networks 2010). The proposed computational neuro-genetic models for brain knowledge discovery was also proposed.

Integrative connectionist, genetic, and quantum-inspired methods are now being developed and applied to various problems, including personalized medicine. The brain knowledge discovery was also proposed.

INNS AWARD ACCEPTANCE STATEMENT:
Nikola Kasabov - INNS Gabor Awardee

It is my great honor to receive the top INNS Gabor Award for my 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 a novel parallel computational architecture utilizing algebraic theory of transformational groups and semantic 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 methods with the probability theory for the development of novel parallel computational architectures utilising algebraic theory of transformational groups and semantic groups. Later I introduced a hybrid connections production rule-based model and developed connectionist-based expert systems.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

I also developed a series of novel methods for transductive reasoning for personalized medicine that created new opportunities for the application of computational intelligence to personalized medicine.

Recent work on evolving spiking neural networks (eSNN) with applications for spatio-temporal pattern recognition, ultramodal audiovisual information processing, and taste recognition (Proc. ICONIP 2007-2011; IEEE WCCI, 2010; Neural Networks 2010). The proposed computational neuro-genetic models for brain knowledge discovery was also proposed.

Integrative connectionist, genetic, and quantum-inspired methods are now being developed and applied to various problems, including personalized medicine. The brain knowledge discovery was also proposed.

INNS AWARD ACCEPTANCE STATEMENT:
Nikola Kasabov - INNS Gabor Awardee

It is my great honor to receive the top INNS Gabor Award for my 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 a novel parallel computational architecture utilizing algebraic theory of transformational groups and semantic 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 methods with the probability theory for the development of novel parallel computational architectures utilising algebraic theory of transformational groups and semantic groups. Later I introduced a hybrid connections production rule-based model and developed connectionist-based expert systems.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

I also developed a series of novel methods for transductive reasoning for personalized medicine that created new opportunities for the application of computational intelligence to personalized medicine.

Recent work on evolving spiking neural networks (eSNN) with applications for spatio-temporal pattern recognition, ultramodal audiovisual information processing, and taste recognition (Proc. ICONIP 2007-2011; IEEE WCCI, 2010; Neural Networks 2010). The proposed computational neuro-genetic models for brain knowledge discovery was also proposed.

Integrative connectionist, genetic, and quantum-inspired methods are now being developed and applied to various problems, including personalized medicine. The brain knowledge discovery was also proposed.

INNS AWARD ACCEPTANCE STATEMENT:
Nikola Kasabov - INNS Gabor Awardee

It is my great honor to receive the top INNS Gabor Award for my 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 a novel parallel computational architecture utilizing algebraic theory of transformational groups and semantic 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 methods with the probability theory for the development of novel parallel computational architectures utilising algebraic theory of transformational groups and semantic groups. Later I introduced a hybrid connections production rule-based model and developed connectionist-based expert systems.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

In the late 1990s I developed a number of published models and systems for higher brain functions, such as Evolving Fuzzy Neural Networks (EFuNN, 2001) and their applications were published in a number of applications going beyond usual neuro-fuzzy models of that time.

I also developed a series of novel methods for transductive reasoning for personalized medicine that created new opportunities for the application of computational intelligence to personalized medicine.

Recent work on evolving spiking neural networks (eSNN) with applications for spatio-temporal pattern recognition, ultramodal audiovisual information processing, and taste recognition (Proc. ICONIP 2007-2011; IEEE WCCI, 2010; Neural Networks 2010). The proposed computational neuro-genetic models for brain knowledge discovery was also proposed.

Integrative connectionist, genetic, and quantum-inspired methods are now being developed and applied to various problems, including personalized medicine. The brain knowledge discovery was also proposed.
systems; connectionist-based models for bioinformatics gene and protein data analysis; methods for cancer drug target discovery; ecological data modeling; neuroinformaton 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 thank Dr. Robert E. Ozma, who served as my advisor to write this letter. Dr. Leonid Perlovsky, who was my postdoctoral advisor at the University of Memphis and Dr. Leonid Perlovsky, who was my postdoctoral advisor 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 into 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 developed by Dr. Walter J. Freeman. K sets are a hyperarchical 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 this multi-sensory processing and thus can be applied to adaptive control problems. The K sets are described by non-linearly coupled second order differential equations. The systemic presence of independent parameters which affect its dynamic properties. I conducted analytical studies to identify structural stability regions of the K sets and various bifurcation types occurring in the behavior of the K sets. Such studies contribute to the understanding of mechanisms used to generate a periodic 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 concerns the Model-Action-Critic networks which are the solution of the challenging part of control as it has to approximate the autonomous agents. The Critic network is a model advocated by Dr. Paul Werbos and used for control of equations. The system contains hundreds of independent nonlinearly coupled second order differential equations. The results have been used to design a simple K model with a chaotic nature.

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 a magnitude. 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 brain processes information in efficient way. Its main feature is the process of transition from explicit to implicit data associations and parameter values. This logic 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 a long the lines of investigating the use of Dynamic Logic in the 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 my Ph.D. advisor, Dr. Denis C. O’Keeffe 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 Asby 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., be havioral, fMRI, genetics) and neural network modeling. My 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 a neural network model 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 theories of creative problem solving to be formulated with sufficient precision to allow for a neural network implementation based on the CLARION cognitive architecture.
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 a broad topic 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 of the abstracts program. Authors of 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 conference book) online 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 cooperation with the International Institute for Advanced Scientific Studies (IIASS) of Vietri sul Mare (Italy), and is a traditional event devoted to the discussion of novelties and innovations related to the 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. 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
**WIRN 2012 - First Call for Papers**

**22nd Italian Workshop on Neural Networks**

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

**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, Opportunistic Networks, Metabolic Networks, Bi-o-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 Ns, Intelligent Interactive Dialogue Systems, Embodied Conversational Agents, 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 a pdf format to “Premio Caianiello-WIRN 2012” c/o IASS before April 20, 2012 to the addresses (wirn2012@associazionesiren.org, iass.segreteria@tin.it). To participate, the Ph.D. degree has to be obtained before 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 website). Submissions forms can be downloaded from the SIREN website. For more information, contact the conference secretariat at 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, Università di Milano
- Simone Bassis, Università di Milano
- Anna Esposito, Seconda Università di Napoli
- Francesco Masulli, Università di Genova
- Francesco Carlo Morabito, Università Mediterranea di Reggio Calabria
- Francesco Palmieri, Seconda Università di Napoli
- Eros Pasero, Politecnico di Torino
- Stefano Squartini, Università Politecnica delle Marche
- Roberto Tagliaferri, Università di Salerno
- Aurelio Uncini, Università di Roma “La Sapienza”
- Salvatore Vitabile, Università 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
- Jose Principe, University of Florida at Gainesville, USA
- John G. Taylor, King's College London
- Alessandro Villa, Università 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