Chapter 5. Building the comparation network and generation of equivalence relation

Qualitative Modelling (Comparation Network showing a relationship)

One view of building a neural network performing the process of **comparation** is to find a network exhibiting the relation properties: reflexive, symmetric and transitive. This is the view we take. This is because at the stage of the process of comparation in the **Verstandes Actus** it performs logical comparison with obscure (not conscious) parástase. Comparison made by identifying particular equivalence classes implies the process is working with conscious parástase. Thus the first postulate is,

Proposition 1: Comparation in the synthesis of equivalence relation exhibits the reflexive and symmetric relations.

Such a network is said to synthesize a compatibility relation. As was discussed earlier, comparison always involves a minimum of two items or objects. Thus,

Proposition 2: The comparation process in Verstandes Actus involves a minimum of two inputs.

Thus, the basic model will involve logical comparison of two inputs. For convenience of illustration let each input appear as a retinal-map of pixels as shown below.

Figure 5.1. The basic model (a) receiving two comparands, each represented as retinal-map matrix sized 5 x 5.

The comparation process is in facet-B and hence is in the mathematical domain. The neural network behavior determines the relationship but does so without any a priori

objective knowledge. Hence, the determination (relationship or not) is made by feature definer and detector which is automatically generated. We do not build them into the model. Hence the third postulate is,

Proposition 3: Comparation compares features that are dynamically generated axiomatic features.

It was mentioned earlier that the comparation process works with obscure parástase (input) and its determination (output) is also an obscure parástase. **TRJ** (teleological reflective judgment) judges **expedience**. This gives us,

Proposition 4: The process of comparation is judged expedient when the act is not contrary to the **categorical imperative**, which regulated to achieve a state of equilibrium.

During the process of comparison when the neural network model reaches a reverberating or resonant state, we will say the model is in steady-state condition. We will consider a neural network in such a state to be in equilibrium and hence may be judged expedient.

In embedding field theory (EFT) there is a family of known and also yet to be discovered neural networks which can reach a resonant state [Wells, 2010]. They are called adaptive resonance theory (ART) networks. A typical anatomy of an ART resonator (ART-R) is shown in figure 5.2a. Based upon this basic anatomy the network model has two resonators, such that each receives a respective input element. Hence, the model receives two comparands (Fig.5.2b). However, for the process of comparation the two basic anatomies must interact and hence cannot be isolated subsystems (Fig.5.2c). Thus,

Proposition 5: The sub-processes for respective comparands are united within the process of comparation.

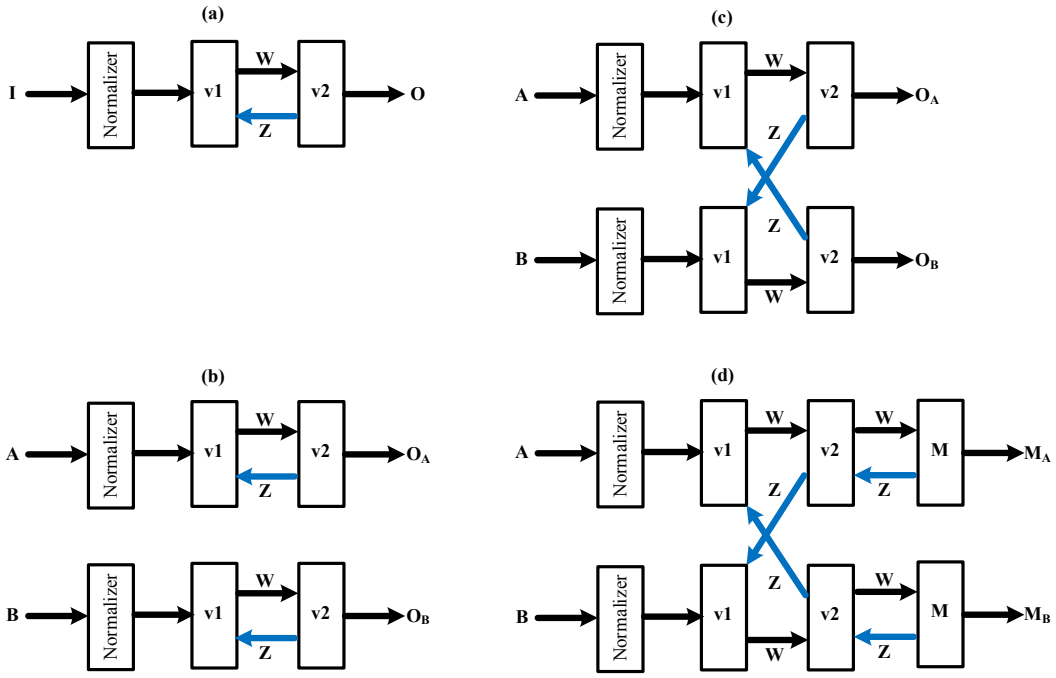


Figure 5.2. Illustration of the evolution from the basic ART-R (a) to the minimal neural network (d).

The basic ART-R (a) has a normalizer layer which then sends normalized inputs to the v1 layer. The v1 and v2 layers interact such that the nodes in v2 layers receives instar weights (W) and send out outstar weights to v1 layer. The interaction therefore plays a major role in achieving resonance.

Each ART just receiving respective comparand input (b) is insufficient. (c) Shows one approach to interaction between the two sub-networks. Here, the outstar-weighted output of the v2 layer of a sub-network feeds into the v1 layer of the other sub-network.

The comparation network is judged for expediency by interacting with M layer. Thus, this addition in the network proxies for part of TRJ. This also means that **nous-soma** connection is made. It should be noted that M_A and M_B are not motor responses but thought of as pre-motor images.

An obvious question for the above model is "how do we know that the process of comparation is expedient?" The model must have this property for at least two reasons. Firstly, since expediency is a **judicial imperative** the determination (relationship) by comparation will have no *practical* purpose if it is not expedient. Secondly, this is judged by TRJ which is a mediator between **nous** and **psyche** co-organizations. This means that the model must have link to **motoregulatory expression**. Thus,

Proposition 6: The process of comparation is judged by TRJ (teleological reflective judgment).

In the model (Fig.5.2c) the judgment for resonance and hence expediency is evaluated by the addition to two more fields (Fig.5.2d). These additions are proxies for motoregulatory *emotivity* such that when they reach steady-state the process of comparation may be judged expedient. It should be noted that there is no longer a clear distinction between the process of comparation and the ability to judge expediency (a functional part of TRJ).

Comparation is a process within **sensibility** but sensibility does not judge, does not deceive and does not confuse[1] [Kant, 1798]. Recall that the **synthesis in sensibility** is transcendental, i.e., necessary for the possibility of experience. Thus the activities of the motor end of the network do not represent motor response of the **OB** but are akin to pre-motor activities. We may therefore consider that if the steady-state values of the pre-motor responses are within a solution set, the comparands have a relationship. Thus,

Proposition 7: The relationship determined by comparation is *practical* in the sense that it is not contrary to the OB's categorical imperative.

Thus meaning of the determined relationship is purely *practical*.

[1] The three anthropological characteristic of the senses may be summarized as:
senses do not confuse :- Sensuous representations comes before representations for understanding and relay themselves all together. Thus, sense perceptions (empirical parástase with consciousness) are 'inner appearances'. *senses do not control* :- Sensuous parástase offer themselves to understanding and do not have control over understanding. *senses do not deceive* :- They do not deceive because they do not judge.

Note that, Kantian anthropology is the science of man's actual behavior and has for its topic 'the subjective laws of free choice'. Free choice (arbitrum liberum) is the power of choice in an OB, which is the choice of freedom from having its (OB's) actions necessarily bound to determination by sensuous representations. This freedom is called 'practical autonomy'. Because actions are subject to conditions set in the OB's self-constructed manifold of practical rules the determined actions can opposes sensuous desire.

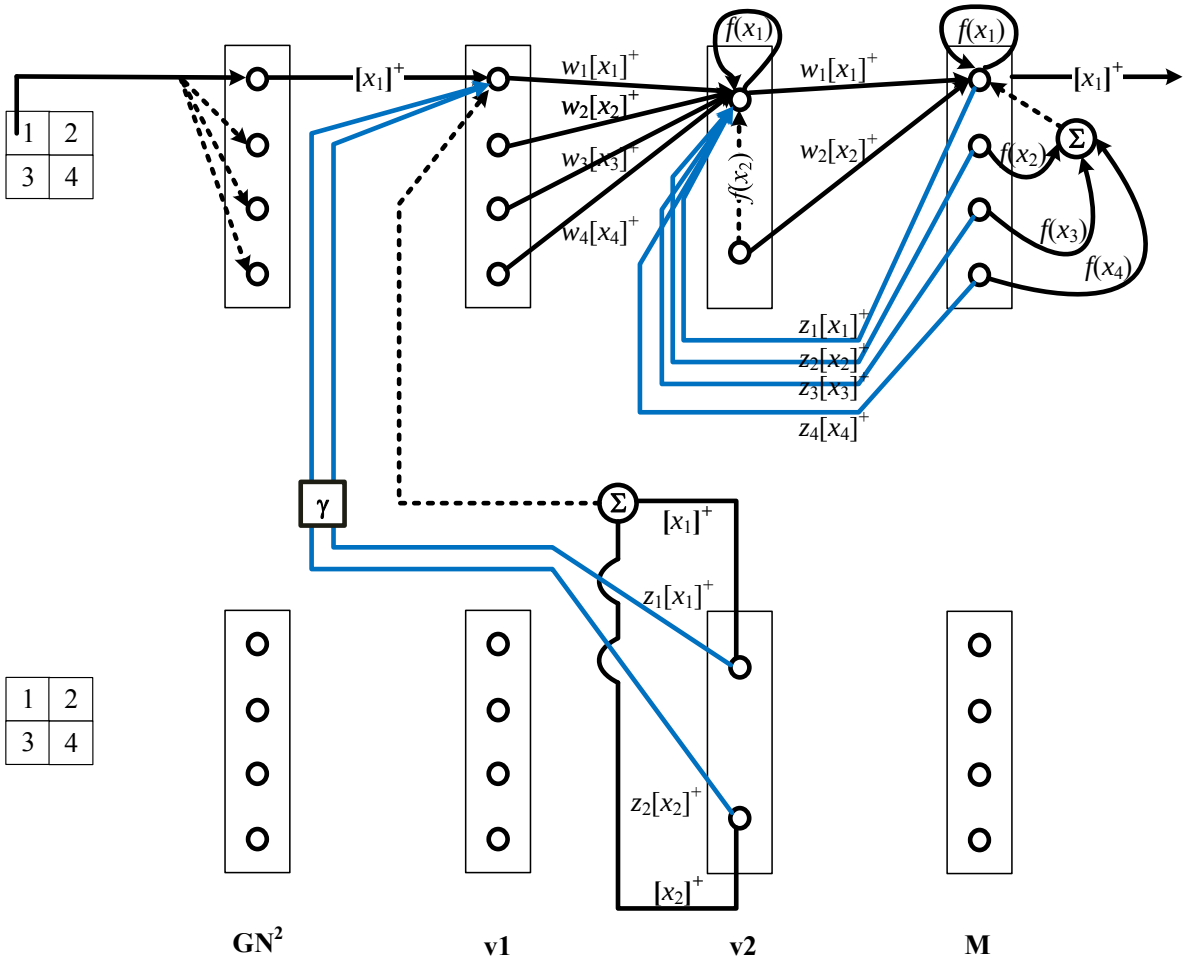


Figure 5.3. A more detailed view of the minimal neural network (Fig.5.2d). The illustration shows the connection of a single node in each layer. Given a retinal-map matrix sized 2 X 2, the first pixel excites (solid line) the first GN^2 node but inhibits (dashed lines) other nodes (of respective sub-network). The first v1 node (SNI^4) then receives excitatory inputs from the first normalized input and outstar-weighted output of all v2 nodes (other sub-network) with gain γ (solid blue line). The inhibition (dashed line) is the sum of these un-weighted v2node outputs.

The first v2 node (SNI^3) receives excitatory inputs from instar-weighted output of all v1 nodes (same sub-network), outstar-weighted output of all M nodes within the same sub-network (solid blue line) and from itself passed through the activation function ($f(\cdot)$). The rest of the v2 nodes within the same sub-network passed through $f(\cdot)$ are added to form the inhibitory input to the first v2 node. 0-1 distribution is enforced for v2 layer.

The first M node is the same kind of SNI^3 as the v2 nodes but without 0-1 distribution. Thus, its excitatory and inhibitory inputs follows a similar pattern as above but with different connectivity. The node does have any excitatory input from outstar-weighted output.

Note that x's and weights w's or z's are labeled the same for each layer for simplicity. However, they are quantitatively different and hence don't represent for instance the same excitation.

Quantitative Model (Comparation)

The quantified (detailed) view of the figure 5.2d is shown in figure 5.3. The ART network has nodes which in general are called shunting node instars (SNI). There are numerous types of SNI's, each having a different behavior from another [Wells, 2010]. This is consistent with EFT and MMA. The general description is given above (Fig. 5.3). Below are the mathematical expressions specific to the model.

The inputs or comparands are first normalized by normalizers called a Grossberg Normalizer-2 (GN^2) [Wells, 2010]. The practical importance of having a normalizer in ART networks is elaborated in Wells' text [Wells, 2010]. The i^{th} GN^2 node is given by the differential equation,

$$\dot{x}_i^{\mathrm{GN}} = -A_o x_i^{\mathrm{GN}} + (B_o - x_i)I_i - (C_o + x_i^{\mathrm{GN}}) \sum_{\forall k \neq i} I_k,$$

where, A_o, B_o and C_o are non-zero parameters such that, $B_o = (n-1)C_o$. As mentioned above the comparand is a retinal-map and hence they are matrix of I_i pixels. There is a corresponding GN^2 node for each pixel. This means there will be a total of n GN^2 nodes for a retina matrix of a rows and b columns ($n = a \cdot b$). In the above equation, $I = \sum_{\forall k} I_k$ and $\omega_i = \frac{I_i}{I}$.

The normalized inputs are then received by respective v1-layer. Like the normalizer layer the v1-layer has n nodes. This layer is comprised of SNI's called SNI^4 [Wells, 2010]. It's i^{th} node is given by,

$$\dot{x}_i^{\mathrm{v1}} = -A_1 x_i^{\mathrm{v1}} + (B_1 - x_i^{\mathrm{v1}})J_i^{+} - (D_1 + x_i^{\mathrm{v1}})J^{-},$$

where, A_1, B_1 and D_1 are non-zero parameters. The excitatory $J_i^+$and inhibitory J^- inputs are given by,

$$J_i^+ = x_i^{GN} + \gamma Z_{\text{v2 to v1}}[\vec{\mathbf{x}}^{\text{v2}}]^+ \quad \text{and} \quad J^- = \sum_{\forall k}[x_k^{\text{v2}}]^+,$$

where, γ is a non-zero parameter. The Heaviside extractor $[\cdot]^+$is given by,

$$[H]^+ = \begin{cases} H \; if \; H \geq 0 \\ 0 \; else \end{cases}.$$

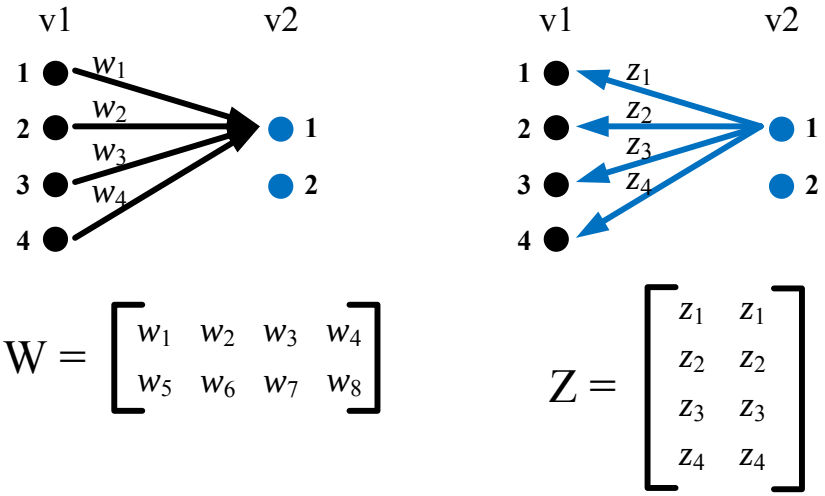


Figure 5.4. Illustration of instar (W) and oustar (Z) weights. Notice that the first row elements of W matrix corresponds to instar weights going into first v2 node (blue). But first column elements of Z matrix corresponds to outstar weights going out of the first v2 node.

$Z_{\text{v2 to v1}}$ is the outstar weight matrix from v2-layer to v1. They go hand in hand with the instar weight matrix from v1 to v2, $W_{\text{v1 to v2}}$ (Fig.5.4). They are given by the form,

$$W_{\text{v1 to v2}}(t+h) = W_{\text{v1 to v2}}(t) - \eta \cdot \left[(W_{\text{v1 to v2}}(t) - [\vec{\mathbf{x}}^{\text{v1}^{\text{T}}}]^+) \cdot [\vec{\mathbf{x}}^{\text{v2}}]^+\right] \text{ and}$$
$$Z_{\text{v2 to v1}}(t+h) = Z_{\text{v2 to v1}}(t) - \eta \cdot [(Z_{\text{v2 to v1}}(t) - [\vec{\mathbf{x}}^{\text{v1}}]^+) \cdot [\vec{\mathbf{x}}^{\text{v2}}]^+],$$

where, η is the adaptation parameter and h is step size. The property of outstar weights in an ART network is that, the weights from the winning v2 nodes is an exemplar of the input.

The nodes in the v1-layer interact with v2-layer nodes. But unlike the v1-layer, the number of nodes in v2 is not restricted by the size of the retina matrix. This also means that the number of nodes in the succeeding M-layer is unrestricted. SNI[3] was chosen for the v2 nodes. They are expressed by,

$$\dot{x}_i^{\mathrm{v2}} = -A_2 + (B_2 - x_i^{\mathrm{v2}})\xi_i^{Ex} - (x_i^{\mathrm{v2}} + D_2)\xi_i^{In},$$

where, A_2, B_2 and D_2 are non-zero parameters. The excitatory ξ_i^{Ex} and inhibitory ξ_i^{In} inputs are given by,

$\xi_i^{Ex} = W_{\mathrm{v1\ to\ v2}} \cdot [\vec{\mathbf{x}}^{\mathrm{v1}}]^+ + Z_{\mathrm{M\ to\ v2}} \cdot [\vec{\mathbf{x}}^{\mathrm{M}}]^+ + f(x_i^{\mathrm{v2}})$ and
$\xi_i^{In} = \sum_{\forall k \neq i} f(x_k)$.

It has been observed that adaptation in most ART networks necessitates a 0-1 distribution in the v2-layer [Wells, 2010]. Therefore, a 0-1 distribution is enforced for the vector, $\vec{\mathbf{x}}^{\mathrm{v2}}$ and the winning v2 node controls network adaptation.

The function $f(\cdot)$ is called the activation shape function. It is given by,

$$f(y) = \begin{cases} 0 & if\ y < 0 \\ \lambda \cdot y & if\ y \leq u^{(1)} \\ g_{max} \cdot y & if\ u^{(1)} < y \leq u^{(2)} \\ g_{max} \cdot u^{(2)} & if\ y > u^{(2)} \end{cases}.$$

where $u^{(1)}$ and $u^{(2)}$ are parameters for the activation shape function. The contrast enhancement capability of some ART networks may attenuate the absolute levels of x_i signals. Proper scaling will not alter the performance of the network apart from limiting the attenuation of the absolute levels [Wells, 2010]. Using the B parameter as the reference this is done here as,

$g_{max} = B \cdot g_{max}^{unscaled}$,

$u^{(1)} = \frac{\left(B - \frac{A}{g_{max}}\right)}{\left(1 - \frac{A}{g_{max}}\right)} \cdot u_{unscaled}^{(1)}$,

$u^{(2)} = B \cdot u_{unscaled}^{(2)}$ and

$\lambda = \frac{g_{max}}{u^{(1)}}$.

where, $g_{max}^{unscaled}$, $u_{unscaled}^{(1)}$ and $u_{unscaled}^{(2)}$ are the unscaled parameters.

The nodes for the M-layer use SNI[3] and hence is similar to the v2-layer. However, M-layer has fair distribution since 0-1 distribution is not enforced. Also, the parameters may be different and because the connections are different the expression of excitatory and inhibitory input are different.

Therefore, the minimal neural network is given by,

$x_i^{\mathrm{M6}}(t+h) = \left[1 - h \cdot \left(A_{\mathrm{M}} + \xi_i^{Ex\_\mathrm{M6}} + \xi_i^{In\_\mathrm{M6}}\right)\right] \cdot x_i^{\mathrm{M6}} + h \cdot \left(B_{\mathrm{M}} \cdot \xi_i^{Ex\_\mathrm{M6}} - D_{\mathrm{M}} \cdot \xi_i^{In\_\mathrm{M6}}\right)$,

$x_i^{\mathrm{M5}}(t+h) = \left[1 - h \cdot \left(A_{\mathrm{M}} + \xi_i^{Ex\_\mathrm{M5}} + \xi_i^{In\_\mathrm{M5}}\right)\right] \cdot x_i^{\mathrm{M5}} + h \cdot \left(B_{\mathrm{M}} \cdot \xi_i^{Ex\_\mathrm{M5}} - D_{\mathrm{M}} \cdot \xi_i^{In\_\mathrm{M5}}\right)$,

$x_i^{\mathrm{v2\_4}}(t+h) = \left[1 - h\left(A_2 + \xi_i^{Ex\_4} + \xi_i^{In\_4}\right)\right] x_i^{\mathrm{v2\_4}} + h\left(B_2 \xi_i^{Ex\_4} - D_2 \xi_i^{In\_4}\right)$,

$x_i^{\mathrm{v2\_3}}(t+h) = \left[1 - h\left(A_2 + \xi_i^{Ex\_3} + \xi_i^{In\_3}\right)\right] x_i^{\mathrm{v2\_3}} + h\left(B_2 \xi_i^{Ex\_3} - D_2 \xi_i^{In\_3}\right)$,

$x_i^{\mathrm{v1\_2}} = \frac{B_1 J_i^{\mathrm{v1\_2+}} - D_1 J^{\mathrm{v1\_2-}}}{A_1 + J_i^{\mathrm{v1\_2+}} + J^{\mathrm{v1\_2-}}}$ and

$x_i^{\mathrm{v1\_1}} = \frac{B_1 J_i^{\mathrm{v1\_1+}} - D_1 J^{\mathrm{v1\_1-}}}{A_1 + J_i^{\mathrm{v1\_1+}} + J^{\mathrm{v1\_1-}}}$.

Rather than numerically solving difference equations, steady-state solution is used for the v1 layer. This is because the v1 layer in ART is considered to be faster than v2 layer. The weights (instar & outstar) for interaction between v1 – v2 and v2 – M are adaptive. The respective expressions for the excitation and inhibitory inputs are,

$\xi_i^{Ex\_\mathrm{M6}} = W_{\mathrm{v2\_4\ to\ M6}} \cdot [\vec{\mathbf{x}}^{\mathrm{v2\_4}}]^+ + f(x_i^{\mathrm{M6}}), \quad \xi_i^{In\_\mathrm{M6}} = \sum_{\forall k \neq i} f(x_k^{\mathrm{M6}});$

$\xi_i^{Ex\_\mathrm{M5}} = W_{\mathrm{v2\_3\ to\ M5}} \cdot [\vec{\mathbf{x}}^{\mathrm{v2\_3}}]^+ + f(x_i^{\mathrm{M5}}), \quad \xi_i^{In\_\mathrm{M5}} = \sum_{\forall k \neq i} f(x_k^{\mathrm{M5}});$

$\xi_i^{Ex\_4} = W_{\mathrm{v1\_2\ to\ v2\_4}} \cdot [\vec{\mathbf{x}}^{\mathrm{v1\_2}}]^+ + Z_{\mathrm{M6\ to\ v2\_4}} \cdot [\vec{\mathbf{x}}^{\mathrm{M6}}]^+ + f(x_i^{\mathrm{v2\_4}}), \quad \xi_i^{In\_4} = \sum_{\forall k \neq i} f(x_k^{\mathrm{v2\_4}});$

$\xi_i^{Ex\_3} = W_{\mathrm{v1\_1\ to\ v2\_3}} \cdot [\vec{\mathbf{x}}^{\mathrm{v1\_3}}]^+ + Z_{\mathrm{M5\ to\ v2\_3}} \cdot [\vec{\mathbf{x}}^{\mathrm{M5}}]^+ + f(x_i^{\mathrm{v2\_3}}), \quad \xi_i^{In\_3} = \sum_{\forall k \neq i} f(x_k^{\mathrm{v2\_3}});$

$J_i^{\mathrm{v1\_2}+} = x_i^{\mathrm{GN\_P2}} + \gamma Z_{\mathrm{v2\_3\ to\ v1\_2}} [\vec{\mathbf{x}}^{\mathrm{v2\_3}}]^+, \quad J^{\mathrm{v1\_2}-} = \sum_{\forall k} \left[x_k^{\mathrm{v2\_3}}\right]^+$ and

$J_i^{\mathrm{v1\_1}+} = x_i^{\mathrm{GN\_P1}} + \gamma Z_{\mathrm{v2\_4\ to\ v1\_1}} [\vec{\mathbf{x}}^{\mathrm{v2\_4}}]^+, \quad J^{\mathrm{v1\_1}-} = \sum_{\forall k} \left[x_k^{\mathrm{v2\_4}}\right]^+.$

The instar (W) and outstar (Z) weights are,

$W_{\mathrm{v2\_4\ to\ M6}}(t+h) = W_{\mathrm{v2\_4\ to\ M6}} - \eta \cdot \left[(W_{\mathrm{v2\_4\ to\ M6}} - [\vec{\mathbf{x}}^{\mathrm{v2\_4}^{\mathrm{T}}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{M6}}]^+\right],$

$Z_{\mathrm{M6\ to\ v2\_4}}(t+h) = Z_{\mathrm{M6\ to\ v2\_4}} - \eta \cdot \left[(Z_{\mathrm{M6\ to\ v2\_4}} - [\vec{\mathbf{x}}^{\mathrm{v2\_4}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{M6}}]^+\right],$

$W_{\mathrm{v2\_3\ to\ M5}}(t+h) = W_{\mathrm{v2\_3\ to\ M5}} - \eta \cdot \left[(W_{\mathrm{v2\_3\ to\ M5}} - [\vec{\mathbf{x}}^{\mathrm{v2\_3}^{\mathrm{T}}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{M5}}]^+\right],$

$Z_{\mathrm{M5\ to\ v2\_3}}(t+h) = Z_{\mathrm{M5\ to\ v2\_3}} - \eta \cdot \left[(Z_{\mathrm{M5\ to\ v2\_3}} - [\vec{\mathbf{x}}^{\mathrm{v2\_3}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{M5}}]^+\right],$

$W_{\mathrm{v1\_2\ to\ v2\_4}}(t+h) = W_{\mathrm{v1\_2\ to\ v2\_4}} - \eta \cdot \left[(W_{\mathrm{v1\_2\ to\ v2\_4}} - [\vec{\mathbf{x}}^{\mathrm{v1\_2}^{\mathrm{T}}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{v2\_4}}]^+\right],$

$W_{\mathrm{v1\_1\ to\ v2\_3}}(t+h) = W_{\mathrm{v1\_1\ to\ v2\_3}} - \eta \cdot \left[(W_{\mathrm{v1\_1\ to\ v2\_3}} - [\vec{\mathbf{x}}^{\mathrm{v1\_1}^{\mathrm{T}}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{v2\_3}}]^+\right],$

$Z_{\mathrm{v2\_4\ to\ v1\_1}}(t+h) = Z_{\mathrm{v2\_4\ to\ v1\_1}} - \eta \cdot \left[(Z_{\mathrm{v2\_4\ to\ v1\_1}} - [\vec{\mathbf{x}}^{\mathrm{v1\_1}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{v2\_4}}]^+\right]$ and

$Z_{\mathrm{v2\_3\ to\ v1\_2}}(t+h) = Z_{\mathrm{v2\_3\ to\ v1\_2}} - \eta \cdot \left[(Z_{\mathrm{v2\_3\ to\ v1\_2}} - [\vec{\mathbf{x}}^{\mathrm{v1\_2}}]^+) \cdot [\vec{\mathbf{x}}^{\mathrm{v2\_3}}]^+\right].$

The weights are initialized such that, outstar weights (Z's) are set at zero. However, instar weights (W's) are set as $w_{ij} = \sigma \bullet \mathrm{rand}(0,1) + \sigma$, where rand(0,1) is a random number from uniform distribution [0,1] and σ is the initialization scale.

Finally the normalization is done only once for a given pattern. Thus its steady-state form is,

$x_i^{\mathrm{GN\_P1}} = \frac{nC_oI}{A_o+I}\left(\omega_i - \frac{1}{n}\right)$ and $x_i^{\mathrm{GN\_P2}} = \frac{nC_oI}{A_o+I}\left(\omega_i - \frac{1}{n}\right).$

The synthesis of sensibility is the noetic process that synthesizes **apprehension** in **consciousness**. Thus the OB has no conscious experience of these acts of synthesis. In other words, comparation does not have memory. Weight changes in W and Z are elastic modulations. This implies,

Proposition 8: The process of comparation does not remember its past actions.

Hence, the states of the instar and outstar weights of the above minimal neural network are not stored for succeeding acts of logical comparison. Therefore this network does not have the problem of stability-plasticity trade-off encountered in some ART networks [Wells, 2010].

Behavior of the proposed minimal neural network

The behavior of the network (Fig.5.3) was observed from simulation experiments for various set of patterns. The pattern sets used for the experiments are: uppercase English letters (A to Z), uppercase English letters versus its modified patterns (translation, rotation), 2x resolution uppercase English letters and finally Arabic numerals (0 to 9). Regardless of pattern set (experiment case), the simulations were done using the same parameters (Fig.5.5).

Unlike v2-layers, the M-layers do not have 0-1 distribution. Thus, for analysis, the activity of the average (median) node of M-layer is considered. The set-membership paradigm was employed to consider whether activities of the two M-layers differ and hence whether the network determines a relationship between the input patterns. Thus identification of feasible set solutions rather than a single “point” solution is considered.

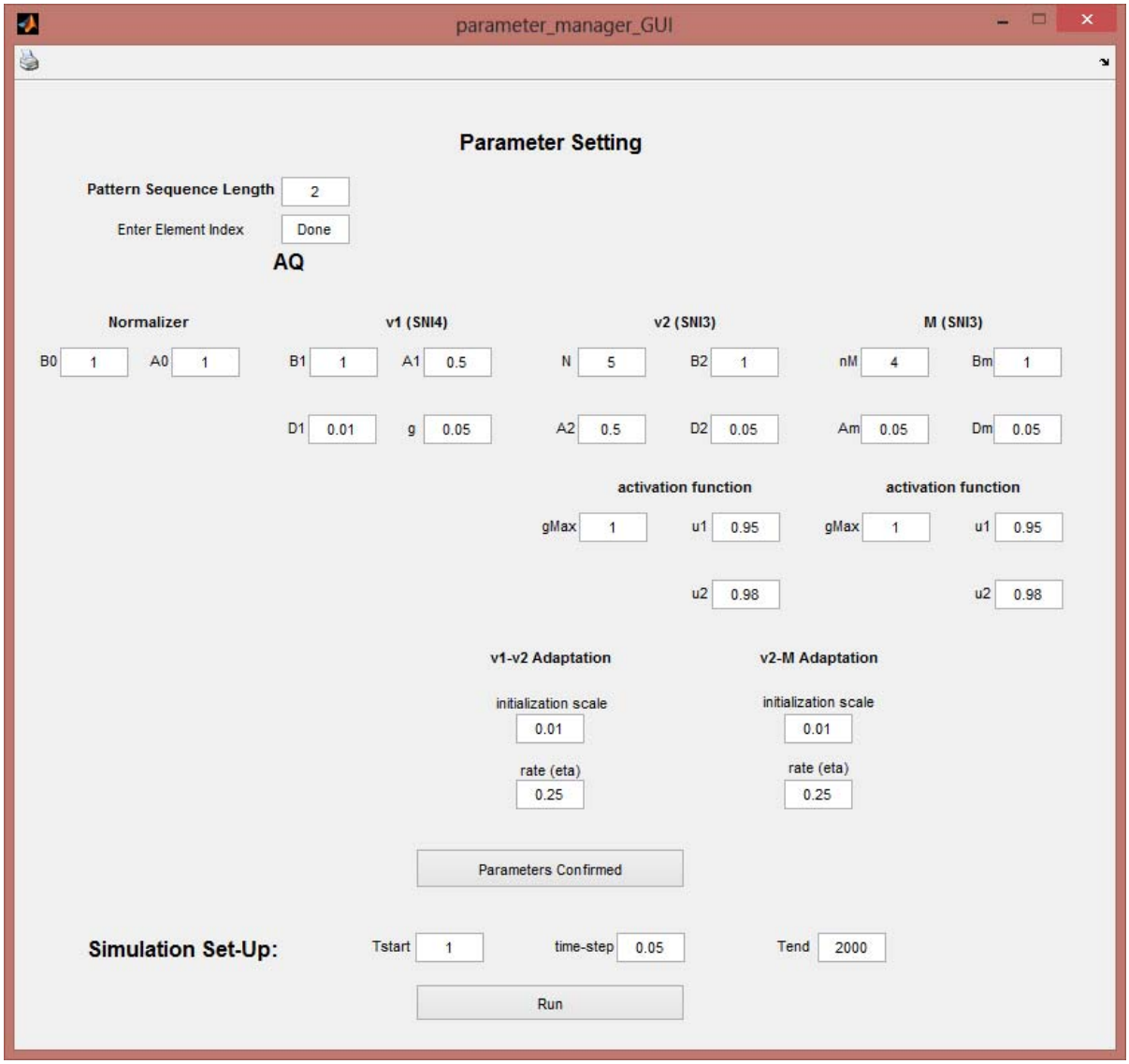


Figure 5.5. Parameters for the simulation. For a particular pattern-combo (say, AQ) in Case-1 (uppercase English letters), two (or more) desired patterns are chosen by entering their indices ranging from [1, 26] (Fig.5.6). The respective parameters are then entered. Note that $g = \gamma$ (Fig.5.3) and the activation function parameters are the unscaled values. Scaling is done as described above within the main code. Both sub-networks use same parameter values.

Though the weights of the network are not stored they are initialized for each run of pattern-combo. This is done by setting the outstar weight matrix elements at zero but the elements of the instar weight matrices are set at some very small (non-zero) random real number. The latter amount is given by initialization scale.

As governed by the functional principles of ART networks [Wells, 2010], weight adaptation occurs after resonance of the network.

Finally, the time-step (h) for the difference equations was chosen to be 0.05 but iterations are variable (usually, 2000, 3000 or 6,000 iterations).

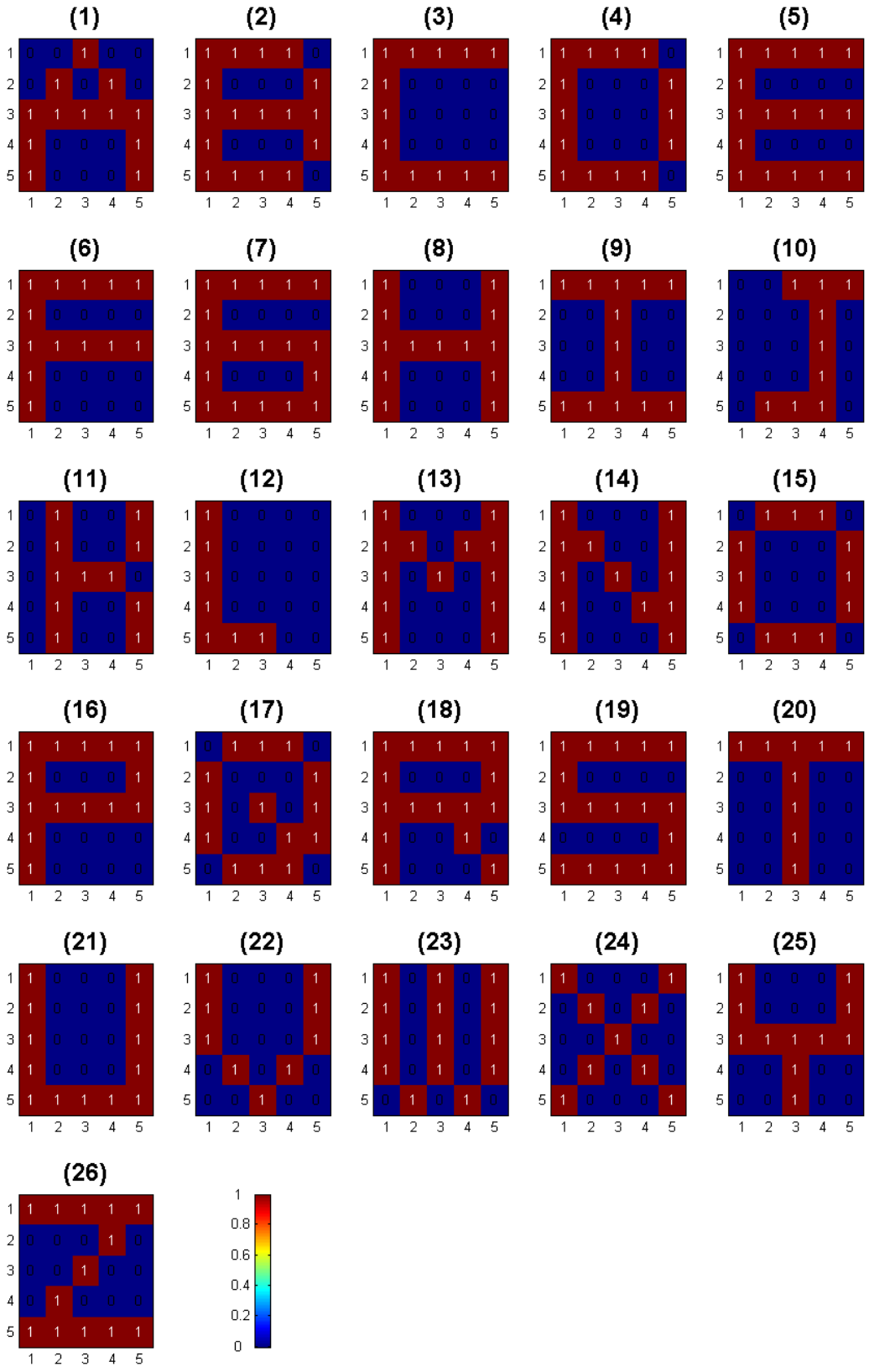


Figure 5.6. Set of upper-case English letters used for observing the network behaviors. Each alphabet is represented by the respective placement of 1's in a background of 0's.

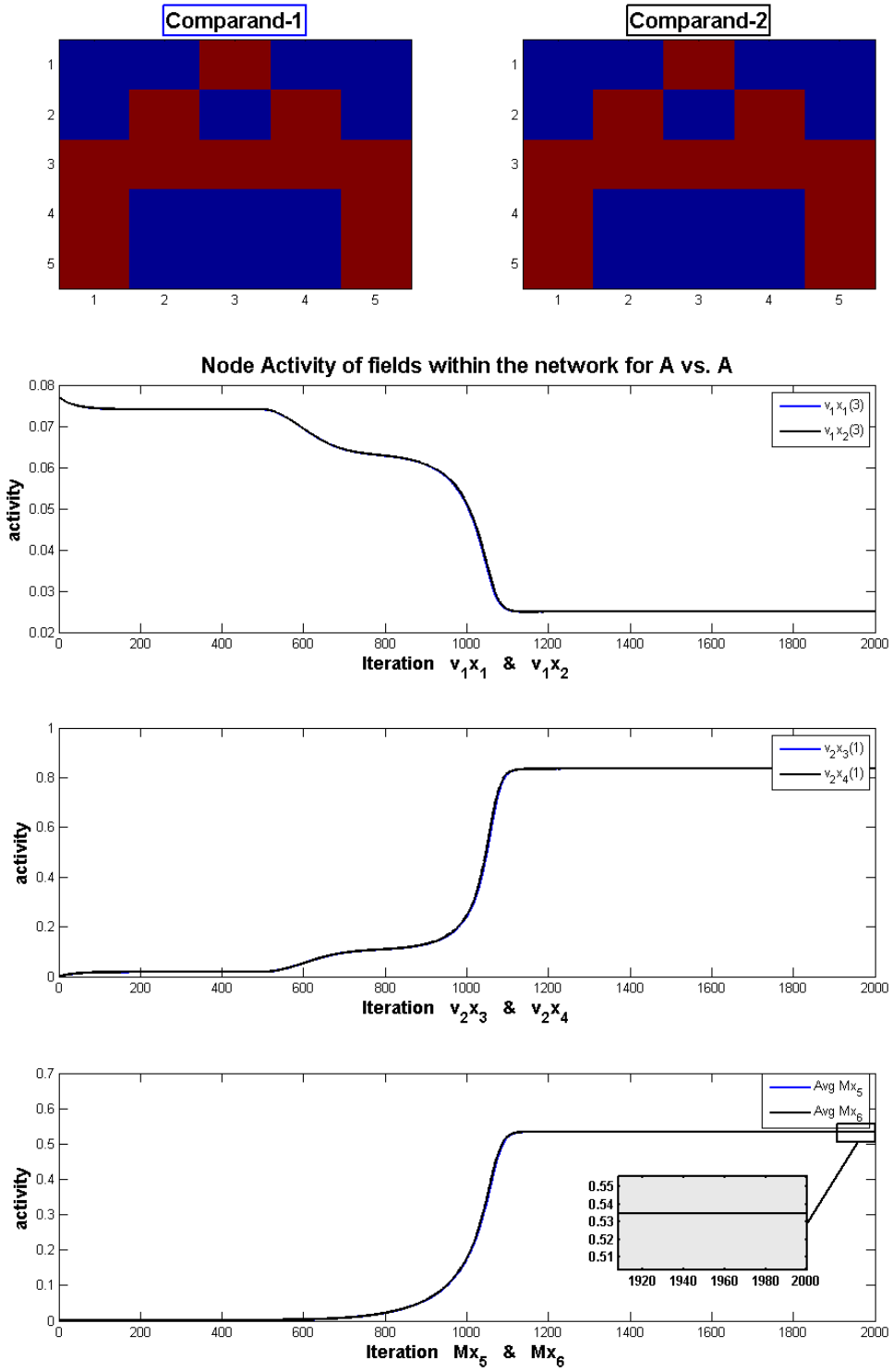


Figure 5.7. Time-series plot for patterns A versus A. The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. The subplots shows the activities of node-1 in respective v1-layers and activities of winning nodes in respective v2-layers. The bottom subplot shows the activities of average (median) nodes of respective M-layers. The inset magnifies the activities of the two M-layers overlapping (≈ 0.5347) each other at final iteration.

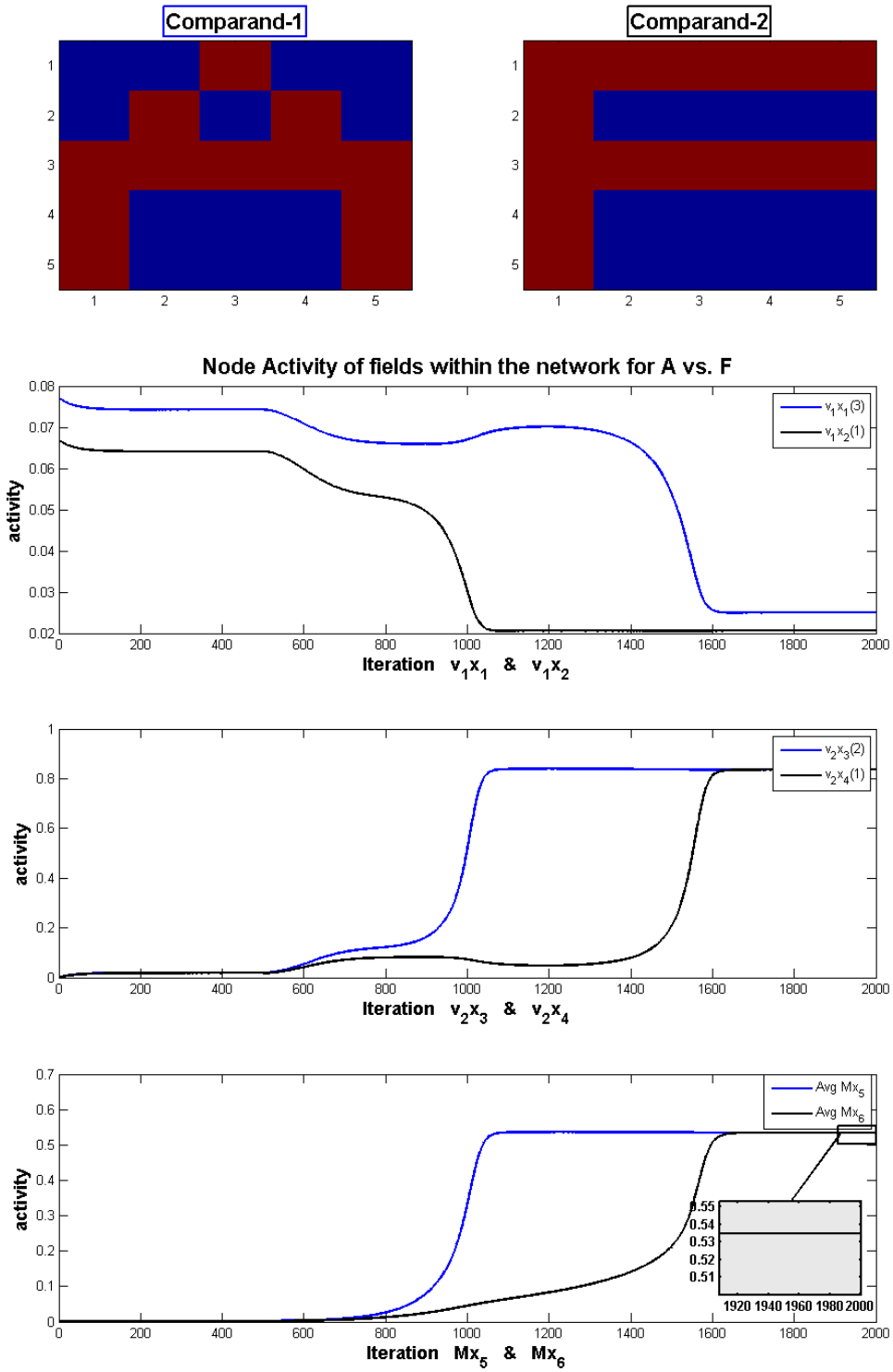


Figure 5.8. Time-series plot for patterns A versus F. The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. The subplots shows the activities of node-1 in respective v1-layers and activities of winning nodes in respective v2-layers. The bottom subplot shows the activities of average (median) nodes of respective M-layers. The inset magnifies the activities of the two M-layers at final iteration which appears to overlap (0.5346 ≈ 99.9% of 0.5347).

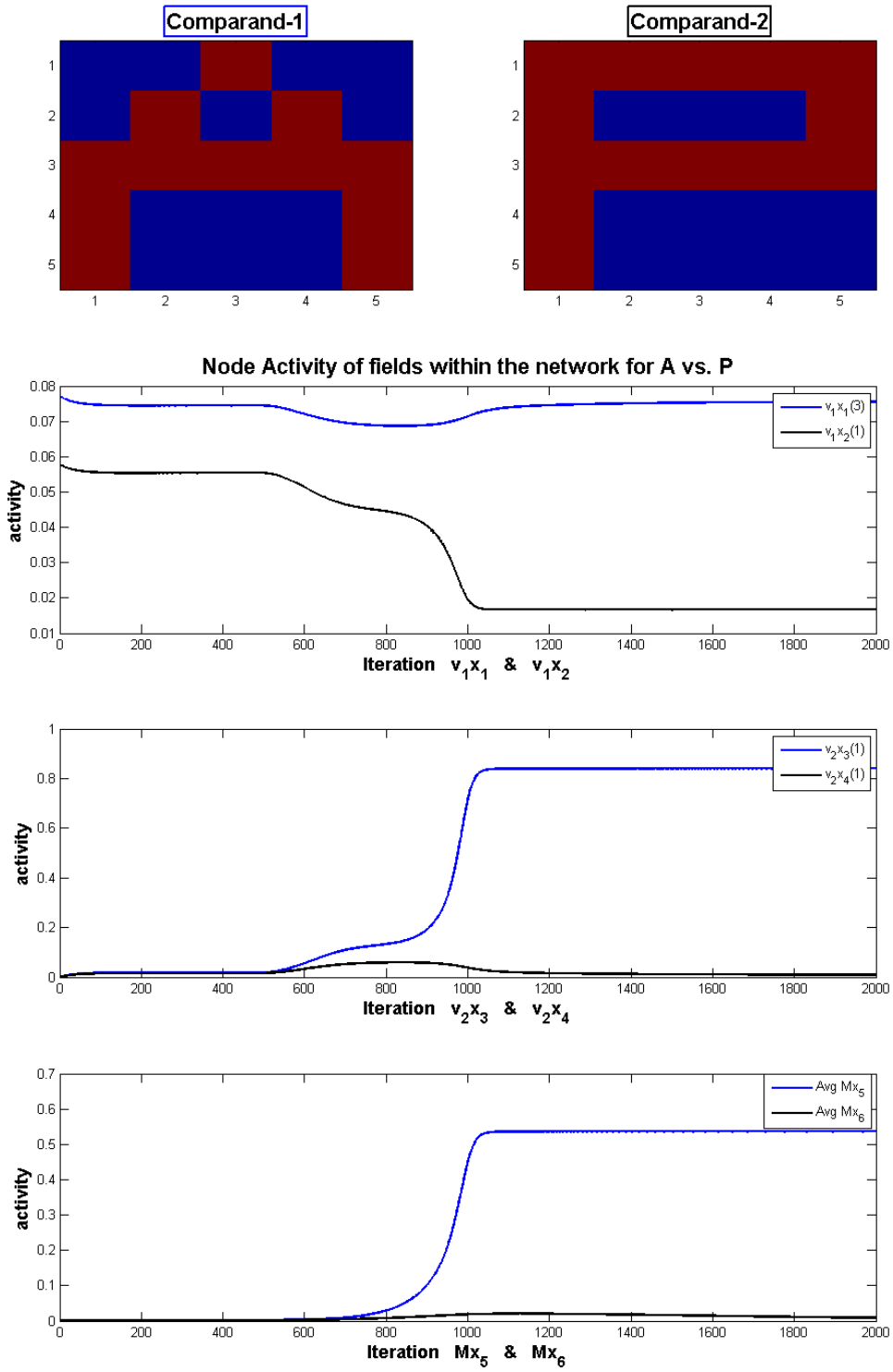


Figure 5.9. Time-series plot for patterns A versus P. The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. The subplots shows the activities of node-1 in respective v1-layers and activities of winning nodes in respective v2-layers. The bottom subplot shows the activities of average (median) nodes of respective M-layers. Notice that the M-layer activities at final iteration are clearly far apart (0.0016 ≈ 0.3% of 0.5358).

Case-1, uppercase English letters (A to Z):

Network testing involved pattern-combos with two patterns in a combo such that each pattern is picked from the set of letters (Fig.5.6). It was found that the network determined relationships between some patterns and not for others. Determination of relationship depends on the expediency judged by TRJ and so the region of interest is the steady-state levels (at final iteration) of the M-layer outputs.

We say that the network determines a relationship between the patterns if the steady-state of the M-layer output falls within the same solution set. The solution set is defined such that the activity with smaller magnitude (if unequal) is ≥ 90% of the larger. In other words, the solution set is defined to contain all the "not-noticeably" different activities and hence indistinguishable patterns.

Figures 5.7 and 5.8 show that relationship was found for pattern A with itself and between patterns A and F. The steady states of M-layer average activities indicate that they fall within the solution set. On the other hand, figure 5.9 demonstrates the network's ability of finding no relationship between patterns A and P. The determination of relationship or no relationship for all the pattern-combos is shown in table 5.1.

The table shows that the network finds relationship for a pattern versus itself. It also shows that in some cases multiple patterns demonstrate relationship with the same collection of patterns. For instance, patterns A and O demonstrate relationship with any patterns within the set {A, C, F, H, I, K, M, N, O, U, Y, Z}. Notice that the determined outcome (relationship or no) for a particular pattern pair is unchanged when the pattern inputs (comparands 1 & 2) are reversed. Thus, relationship is seen in both A vs. F and F vs. A.

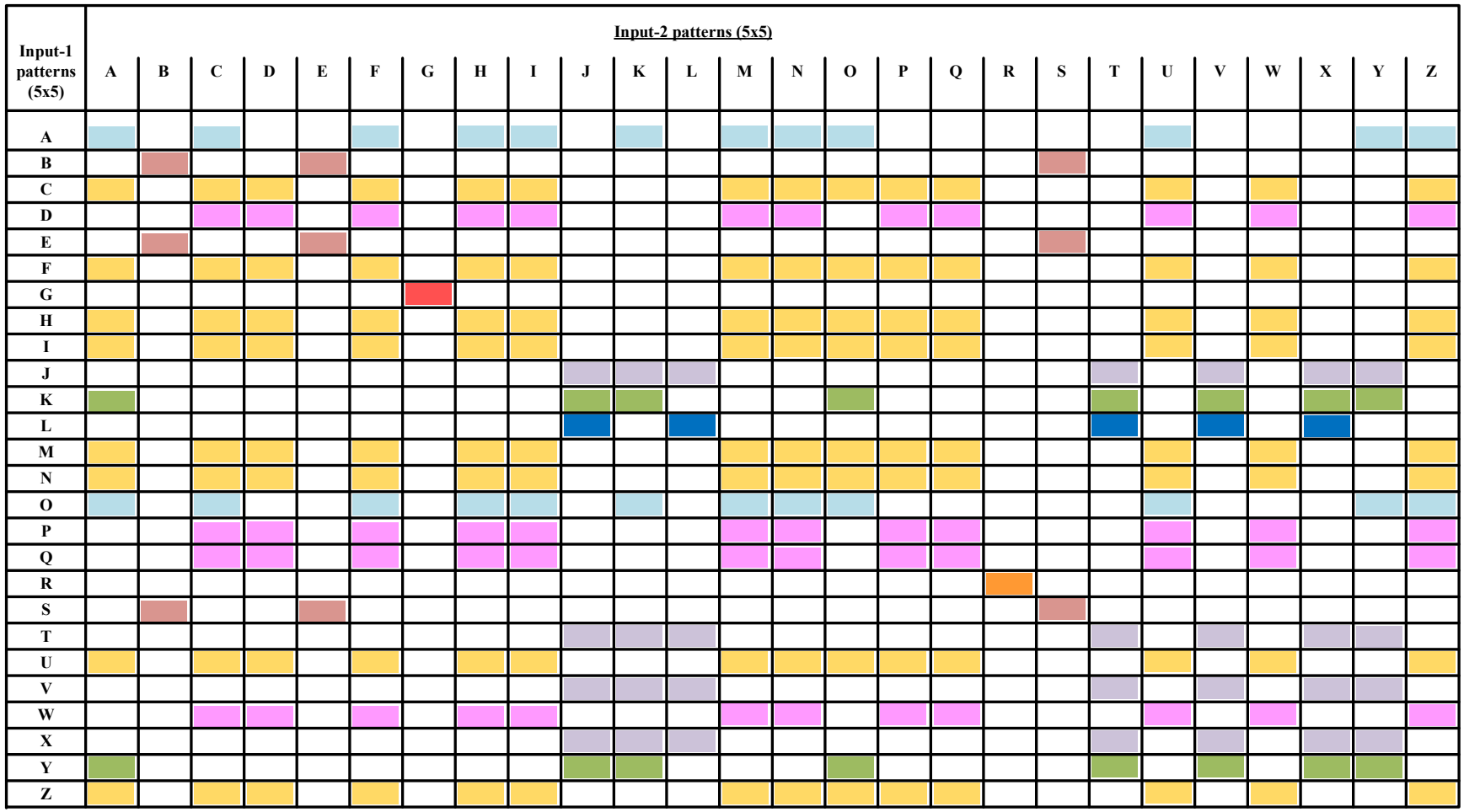


Table 5.1. Table illustrating similarities for Capital English letters, A to Z vs. A to Z. For a particular row, the colored (same) boxes represent similarity (for e.g. row for A). When compared to A to Z letters, if two or more rows have the same pattern of similarity, they are depicted having same colored boxes (for e.g. row for A & row for O). Two letters are defined similar if the smaller average (median) node activity of the two M-fields is ≥ 90% of the larger. This is equivalent to < 10% vigilance percentage. Note that the uncolored (white) boxes implies no similarity. The raw numerical data is given in appendix, Tables A1-1 to A1-8.

Observe that relationship is seen for A vs. C and C vs. D but not for A vs. D (table 5.1). Transitivity is therefore absent. Such state of affairs has been empirically observed in children by Piaget [Piaget, 1953a]. For instance, child (aged around 8 – 9 years) may say that a metal bar A weighs the same as another bar C, and that bar B weighs the same as metal ball D, A = C and C = D. But from past experience the child expects the relation, A < D. The child therefore says "bar B is definitely as heavy as ball D, but it will be different with A" and concludes that A ≠ D.

On the other hand, relationship is seen for A vs. C, C vs. F and A vs. F (table 5.1). Piaget has also accounted such transitive properties in children [Piaget, 1953b]. Given a tower of blocks on a table the child (aged around 6 years) is asked to build a second tower of the same height on another (lower or higher) table with block of different size. The rule of the game forbids the child from moving the tower. The child having built the tower F, builds another (third) tower C and moves it (which is allowed by the rules) over to the reference tower A. The child moves the tower C back and forth adjusting its height to A and then F. The child presupposes that C = F and C = A therefore A = F.

The network can therefore demonstrate relationship between some patterns and non-relationship for others. This ability of the network has been achieved without a priori definition of what a feature is. In other words, the network self-defines feature sets. This then begets the question, what is it that the network considers "features" and thereby recognizes a common feature among patterns shown to have a relationship? Though, this question would be a topic for future research, based on empirical observations a possible explanation would be suggested here.

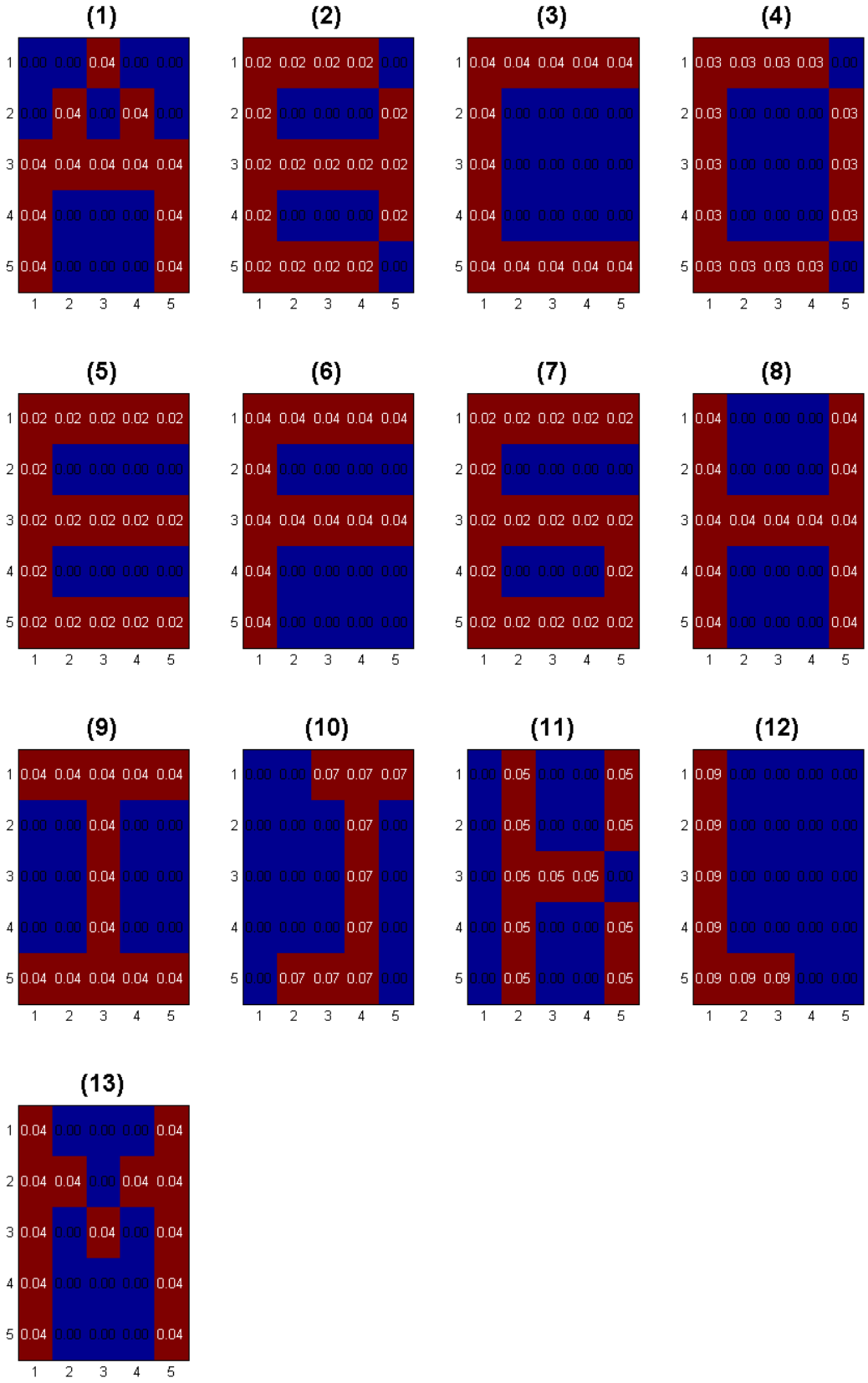


Figure 5.10. Normalized patterns (A to M) of Fig.5.6.

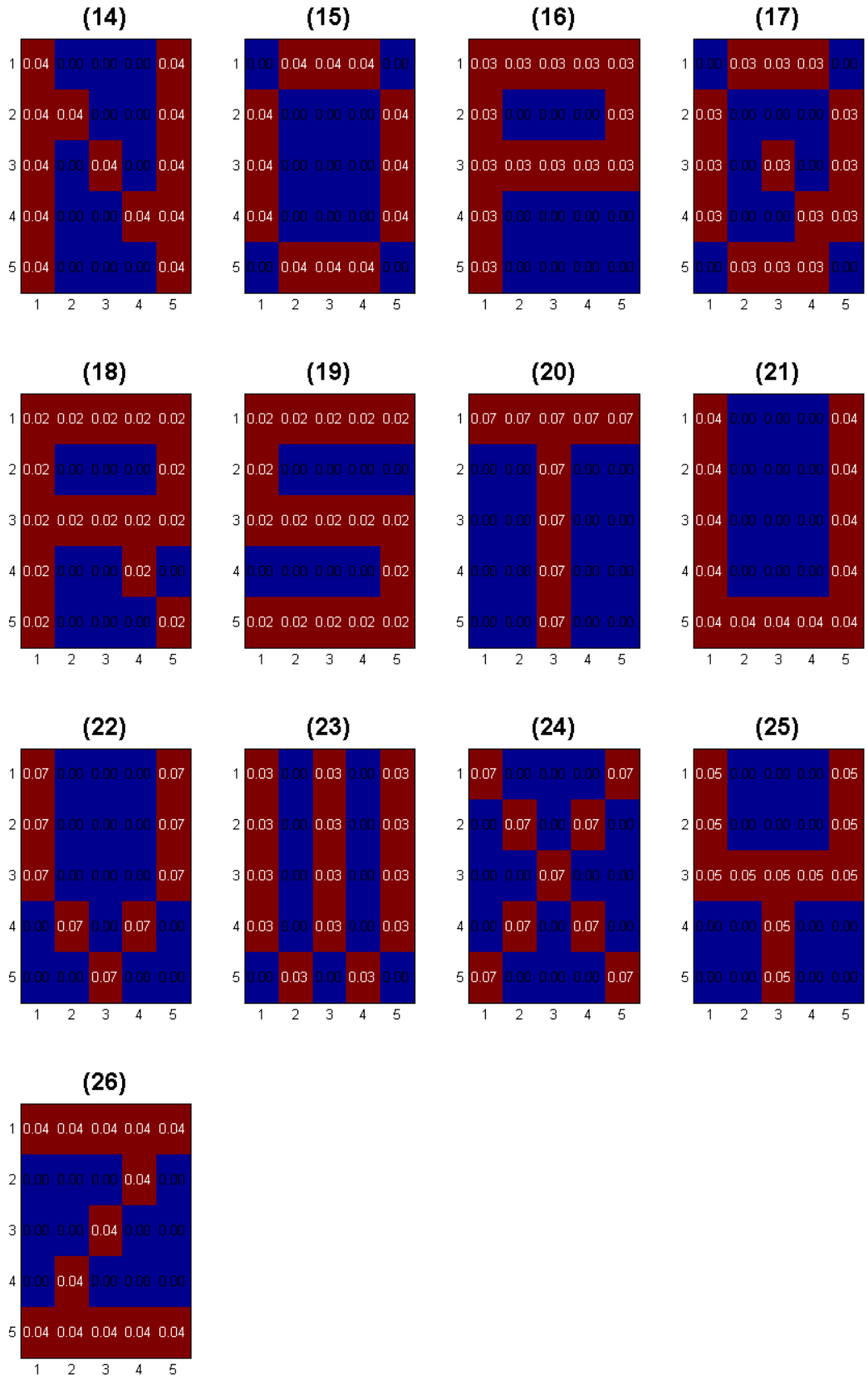


Figure 5.11. Normalized patterns (N to Z) of Fig.5.6.

Number of non zero pixels	Non zero pixel value	Cumulative pixel value	Patterns (5x5)
18	0.0154	0.2772	G
17	0.0185	0.3145	B, E, S
16	0.0221	0.3536	R
14	0.0306	0.4284	D, P, Q, W
13	0.0357	0.4641	C, F, H, I, M, N, U, Z
12	0.0417	0.5004	A, O
11	0.0486	0.5346	K, Y
9	0.0667	0.6003	J, T, V, X
7	0.0937	0.6559	L

Table 5.2. Table showing the 26 uppercase English letters arranged with respect to their number of non-zero pixels (out of 25 pixels, 5 x 5 matrix), their normalized value (2nd column) and their total (3rd column, CPV or cumulative pixel value). Note that for a particular pattern all its non-zero pixels have the same magnitude of normalized pixel value.

Subset of comparand-1 patterns (5x5)	Subset of comparand-2 patterns (5x5)
{G}	{G}
{B, E, S}	{B, E, S}
{R}	{R}
{D, P, Q W}	{C, D, F, H, I, M, N, P, Q, U, W, Z}
{C, F, H, I, M, N, U, Z}	{A, C, D, F, H, I, M, N, O, P, Q, U, W, Z}
{A, O}	{A, C, F, H, I, K, M, N, O, U, Y Z}
{K, Y}	{A, J, K, O, T, V, X, Y}
{J, T, V, X}	{J, K, L, T, V, X, Y}
{L}	{J, L, T, V, X}

Table 5.3. Table of subset of patterns derived from results in Table 5.1. Relationship is found for a given comparand-1 pattern with any comparand-2 pattern within a particular (same row) subset.

Comparand-1 Patterns (5x5)	Comparand-2 Patterns (5x5)	Largest difference w/ α	Smallest difference w/o α
G	{G}	{G} α {G} : 0(0) ⟹ 0.0000	{G} α̸ {B, E, S} : 0.1186(1) ⟹ 0.0474
B, E, S	{B, E, S}	{B, E, S} α {B, E, S} : 0(0) ⟹ 0.0000	{B, E, S} α̸ {R} : 0.1105(1) ⟹ 0.0442
R	{R}	{R} α {R} : 0(0) ⟹ 0.0000	{R} α̸ {B, E, S} : 0.1105(1) ⟹ 0.0442
D, P, Q, W	{C, D, F, H, I, M, N, P, Q, U W, Z}	{D, …, W} α {C, …, Z} : 0.0769(1) ⟹ 0.0307	{D, …, W} α̸ {A, O} : 0.1439(2) ⟹ 0.0460
C, F, H, I, M, N, U, Z	{A, C, D, F, H, I, M, N, O, P, Q, U, W, Z}	{C, …, Z} α {D, …, W} : 0.0769(1) ⟹ 0.0307	{C, …, Z} α̸ {K, Y} : 0.1319(2) ⟹ 0.0422
A, O	{A, C, F, H, I, K, M, N, O, U, Y, Z}	{A, O} α {C, …, Z} : 0.0725(1) ⟹ 0.0290	{A, O} α̸ {J, …, X} : 0.1664(3) ⟹ 0.0416
K, Y	{A, J, K, O, T, V, X, Y}	{K, Y} α {J, …, X} : 0.1094(2) ⟹ 0.0350	{K, Y} α̸ {C, …, Z} : 0.1319(2) ⟹ 0.0422
J, T, V, X	{J, K, L, T, V, X, Y}	{J, …, X} α {K, Y} : 0.1094(2) ⟹ 0.0350	{J, …, X} α̸ {A, O} : 0.1664(3) ⟹ 0.0416
L	{J, L, T, V, X}	{L} α {J, …, X} : 0.0847(2) ⟹ 0.0271	{L} α̸ {K, Y} : 0.1849(4) ⟹ 0.0425

Table 5.4. Table demonstrating the diff-pc metric for the case of comparing patterns from 26 Capital English letters of retinal map size, 5x5. The first two columns shows the relationship (α) found for a given comparands-1 pattern (column-1) with any comparands-1 pattern (column-2) within a particular (same row) subset (table 5.3).

Every patterns has a corresponding cumulative pixel value, CPV (table 5.2). If two patterns have different CPV's, U and L such that U > L. Then difference-factor, diff-fact = (U – L) / U and difference-percentage, diff-pc = diff-fact • ppx. Every unique difference of number of non-zero pixels, px has an empirically determined parameter, ppx. Thus (px) → ppx. The ppx values are: (1) → 0.4, (2) → 0.32, (3) → 0.25, (4) → 0.23, (5) → 0.22, (6 to 10) → 0.8, (11 to 20) → 0.8•2 = 1.6, (21 to 30) → 0.8•3 = 2.4, (31 to 40) → 0.8•4 = 3.2, (41 to 50) → 0.8•5 = 4, (51 to 60) → 0.8•6 = 4.8 and so on…

The third column shows the diff-fact, px (bracket) and resultant largest diff-pc amongst comparands-1 patterns and subset of comparands-2 patterns that has demonstrated α is < 4%. The fourth column shows the smallest diff-pc amongst those that has not demonstrated α is > 4%. The diff-pc for patterns that demonstrates α and α̸ was determined to be largest or smallest from the table containing all the computed diff-pc values (tables A2-1 & A2-2).

Figures 5.10 and 5.11 shows the outputs of the Grossberg normalizers (GN^2, Fig.5.3) receiving respective pattern (Fig.5.6). They are therefore the normalized input to the network. Based on these normalized patterns, table 5.2 is constructed.

The table shows the grouping of patterns based upon their number of pixels (out of 25 pixels) that are not zero magnitude. The non-zero magnitudes are the normalized values (2^{nd} column of the table). It also shows a column whose entry is the sum of normalized pixel values. We shall call this, the cumulative pixel value (CPV). Thus, CPV is a function of the number of non-zero pixels. Notice that some patterns (like patterns: B, E, S) have equal number of non-zero pixels (17) and also equal CPV (0.3145).

Table 5.3 is however a summarized or condensed table 5.1. It shows pattern groups (comparands-1) that has demonstrated relationship with any patterns within respective set (comparands-2). Thus pattern D demonstrates relationship with any pattern from the pattern set for comparands-2, {C, D, F, H, I, M, N, P Q, U, W, Z}. Furthermore, this relationship is also found for any pattern for the pattern set containing D for comparands-1, {D, P, Q, W}. Notice that the pattern sets for comparands-1 correspond to patterns having equal number of non-zero pixels.

Let us assume that the number of non-zero pixels plays an important role on how the network defines feature sets. We shall therefore consider CPV since it is a function of this number. Let us now define difference-factor (diff-fact) to be the factor of U, such that diff-fact • U = U – L where, U and L are the upper (larger magnitude) and lower CPV respectively. Finally, we define difference-percentage (diff-pc) between two patterns to be the product, diff-fact • ppx where, ppx is the parameter corresponding to px, difference of number of non-zero pixels. The ppx value is empirically determined (table 5.4).

If we consider diff-pc to be a possible metric for feature, we observe that the network finds relationship among patterns such that diff-pc < 4% (table 5.4). On the other hand, among patterns with non-relationship, the smallest (magnitude) diff-pc appears to be always > 4%.

We shall continue to consider diff-pc as the possible measure for what the network considers as "feature" for the remaining cases. Thus discussions on the diff-pc will continue for the subsequent cases.

Case-2, uppercase English letters (A to Z) versus its modified patterns:

Let us now consider the case of patterns taken from uppercase English letters (case-1) versus its modifications. The modifications considered are translation and rotation.

Figures 5.12 and 5.13 respectively show that relationship was found for pattern A with itself (translated, rotated or both) and between A and F (translated, rotated or both). This outcome of finding relationships is similar to the earlier case (Fig.5.7 & 5.8).

On the other hand, figure 5.14 (unlike Fig.5.9) demonstrates relationship between patterns A and P. This observed behavior seems to be linked with the relationship found amongst patterns A and translated-A or translated-F (Fig.5.12b & Fig.5.13b). Note that for test with rotated patterns, the retinal-map size is 5x5 (Fig.5.12a, Fig.5.13a & Fig.5.14a). However the retinal-map is larger (10x10 in Fig.5.12b, Fig.5.12c, Fig.5.13b, Fig.5.13c, Fig.5.14b & Fig.5.14c) when the patterns are translated.

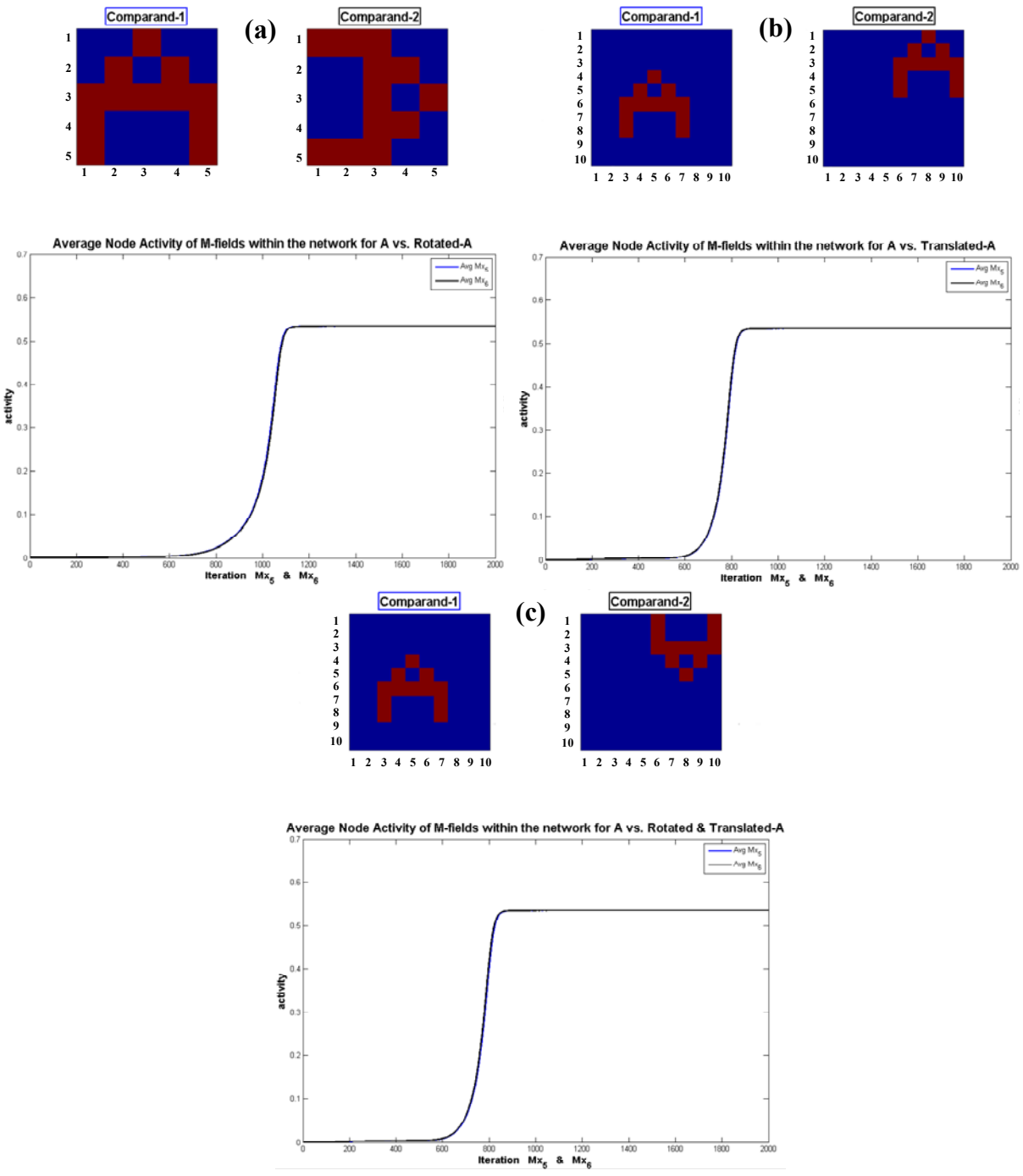


Figure 5.12. Time-series plot for patterns A versus modified-A. The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. Given a pattern-combo (a, b or c) the plots shows the activities of average (median) nodes of respective M-layers. Note that the retinal-map size is 5x5 for (a) and 10x10 for (b) and (c).

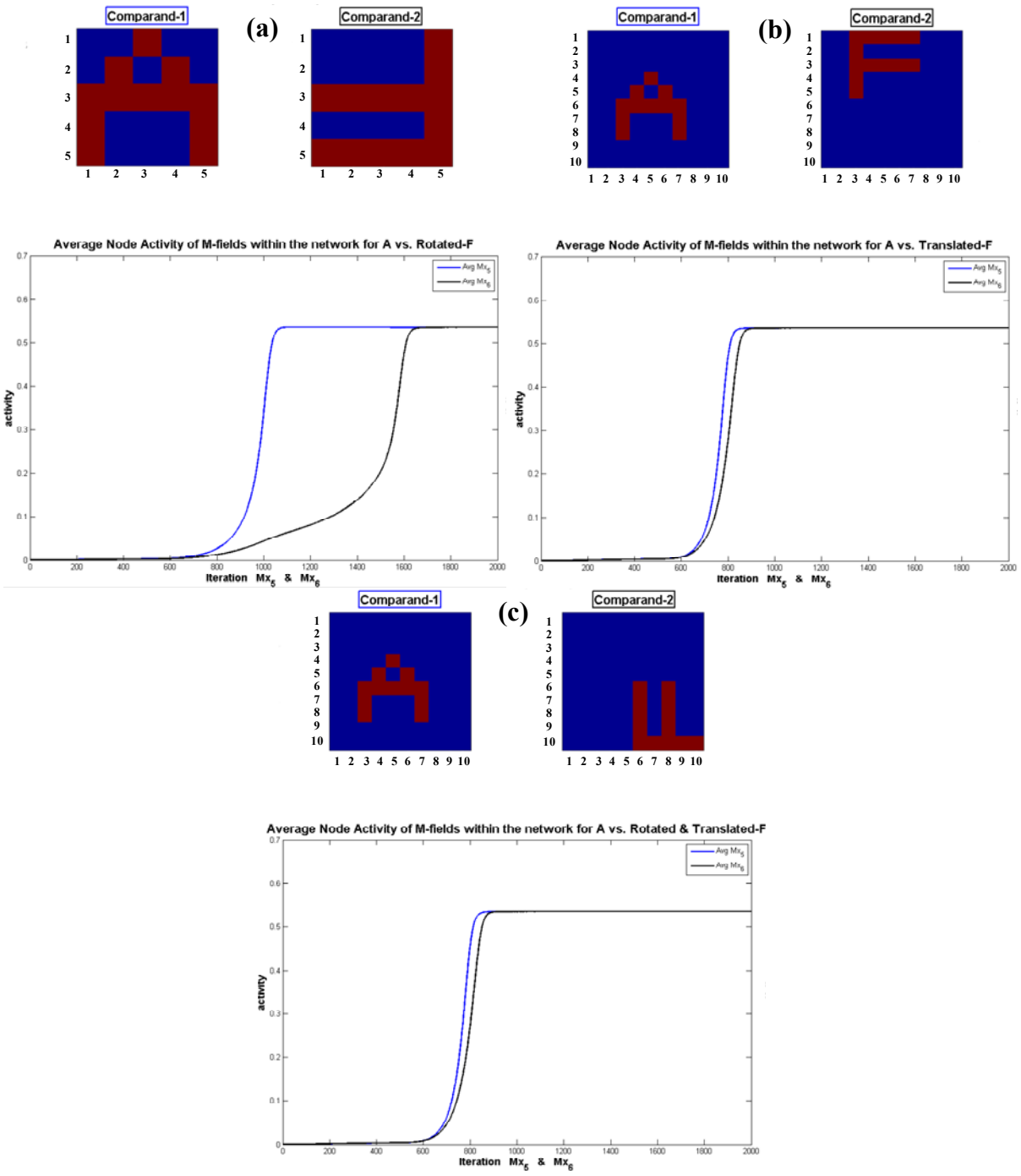


Figure 5.13. Time-series plot for patterns A versus modified-F. The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. Given a pattern-combo (a, b or c) the plots shows the activities of average (median) nodes of respective M-layers. Note that the retinal-map size is 5x5 for (a) and 10x10 for (b) and (c).

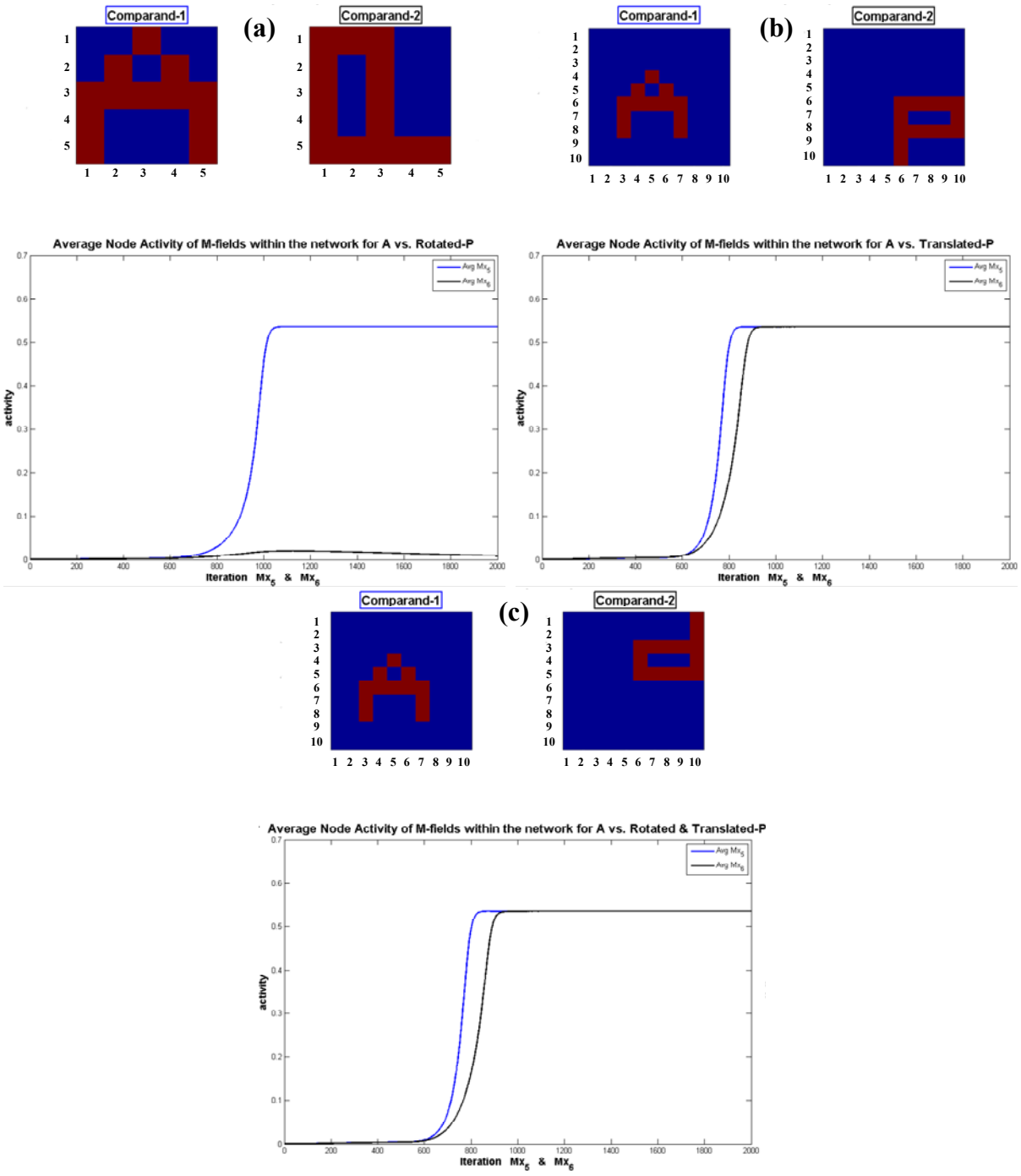


Figure 5.14. Time-series plot for patterns A versus modified-P. The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. Given a pattern-combo (a, b or c) the plots shows the activities of average (median) nodes of respective M-layers. Note that the retinal-map size is 5x5 for (a) and 10x10 for (b) and (c).

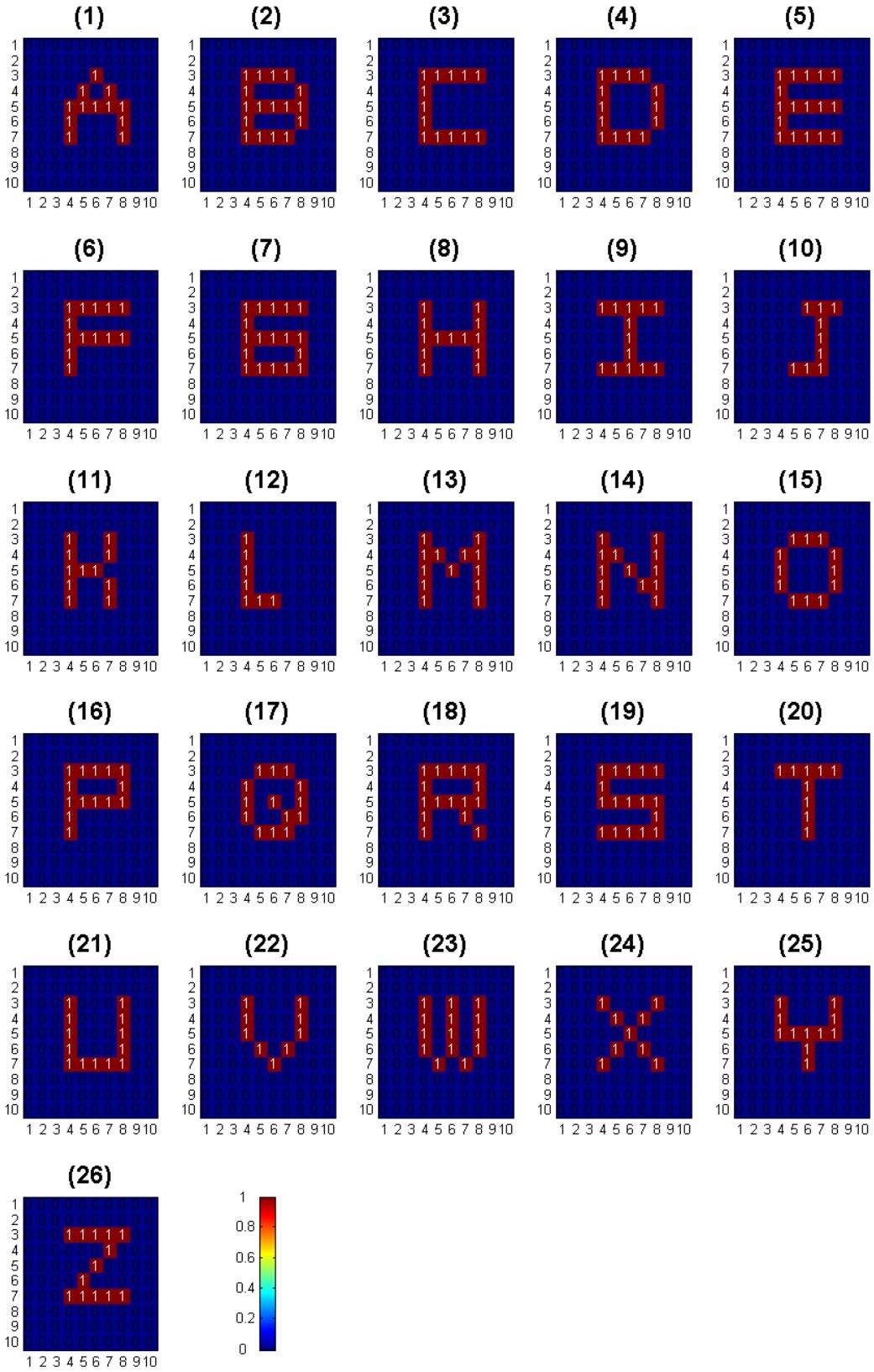


Figure 5.15. Set of upper-case English letters but with 10x10 retinal-map size. Each alphabet is represented by the respective placement of 1's in a background of 0's.

Number of non zero pixels	Non zero pixel value	Cumulative pixel value	Patterns (10x10)
18	0.0436	0.7848	G
17	0.0466	0.7922	B, E, S
16	0.0499	0.7984	R
14	0.0579	0.8685	D, P, Q, W
13	0.0628	0.8164	C, F, H, I, M, N, U, Z
12	0.0684	0.8208	A, O
11	0.0749	0.8239	K, Y
9	0.0919	0.8271	J, T, V, X
7	0.1174	0.8218	L

Table 5.5. Table showing the 26 uppercase English letters arranged with respect to their number of non-zero pixels (out of 100 pixels, 10 x 10 matrix), their normalized value (2nd column) and their total (3rd column, CPV or cumulative pixel value). Note that for a particular pattern all its non-zero pixels have the same magnitude of normalized pixel value.

Subset of comparand-1 patterns (10x10)	Subset of comparand-2 patterns (10x10)
{G}	{A, …, Z}
{B, E, S}	{A, …, Z}
{R}	{A, …, Z}
{D, P, Q W}	{A, …, Z}
{C, F, H, I, M, N, U, Z}	{A, …, Z}
{A, O}	{A, …, Z}
{K, Y}	{A, …, Z}
{J, T, V, X}	{A, …, Z}
{L}	{A, …, Z}

Table 5.6. Table of subset of patterns derived from simulation done with patterns from Capital English Letters of retinal size 10x10. Relationship is found for a given comparand-1 pattern with any comparand-2 pattern within a particular (same row) subset. The raw numerical data is given in appendix, Tables A3-1 to A3-8.

Comparand-1 Patterns (10x10)	Comparand-2 Patterns (10x10)	Largest difference w/ α	Smallest difference w/o α
G	{A, …, Z}	{G} α {J, …, X} : 0.0511(9) ⟹ 0.0102	-
B, E, S	{A, …, Z}	{B, E, S} α {J, …, X} : 0.0422(8) ⟹ 0.0084	-
R	{A, …, Z}	{R} α {J, …, X} : 0.0347(7) ⟹ 0.0069	-
D, P, Q, W	{A, …, Z}	{D, …, W} α {G} : 0.0318(4) ⟹ 0.0073	-
C, F, H, I, M, N, U, Z	{A, …, Z}	{C, …, Z} α {G} : 0.0387(5) ⟹ 0.0085	-
A, O	{A, …, Z}	{A, O} α {G} : 0.0439(6) ⟹ 0.0088	-
K, Y	{A, …, Z}	{K, Y} α {G} : 0.0475(7) ⟹ 0.0095	-
J, T, V, X	{A, …, Z}	{J, …, X} α {G} : 0.0511(9) ⟹ 0.0102	-
L	{A, …, Z}	{L} α {G} : 0.0450(11) ⟹ 0.0090	-

Table 5.7. Table demonstrating the diff-pc metric for the case of comparing patterns from 26 Capital English letters of retinal map size, 10x10. The first two columns shows the relationship (α) found for a given comparands-1 pattern (column-1) with any comparands-1 pattern (column-2) within a particular (same row) subset (table 5.6).

Every patterns has a corresponding cumulative pixel value, CPV (table 5.5). If two patterns have different CPV's, U and L such that U > L. Then difference-factor, diff-fact = (U – L) / U and difference-percentage, diff-pc = diff-fact • ppx. Every unique difference of number of non-zero pixels, px has an empirically determined parameter, ppx. Thus (px) → ppx. The ppx values are: (1) → 0.4, (2) → 0.32, (3) → 0.25, (4) → 0.23, (5) → 0.22, (6 to 10) → 0.8, (11 to 20) → 0.8•2 = 1.6, (21 to 30) → 0.8•3 = 2.4, (31 to 40) → 0.8•4 = 3.2, (41 to 50) → 0.8•5 = 4, (51 to 60) → 0.8•6 = 4.8 and so on…

The third column shows the diff-fact, px (bracket) and resultant largest diff-pc amongst comparands-1 patterns and subset of comparands-2 patterns that has demonstrated α is < 4%. Since all the patterns within the set demonstrated relationship there are no entries in the fourth column for smallest diff-pc amongst $\not\alpha$ is > 4%. The diff-pc for patterns that demonstrates α and $\not\alpha$ was determined to be largest or smallest from the table containing all the computed diff-pc values (tables A4-1 & A4-2).

The normalizer (GN^2) is a function of number of pixels of the retinal-map (n = 25 for 5x5 and n = 100 for 10x10). Thus, the normalized non-zero pixel value and hence CPV differ for respective size of retinal-map. Table 5.5 shows the normalized values and CPV of the translated patterns (Fig.5.12, Fig.5.13 & Fig.5.14) having 10x10 retinal-map size (Fig.5.15). Comparing the table with table 5.2 illustrates the difference.

Regardless of the retinal-map size, the number of non-zero pixels remain unchanged. However, the non-zero pixel values increases with the larger retinal-map size. Thus CPV's also increase for all the patterns. The result is that the spread of CPV's is narrowed with increased retinal-map size (variance = 0.0002 in 10x10 and variance = 0.0165 in 5x5 retinal-map size).

Table 5.6 shows the summarized outcome of the simulation done with pattern set made up of Capital English letters but with retinal-map size, 10x10. It shows that all patterns within the set demonstrates relationship with any pattern within the set. Applying our diff-pc as the possible metric for feature, one notices that all the diff-pc < 4% (table 5.7).

This result is interesting. It suggests a prediction made by the network. This is functionally similar to the situation when fonts "a" and "e" appear similar at a distance.

Various forms of acuity is observed, physiologically and behaviorally. The normal acuity in discriminating two point light source is about 25 seconds of arc [Guyton, pp.621, 2006]. Thus, two point light sources 10 meters away from the eye are distinguishable when they are 1.5 to 2 mm apart. Two point tactile discrimination test is commonly used for testing tactile acuity. Tactile acuity varies with skin region, skin thickness and condition. But in general, two points on skin are discriminated if they are 1 to 2 mm apart on fingers and 30 to

70 mm apart on the back [Guyton, pp.592, 2006]. Another form of acuity is the auditory temporal acuity, which is the sensitivity to amplitude modulation [Purcell & John, 2004]. It is the ability to discriminate rapid changes in sound envelope. Researchers have demonstrated that envelopes of speech signal in different spectral regions play a crucial role in speech understanding [Van Tasell et al., 1987; Shannon et al., 1995]. The above examples illustrates various forms of acuity.

The result that, all Capital English letters (5x5) when placed on a larger retinal-map size of 10x10 results in a relationship amongst all the patterns is therefore a demonstration of loss of acuity. We may refer to this loss of sharpness of discriminating patterns with increasing retinal-map size (but unchanged size of pattern) as the loss of sensory acuity.

Case-3, 2x resolution uppercase English letters (A to Z):

Continuing with the retinal-map size of 10x10 and patterns of Capital English letters, let us consider the case when these patterns fill the retinal-map (unlike Fig.5.16). Each pixels (in Fig.5.6 & Fig.5.15) is now represented by four pixels. In other words, they are 2x resolution of same pattern in 5x5 retinal-map.

Figures 5.17a and 5.17b shows relationship amongst patterns A (2x) with itself and also between A (2x) and O (2x). However no relationship was demonstrated amongst A (2x) and any other patterns within the set of 2x resolution Capital English letters (Fig.5.17c & Fig.5.17d). The determination of relationship or no relationship for all the pattern combos is shown in table 5.8. The condensed form (table 5.10) illustrates only patterns shown to demonstrate relationship.

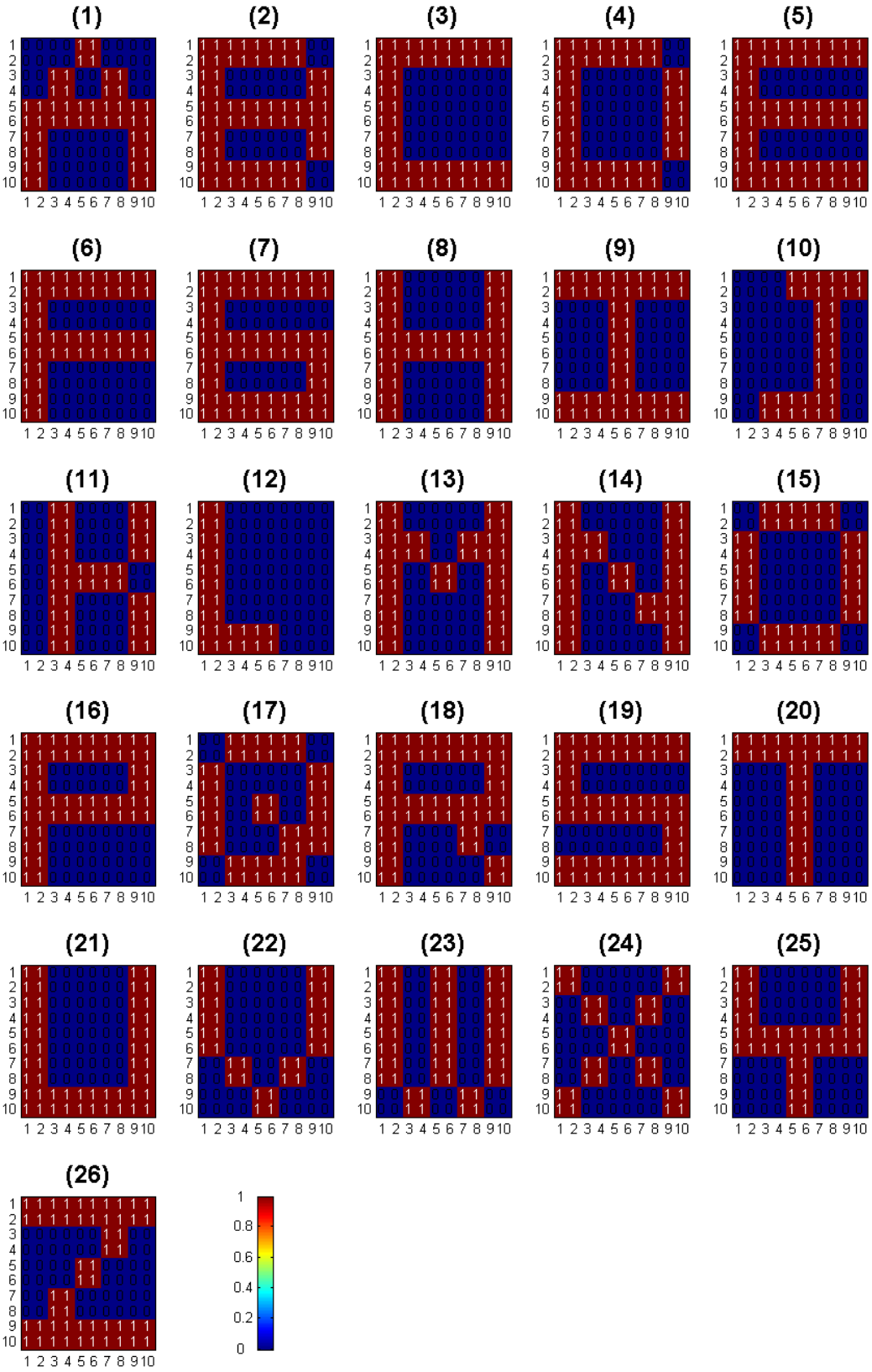


Figure 5.16. Set of 2x resolution upper-case English letters. Each alphabet is represented by the respective placement of 1's in a background of 0's.

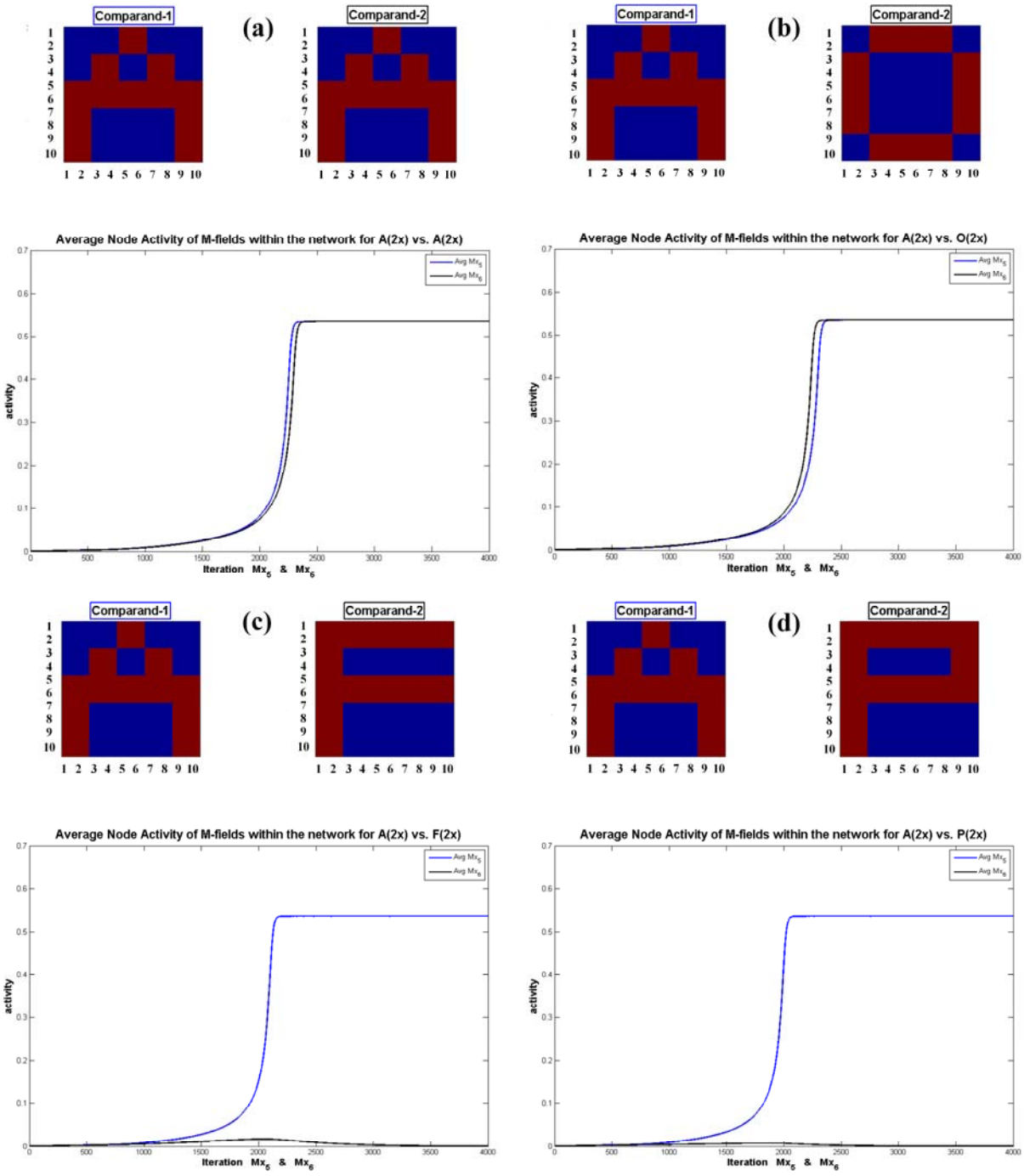


Figure 5.17. Time-series plot for 2x resolution patterns, A versus A(a), O(b), F(c) and P(d). The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. Given a pattern-combo (a, b, c or d) the plots shows the activities of average (median) nodes of respective M-layers. Note that the retinal-map size is 10x10 (Fig.5.16).

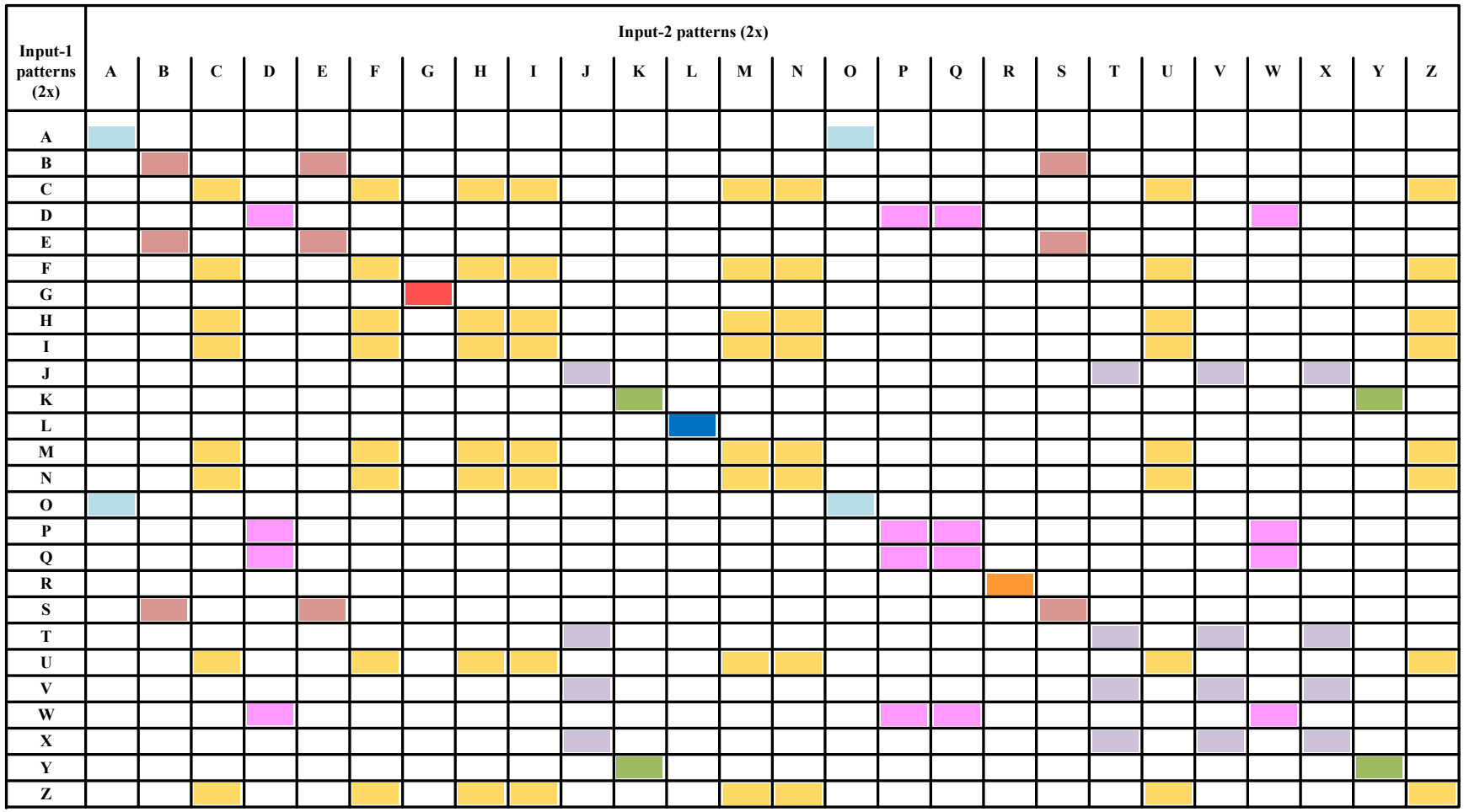


Table 5.8. Table illustrating similarities for 2x resolution Capital English letters, A to Z vs. A to Z. For a particular row, the colored (same) boxes represent similarity (for e.g. row for A). When compared to A to Z letters, if two or more rows have the same pattern of similarity, they are depicted having same colored boxes (for e.g. row for A & row for O). Two letters are defined similar if the smaller average (median) node activity of the two M-fields is ≥ 90% of the larger. This is equivalent to < 10% vigilance percentage. Note that the uncolored (white) boxes implies no similarity. The raw numerical data is given in appendix, Tables A5-1 to A5-8.

Number of non zero pixels	Non zero pixel value	Cumulative pixel value	Patterns (2x English letters)
72	0.0039	0.2808	G
68	0.0047	0.3196	B, E, S
64	0.0056	0.3584	R
56	0.0078	0.4368	D, P, Q, W
52	0.0091	0.4732	C, F, H, I, M, N, U, Z
48	0.0107	0.5136	A, O
44	0.0126	0.5544	K, Y
36	0.0175	0.6300	J, T, V, X
28	0.0251	0.7028	L

Table 5.9. Table showing the 2x resolution of the 26 uppercase English letters arranged with respect to their number of non-zero pixels (out of 100 pixels, 10 x 10 matrix), their normalized value (2nd column) and their total (3rd column, CPV or cumulative pixel value). Note that for a particular pattern all its non-zero pixels have the same magnitude of normalized pixel value.

Subset of comparand-1 patterns (2x)	Subset of comparand-2 patterns (2x)
{G}	{G}
{B, E, S}	{B, E, S}
{R}	{R}
{D, P, Q W}	{D, P, Q, W}
{C, F, H, I, M, N, U, Z}	{C, F, H, I, M, N, U, Z}
{A, O}	{A, O}
{K, Y}	{K, Y}
{J, T, V, X}	{J, T, V, X}
{L}	{L}

Table 5.10. Table of subset of patterns derived from simulation done with patterns from 2x resolution Capital English Letters of retinal size 10x10. Relationship is found for a given comparand-1 pattern with any comparand-2 pattern within a particular (same row) subset.

Comparand-1 Patterns (2x)	Comparand-2 Patterns (2x)	Largest difference w/ α	Smallest difference w/o α
G	{G}	{G} α {G} : 0(0) ⟹ 0.0000	{G} α̸ {B, E, S} : 0.1214(4) ⟹ 0.1116
B, E, S	{B, E, S}	{B, E, S} α {B, E, S} : 0(0) ⟹ 0.0000	{B, E, S} α̸ {R} : 0.1082(4) ⟹ 0.0995
R	{R}	{R} α {R} : 0(0) ⟹ 0.0000	{R} α̸ {B, E, S} : 0.1082(4) ⟹ 0.0995
D, P, Q, W	{D, P, Q, W}	{D, …, W} α {D, …, W} : 0(0) ⟹ 0.0000	{D, …, W} α̸ {C, …, Z} : 0.0769(4) ⟹ 0.0707
C, F, H, I, M, N, U, Z	{C, F, H, I, M, N, U, Z}	{C, …, Z} α {C, …, Z} : 0(0) ⟹ 0.0000	{C, …, Z} α̸ {D, …, W} : 0.0769(4) ⟹ 0.0707
A, O	{A, O}	{A, O} α {A, O} : 0(0) ⟹ 0.0000	{A, O} α̸ {K, Y} : 0.0735(4) ⟹ 0.0676
K, Y	{K, Y}	{K, Y} α {K, Y} : 0(0) ⟹ 0.0000	{K, Y} α̸ {A, O} : 0.0735(4) ⟹ 0.0676
J, T, V, X	{J, T, V, X}	{J, …, X} α {J, …, X} : 0(0) ⟹ 0.0000	{J, …, X} α̸ {L} : 0.1035(8) ⟹ 0.0828
L	{L}	{L} α {L} : 0(0) ⟹ 0.0000	{L} α̸ {J, …, X} : 0.1035(8) ⟹ 0.0828

Table 5.11. Table demonstrating the diff-pc metric for the case of comparing patterns from 2x resolution of the 26 Capital English letters of retinal map size, 10x10. The first two columns are the same as table 5.10.

Every patterns has a corresponding cumulative pixel value, CPV (table 5.9). If two patterns have different CPV's, U and L such that U > L. Then difference-factor, diff-fact = (U – L) / U and difference-percentage, diff-pc = diff-fact • ppx. Every unique difference of number of non-zero pixels, px has an empirically determined parameter, ppx. But unlike the previous case (tables 5.4 & 5.7) the pixels have increased 4x, thus (px) → 4•ppx. The ppx values are: (1) → 4•0.4 = 1.6, (2) → 4•0.32 = 1.6, (3) → 4•0.25 = 1, (4) → 4•0.23 = 0.92, (5) → 4•0.22 = 0.88, (6 to 10) → 4•0.8 = 3.2, (11 to 20) → 4•0.8•2 = 6.4, (21 to 30) → 4•0.8•3 = 9.6 and so on…

The third column shows the diff-fact, px (bracket) and resultant largest diff-pc amongst comparands-1 patterns and subset of comparands-2 patterns that has demonstrated α is < 4%. Since all the patterns within the set demonstrated relationship only with those having same number of non-zero pixels, all entries in the third column for largest diff-pc is zero. The fourth column shows the smallest diff-pc amongst those that has not demonstrated α is > 4%. The diff-pc for patterns that demonstrates α and α̸ was determined to be largest or smallest from the table containing all the computed diff-pc values (tables A6-1 & A6-2).

The normalized non-zero pixel values and their corresponding CPV's are shown in table 5.9. Notice that the number of non-zero pixels have increased 4x in comparison to the same pattern with 5x5 retinal-map size (table 5.2). The non-zero pixel values and CPV's have however decreased by four-folds. The result is a widened spread of CPV's with increased number of non-zero pixels and increased retinal-map size (variance = 0.0201 in 10x10 with 4x increase in non-zero pixels and variance = 0.0165 in 5x5 retinal-map size).

Amongst the 2x resolution patterns, all patterns show relationship only with patterns having same number of non-zero pixels and hence equal CPV's (table 5.9 & table 5.10). The diff-pc amongst all other patterns are > 4% (table 5.11).

Unlike the results for the case of Capital English letters sized 5x5 placed on a larger retinal-map size of 10x10 (table 5.6), the above results for 2x resolution patterns suggests increased sensory acuity. Thus, we may say that the network has become more selective in what it considers a particular pattern-combo to have a relationship (or non-relationship).

Recall that v1-layers in both sub-networks (Fig.5.3) have equal number of nodes. Furthermore, the number of nodes in a v1-layers is equal to the size of retinal-map. Thus, to compare regular resolution Capital English letters against its higher 2x resolution patterns, both the input patterns have 10x10 retinal-map size.

Figures 5.18a and 5.18b shows no relationship amongst patterns A (1x) versus A (2x) and A (1x) versus O (1x). Since the 1x patterns have increased CPV's (table 5.5) but decreased CPV's for 2x patterns (table 5.9), none of the 1x patterns demonstrates relationship with any of the 2x patterns (Fig.5.18). The diff-pc amongst all the patterns are > 4% (table 5.12).

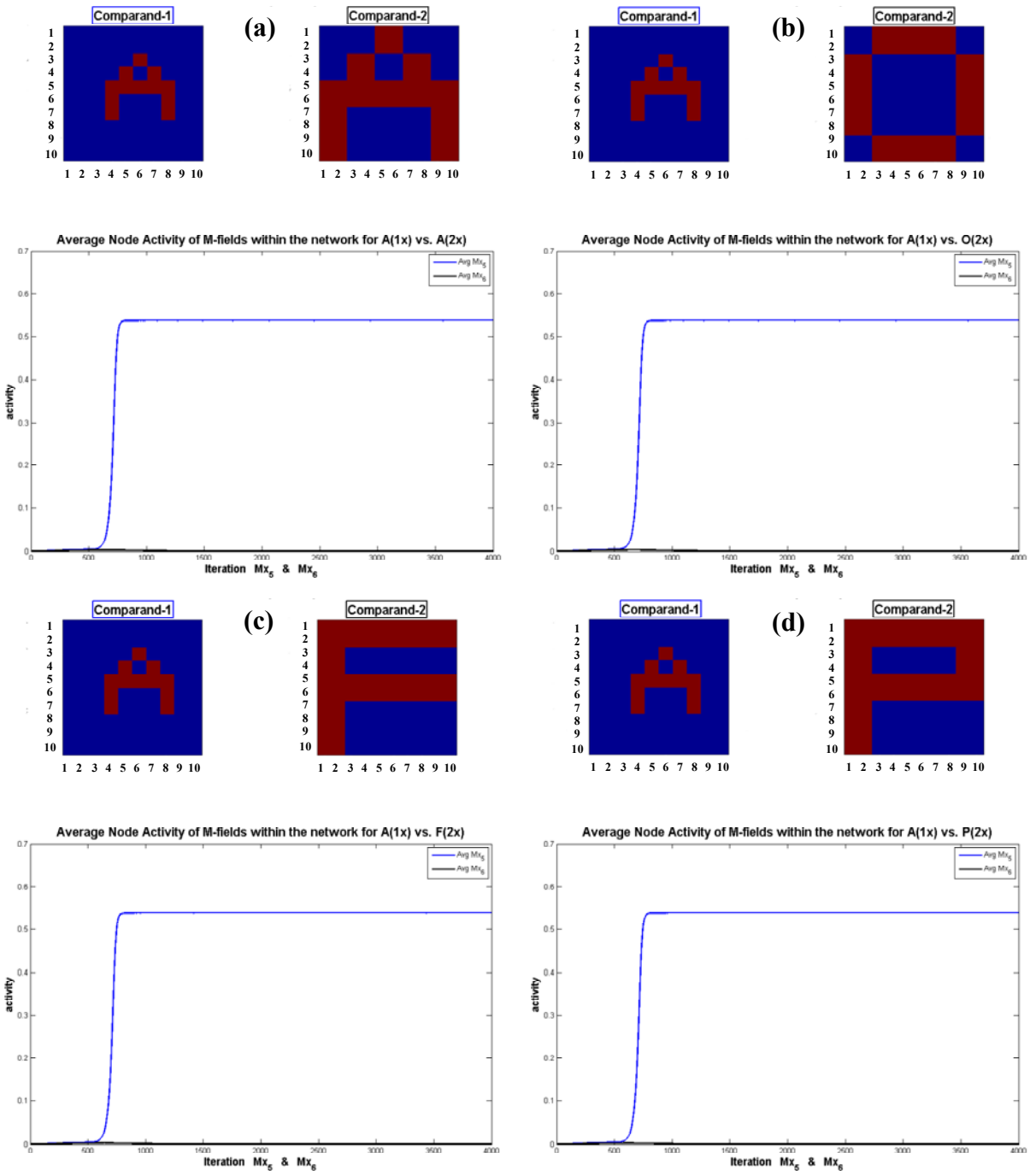


Figure 5.18. Time-series plot for 1x versus 2x resolution patterns, A versus A(a), O(b), F(c) and P(d). The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. Given a pattern-combo (a, b, c or d) the plots shows the activities of average (median) nodes of respective M-layers. Note that the retinal-map size is 10x10 (Fig.5.15 & Fig.5.16).

Comparand-1 Patterns (1x)	Comparand-2 Patterns (2x)	Largest difference w/ α	Smallest difference w/o α
G	{-}	-	{G} ⍺ {L} : 0.1045(10) ⟹ 0.0836
B, E, S	{-}	-	{B, E, S} ⍺ {L} : 0.1129(11) ⟹ 0.0903
R	{-}	-	{R} ⍺ {L} : 0.1197(12) ⟹ 0.0958
D, P, Q, W	{-}	-	{D, …, W} ⍺ {L} : 0.1330(14) ⟹ 0.1064
C, F, H, I, M, N, U, Z	{-}	-	{C, …, Z} ⍺ {L} : 0.1391(15) ⟹ 0.1113
A, O	{-}	-	{A, O} ⍺ {L} : 0.1438(16) ⟹ 0.1150
K, Y	{-}	-	{K, Y} ⍺ {L} : 0.1470(17) ⟹ 0.1176
J, T, V, X	{-}	-	{J, …, X} ⍺ {L} : 0.1503(19) ⟹ 0.1202
L	{-}	-	{L} ⍺ {J, …, X} : 0.1448(21) ⟹ 0.1158

Table 5.12. Table demonstrating the diff-pc metric for the case of comparing 1x patterns against 2x resolution of the 26 Capital English letters of retinal map size, 10x10. Since none of the 1x patterns showed any relationship with any of the 2x patterns, the entries in the comparands-2 column (2^{nd}) are empty set.

Every patterns has a corresponding cumulative pixel value, CPV (table 5.9). If two patterns have different CPV's, U and L such that U > L. Then difference-factor, diff-fact = (U – L) / U and difference-percentage, diff-pc = diff-fact • ppx. Every unique difference of number of non-zero pixels, px has an empirically determined parameter, ppx. But unlike the previous case (tables 5.4 & 5.7) the pixels have increased 4x, thus (px) → 4•ppx. The ppx values are: (1) → 4•0.4 = 1.6, (2) → 4•0.32 = 1.6, (3) → 4•0.25 = 1, (4) → 4•0.23 = 0.92, (5) → 4•0.22 = 0.88, (6 to 10) → 4•0.8 = 3.2, (11 to 20) → 4•0.8•2 = 6.4, (21 to 30) → 4•0.8•3 = 9.6 and so on…

Since no relationship was demonstrated, there are no entries in the third column. The fourth column shows the diff-fact, px (bracket) and resultant smallest diff-pc amongst comparands-1 patterns and subset of comparands-2 patterns that has not demonstrated α is >4%. The diff-pc for patterns that demonstrates α and ⍺ was determined to be largest or smallest from the table containing all the computed diff-pc values (tables A8-1 & A8-2).

Case-4, Arabic numerals (0 to 9):

Returning to retinal-map size of 5x5, let us consider the pattern-set of Arabic numerals, 0 to 9 (Fig.5.19). Figure 5.20 shows relationship amongst 0 and 2 but not between 0 and 1. Experiment done amongst patterns from this set demonstrating relationship and no relationship is shown in table 5.12. Those patterns that showed relationships (table 5.14) has a diff-pc < 4% (table 5.15).

Finally, let us consider patterns for comparands-1 from the pattern-set of Capital English letters (Fig.5.6) and comparands-2 patterns from Arabic numerals (Fig.5.19). Figure 5.21 shows relationship for A versus 2 but not for A and 1. The determination for all the pattern-combos is shown in table 5.17. As with earlier experiments, those pattern-combos that showed relationships has a diff-pc < 4% (table 5.19).

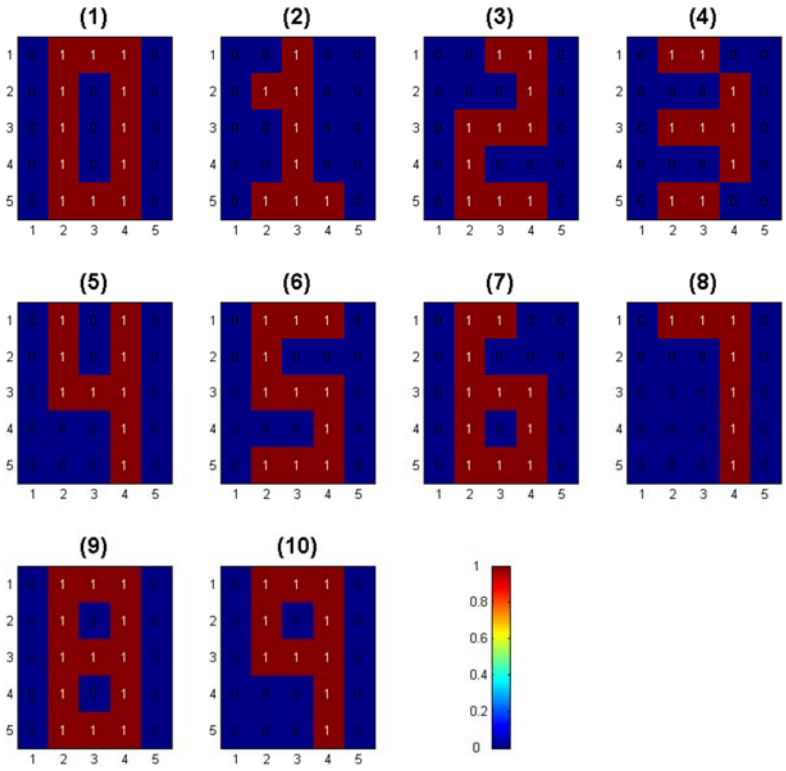


Figure 5.19. Set of Arabic numerals (0 – 9). Each alphabet is represented by the respective placement of 1's in a background of 0's.

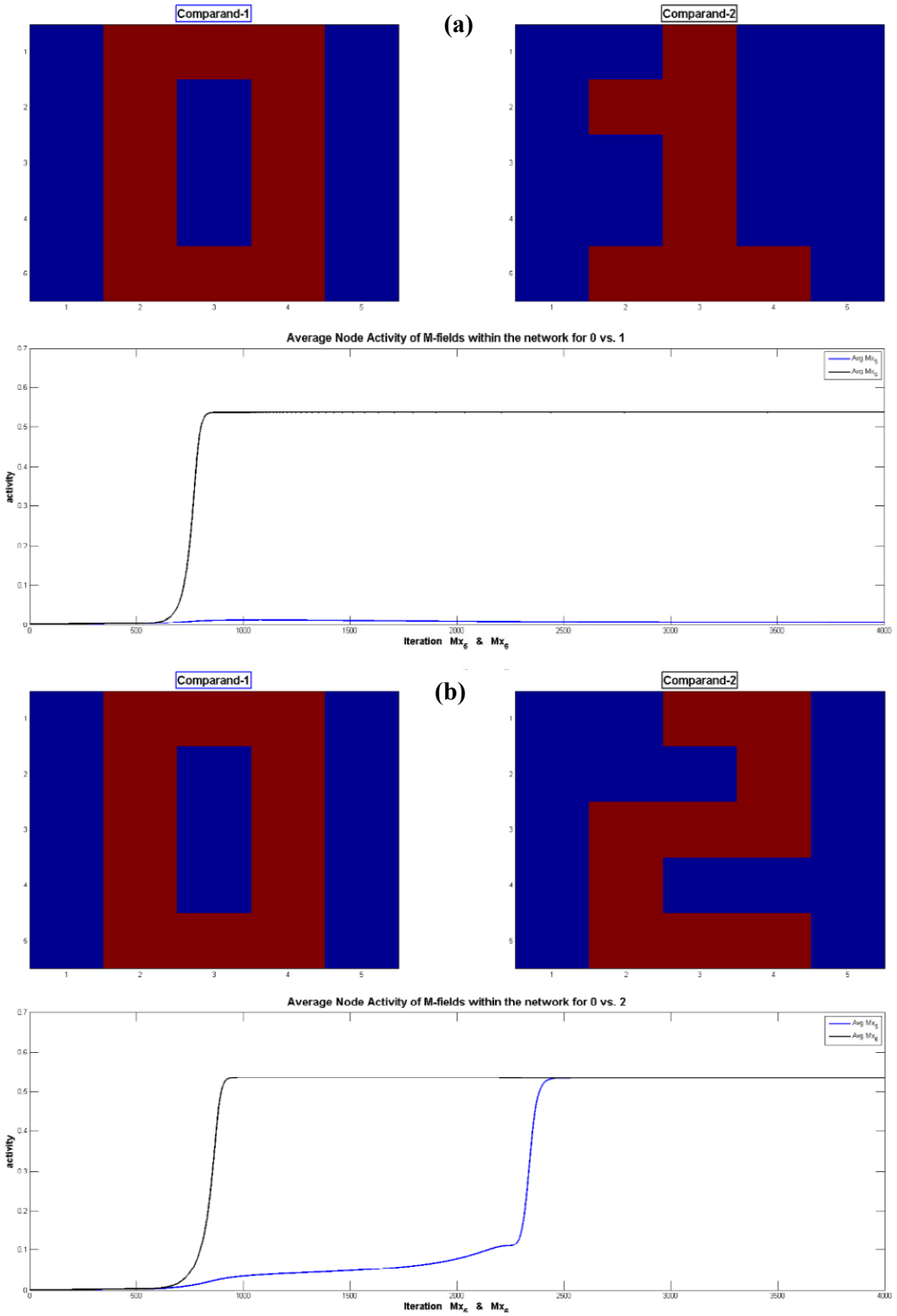


Figure 5.20. Time-series plot for numbers 0 vs. 1 (a) and 0 vs. 2 (b). The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. Given a pattern-combo (a or b) the plots shows the activities of average (median) nodes of respective M-layers. Note that the retinal-map size is 5x5.

Input-1 patterns (5x5)	Input-2 patterns (5x5) 0	1	2	3	4	5	6	7	8	9
0										
1										
2										
3										
4										
5										
6										
7										
8										
9										

Table 5.13. Table illustrating similarities for pattern-set of Arabic numerals, 0 to 9. For a particular row, the colored (same) boxes represent similarity (for e.g. row for 2). When compared to 0 to 9, if two or more rows have the same pattern of similarity, they are depicted having same colored boxes (for e.g. row for 2 & row for 9). Two numeral patterns are defined similar if the smaller average (median) node activity of the two M-fields is ≥ 90% of the larger. This is equivalent to < 10% vigilance percentage. Note that the uncolored (white) boxes implies no similarity. The raw numerical data is given in appendix, Table A9.

Number of non zero pixels	Non zero pixel value	Cumulative pixel value	Patterns (5x5)
13	0.0357	0.4641	8
12	0.0417	0.5004	0
11	0.0486	0.5346	5, 6
10	0.0568	0.5680	2, 9
9	0.0667	0.6003	3, 4
8	0.0787	0.6296	1
7	0.0937	0.6559	7

Table 5.14. Table showing the pattern-set of Arabic numerals (0 to 9) arranged with respect to their number of non-zero pixels (out of 25 pixels, 5 x 5 matrix), their normalized value (2nd column) and their total (3rd column, CPV or cumulative pixel value). Note that for a particular pattern all its non-zero pixels have the same magnitude of normalized pixel value.

Subset of comparand-1 patterns (5x5)	Subset of comparand-2 patterns (5x5)
{8}	{0, 8}
{0}	{0, 2, 5, 6, 8, 9}
{5, 6}	{0, 1, 2, 3, 4, 5, 6, 9}
{2, 9}	{0, 1, 2, 3, 4, 5, 6, 7, 9}
{1, 3, 4}	{1, 2, 3, 4, 5, 6, 7, 9}
{7}	{1, 2, 3, 4, 7, 9}

Table 5.15. Table of subset of patterns derived from simulation done with patterns from Arabic Numerals, 0 to 9 of retinal size 5x5. Relationship is found for a given comparand-1 pattern with any comparand-2 pattern within a particular (same row) subset.

Comparand-1 Patterns (5x5)	Comparand-2 Patterns (5x5)	Largest difference w/ α	Smallest difference w/o α
8	{0, 8}	{8} α {0} : 0.0725(1) ⟹ 0.0290	{8} α̸ {5,6} : 0.1318(2) ⟹ 0.0421
0	{0, 2, 5, 6, 8, 9}	{0} α {2, 9} : 0.1190(2) ⟹ 0.0380	{0} α̸ {3, 4} : 0.1664(3) ⟹ 0.0416
5, 6	{0, 1, 2, 3, 4, 5, 6, 9}	{5, 6} α {1} : 0.1508(3) ⟹ 0.0377	{5, 6} α̸ {7} : 0.1849(5) ⟹ 0.0406
2, 9	{0, 1, 2, 3, 4, 5, 6, 7, 9}	{2, 9} α {0} : 0.1190(2) ⟹ 0.0380	{2, 9} α̸ {8} : 0.1829(3) ⟹ 0.0457
3, 4	{1, 2, 3, 4, 5, 6, 7, 9}	{3, 4} α {5, 6} : 0.1094(2) ⟹ 0.0350	{3, 4} α̸ {0} : 0.1664(3) ⟹ 0.0416
1	{1, 2, 3, 4, 5, 6, 7, 9}	{1} α {5, 6} : 0.1508(3) ⟹ 0.0377	{1} α̸ {0} : 0.2052(4) ⟹ 0.0471
7	{1, 2, 3, 4, 7, 9}	{7} α {2, 9} : 0.1340(3) ⟹ 0.0335	{7} α̸ {5,6} : 0.1849(4) ⟹ 0.0425

Table 5.16. Table demonstrating the diff-pc metric for the case of comparing patterns from Arabic Numerals (0 to 9) of retinal map size, 5x5. The first two columns shows the relationship (α) found for a given comparands-1 pattern (column-1) with any comparands-1 pattern (column-2) within a particular (same row) subset (table 5.15).

Every patterns has a corresponding cumulative pixel value, CPV (table 5.14). If two patterns have different CPV's, U and L such that U > L. Then difference-factor, diff-fact = (U – L) / U and difference-percentage, diff-pc = diff-fact • ppx. Every unique difference of number of non-zero pixels, px has an empirically determined parameter, ppx. Thus (px) → ppx. The ppx values are: (1) → 0.4, (2) → 0.32, (3) → 0.25, (4) → 0.23, (5) → 0.22, (6 to 10) → 0.8, (11 to 20) → 0.8•2 = 1.6, (21 to 30) → 0.8•3 = 2.4, (31 to 40) → 0.8•4 = 3.2, (41 to 50) → 0.8•5 = 4, (51 to 60) → 0.8•6 = 4.8 and so on…

The third column shows the diff-fact, px (bracket) and resultant largest diff-pc amongst comparands-1 patterns and subset of comparands-2 patterns that has demonstrated α is < 4%. The fourth column shows the smallest diff-pc amongst those that has not demonstrated α is > 4%. The diff-pc for patterns that demonstrates α and α̸ was determined to be largest or smallest from the table containing all the computed diff-pc values (table A10).

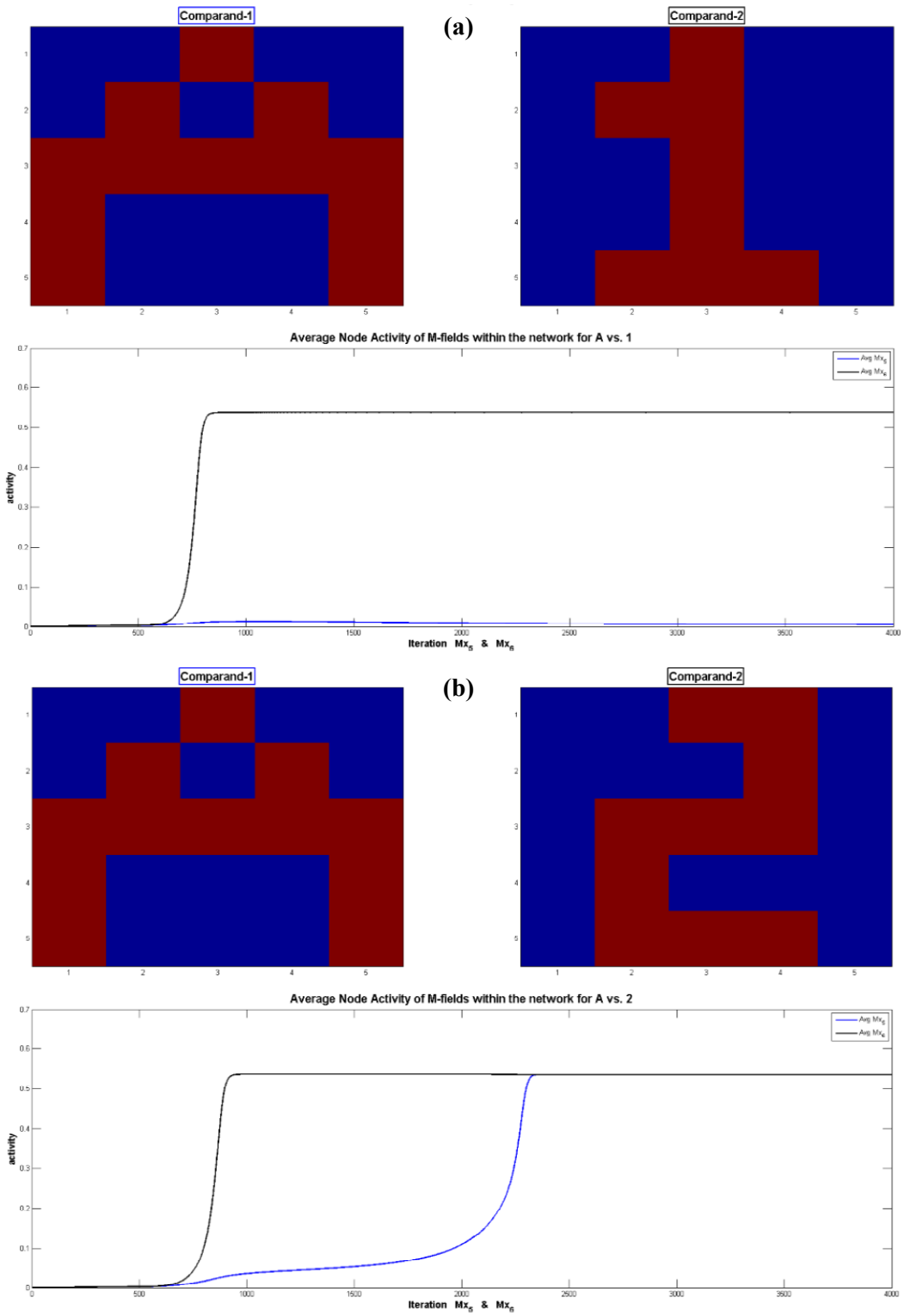


Figure 5.21. Time-series plot for Capital letter A vs. number 1 (a) and A vs. 2 (b). The top sub-network in Fig.5.3 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.3 receives comparand-2, color coded as black. Given a pattern-combo (a or b) the plots shows the activities of average (median) nodes of respective M-layers. Note that the retinal-map size is 5x5.

Input-1 patterns (5x5)	Input-2 patterns (5x5) 0	1	2	3	4	5	6	7	8	9
A										
B										
C										
D										
E										
F										
G										
H										
I										
J										
K										
L										
M										
N										
O										
P										
Q										
R										
S										
T										
U										
V										
W										
X										
Y										
Z										

Table 5.17. Table illustrating similarities for the case of comparing Capital English letters (A to Z) against Arabic Numerals (0 to 9). For a particular row, the colored (same) boxes represent similarity (for e.g. row for A). When compared to 0 to 9 numerals, if two or more rows have the same pattern of similarity, they are depicted having same colored boxes (for e.g. row for A & row for O). A letter-numeral pair is defined similar if the smaller average (median) node activity of the two M-fields is ≥ 90% of the larger. This is equivalent to < 10% vigilance percentage. Note that the uncolored (white) boxes implies no similarity. The raw numerical data is given in appendix, Tables A11-1 to A11-3.

Number of non zero pixels	Non zero pixel value	Cumulative pixel value	Patterns (Letters & Numbers)
18	0.0154	0.2772	G
17	0.0185	0.3145	B, E, S
16	0.0221	0.3536	R
14	0.0306	0.4284	D, P, Q, W
13	0.0357	0.4641	C, F, H, I, M, N, U, Z, 8
12	0.0417	0.5004	A, O, 0
11	0.0486	0.5346	K, Y, 5, 6
10	0.0568	0.5680	2, 9
9	0.0667	0.6003	J, T, V, X, 3, 4
8	0.0787	0.6296	1
7	0.0937	0.6559	L, 7

Table 5.18. Table showing the 26 uppercase English letters and 9 Arabic numerals arranged with respect to their number of non-zero pixels (out of 25 pixels, 5 x 5 matrix), their normalized value (2nd column) and their total (3rd column, CPV or cumulative pixel value). Note that for a particular pattern all its non-zero pixels have the same magnitude of normalized pixel value. Note that this table is a combination of both table 5.2 and 5.14.

Subset of comparand-1 patterns (5x5)	Subset of comparand-2 patterns (5x5)
{D, P, Q, W}	{8}
{C, F, H, I, M, N, U, Z}	{0, 8}
{A, O}	{0, 2, 5, 6, 8, 9}
{K, Y}	{0, 1, 2, 3, 4, 5, 6, 9}
{J, T, V, X}	{1, 2, 3, 4, 5, 6, 7, 9}
{L}	{1, 2, 3, 4, 7, 9}

Table 5.19. Table of subset of patterns derived from simulation done for the case of comparing Capital English letters (A to Z) against Arabic Numerals (0 to 9) with retinal-map size, 5x5. Relationship is found for a given comparand-1 pattern with any comparand-2 pattern within a particular (same row) subset.

Comparand-1 Patterns (5x5)	Comparand-2 Patterns (5x5)	Largest difference w/ α	Smallest difference w/o α
G	{-}	-	{G} ⍺ {8} : 0.4027(5) ⟹ 0.0885
B, E, S	{-}	-	{B, E, S} ⍺ {8} : 0.3223(4) ⟹ 0.0741
R	{-}	-	{R} ⍺ {8} : 0.2380(3) ⟹ 0.0595
D, P, Q, W	{8}	{D, …, W} α {8} : 0.0769(1) ⟹ 0.0307	{D, …, W} ⍺ {0} : 0.1439(2) ⟹ 0.0460
C, F, H, I, M, N, U, Z	{0, 8}	{C, …, Z} α {0} : 0.0725(1) ⟹ 0.0290	{C, …, Z} ⍺ {5, 6} : 0.1319(2) ⟹ 0.0422
A, O	{0, 2, 5, 6, 8, 9}	{A, O} α {2, 9} : 0.1190(2) ⟹ 0.0380	{A, O} ⍺ {3, 4} : 0.1664(3) ⟹ 0.0416
K, Y	{0, 1, 2, 3, 4, 5, 6, 9}	{K, Y} α {1} : 0.1508(3) ⟹ 0.0377	{K, Y} ⍺ {8} : 0.1319(2) ⟹ 0.0422
J, T, V, X	{1, 2, 3, 4, 5, 6, 7, 9}	{J, …, X} α {5, 6} : 0.1094(2) ⟹ 0.0350	{J, …, X} ⍺ {0} : 0.1664(3) ⟹ 0.0416
L	{1, 2, 3, 4, 7, 9}	{L} α {2, 9} : 0.1340(3) ⟹ 0.0335	{L} ⍺ {5,6} : 0.1849(4) ⟹ 0.0425

Table 5.20. Table demonstrating the diff-pc metric for the case of comparing Capital English letters (A to Z) against Arabic Numerals (0 to 9) with retinal-map size, 5x5. Since none of the G, B, E, S, and R patterns showed any relationship with any of the numeral patterns, the entries in the comparands-2 column (2^{nd}) are empty set.

Every patterns has a corresponding cumulative pixel value, CPV (table 5.18). If two patterns have different CPV's, U and L such that U > L. Then difference-factor, diff-fact = (U – L) / U and difference-percentage, diff-pc = diff-fact • ppx. Every unique difference of number of non-zero pixels, px has an empirically determined parameter, ppx. Thus (px) → ppx. The ppx values are: (1) → 0.4, (2) → 0.32, (3) → 0.25, (4) → 0.23, (5) → 0.22, (6 to 10) → 0.8, (11 to 20) → 0.8•2 = 1.6, (21 to 30) → 0.8•3 = 2.4, (31 to 40) → 0.8•4 = 3.2, (41 to 50) → 0.8•5 = 4, (51 to 60) → 0.8•6 = 4.8 and so on…

The third column shows the diff-fact, px (bracket) and resultant largest diff-pc amongst comparands-1 patterns and subset of comparands-2 patterns that has demonstrated α is < 4%. The fourth column shows the smallest diff-pc amongst those that has not demonstrated α is > 4%. The diff-pc for patterns that demonstrates α and ⍺ was determined to be largest or smallest from the table containing all the computed diff-pc values (table A12).

Pattern-combos	$\bar{x}_1$	$\bar{x}_2$	C.I	p-value
A vs. A	0.5347	0.5357	[0, 0]	-
A vs. F	0.5347	0.5356	$[3.6\cdot10^{-5}, 3.6\cdot10^{-5}]$	0
A vs. P	0.5358	0.0025	[0.5333, 0.5333]	$1.69\cdot10^{-141}$

Table 5.21. Summarized statistical (paired t-test) result. For a given pattern-combo (comparands-1 vs. comparand-2), results includes: mean ($\bar{x}_1$ & $\bar{x}_2$) of the average (median) activity of all nodes in respective M-field at final iteration (3000) over n-samples (n=35), confidence interval (CI) and p-value. The results are for 95% confidence and hence t*multiplier = 3.6 and alpha = 0.05. Note that p-value is not computed for cases with CI = [0, 0]. The raw numerical data is given in appendix, Tables A13-1 to A13-11.

Recall that the initial instar (W's) weights were set to random non-zero weights (and outstar weights Z's were set to zero). To test the reproducibility (consistency) of the networks determination of relationship or non-relationship, statistical tests (t-test and ANOVA f-test) were done. Considering the case of patterns from Capital English letters (case-1, Fig.5.6), data was collected from randomly independent sample of sample size = 35.

Since the initial W's are non-zero random values, one question is: Does the determined outcome (of the network, Fig.5.3) remain unchanged? For instance, does patterns A vs. A (Fig.5.7) and A vs. F (Fig.5.8) always show relationship and patterns A vs. P (Fig.5.9) always show non-relationship?

For A vs. A, the 95% confidence interval estimate for the difference in means (over n = 35) of the median M-field node activities was [0, 0] (table 5.21). The interval covers 0. Therefore, we can be fairly confident that the population means of the median M-field node activities are equal.

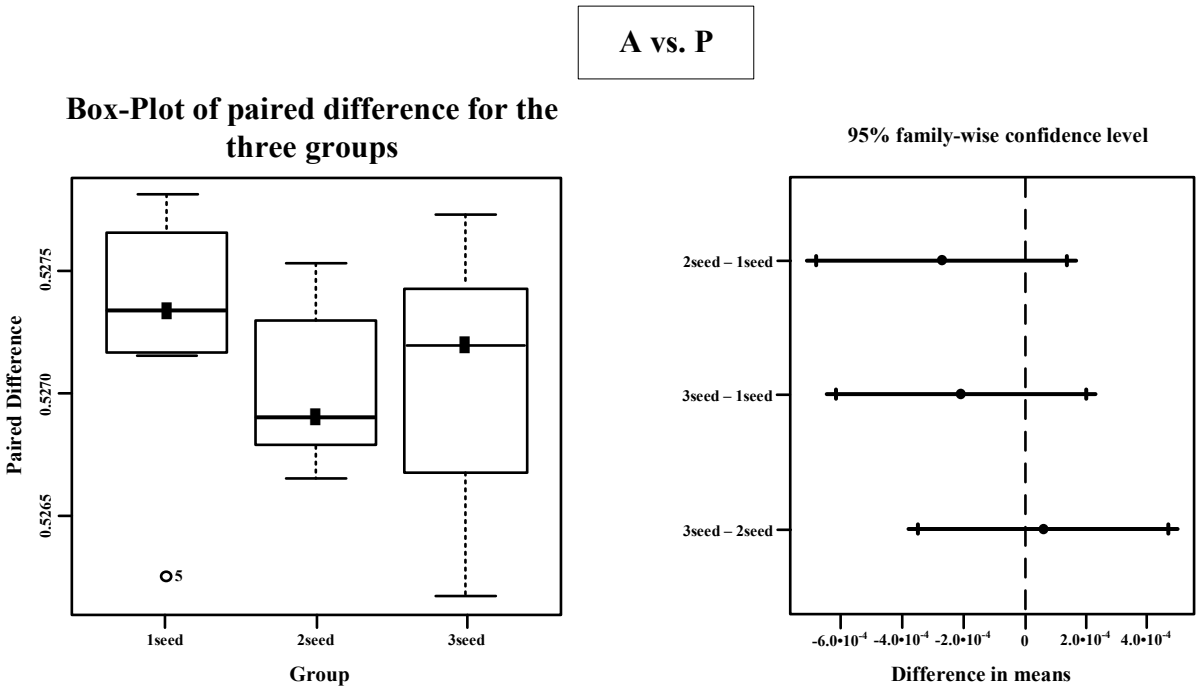
For A vs. F and A vs. P, the 95% confidence interval estimate for the paired-difference was $[3.6 \bullet 10^{-5}, 3.6 \bullet 10^{-5}]$ and [0.5333, 0.5333] respectively (table 5.21). Both their null hypothesis was rejected (p-values < 0.05). The difference in means for both cases was therefore statistically significant.

For A vs. F, the mean of 0.5346 (Mx6) is about 95% of 0.5347 (Mx5). Thus, the difference in means has statistical but not practical significance[2]. On the other hand, 0.0025 (Mx6) is about 0.4% of 0.5358 (Mx5) for A vs. P. Thus difference in means for this case is both statistically and practically significant.

To compare response variable for different initial W weights, one-way ANOVA f-test is chosen. Three (k) groups were chosen. The groups were defined with respect to the seed (= 1, 2, 3) of the random number generator. All the groups have equal sample size, $n_k = 10$. Thus, total sample size N = 30. The response variable is the paired-difference of the means of M5 and M6-field activities at final iteration (2000) over n_k samples for respective k-group.

Figure 5.22 shows the results for A vs. P. The box-plot show more variation between the groups than variations within groups. The 95% pairwise comparison shows all three intervals crossing 0. The p-value ≈ 0.3 implies that the F-statistic (= 1.26) is not in the rejection region. Thus, with 95% confidence we can conclude that there is not enough evidence of difference in M-field outputs for different initial W weights.

[2] Practical significance refers to the magnitude of relationship between variables and whether or not that magnitude is important. Though statistical significance is informative, it does not necessarily mean that the relationship between the two variables has practical significance [Utts & Heckard, 2012].
Statistical significance means that the data are strong enough to reject null-hypothesis but p-value does not provide information about the magnitude of the effect. Since there is almost always at least a slight relationship between two variables, almost any null hypothesis can be rejected if the sample size is large enough.
For example, in "Height depends on month of birth" [Weber et al., 1998], the authors concluded that men born in spring is 0.6cm (1/4 inch) taller than those born in fall. Their sample size was 507,125 military recruits. Thus, statistical significance does not guarantee practical significance.



ANOVA Table

Source	df	Sum Sq.	Mean Sq.	F	P
Groups	2	$3.9702 \cdot 10^{-7}$	$1.9851 \cdot 10^{-7}$	1.261	0.2996
Error	27	$4.2511 \cdot 10^{-6}$	$1.5745 \cdot 10^{-7}$		
Total	29	$4.6482 \cdot 10^{-6}$			

Figure 5.22. Comparison of paired-difference (pattern-combo, A vs. P) for three (k) groups. The patterns have retinal-map size 5x5. For the ANOVA f-test the groups are defined with respect to the seed (= 1, 2, 3) of the random number generator. Each group has sample size, $n_k = 10$. Thus, $N = n_1 + n_2 + n_3 = 30$.

Paired-difference is the difference of the means of M5 and M6-field activities at final iteration (2000) over n_k samples for respective k-group. The ▉ within the box indicates sample mean.

The sample means (approximate) are: 0.5273, 0.5270 and 0.5271 for groups; 1seed, 2seed and 3seed respectively. The sample standard deviations (sd) are: $4.033 \cdot 10^{-4}$, $2.785 \cdot 10^{-4}$ and $4.818 \cdot 10^{-4}$. Thus, the largest, $4.81 \cdot 10^{-4}$ > 2x smallest, $2.78 \cdot 10^{-4}$.

The 95% simultaneous confidence intervals (Tukey's procedures) indicate that all the three intervals cover 0. The pairwise intervals are; 2seed – 1 seed: $[-7.093 \cdot 10^{-4}, 1.707 \cdot 10^{-4}]$, 3seed – 1seed: $[-6.465 \cdot 10^{-4}, 2.335 \cdot 10^{-4}]$ and 3seed – 2seed: $[-3.772 \cdot 10^{-4}, 5.028 \cdot 10^{-4}]$. The mean (●) of the respective intervals are: $-2.693 \cdot 10^{-4}$, $-2.065 \cdot 10^{-4}$ and $6.280 \cdot 10^{-5}$.

This ends the demonstration of ability of the proposed minimal neural network (Fig.5.3) in determining relationship or non-relationship between some patterns and not for others. From here onwards all the simulation results will be for the pattern-set of Capital English letters or 5x5 retinal-map size (Fig.5.6).

The results demonstrated above make a novel contribution of significance in ART network theory. The behaviors just described are emergent network properties. Nothing in the design of the networks explicitly inserted the relationship vs. no relationship characteristics just described. This means the theory did not introduce any a priori objective criterion for defining what the network treats as a feature for pattern matching. The experiments just reported are the first-ever demonstration that ART is capable of self-defining feature sets.

The importance of this finding must not be minimized. If a theorist deliberately introduces any objective character of a feature into the design of a network, this amounts to building objective knowledge a priori into the synthesis of apprehension. However doing so is a violation of an epistemological law of mental physics, which holds that human beings are born with no objective knowledge a priori whatsoever. It is therefore a real necessity that any network used in modelling the synthesis of objective perceptions (intuitions) must be capable of self-determining what will or will not constitute an objective feature. It can be conjectured that this ability for a network to auto-determine features is a possible explanation of how human beings come to be capable of making metaphors and analogies. This work is the first time any neural network system has demonstrated this capacity, and this is an original contribution to knowledge from this research project.

Above results describes the basic behavior of the network in terms of how pattern-combos either have relationship or not. We must now consider the three properties required for generating equivalence relations, that is, reflexive, symmetric and transitive.

Additions to the above minimal neural network

It was observed that every pattern received by the network was shown to be reflexive. In other words, when the left and right sub-networks received the same pattern the M layer output always fell within the solution-set. This behavior is due to the chosen network configuration. Referring back to the model (Fig.5.2d & Fig.5.3), recall that the sub-networks have the same architecture and parameters (Fig.5.6). Since the two similar sub-network are coupled by reciprocal inhibition, it is understandable that the network consistently shows that the relationship between patterns are reflexive. By the same argument, if a relationship is shown between two different patterns then this relationship is always symmetric. This property of the network dynamics is not a hindrance to the act of comparation because generation of equivalence relation is the mathematical act of comparation.

Therefore, if two patterns are shown by the network to have a relationship then the relationship satisfies both reflexive and symmetric properties. These two properties can be therefore be realized from a single pattern-combo. In other words, the network does not have to process another pattern-combo such that they are of same patterns or flipped patterns. If the network had to process other pattern-combos composed of different patterns, it would mean that the realization of the two relation properties would involve a logical ordering process. The network however requires an ordering processing for transitivity. The above minimal neural network does not have this capability.

Comparation is a process in synthesis in sensibility which in turn does not have memory because the obscure parástase from sensory data has not yet become a conscious or objective parástase. The obscure parástase exists however. Therefore, comparation has a 'state'. This means then, if an obscure parástase is a sequence of patterns, the comparation network will be in a 'state' whose consequence would be, a realization (or not) of equivalence relation.

One of the functions of the **pure intuition of time** (**PIT**) within the OB (Fig.3.13) is that it determines 'content' in time. In practice, this means that the **PIT** can determine which pair of elements (in a sequence) is to be processed by the comparation network. Thus we have,

Proposition 9: The comparation network interacts with the pure intuition of time (PIT) such that the PIT determines the order of sequence of pattern pairs to be processed.

Since the aim of the project is to build a comparation network generating equivalence relation and not build other nous processes of the OB, a PIT proxy was built (Fig.5.23). The PIT proxy determines the content for the comparation process by picking pattern-combos one at a time from the pattern sequence. Using results from the above observations that a pattern-combo shown to have a relationship is reflexive and also symmetric, the PIT proxy picks L unique pattern-combos, where L is the length of a pattern sequence.

If one of the pattern-combo in the sequence is shown by the comparation network not to have a relationship, then there is no point for comparation to process the remaining pattern-combos. This would mean that the M layer outputs do not fall inside the solution-set. In other words, the M layer outputs are not in equilibrium and hence not expedient.

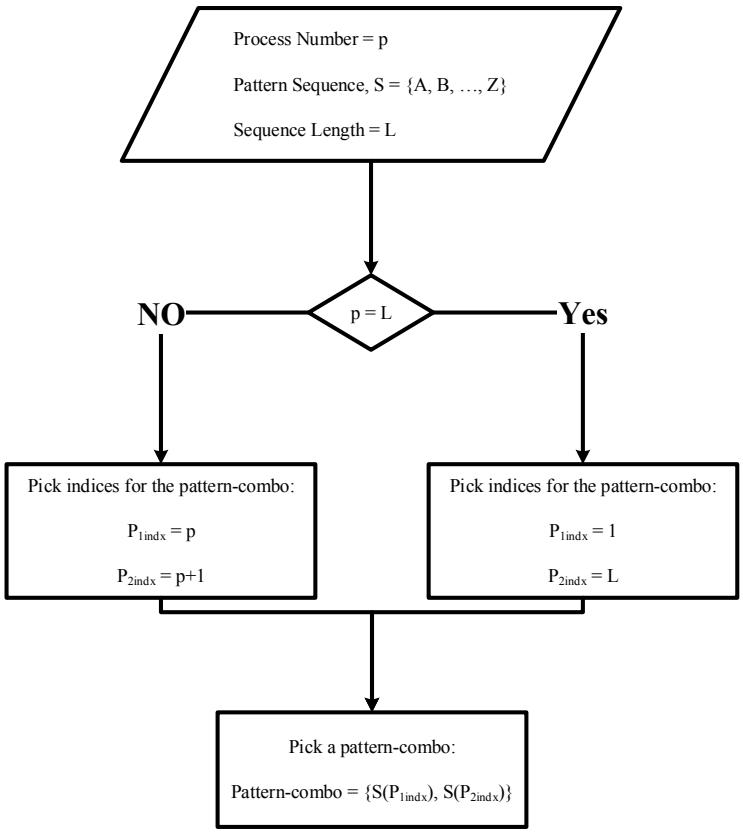


Figure 5.23. The pure intuition of time (PIT) proxy for determining content (pattern-combo) for the process of comparation network (Fig.5.3).

Since the PIT is a sub-process within the synthesis in sensibility (Fig.3.11 & Fig.3.13), like comparation the PIT is also linked to the reflective judgment (TRJ). Thus, the TRJ judge expediency based on the two pre-motor images and then stops the comparation process if one of the pattern-combo in the pattern sequence is not expedient. TRJ does not have the capability to directly stop comparation but it can indicate to PIT that a pattern-combo is not expedient. Hence,

Proposition 10: The function of PIT determining the content of the comparation process can short-circuit the order of pattern-combos based on non-expediency judged by TRJ.

The judging of M layer outputs for expediency was performed by a TRJ proxy for this particular function. Since Weber and Fechner, the concept of 'just noticeable difference' has

been well known psychological phenomenon [Kandel, 2000]. Based upon this notion, the TRJ proxy was built by comparing the mismatch between the M-layer outputs against the vigilance percentage parameter, δ which is akin to difference limen or just notifiable difference.

If vectors, $\mathbf{x}_5$ and $\mathbf{x}_6$ are steady-state (last iteration) M-layer outputs then the mismatch $= \sum_{\forall i} |m_i^{(5)} - m_i^{(6)}|$ where $\mathbf{M}^{(5)}$ and $\mathbf{M}^{(6)}$ are the normalized $\mathbf{x}_5$ and $\mathbf{x}_6$ ($\mathbf{M}^{(k)} = \mathbf{x}_k$ / d, where d = max-element ($\mathbf{x}_5$, $\mathbf{x}_6$)).

Figure 5.24 shows the new minimal network model based upon the previous minimal anatomy (Fig.5.2d) incorporated with the PIT and TRJ proxies.

Behavior of the proposed minimal anatomy.

The parameters used were the same set as earlier (Fig.5.5) with the addition of vigilance percentage, δ = 0.1. The patterns comprising any desired sequence was picked from the same set of upper-case English alphabets (Fig.5.6). Because normalizer parameters are the same, the resulting normalized inputs are also unchanged (Fig.5.10 & 5.11).

The minimal network showed equivalence relationship for some pattern sets and not for others. For a pattern sequence {C, F, A, O}, the network process finds relationship between C & F (C α F) and so it continues for {F, A}, {A, O} and finally {C, O} (Fig.5.25). On the other hand, for {C, A, Y, G}, the network finds C α A and proceeds to {A, D} but finds A $\not\alpha$ Y and hence stops the process (Fig.5.26). This short-circuiting of the process is also seen in {B, D, E, P, S, G}, where the process stops at {B, D} (Fig.5.27). Notice that, if the pattern sequence was composed only of either {D, P} or {B, E, S} the network would continue the process and find relationship among the respective patterns (Fig.5.28 & 5.29).

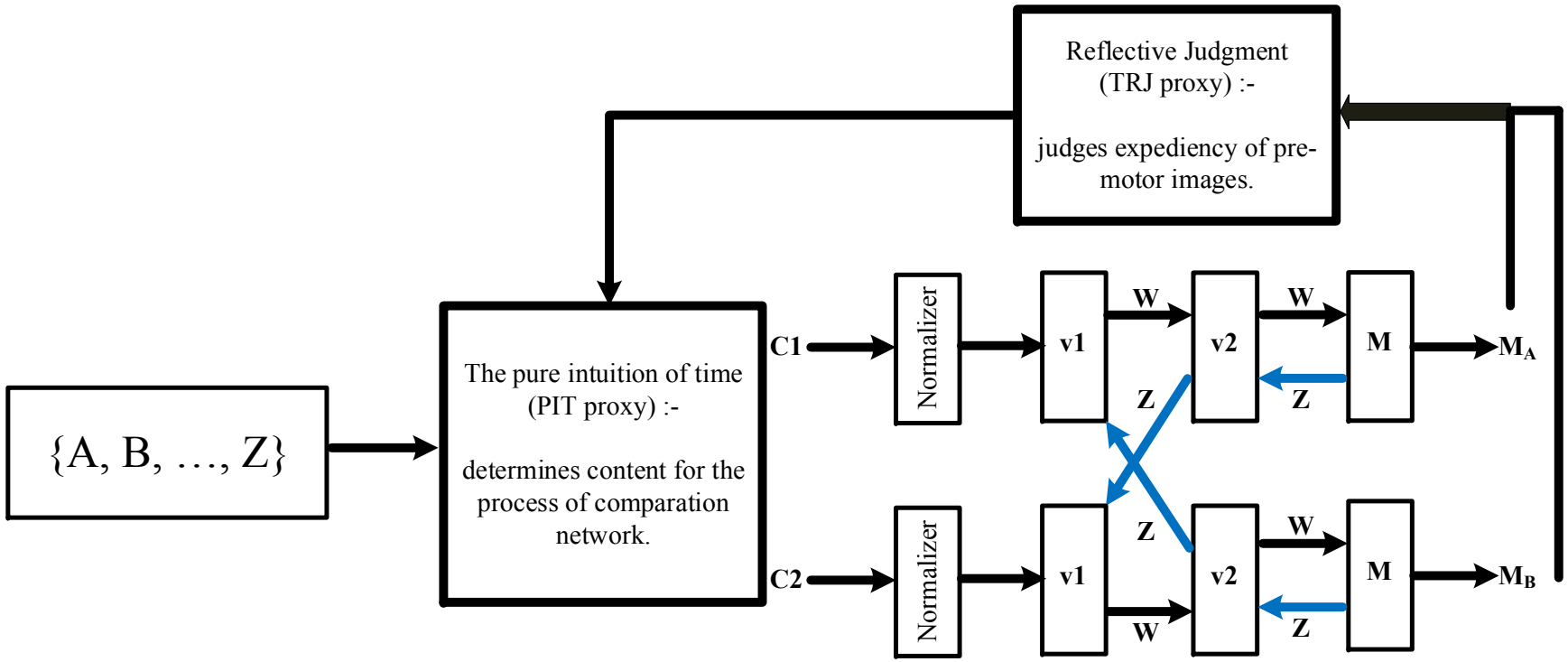


Figure 5.24. The minimal anatomy for generating equivalence relation. Notice that this is based on the previous minimal network (Fig.5.2d) with the addition of PIT proxy (Fig.5.23) determining the content (comparands, C1 & C2) from a sequence of patterns and also receives report of expediency from TRJ proxy.

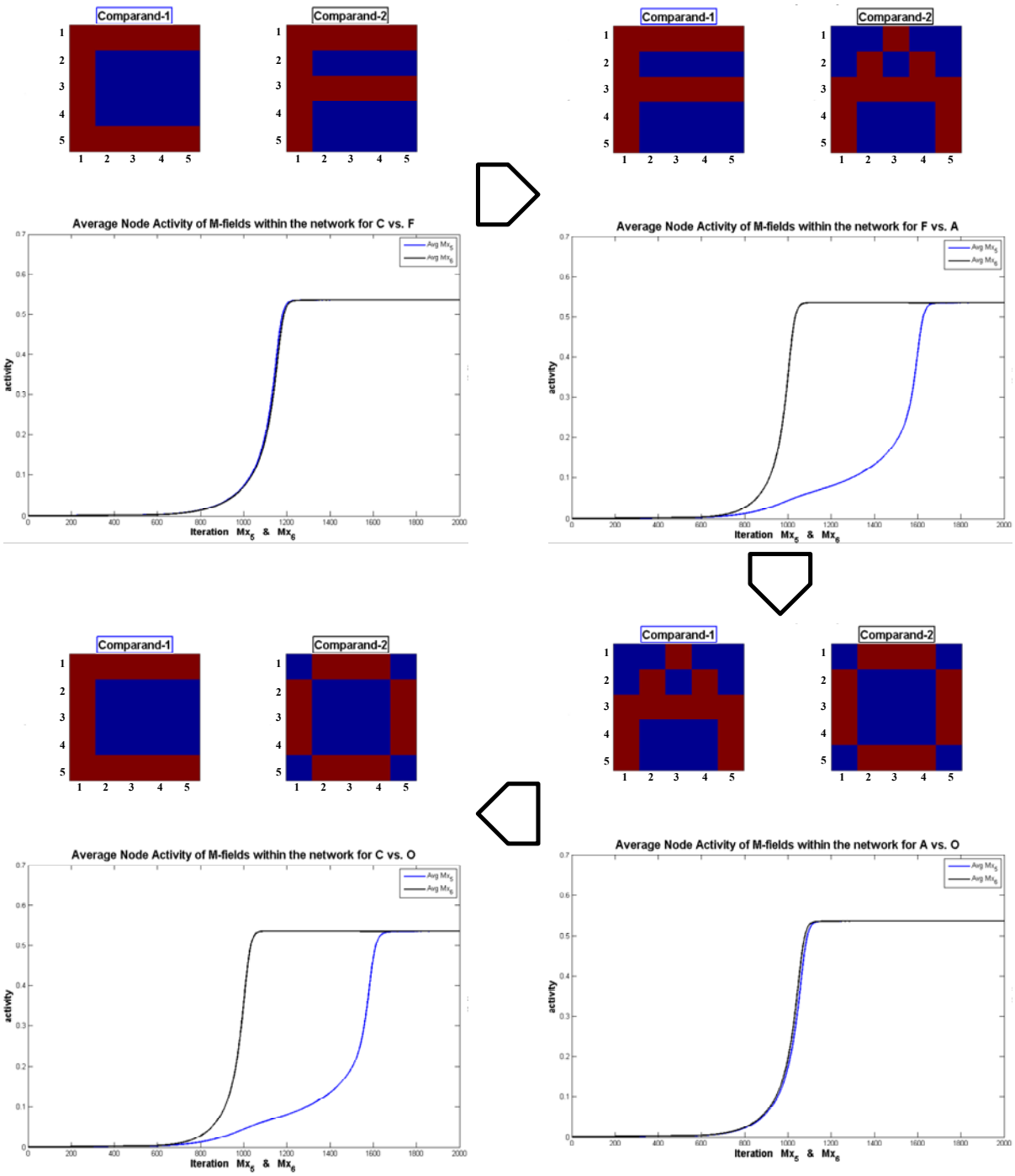


Figure 5.25. Time-series plot for the pattern sequence {C, F, A, O}. The top sub-network in Fig.5.24 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.24 receives comparand-2, color coded as black. Given a pattern-combo the plots shows the activities of average (median) nodes of respective M-layers. Clockwise from the top-left, the network proceeds from {C, F} and then {F, A}, {A, O} and finally {C, O}. Here all the patterns are shown to have a relationship.

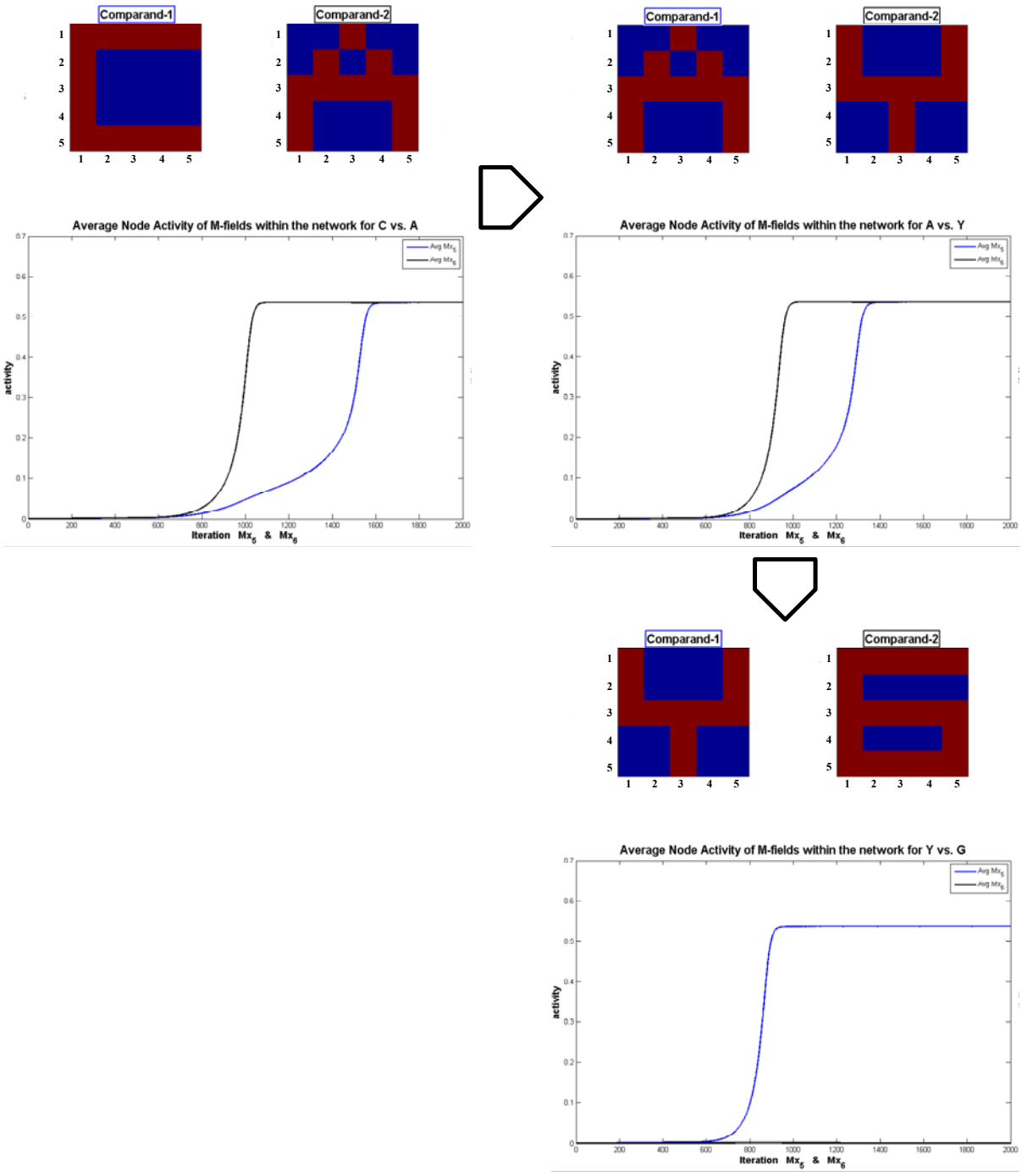


Figure 5.26. Time-series plot for the pattern sequence {C, A, Y, G}. The top sub-network in Fig.5.24 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.24 receives comparand-2, color coded as black. Given a pattern-combo the plots shows the activities of average (median) nodes of respective M-layers. Clockwise from the top-left, the network finds relationship in {C, A} and {A, Y} but not in {Y, G} and hence stopped for {C, G}.

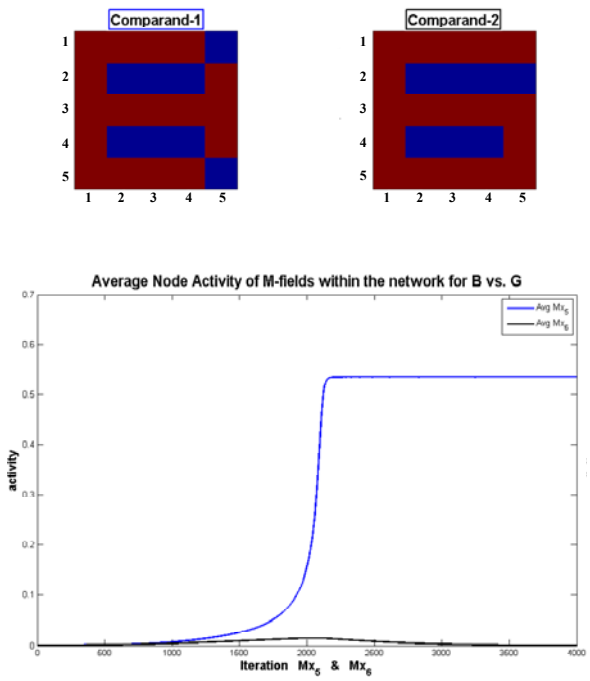


Figure 5.27. Time-series plot for the pattern sequence {B, G, D, E, P, S}. The plot shows the activities of average (median) nodes of respective M-layers for pattern-combo, {B, G}. Since there is no relationship for {B, G} it is stopped for {G, D} and succeeding pattern-combos.

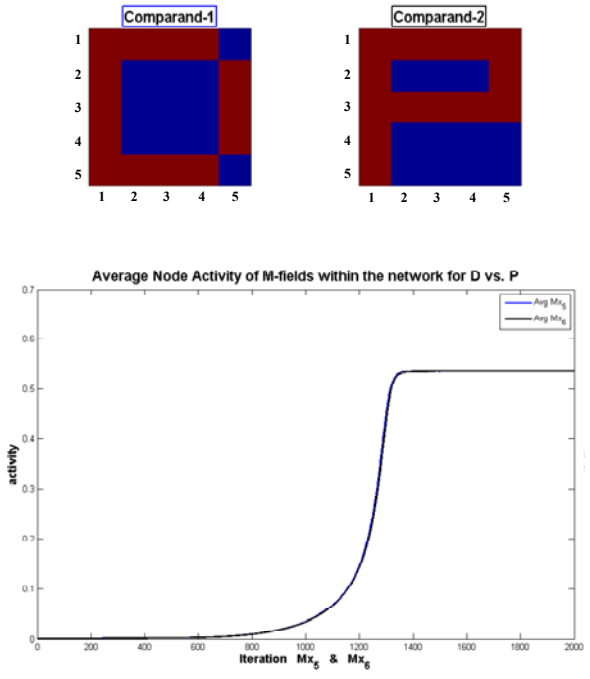


Figure 5.28. Time-series plot for the pattern sequence {D, P}. The plot shows the activities of average (median) nodes of respective M-layers for pattern-combo, {D, P}. Here all the patterns are shown to have a relationship.

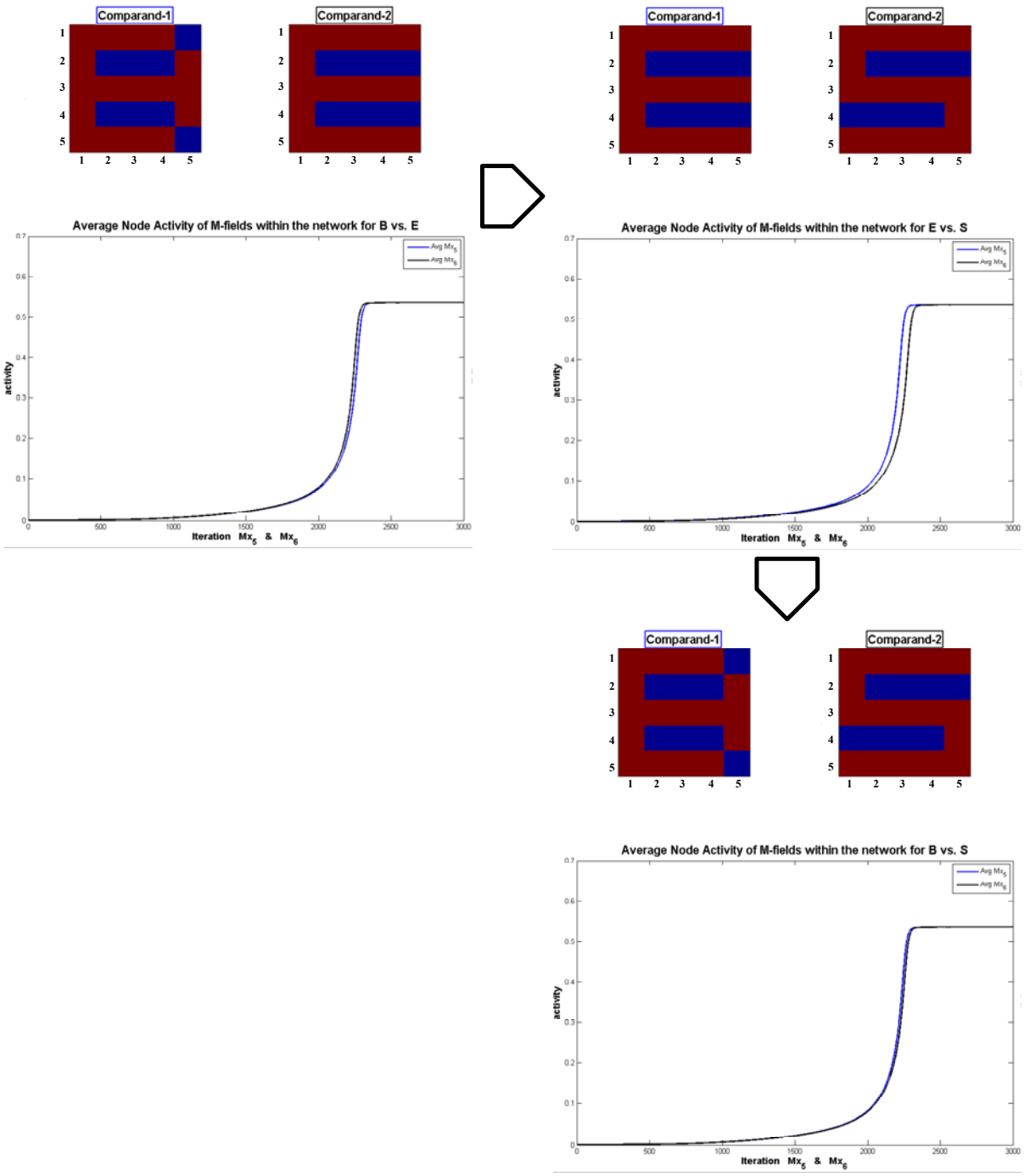


Figure 5.29. Time-series plot for the pattern sequence {B, E, S}. The top sub-network in Fig.5.24 receives comparand-1, color coded as blue. The bottom sub-network in Fig.5.24 receives comparand-2, color coded as black. Given a pattern-combo the plots shows the activities of average (median) nodes of respective M-layers. Clockwise from the top-left, the network proceeds from {B, E} and then {E, S} and finally {B, S}. Here all the patterns are shown to have a relationship.

These demonstrations do not strictly show the complete synthesis of an equivalence relation. To do that transitivity would have to be demonstrated over the whole of the candidate set of patterns. For instance, demonstrating transitivity for the set {C, F, A, O} requires, in addition to the demonstration of figure 5.25, that the following pairs have the same relationship:

$C \rightarrow A$,
$F \rightarrow O$.

However, recall that if a relationship is found between patterns then as a dynamic property of the network, the relationship is reflexive and also symmetric. Thus for the pairs that have been shown to have the same relationship (Fig.5.25), due to the property of the network dynamics, the following pairs will also have the same relationship:

$C \leftarrow F$ (since, $C \rightarrow F$),
$F \leftarrow A$ (since, $F \rightarrow A$),
$O \leftarrow A$ (since, $A \rightarrow O$), and
$C \leftarrow O$ (since, $C \rightarrow O$).

In other words, the relationship in the pattern-combos implies that the patterns are interchangeable, i.e., F & A are interchangeable and A & O are interchangeable. Thus, the following pairs will have the same relationship:

$C \rightarrow A$ (since, $C \rightarrow F$ & $F \leftrightarrow A$), and
$F \rightarrow O$ (since, $F \rightarrow A$ & $A \leftrightarrow O$).

Simulations (refer table 5.1) of these pattern pairs confirm this.

The demonstrations provided here shows that the generation of equivalence relation by means of the minimal network of figure 5.24 is possible, and this is sufficient to demonstrate the Verstandes Actus of comparation. In this, the particular matter of equivalence is irrelevant, as it must be according to the laws of mental physics.

In the OB, determination of the process of the PIT is regulated by ratio-expression from practical Reason acting through determining judgment (Fig.3.11). Therefore, the complete synthesis of objective equivalence is realized by means of the overall synthesis of judgmentation.

In conclusion, the above minimal neural network (Fig.5.24) has the ability to generate equivalence relations. With the exception of pattern sequence length L = 2, if the network demonstrates relationship for the L^{th} process then the patterns within that sequence are candidates for an equivalence relation. It should be emphasized that the generation of equivalence relations is based on network dynamics without any a priori knowledge. In other words, the relationships are generated without introducing any a priori information about the patterns.