

Sensitivity from Short-Term Memory vs. Stability from Long-Term Memory in Visual Attention Method

María T. López¹, Antonio Fernández-Caballero¹,
Miguel A. Fernández¹, and Ana E. Delgado²

¹ Universidad de Castilla-La Mancha,
Escuela Politécnica Superior, 02071 - Albacete, Spain

{mlopez,caballer,miki}@info-ab.uclm.es

² Universidad Nacional de Educación a Distancia,
E.T.S.I. Informática, 28040 - Madrid, Spain
adelgado@dia.uned.es

Abstract. In this paper a special focus on the relationship between sensitivity and stability in a dynamic selective visual attention method is described. In this proposal sensitivity is associated to short-term memory and stability to long-term memory, respectively. In first place, all necessary mechanisms to provide sensitivity to the system are included in order to succeed in keeping the attention in our short-term memory. Frame to frame attention is captured on elements constructed from image pixels that fulfill the requirements established by the user and gotten after feature integration. Then, stability is provided by including mechanisms to reinforce attention, in such a way that elements that accept the user's predefined requirements are strengthened up to be configured as the system attention centre stored in our long-term memory.

1 Introduction

The name dynamic selective visual attention (DSVA) embraces a set of image processing mechanisms for focusing vision on those regions of the image where there are relevant local space-time events. These DSVA mechanisms help find, using an active search process, the relevant information at each moment to perform the interaction task with the system [1], [2]. In this paper a special focus on the behavior of sensitivity and stability in our visual attention method is pursued. Sensitivity and stability are terms widely expressed in dynamic systems [3]. In systems associated to image sequences sensitivity and stability have also been explored due to their importance [4], [5]. Our intention is to introduce these concepts in dynamic visual attention, associating sensitivity to short-term memory and stability to long-term memory, respectively. Fig. 1 shows the block diagram that illustrates the two components of sensitivity and stability of the DSVA task as studied in this paper.

As also shown in Fig. 1, our solution to DSVA defines a model with two types of processes: bottom-up processes (based on the scene), which enable to

extract features from the input image and allow to create the elements of interest; and top-down processes (based on the object) by means of which the features are integrated. The selection of the interest elements of the scene starts with setting some criteria based on the features extracted from the elements (Feature Extraction). This way, in first place, all necessary mechanisms to provide sensitivity to the system are included in order to succeed in capturing the attention. Frame to frame attention is derived (Attention Building) to elements constructed from image pixels that fulfill the requirements established by the user and gotten after a Feature Integration. On the other hand, stability has to be provided to the system. This has been achieved by including mechanisms to reinforce attention (Attention Reinforcement), in such a way that elements that accept the user's predefined requirements are strengthened up to be configured as the system attention centre. Thus, the relationship between sensitivity gotten in the Short-Term Memory (Attention Building) and stability obtained in the Long-Term Memory (Attention Reinforcement) is developed in our proposal.

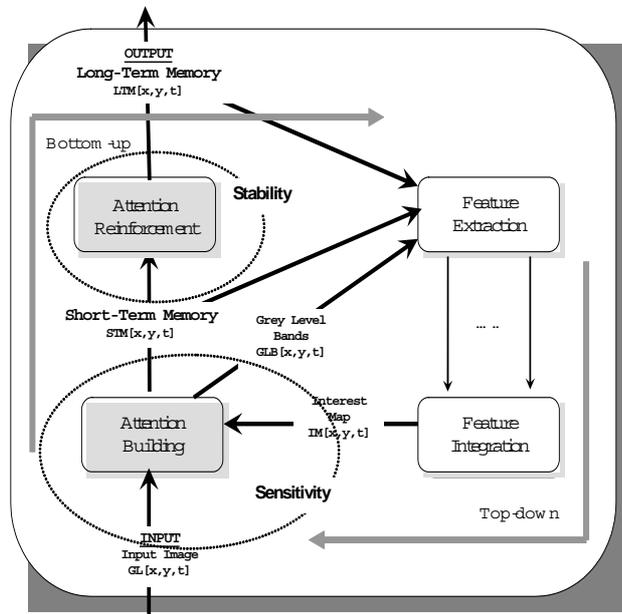


Fig. 1. DSVA block diagram with special emphasis on Attention Building and Attention Reinforcement

In previous works of our research team some methods based on image segmentation from motion have already been used. These methods are the permanency effect and the lateral interaction [6]. Based on the satisfactory results of these algorithms [7], [8], in this paper we propose to use mechanisms of charge and discharge together with mechanisms of lateral interaction to solve the fundamental aspects of sensitivity and stability in the DSVA task.

2 Short-Term and Long-Term Memory in DSVA Method

Short-term memory and long-term memory are expressions taken from cognitive psychology. Short-term memory (STM) -also called working memory or functional memory- is the cognitive system that allows keeping active a limited amount of information for a brief period of time [9], [10], [11], [12]. It was thought to have two functions: storing material that we have to recall in a few seconds and providing a gateway to long-term memory (LTM) [13]. LTM contrasts with STM in that information can be stored for extended periods of time. In standard theories of memory [13], [14], information can be stored in LTM only after it has been stored in STM, and even then, storage in LTM is a probabilistic event. Originally, it was proposed that the probability of storage in LTM is a function of the time an item was maintained in STM. More recently, Anderson [15] suggested that the probability of storage is a function of the number of times an item enters STM. LTM has a strong influence on perception through top-down processing. This is the process by which our prior knowledge affects how we perceive sensory information. LTM influences what aspects of a situation we pay attention to -allowing us to focus on relevant information and disregard what is not important [16].

In the DSVA method proposed in this paper all necessary mechanisms necessary to obtain a Short-Term Memory and a Long-Term Memory are explained. The mechanisms used to generate the Short-Term Memory endow the system of sensitivity, as it includes elements associated to interest points in the memory at each frame. But, the Short-Term Memory introduced is noisy, as blobs that are not of a real interest to the user may appear. In order to generate the Long-Term Memory, that is to say, in order to provide stability, some cues are included for inserting into the Long-Term Memory all elements reinforced in the Short-Term Memory through a persistency measure.

3 Sensitivity through Attention Building

The purpose of Attention Building is to select and to label zones (blobs) of the objects (figures) to pay attention on. See, therefore, that after processing Attention Building, not the complete figures are classified, but each one of the blobs, understood as homogeneous connected zones that form the figures, are marked with different labels. Obviously, the blobs are built from image points that fulfill the requisites established by the guidelines of the observer (points of interest). Fig. 2 shows a process scheme for Attention Building. The output of Attention Building is precisely called Short-Term Memory. In our case, only blobs constructed in the Short-Term Memory will potentially form the figures of the system's Long-Term Memory.

In our proposal the blobs of the Short-Term Memory are built from the information provided through the so called Interest Map and from the input image divided into Grey Level Bands. The Interest Map is obtained by performing feature integration, both of motion and shape features. For each image pixel, in

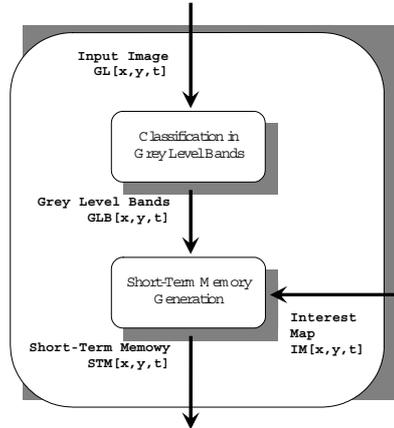


Fig. 2. Attention Building process scheme

the Interest Map the result of a comparison among three classes - "activator", "inhibitor" and "neutral"- is stored. The interest points are those points of the Interest Map labeled as "activator" points.

3.1 Classification in Grey Level Bands

Classification in Grey Level Bands transforms the 256 grey level input images into images with a minor number of levels. These new images are called images segmented into Grey Level Bands (*GLB*). The reason to working with grey level bands is twofold. (1) Some traditional methods of motion detection are based on image differencing. The noise level diminishes for little changes in grey level (or luminosity) of a same object between two consecutive images, when joining a range of grey levels into a single band. (2) On the other hand, a decrease of the computational complexity is achieved, bearing in mind the great parallelism used in the algorithms of the proposed model. We now calculate in parallel in the order of magnitude of grey level bands n , and not of grey levels N , where $N > n$.

The calculus of the grey level band of pixel $[x, y]$ at t , $GLB[x, y, t]$, is expressed in Equation 1. As you may notice, this is just an easy scale transformation.

$$GLB[x, y, t] = \frac{GL[x, y, t] \cdot n}{GL_{max} - GL_{min} + 1} + 1 \quad (1)$$

where n is the number of grey level bands in which the image is split, GL_{max} is the maximum and GL_{min} are the minimum grey levels, respectively, of the input image.

3.2 Short-Term Memory Generation

The objective of Short-Term Memory Generation is firstly to select and to label (to classify numerically) image blobs associated to pixels of interest -pixels that

possess dynamic features in predefined numerical intervals. Secondly, it eliminates the blobs whose shape features do not correspond with the pre-established ones. In order to achieve these aims, the images in Grey Level Bands are segmented into regions composed of connected points whose luminosity level belongs to a same interval (or grey level band) and to select only connected regions that include some "activator" point (or, point of interest) in the Interest Map. Each region or zone of uniform grey level is a blob of potential interest in the scene.

The idea consists in overlapping, as with two superposed transparencies, the Grey Level Bands image of the current frame (t) with the Interest Map image built at the previous frame ($t - 1$). At t , only blobs of the Grey Level Bands image are selected where at least one point of interest fell at $t - 1$ in the Interest Map. Nevertheless, not the total blob is taken; pixels that coincide with points of the Interest Map classified as "inhibitors" are eliminated. The computational model used to perform the preceding steps incorporates the notion of lateral interaction, which enables that the points of interest flood their zones of uniform grey levels whilst eliminating all points classified as "inhibitors". In order to achieve the aims of Short-Term Memory Generation, the processes shown in Fig. 3 are performed.

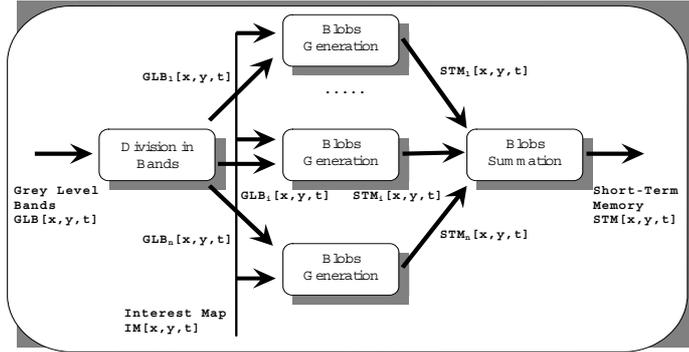


Fig. 3. Short-Term Memory Generation process scheme

Division in Bands. Division in Bands obtains from an image in grey level bands, $GLB[x, y, t]$, n binary images $GLB_i[x, y, t]$ (one image for each band). Each one of these images, $GLB_i[x, y, t]$, stores a value of 1 for a pixel whose grey level band is i and a 0 in the opposite case. That is to say (Equation 2):

$$GLB_i[x, y, t] = \begin{cases} 1, & \text{if } GLB[x, y, t] = i \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Blobs Generation. For each $GLB_i[x, y, t]$, the different connected regions that include an "activator" point in the Interest Map and that do not correspond to "inhibitor" points in the Interest Map are labeled. Thus, Blobs Generation gets

and labels for each grey level band, pixels belonging to connected regions that include any "activator" point in the Interest Map but do not correspond with "inhibitor" points of the Interest Map. Its output, Short-Term Memory for Grey Level Band i , $STM_i[x, y, t]$, stores for each pixel the label corresponding to the generated blob if it belongs to the blob, or the value 0. Let us define $v_{activator}$ as the value given to the points of interest ("activators") of the Interest Map, $v_{neutral}$ as the value for the "neutral" points of the Interest Map, and $v_{inhibitor}$ as the value for the "inhibitor" points of the Interest Map. Let also NR be the number of rows of the image, and NC the number of columns of the image. Firstly, all points where $GLB_i[x, y, t] = 1$ are assigned an initial and provisional label value as shown in Equation 3:

$$STM_i[x, y, t] = \begin{cases} x * NC + y + 1, & \text{if } GLB_i[x, y, t] = i \wedge IM[x, y, t] = v_{activator} \\ NR * NC + 1, & \text{if } GLB_i[x, y, t] = i \wedge IM[x, y, t] = v_{neutral} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $IM[x, y, t]$ is the value of the Interest Map at pixel $[x, y]$. This value corresponds to $(x * NC + y + 1)$ when $IM[x, y, t] = v_{activator}$, to a greater value than $NR * NC$ when $IM[x, y, t] = v_{neutral}$ and to value 0 in the rest of the cases. In other words, if pixel $[x, y]$ belongs to the grey level band and corresponds to a point of interest of the Interest Map, the tentative value for it is a function of its proper coordinate. Now, if the pixel belongs to the grey level band but corresponds to a "neutral" point of the Interest Map, the provisional value given to it is a value greater than any possible value of the coordinate function. In any other case, the value is 0. This initial value assignment to all pixels serves to get an agreement in the labels of the blobs after a negotiation (consensus) period. The label value for each pixel $[x, y]$ is iteratively calculated as the minimum value of the proper value of the pixel and the value of its 8 surrounding pixels. Of course, there will only be collaboration among neighboring pixels that possess an initial value greater than 0. The iterative calculus up to obtaining a common value for all pixels of a same blob is shown in Equation 4.

$$STM_i[x, y, t] = \min(STM_i[\alpha, \beta, t]), \forall [\alpha, \beta] \in [x \pm 1, y \pm 1] \quad (4)$$

whenever $0 \leq STM_i[\alpha, \beta, t] \leq NR * NC + 1$.

Thus, blobs are labelled with the ordinal corresponding to the point with the lowest coordinate (if taking as origin the superior left image pixel).

Blobs Summation. Lastly, Blobs Summation gets the Short-Term Memory, $STM[x, y, t]$, as the result of summing up all blobs computed at each of the n Short-Term Memories for Grey Level Band i , $STM_i[x, y, t]$, where $i = 1, 2, \dots, n$. The final value for each pixel $[x, y]$ is the maximum value of the n $STM_i[x, y, t]$. Only values that possess a label value less than $NC \cdot NR + 1$ are considered, as shown in Equation 5.

$$STM[x, y, t] = \max_i STM_i[x, y, t], \forall i \in [1..n] | STM_i[x, y, t] < NR * NC + 1 \quad (5)$$

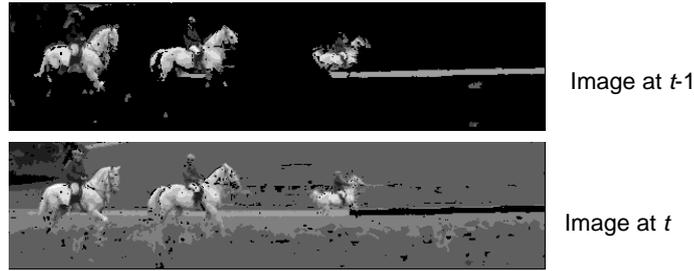


Fig. 4. Short-Term Memory for a couple of images

Notice that this maximum selection operation has to be performed for all elements of matrixes $STM_i[x, y, t]$ to obtain the corresponding element in a single matrix of blobs, $STM[x, y, t]$. This way all blobs of all grey level bands have been united and labeled with a common value. Fig. 4 shows the contents of the Short-Term Memory after processing all steps of Attention Building.

See also the sensitivity of the task through the contents of the Short-Term Memory in two consecutive frames as shown Fig. 4 taken from the picture "The Living Daylights". The input sequence has been captured by a camera in constant translational movement following the motion of the horse riders. Elements of the Short-Term Memory are composed of connected pixels that are not drawn in black color. The attention focus pursued in this case is the set of horses and horsemen. As you may notice, all interest elements are really detected. Nevertheless, other elements appear that neither are of the user's interest. The example shows the necessity for stability.

4 Stability through Attention Reinforcement

The mechanisms used to generate the Short-Term Memory endow the system of sensitivity, as it includes elements associated to interest points ("activators") in the memory at each frame. Unfortunately, in the Short-Term Memory scene blobs whose shape features do not correspond to those defined by the observer may appear at a time instant t . This is precisely because their shape features have not yet been studied. But, if these blobs shape features really do not seem to be interesting for the observer, they will appear as "inhibitors" in $t + 1$ in the Interest Map (now, in $t + 1$, their shape features will have been obtained). And, this means that in $t + 1$ they will disappear from the Short-Term Memory.

In order to obtain at each frame only blobs with the desired features, Attention Reinforcement performs an accumulative mechanism followed by a threshold. Accumulation is realized on pixels that have a value different from 0 (pixels that do not belong to labeled blobs) in the Short-Term Memory. The result of this accumulative process followed by a threshold offers as output the Long-Term Memory, $LTM[x, y, t]$. More concretely, pixels that appear with a value different from 0 in the Short-Term Memory reinforce attention, whilst those that appear

with a value 0 diminish the attention value. The process manages to keep atti-

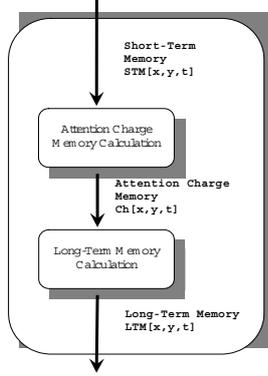


Fig. 5. Scheme for "Attention Reinforcement"

ated in a stable way a set of pixels that belong to a group of objects (figures) of the scene that are interesting for the observer. Fig. 5 shows the decomposition of Attention Reinforcement into Attention Charge Memory Calculation and Long-Term Memory Calculation.

4.1 Attention Charge Memory Calculation

Attention Charge Memory Calculation performs an accumulative computation on the Short-Term Memory to get the Attention Charge Memory $Ch[x, y, t]$. The idea underlying Attention Charge Memory Calculation is that pixels that belong to a blob of the Short-Term Memory through time reinforce attention whilst all other ones decrease attention. The accumulative computation [1], [6], [7] takes the form of Equation 6, based on the more general charge/discharge accumulative computation mode [17].

$$Ch[x, y, t] = \begin{cases} \max(Ch[x, y, t-1] - D, Ch_{min}), & \text{if } STM[x, y, t] = 0 \\ \min(Ch[x, y, t-1] + C, Ch_{max}), & \text{if } NC * NR + 1 > STM[x, y, t] > 0 \end{cases} \quad (6)$$

where Ch_{min} is the minimum and Ch_{max} is the maximum value, respectively, that the values stored in the Attention Charge Memory can reach, and C and D are the charge increment and decrement, respectively, in the memory computation. The charge value $Ch[x, y, t]$ goes incrementing up to Ch_{max} , if pixel $[x, y]$ belongs to a blob of the Short-Term Memory, and goes decrementing down to Ch_{min} if the pixel does not. Charge value $Ch[x, y, t]$ represents a measure of the persistency of a blob in the Short-Term Memory on each image pixel $[x, y]$.

4.2 Long-Term Memory Calculation

Long-Term Memory Calculation produces, starting from the Attention Charge Memory, the points that configure the Long-Term Memory, labeling the figures

obtained. The focus is in form of figures, obtained by the union of the connected blobs that have appeared successively in the Short-Term Memory and whose value in the Attention Charge Memory is greater or equal to a given threshold, θ . In the output, the label corresponding to the figure is stored; value 0 is assigned to all pixels that do not belong to any figure. Firstly, the Long-Term Memory at pixel $[x, y]$ is assigned an initial and provisional value (yet not agreed with the neighbours) corresponding to a function of the coordinate of the pixel, if the charge value overcomes threshold θ (see Equation 7):

$$LTM[x, y, t] = \begin{cases} x * NC + y + 1, & \text{if } Ch[x, y, t] > \theta \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Next, in an iterative way up to reaching a common value for all pixels of a same figure (by calculating the minimum value of each pixel and its 8 surrounding neighbours), a calculation is performed according to Equation 8:

$$LTM[x, y, t] = \min(LTM[\alpha, \beta, t]), \forall [\alpha, \beta] \in [x \pm 1, y \pm 1] | 0 < LTM[\alpha, \beta, t] \quad (8)$$

Fig. 6 now shows the contents of the Long-Term Memory after processing all steps of Attention Reinforcement. Notice that we got the desired stability.

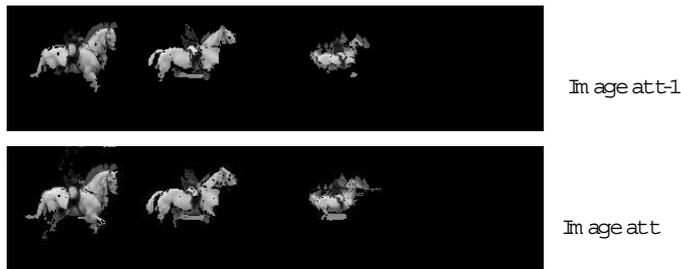


Fig. 6. Long-Term Memory for a couple of images

5 Conclusions

In this paper the relationship between sensitivity and stability in our particular DSVa method has been described. In the proposal sensitivity has been associated to Short-Term Memory and stability to Long-Term Memory, respectively. The generation of the Short-Term Memory, and hence the coming out of sensitivity, is related to task Attention Building, whereas Long-Term Memory is obtained after Attention Reinforcement, getting the desired stability to visual attention.

As described, Attention Building is achieved by means of two main steps, namely Classification in Grey Level Bands and Short-Term Memory Generation, getting as output figure blobs in the Short-Term Memory in a noisy way. On the other hand, Attention Reinforcement is divided into Attention Charge Memory Calculation and Long-Term Memory Calculation, and obtains persistent figures through time in a stable Long-Term Memory.

6 Acknowledgements

This work is supported in part by the Spanish CICYT TIN2004-07661-C02-01 and TIN2004-07661-C02-02 grants.

References

1. Fernández-Caballero, A., López, M.T., Fernández, M.A., Mira, J., Delgado, A.E., López-Valles, J.M.: Accumulative computation method for motion features extraction in active selective visual attention. *LNCS* **3368** (2004) 206–215
2. López, M.T., Fernández, M.A., Fernández-Caballero, A., Delgado, A.E.: Neurally inspired mechanisms for the active visual attention map generation task. *LNCS* **2686** (2003) 694–701
3. Oppenheim, A.V., Willsky, A.S., Nawab, S.H.: *Signals and Systems*, 2nd edition. Prentice-Hall Inc (1997)
4. Daniilidis, K., Spetsakis, M.: Understanding noise sensitivity in structure from motion. Aloimonos, Y. (ed.), *Visual Navigation* (1996) 61–88
5. Fermüller, C., Aloimonos, Y.: Algorithm-independent stability analysis of structure from motion. University of Maryland TR 3691 (1996)
6. Fernández-Caballero, A., Mira, J., Delgado, A.E., Fernández, M.A.: Lateral interaction in accumulative computation: A model for motion detection. *Neurocomputing* **50** (2003) 341–364
7. Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Spatio-temporal shape building from image sequences using lateral interaction in accumulative computation. *Pattern Recognition* **36**:5 (2003) 1131–1142
8. Fernández-Caballero, A., Mira, J., Fernández, M.A., Delgado, A.E.: On motion detection through a multi-layer neural network architecture. *Neural Networks* **16**:2 (2003) 205–222
9. Baddeley, A.D., Hitch, G.J.: Short-Term Memory. Bower, G., (ed.), *Recent Advances in Learning and Motivation* **8** (1974)
10. O'Reilly, R.C., Braver, T.S., Cohen, J.D.: A biologically-based computational model of working memory. Miyake, A., Shah P. (eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (1999) 375–411
11. Awh, E., Anllo-Vento, L., Hillyard, S.A.: The role of spatial selective attention in working memory for locations: evidence from event-related potentials. *Journal of Cognitive Neuroscience* **12** (2000) 840–847
12. Awh, E., Jonides J.: Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences* **5**(3) (2001) 119–126
13. Atkinson, R.C., R. M. Shiffrin, R.M.: Human memory: A proposed system and its control processes. Spence, K.W., Spence, J.T. (eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory* **2** (1968)
14. Waugh, N., Norman, D.A.: Primary memory. *Psychological Review* **72** (1965) 89–104
15. Anderson, J.R.: *The Architecture of Cognition*. Harvard University Press (1983)
16. Winn, W., Snyder, D.: Cognitive perspectives in psychology. Jonassen, D.H. (ed.), *Handbook of Research for Educational Communications and Technology* (1996) 115–122
17. Mira, J., Fernández, M.A., López, M.T., Delgado, A.E., Fernández-Caballero, A.: A model of neural inspiration for local accumulative computation. *LNCS* **2809** (2003) 427–435