Modeling Simultanagnosia

Anna Belardinelli (belardinelli@uni-tuebingen.de)*

Johannes M. Kurz (johannes.kurz@student.uni-tuebingen.de)*

Esther F. Kutter (esther.kutter@student.uni-tuebingen.de)*

Heiko Neumann (heiko.neumann@uni-ulm.de)
Institute of Neural Information Processing, Ulm University,
James-Franck-Ring 89081, Ulm, Germany

Hans-Otto Karnath (karnath@uni-tuebingen.de)
Neuropsychology Section, Dept. of Cognitive Neurology, Hertie Institute for Clinical Brain Research,
Hoppe-Seyler-Str. 3 72076 Tübingen, Germany

Martin V. Butz (martin.butz@uni-tuebingen.de)*
* Cognitive Modeling, Dept. of Computer Science, University of Tübingen,
Sand 14, 72076, Tübingen, Germany

Abstract

Simultanagnosia is a visual cognitive disorder following a bilateral lesion in parieto-occipital brain areas. It affects the patients integrative perception so that scenes or hierarchically organized objects cannot be perceived as a whole but just in a piecemeal fashion. Qualitative explanations consider impairments of attentional selection mechanisms, feature binding, or global shape processing. Until now, however, no computational model has been used to quantitatively reproduce the performance patients suffering from simultanagnosia. We focus on modeling the impairment of recognizing hierarchical stimuli (Navon letters), which have been used in several studies with patients and healthy subjects. In particular, we apply the established HMAX model of object recognition, specifically trained on letter recognition, to investigate the role of low-level, bottom-up processing, salience, and the selection of a window of attention. We also assess to which extent a top-down modulatory mechanism is necessary to quantitatively reproduce the global letter recognition performance in patients. Our results indicate that as long as a bottom-up segmentation of the global shape from local elements is possible, global shape recognition succeeds. However, when top-down form completion appears necessary to identify the global shape, the current, bottom-up processing model fails. Moreover, we replicate training effects in the global task, which are comparable with patients performance in similar tasks. The present results suggest several promising future research directions to extend the model for modeling the mechanisms underlying global shape recognition in healthy subjects as well as the impairments in patients suffering from simultanagnosia.

Keywords: simultanagnosia; cognitive modeling; global shape perception; models of the visual cortex.

Introduction

Simultanagnosia, sometimes described as the inability to see the forest for the trees, is a visual deficit, which is part of the Bálint syndrome (Bálint, 1909). Patients suffering from simultanagnosia experience severe difficulties in perceiving more than one object at a time. They are usually unable to recognize complex scenes or composed objects (from their global shape) (Wolpert, 1924; Luria, 1959). The symptoms generally appear after a bilateral lesion within and in the vicinity of parieto-occipital cortical areas. Other typical symptoms include the inability to recognize two distinct objects at once, but to perceive them as a single unit if joined by a line or when combinable into a semantic unit (Luria, 1959; Coslett & Saffran, 1991; Dalrymple, Kingstone, & Barton, 2007).

Since the recognition of single objects per se is not hampered, simultanagnosia is considered to impair attention or substantially reduce its resources. Attention and recognition (and conscious perception) indeed may interact at multiple points in the visual hierarchy, and in different waves of feedforward and feed-back processing (Lamme & Roelfsema, 2000). To add to this multifaceted picture, evidences were presented both in favor of a space-based and of an object-based account of the disorder (Dalrymple, Barton, & Kingstone, 2013). It is hence not clear which and how many cognitive functions are actually damaged. Different types of symptoms could be due to different subtypes of simultanagnosia, impairing concurrently or exclusively early vision, feature binding, and object and location binding (Coslett & Lie, 2008). Research efforts have investigated to which extent the spatial relationships between figure elements are critical and whether the problem would be explicable as a restricted window of attention, or the inability to shift from one item to another (Luria, 1959; Farah, 1990) as a result of an extreme winner-takes-all competition (Jackson et al., 2009). Alternative explanations point to a reduced processing speed (Duncan et al., 2003), or to the inability to integrate single stimuli and even single features into a unity (Robertson, 2003; Coslett & Lie, 2008).

Qualitative explanations cannot reproduce or predict patients' performance quantitatively, test which mechanism exactly produces the observed effects, how selective attention contributes to integrative perception, and how bottom-up saliency and top-down modulation interact in selecting the level of representation of the observed scene. To the best of our knowledge, no neurocognitve model of the visual cortex has been devised or applied to model simultanagnosia so far. To develop a robust and flexible-enough model of such a disorder would have many far-reaching implications: understanding this disorder more deeply would help identify the mechanisms of global shape perception and the influence of attention in healthy subjects and improve current computer vision systems in a cognitive way.

In this paper we put forward an experimental testbed that consists of a simple set of stimuli, a state-of-the-art model of object recognition (Serre et al., 2007), and possible extensions thereof. The purpose of this testbed is to account for both healthy and brain-lesioned subjects' performance in order to investigate which mechanism is responsible for which symptom. The presented model is able to cope with the stimuli used in a real study (Huberle & Karnath, 2006) and to reproduce the performance of patients. We finally discuss gained insights and future developments of the model in order to reproduce healthy subjects' performance and be able to selectively test which modules are responsible for the impairments present in simultanagnosia.

A testbed for modeling simultanagnosia

Many studies reported that patients suffering from simultanagnosia suffer from 'local capture' effects, that is, the tendency to focus attention on atomic elements of a larger, hierarchical picture, and the inability to switch between the local and global level of the image (Karnath, Ferber, Rorden, & Driver, 2000; Huberle & Karnath, 2006). This is typically demonstrated via Navon hierarchical stimuli: large letters that consists of many much smaller letters (Navon, 1977). Healthy subjects usually present a "global precedence" effect, being faster at recognizing the global letter, while patients suffering from simultanagnosia are typically stuck on the local letter. Yet, this is not a completely binary effect, and in some cases patients are able to recover the global shape (or to see just the global one, as achieved in an experiment by Thomas, Kveraga, Huberle, Karnath, and Bar (2012)). Performance in the global task with Navon stimuli is strongly modulated by the inter-element distance (Huberle & Karnath, 2006), by the size of the global stimulus (Dalrymple et al., 2007), and by the salience of the local elements with respect to the salience of the global letter (Huberle & Karnath, 2010). In order to model simultanagnosia, these stimuli present the advantage of being rather simple (just black and white) and devoid of background clutter and semantic implications, thus enabling an object recognition system to be trained on them rather straightforwardly, without any pre-processing steps. These stimuli, nevertheless, tackle the global processing issue in patients in a critical way, and allow for a fair comparison of performance between patients and model.

Stimulus material and human experimental results

As a first set of data and stimuli, we used the material and results from Huberle and Karnath (2006). In that study, five different letters (A,B,E,H,N) were used for both the local and the global scale, with each global letter made up of one of the other 4 letters (only inconsistent stimuli). The stimuli were 700x700 pixel in size, where the global letters were centered in a bounding box of 600x600 px (10.9°x10.9°) while local letters were 16x16 px in size (0.35°x0.35°). The inter-element distance was varied fivefold (see Figure 1): 145 px (Distance 1, 2.55°), 96 px (Distance 2, 1.70°), 73 px (Distance 3, 1.28°), 48 px (Distance 4, 0.85°), 36 px (Distance 5, 0.64°). Hence the whole dataset consisted of 100 stimuli.

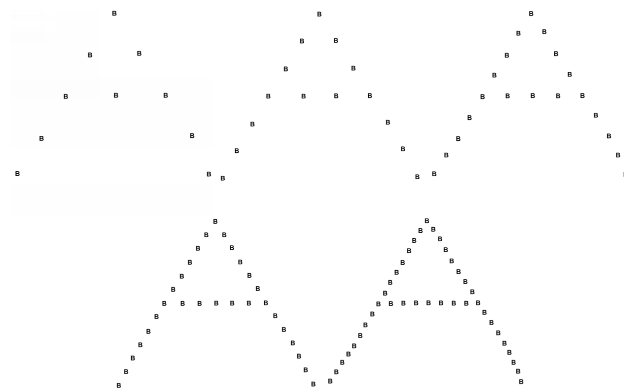

Figure 1: One example (A made of Bs) of the Navon stimuli used in (Huberle & Karnath, 2006) and in this study. Versions for the five inter-element distance are shown (Distance 1, top left, to 5, bottom right).

Two patients and five healthy subjects were presented with each of these stimuli repeatedly, one block with the task of naming the letter at the local scale and another with the task of naming the global letter. Before starting, all subjects were familiarized with the stimuli and the tasks. Control subjects performed 100% correctly on the local task and 99.5% correctly on the global task. Both patients responded correctly well above chance in the local scale task (between 75% and 100% correct responses). Conversely, patients performed rather poorly on the global task. In general, the larger the inter-element distance was, the worse was the performance. One patient performed well above chance with denser stimuli (Distance 4 and 5 around 85% and 75% respectively), while the other could give at most 40% correct answers even on Distance 5 stimuli. Both were at chance level (20%) or around the 95% confidence interval for chance level (33%) for stimuli at distances 1 and 2.

The HMAX model for a letter recognition task

In the computer vision literature a lot of object recognition approaches have been proposed (Andreopoulos & Tsotsos, 2013), with some more grounded in computational neuroscience and inspired by the hierarchical structure of the visual cortex. To be able to generate testable predictions and rely on a robust and biologically plausible model, we chose the well-established HMAX model (Riesenhuber & Poggio, 1999; Serre et al., 2007), in the publicly available implementation by Mutch and Lowe (2006). The model is hierarchically organized in functional layers that can be linked to corresponding cortical areas. It relies on the purely feed-forward, interleaved application of a cascade of Simple ("S") layers, computing the convolution of local filters with the output of the previous layer, and Complex ("C") layers, applying a max-like operator pooling units from the previous layer over a local neighborhood.

In the implementation used, the first layer is the image layer, where the input image is converted to grayscale and the shorter edge is scaled to 140 pixels. A pyramid of 9 levels is created, reducing the image at each scale by a factor $2^{1/4}$. Layer S1 filters the pyramid via a Gabor filter bank (emulating V1 simple cell receptive fields) to gain a multi-scale representation of edge orientation features present at every location. Layer C1 yields a certain position invariance by pooling nearby S1 units with same orientations, extracting the maximal response and subsampling the pyramid (functionally simulating V1 complex cells). The intermediate feature layer S2 uses a learned dictionary of prototype patches (considered as higher level features) to match with C1 unit responses via a Gaussian radial basis function. This step is believed to correspond with processing in V4 or posterior IT. Prototypes (square patches of side 4,8,12 and 16 px) are learned as features characterizing the classes of objects to be recognized. Their structure is critical for discriminating one class from the other. The C2 layer finally reaches higher invariance by computing the maximal response to every prototype and storing it in a vector. The last stage feeds this vector to a classifier system for multi-class classification. In Mutch and Lowe (2006) this is done via Support Vector Machines, or in the Matlab implementation via a one-vs-rest regularized least-squares classifier. This is admittedly the least biologically plausible step, but it represents a decision-making process, and could be, for example, replaced by a Hebbian learning rule based classifier.

As mentioned, this is a feed-forward recognition system, relying on the assumption that the object to recognize is more or less centered in the picture with minor scale variations. Since the Navon stimuli present a ratio of about 1:30 between local and global scale, the huge scale difference poses an additional challenge, specifically regarding invariance in terms of letter size, location, and font thickness. Hence, the model was first tested with parameters as in Mutch and Lowe (2006). The size of the dictionary was set to 500 prototypes. Prototypes are uniformly randomly picked from every given stimulus.To avoid non-meaningful features from the white background, only features with a minimal variance of $\sigma = 10^{-4}$ were allowed in the dictionary. By experimenting with test pictures of the five letters mentioned above (solid letters with the same font as in the Navon stimuli but obtained by varying font thickness, position and size), we assessed how robust and general the resulting model is in, for example, recognizing a letter of a size that is different from the letters used to generate the feature dictionary and to train the classifier. During exploratory testing, we identified a suitable dictionary for the discrimination of the five letters, to be used in the experiments with the Navon letters. For the following experiments, the dictionary was established by learning 500 prototypes from images that contained centered, 6px thick letters of size 62x62 px. The training set for the classifier allowed for some variability in the set by using examples of the five letters in four sizes (14, 26, 62, 108 px), where for the first three sizes the letter was placed at one position in a 10x10, 5x5 or 2x2 grid, respectively, within the 140x140 image. This is the baseline training set (130 images for every letter class) used in the following experiments.

Model results on Navon Stimuli

After tuning and testing the model for the baseline task (letter recognition), we assessed its performance on the Navon stimuli set illustrated above. We first tested the basic model, which does not include an attentional focus, that is, it assumes the whole image as the 'spotlight' of attention. Performance was computed as correct classification of the global letter. The model was tested both on normal letters (control task) and on all Navon stimuli. The dictionary used was always the same, while the training set for the classifier was tested in four different configurations (see grouped bar charts in Figure 2).

In the first case ('baseline'), only the baseline training set was used to train the classifier. This training condition produced a perfect performance in the control task (100% correct classification of all normal letters), but failed with any of the Navon letters at any distance (chance level performance or slightly above for Distance 5). In the second training set, we tested the use of sole Navon letters. Yet, to limit overfitting only samples from one distance for every class were used, either Distance 5 (Figure 2, top plot, 'only Navon'), which was considered more similar to the dictionary set, or Distance 3 (Figure 2, bottom plot), which was expected to be more capable to generalize to other distances. For each of the five letters, only one sample at the chosen distance was shown to train the classifier (5 Navon stimuli in total). Training sets of the five global letters made of the remaining four local letters were produced (4^5 sets). During testing, however, the complete Navon set was presented (average performance is shown). In this case, Navon stimuli were classified better, but mostly those at the trained distance. Normal letters on the other hand were hardly correctly classified. The feature dictionary is, thus, apt to represent also Navon stimuli. However, to be able to correctly classify both types of stimuli, it seems critical to train on both stimulus types (normal and Navon letters). Thus, in the third training set tested ('normal+Navon'), the best-performing Navon letters from the previous test were used for training along with normal 108x108 px letters – one for each class (10 training examples). The performance was much better, with 80% correct classification even at Distance 1. Again, yet, on normal letters performance was rather poor. This suggests that the normal letter presents important features of each class to the classifier, which is hardly available

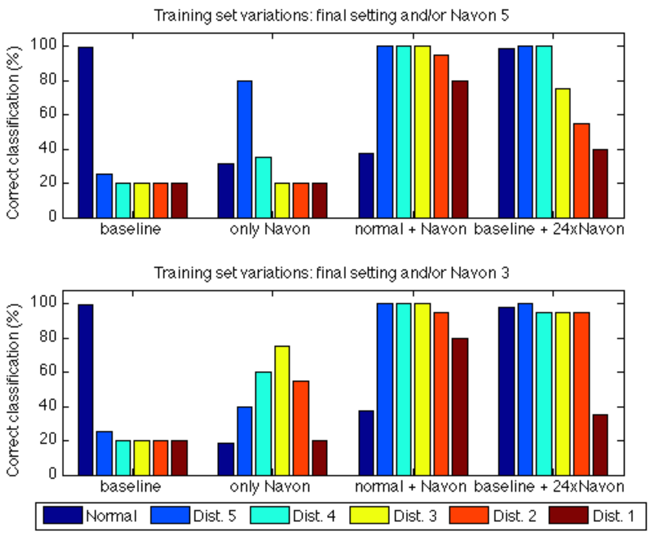

Figure 2: Performance of the HMAX model on normal letters and Navon Stimuli (grouped on the abscissae with respect to inter-element distance). The different color bars represent the different stimuli (normal letters and Navon at five distances). Each group of bars represents a different training set.

in the corresponding Navon stimulus. Hence, training on both helps the classifier to identify the most discriminative features and weigh them more strongly.

The second and third training sets consisted of a very limited number of examples . In the final, fourth test, we thus used the training set along with a more conspicuous number of Navon samples. To still be able to test the capability to generalize over diverse Navon letters, 24 copies of the very same Navon letter (one for each class) at the same distance (either 3 or 5) were put in the training set (hence, 130+24 training images for each letter class). The hypothesis behind this test was that by training with enough varied examples of both normal and Navon letters the classifier would be able to learn a suitable set of discriminative features and their intensity spectrum. Note that this last procedure may be considered as simulating the familiarization phase participants had before the experiment. In this case ('baseline+24xNavon' in Figure 2), performance was very good both on normal and Navon stimuli, especially for the case of Distance 3 for the training, being able to generalize to neighboring distances. Both training sets (with Distance 3 and 5), however, still had problems with the sparsest stimuli (Distance 1).

These simulations have shown that Navon letters can be recognized by the HMAX model particularly well when Navon stimuli were also used during training and even better when normal and Navon stimuli were used and better in the cases where perceptual grouping occurs more easily (Distance 5 to 3, due to fusion of the letters in the downscaling of the image, as when looking at a composite object squinting). Such training effects were demonstrated also in simultanagnosic patients (Shalev, Mevorach, & Humphreys, 2007).

So far, the model was tested on the whole Navon stimuli. As said, the image layer in the HMAX model rescales the input at 140x140. For the original Navon stimuli (700x700) this means that local letters (of 16 px side) end up covering a 2x2 px patch. Thus, obviously the system cannot perform the local letter recognition task on the same image resolution. As for humans, an 'attentional spotlight', that is, a focus mechanism is needed to select the location and image resolution of interest. To implement a first resolution selection mechanism, we experimented with a very simple attentional filter of fixed size, which was arbitrarily imposed on the image. The filter is a mask extracting an image patch in three possible sizes: 700x700, 140x140, or 36x36 px – either entailing the entire image (global focus), multiple local letters (medium focus), or just one local letter (local focus). The assumption is that an attention mechanism delivers one of the three scales of processing: the global object, part of it (such as a T-junction), or a sub-element. Figure 3 shows on the left how the three patches look on the original image and on the right how they look to the model, once cropped and rescaled to 140x140 px in the image layer of the HMAX model.

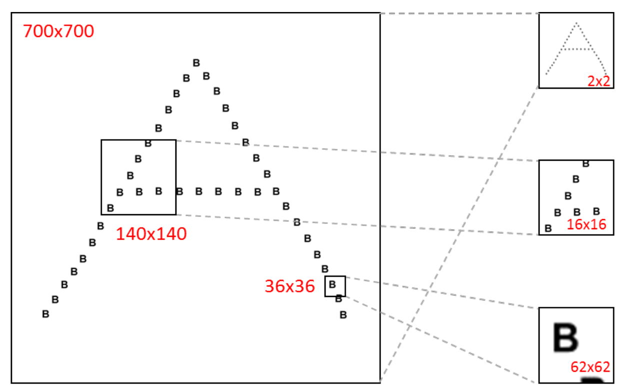

Figure 3: Attention scale selection: The system is given a portion of the image selected by a focus mechanism with three sizes. On the left, the three sizes are shown in the original image size (patch sizes in red). On the right, the patches were rescaled to 140x140 px, which represents the information that is fed into the HMAX model (rescaled size of the local letters in red).

The feature dictionary stayed the same as for the experiments above. The classifier was trained with the baseline training set plus the 24 replications of the best-performing Navon letter for each class (either at Distance 3 or 5). Hence the training set consisted of a total of 270 stimuli (130 normal and 24 Navon x 5 classes). During testing, all 100 original stimuli were used for the 700x700 mask size. For the smaller mask sizes, patches were sampled by scanning all 100 stimuli at 35 pixel steps in the medium focus and at 20 pixel steps in the small focus, retaining the 10 and 20 patches that contained the most non-white pixels, respectively, thus avoiding empty patches. The test set consisted of 3100 stimuli (1+10+20 patches x 100 Navon stimuli). Results for these three foci are shown in Figure 4 along with the results of patients, averaged across the two subjects reported in Huberle and Karnath (2006). Model performance in the local task (small and medium focus) was very good for each of the inter-element distances (above 80%) as for the patients. Performance in the global task (global focus) is again very good on the trained distances and drops dramatically at Distance 2 and Distance 1, which was indeed also the case for the tested patients, even if to a different extent.

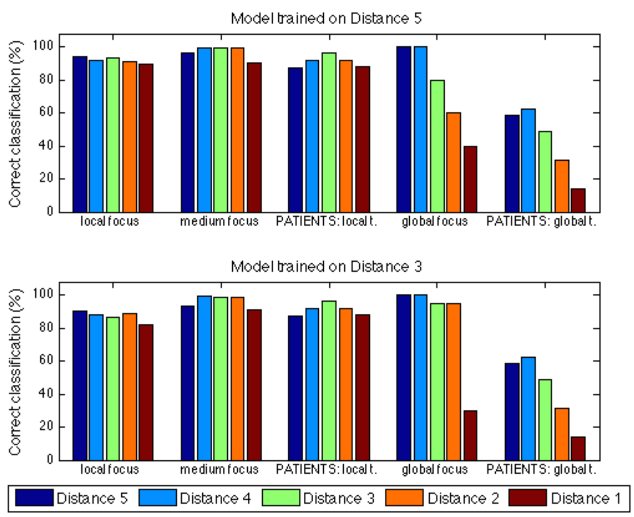

Figure 4: Model and patient performance: color bars indicate performance on Navon stimuli at different distances. The first 2 bar groups on the left simulate selection of local foci of attention of different size. The third group represent the patients performance in the local task. The last two groups show model and patients performance in the global task.

Discussion

The experiments presented in the previous section showed that the HMAX model is capable of flexibly and robustly recognizing letters in normal and Navon configurations. Moreover it is capable of reproducing the patients' performance and training effects in the global task. In the present state, the model is completely bottom up and feedforward. It cannot simulate healthy subjects' classification scores in recovering the global Navon shape, specifically when this does not easily emerge from low-level perceptual grouping. In that case a top-down effort is likely to be required for form completion. Nonetheless, the current model shows that the system needs a focus mechanism to succeed in both the global and the local task. This mechanism determines the scale of processing for the current task. When the whole image is considered, the resolution of the local letters is too low for recognition. On the other hand, when focusing on some local letters or just one letter, the bigger picture is lost. In the following, we discuss which modules and mechanisms should be added for succeeding in modeling the performance of healthy participants and for revealing the actual processes impaired in patients suffering from simultanagnosia.

In the presented experiments the focus size and location was arbitrarily and manually set. In the future a more flexible mechanism based on saliency grouping should be added. By pre-processing the whole image with a bottom-up saliency model that (e.g. Itti, Koch, & Niebur, 1998), which also makes use of a multiresolution representation, the areas or scales containing most saliency can be identified and used to select a sub-image for further recognition processing. For Navon letters at Distances 1 and 2 the saliency map would contain multiple activation peaks, centered on each letter, while for Distance 4 and 5 a single salient area would emerge – simulating grouping by proximity and similarity, particularly fast in humans (Ben-Av & Sagi, 1995). Areas of connected saliency would be segmented and the patch (or proto-object) containing the maximum saliency would be fed to the HMAX model. Alternatively, the saliency mask could directly modulate processing in the S1 and S2 layers (see Walther & Koch, 2006). Higher saliency of the global form with respect to local items may explain the correct classification of denser Navon letters. Further tests should investigate the replication of the results in Huberle and Karnath (2010), where the local/global saliency ratio was manipulated by varying local item size, global item size, or both.

Although enhanced with an attention mechanism, the resulting system would still work in a purely bottom-up fashion. It still would not be able to reproduce the performance of healthy subjects. Grouping effects and figure-background discrimination appear to be accomplished by additional top-down feedback mechanisms to primary visual areas (Roelfsema, 2006). Such top-down modulations can be related to the principle of predictive coding (Rao & Ballard, 1999), according to which endogenous expectations are recurrently compared with the sensory stimulation and differences are propagated up in the hierarchy to adapt internal states and expectations.

Chikkerur, Serre, Tan, and Poggio (2010) extended HMAX with a Bayesian attentional process to model expectations regarding multiple features or location, producing space- and feature-based selectivity. Such an architecture may allow the modeling of the top-down, task-related information expectation at a specific scale, hence top-down biasing the competition between a global and a local focus of attention, even when saliency at the local level is stronger. This would help to assess to what extent focus size selection is due to scale competition and how this competition is disrupted in patients.

Besides static top-down mechanisms, temporal dynamics need to be taken into consideration to account for multiple feed-forward and feed-back sweeps, from selection, through completion, to recognition. Such interactive dynamics seem particularly necessary to recognize sparser Navon letters, as indicated by longer reaction times in healthy subjects when reporting the global letter in those configurations in which patients failed (Dalrymple et al., 2007). A neuro-dynamic model has been indeed shown to achieve shape-completion on challenging stimuli, such as the Kanizsa figures with illusory contours (Weidenbacher & Neumann, 2009).

In conclusion, the present study lays the basis for an experimental framework to model simultanagnosia. First tests succeeded in reproducing results of patients suffering from simultanagnosia on typical stimuli used to investigate factors influencing local versus global capture. A complete model should be able to account also for healthy subjects' results, by entailing the main mechanisms known to be at play in such tasks. This would allow the identification of the functional modules (and responsible areas) that are affected by the lesion. We expect that a neuro-dynamic model, operating on a multiscale representation, will be essential to identify the cognitive aspects impaired by simultanagnosia.

References

Andreopoulos, A., & Tsotsos, J. K. (2013). 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding*, *117*(8), 827 - 891.

Bálint, R. (1909). Seelenlähmung des 'Schauens', optische Ataxie, räumliche Störung der Aufmerksamkeit. *Monatsschrift für Psychiatrie und Neurologie*, *25*, 51–81.

Ben-Av, M. B., & Sagi, D. (1995). Perceptual grouping by similarity and proximity: experimental results can be predicted by intensity autocorrelations. *Vision Res*, *35*(6), 853–866.

Chikkerur, S., Serre, T., Tan, C., & Poggio, T. (2010). What and where: A bayesian inference theory of attention. *Vision Research*, *50*(22), 2233 - 2247.

Coslett, H. B., & Lie, G. (2008). Simultanagnosia: When a rose is not red. *J. Cognitive Neuroscience*, *20*(1), 36–48.

Coslett, H. B., & Saffran, E. (1991). Simultanagnosia: To see but not two see. *Brain*, *114*(4), 1523-1545.

Dalrymple, K. A., Barton, J., & Kingstone, A. (2013). A world unglued: Simultanagnosia as a spatial restriction of attention. *Frontiers in Human Neuroscience*, *7*(145).

Dalrymple, K. A., Kingstone, A., & Barton, J. J. (2007). Seeing trees or seeing forests in simultanagnosia: Attentional capture can be local or global. *Neuropsychologia*, *45*(4), 871 - 875.

Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Ward, R., Kyllingsbaek, S., et al. (2003). Attentional functions in dorsal and ventral simultanagnosia. *Cogn. Neuropsychol.*, *20*, 675–701.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. MIT Press.

Huberle, E., & Karnath, H.-O. (2006). Global shape recognition is modulated by the spatial distance of local elements—evidence from simultanagnosia. *Neuropsychologia*, *44*(6), 905 - 911.

Huberle, E., & Karnath, H.-O. (2010). Saliency modulates global perception in simultanagnosia. *Experimental Brain Research*.

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

Jackson, G., Swainson, R., Mort, D., Husain, M., & Jackson, S. (2009). Attention, competition, and the parietal lobes: insights from balint's syndrome. *Psychological Research*, *73*(2), 263–270.

Karnath, H. O., Ferber, S., Rorden, C., & Driver, J. (2000). The fate of global information in dorsal simultanagnosia. *Neurocase*, *6*, 295–306.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571 - 579.

Luria, A. (1959). Disorders of "simultaneous perception" in a case of bilateral occipitoparietal brain injury. *Brain*, *82*, 437–449.

Mutch, J., & Lowe, D. G. (2006). Multiclass object recognition with sparse, localized features. In *Computer vision and pattern recognition, 2006* (pp. 11–18). IEEE.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, *2*(1), 79–87.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019-1025.

Robertson, L. C. (2003). Binding, spatial attention and perceptual awareness. *Nature Rev. Neuroscience*, *4*(2), 93–102.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual review of neuroscience*, *29*, 203–227.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with Cortex-Like mechanisms. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, *29*(3), 411–426.

Shalev, L., Mevorach, C., & Humphreys, G. W. (2007). Local capture in balint's syndrome: Effects of grouping and item familiarity. *Cognitive Neuropsychology*, *24*(1), 115-127.

Thomas, C., Kveraga, K., Huberle, E., Karnath, H.-O., & Bar, M. (2012). Enabling global processing in simultanagnosia by psychophysical biasing of visual pathways. *Brain*, *135*(5), 1578-1585.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 1395–1407.

Weidenbacher, U., & Neumann, H. (2009). Extraction of Surface-Related features in a recurrent model of V1-V2 interactions. *PLoS ONE*, *4*(6).

Wolpert, I. (1924). Die Simultanagnosie — Störung der Gesamtauffassung. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, *93*(1), 397-415.