

Evaluation of the perceptual grouping parameter in the CTVA model

Carmen Santisteban^{*}, Jesus M^a Alvarado, and Manuel Cortijo
Universidad Complutense de Madrid

The CODE Theory of Visual Attention (CTVA) is a mathematical model explaining the effects of grouping by proximity and distance upon reaction times and accuracy of response with regard to elements in the visual display. The predictions of the theory agree quite acceptably in one and two dimensions (CTVA-2D) with the experimental results (reaction times and accuracy of response). The difference between reaction-times for the compatible and incompatible responses, known as the response-compatibility effect, is also acceptably predicted, except at small distances and high number of distractors. Further results using the same paradigm at even smaller distances have been now obtained, showing greater discrepancies. Then, we have introduced a method to evaluate the strength of sensory evidence (η parameter), which takes grouping by similarity into account and minimizes these discrepancies.

Many of the models used to explain how attentional resources are allotted can be summarised in space-based models and objects-based models. Results supporting either space-based models or objects-based models have been obtained by several authors.

In the space-based attentional models, attention is directed to a limited, continuous region of visual space. Stimuli within such region are allotted the whole of processing resources, whereas those outside the attention focus are completely excluded. Empirical evidence supporting space-based models comes from experiments deriving from the response-competition model (B.A. Eriksen & C.W. Eriksen, 1974; C.W. Eriksen & Hoffman, 1973), designed to study the width of the attentional field, i.e., the extent to which non-target stimuli are processed together with the target stimulus. The results show that

^{*} **Acknowledgements:** This research was supported by the Spanish Ministerio de Educación y Ciencia (MEC), Grant No. DGICYT PB92-0207, and the Ministerio de Trabajo y Asuntos Sociales Ref: RS/MS2001- 16/01. We are grateful to Dr. John Trout for revising our English text. **Correspondence:** Carmen Santisteban. Facultad de Psicología, Departamento de Metodología de las Ciencias del Comportamiento, Universidad Complutense de Madrid, 28223 Campus de Somosaguas, Madrid, Spain. Tel.: +34 91 3943257; Fax: +34 91 3943245; E-mail: c.santisteban@psi.ucm.es

when the display includes a noise letter associated with the opposite response from the target letter the presence of the incompatible noise letter results in an increase in reaction time, but the effects found on reaction times when the noise letter belongs to the same class of response are either small or non-existent. The distance between the target and noise letters is important, the response-competition produced by incompatible noise letters decreasing concomitantly with an increase in target-noise distance (Alvarado, Santalla & Santisteban, 1998; C.W. Eriksen & Hoffman, 1973; C.W. Eriksen & St. James, 1986; Pan & C.W. Eriksen, 1993; Yantis & Johnston, 1990).

Object-based models assume that attention selects regions of space occupied by objects, defined according to the Gestalt principles of perceptual organization (proximity, good continuation and similarity). Supporting object-based models, some authors have shown that such features as colour, shape, spatial orientation and frequency can influence attentional processing. Baylis and Driver (1992) showed that increased response competition for incongruent distractors was determined not only by the distance to the target, but also by such factors as colour. Kramer & Jacobson (1991) observed that when the target and the distractors were very near (0.25°), grouping factors like colour and/or the borders produced important effects in the capability of the subjects to focusing attention.

The Theory of Visual Attention (TVA; Bundesen, 1990) is a mathematical model in which both visual-recognition and element-selection in the visual field depend upon making *perceptual categorisations*. The COntour DETector theory (CODE; van Oeffelen and Vos, 1982, 1983) is based in the perceptual grouping by similarity. The CODE theory of visual attention (CTVA) proposed by Logan (1996) integrates both theories, explaining the effects of grouping by proximity and distance upon reaction time (RT) and accuracy of response with regard to elements in the visual display distributed across 1D space. The CODE model maintains that grouping occurs when two or more nearby elements cause the strength of grouping for a region of the display to exceed a specific threshold value, in which case all of the elements within the same above-threshold region of the CODE *surface* are perceived as belonging to the same perceptual group or object.

We went on to extend this theory to two dimensions (CTVA-2D) (Alvarado, Santalla & Santisteban, 1999). The theoretical predictions of this extended theory compared fairly well with the experimental results. In particular, the experimental RTs decreased with display size for compatible responses and increased for incompatible responses according to theoretical expectation. Nevertheless, at the smallest distance used in this experiment the difference between RT for the compatible and incompatible responses (known as Response Compatibility Effect, RCE) was underpredicted when we used only two noise letters and overpredicted for higher numbers.

We put these discrepancies down qualitatively to the grouping effect by similarity of processing. Along these lines, Bundesen (1990, page 536) pointed out that "by being embedded in a group, an individual suffers a

general loss in eta values, and the stronger the group, the greater the loss"; he did not, however, offer any quantitative evidence for this. We consider that the strength of the sensory evidence that element x belongs to category i , known as the $\eta(x,i)$ or eta value, should in fact take into account the discrepancies observed in our previous study (Alvarado, Santalla and Santisteban, 1999). Hence, we are performing here a further development of the CTVA-2D theory, which allows the evaluation of eta values as a function of two well-known variables of the experiment: the number of letters presented on the display and the distance among them. It appears to improve the agreement between experimental and theoretical values for RCE and for the probability of responding correctly (PRc). We also report our latest comparisons between old and new experimental results and CTVA-2D predictions.

The Model

The CTVA (Logan, 1996) and CTVA-2D (Alvarado, Santalla and Santisteban, 1999) models predict reaction times for a correct response (RT), as well as for the probability of responding correctly (PRc), by using a counter model with a categorization rate $v(x,i)$ defined by the following equation:

$$\begin{aligned} v(x,i) &= C_x \eta(x,i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} \\ &= C_x \eta(x,i) \beta_i \frac{\sum_{j \in R} \eta(x,j) \pi_j C_x}{\sum_{z \in S} \sum_{j \in R} \eta(z,j) \pi_j C_z} \end{aligned} \quad (1)$$

where $\eta(x,i)$ represents the "strength of the sensory evidence that element x belongs to category i " (Bundesen, 1990 p. 524), β_i is "a perceptual decision bias associated with category i " (Bundesen, 1990 p. 524), w_x and w_z are the "attentional weights of elements x and z respectively, as derived from *pertinence* values, π_j , which represent the importance of attending to elements belonging to category j " (Bundesen, 1990 p. 524) and the feature catch C_x determined in CTVA-2D by

$$C_i = \iint_A h_i(x,y) dx dy \quad (2)$$

where the integration region for threshold T is $A = \{(x,y): h_i(x,y) > T\}$ and $h(x,y)$ is the *surface* for a bi-dimensional display of n elements determined by

$$h(x,y) = \sum_{i=1}^n h_i(x,y) = \sum_{i=1}^n \frac{1}{\sigma_{x_i} \sqrt{2}} e^{-\left(2(r_{x_i}^2 + r_{y_i}^2)/\sigma_{x_i}\right)^{1/2}} \quad (3)$$

where μ_x, μ_y and σ_x, σ_y are the means and the standard deviations respectively of the marginal distributions of element i in the display. $r_x = (x - \mu_x)$ and $r_y = (y - \mu_y)$ are the spatial co-ordinates of the element i , assuming equal standard deviation in the directions of the x and y axes ($\sigma_x = \sigma_y$) and zero covariances $\sigma_{xy} = 0$. Feature catch defined by equation (2) can be understood considering that CTVA theory assumes that attention samples the features of the elements falling within the above-threshold region of the $h(x,y)$ surface, with a probability equal to the volume of the distribution of the element that falls within the above-threshold region (*feature catch*)

In CTVA-2D the selection of a particular categorization for a specific element is conceived as a *counter* model, in which K_i categorizations of type i must be made before the subject responds with i . Thus, let us consider a case in which a person has to make a distinction between two letters: "p" and "q". In this case, we assume two counters, one for "p" (K_p) and another for "q" (K_q). There is a previously decided number of counts for each counter and the process finishes when this number is reached in either one counter or the other. In this example prediction about the reaction times for a correct response, RT , (i.e., responding "p" when the target was "p") is:

$$RT = \left\{ \left[\sum_{j=0}^{K_q-1} \binom{K_p + j - 1}{j} \left(\frac{v(x,p)}{v(x,p) + v(x,q)} \right)^{K_p} \left(\frac{v(x,q)}{v(x,p) + v(x,q)} \right)^j \left(\frac{K_p + j}{v(x,p) + v(x,q)} \right) \right] \frac{1}{PRc} \right\} + b \quad (4)$$

where PRc represents the probability of responding correctly, and b is a constant. The probability of responding correctly, PRc , is:

$$PRc = \sum_{j=0}^{K_q-1} \binom{K_p + j - 1}{j} \left(\frac{v(x,p)}{v(x,p) + v(x,q)} \right)^{K_p} \left(\frac{v(x,q)}{v(x,p) + v(x,q)} \right)^j \quad (5)$$

Evaluation of eta values

Alvarado, Santalla and Santisteban (1999) used $\eta(x,i)$ constant values of 1.0 and 0.01 for both compatible and incompatible letters for the calculation of the categorization rate $v(x,i)$. As we have commented above, the measured RCE (calculated by the difference between RT for the compatible and incompatible responses) did not quantitatively agree with the theoretical predictions at small distances, perhaps because we should have taken grouping by similarity into account in these cases, which might mean that the

values of $\eta(x,i)$ may be smaller than 1, the decline being greater both for smaller distances, α , and for larger numbers of noise letters, n (distractors). It means that $\eta = f(n/\alpha)$ with two restrictions: $\eta \rightarrow 1$ when $(n/\alpha) \rightarrow 0$ and $\eta \rightarrow 0$ when $(n/\alpha) \rightarrow \infty$, because the strength of the sensory evidence, η , should be equal to 0 when there are a lot of noise letters or when the two nearest noise letters are very close to each other. In the same way, it should be equal to 1 for a small number of noise letters or when they are far away from each other. One easy equation fulfilling these requirements is the following one:

$$\eta = \exp(-n/\alpha) \quad (6)$$

Therefore we used the results obtained at the smallest T-N distance (target to noise letters distance) described in Alvarado, Santalla y Santisteban (1999) to test this equation and to choose the best units for the distance between the two nearest noise letters, hoping thus to attain values of η that might help to explain our results. It appears that the use of milliradians to measure these distances might explain the previously published discrepancies between our experimental results and the theoretical predictions. The fact that equation (6) could explain our data should be very worthy, due to the large number of natural phenomena that can be explained by a natural logarithm law (consider that equation (6) can also be written in the form $\log \eta = -n/\alpha$). It could then represent that the increase or decline rate of the eta variable might be proportional to the variable itself. Meanwhile we go deeper with the implications of this equation, let us use it as an empirical one to solve the disagreements above mentioned. Several reasons encourage the choice of milliradians. First, distances among letters are easier represented by only one number (the visual angle among them) than by two numbers: the geometric distance among the letters and the geometric distance between the subject and the display. Second, the use of radians (π radians = 180°) is mathematically sounder than other units to represent an angle. We prefer to use milliradians instead of radians for representing our data to avoid the use of decimal numbers. Finally, we would like to point out that the use of equation (6) avoids the introduction of new parameters for improving the agreement between theory and experience.

Bearing all the above in mind, we made further experiments to see whether the above equation, devised to determine the effect of the number of noise letters and the distance between them upon the η values, might avoid discrepancies between the predicted and experimental values for the response-compatibility effect, as well as for the proportion of correct responses, PRC, observed at small distances or with high number of noise letters. We are now using the same methodology, apparatus, stimuli, design and procedure as those described in Alvarado Santalla and Santisteban (1999), except for the details commented upon below, and we applied the CTVA model extended to 2 dimensions (CTVA-2D).

METHOD

Subjects. Eight trained subjects participated as volunteers. All right-handed and had normal or corrected-to-normal vision.

Apparatus and stimuli. A standard personal computer with a colour monitor and SVGA card was used to present the visual stimuli and collect the subjects' responses. Stimulus presentation and response recording were controlled by the DEVAT program (Alvarado and Santisteban, 2000). The subjects performed the task in a sound-proof room with their heads resting on a chin-rest. They were instructed to respond only to the target letter that appeared in the centre of the display and to respond as quickly as possible whilst avoiding errors.

The subjects' task was to identify the target letter that appeared in the centre of the screen surrounded by distractor letters. The subject had to respond by pressing the left cursor (\leftarrow) if the target was "q" or "d" or the right cursor (\rightarrow) if it was "p" or "b". These four target letters appeared at random with equal probability. The noise letters were also presented at random. The target and noise letters appeared simultaneously.

Variables and design. Three independent variables were considered:

1. Compatibility of noise and target letters with two levels: compatible and incompatible conditions.
2. Display sizes with three, five or nine letters, depending upon the number of noise letters displayed (see figure 1).
3. Target-noise (T-N) distances of 0.65° , 0.97° and 1.30° . All distances were defined between the centres of the letters.

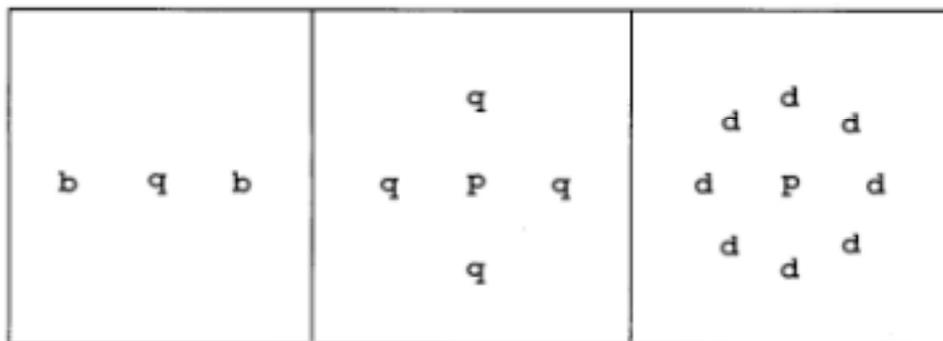


Figure 1. Examples of three display sizes for incompatible conditions.

The dependent variables measured were: RT and accuracy (i.e. correct or incorrect responses, and the mistake of pressing a different key from the two indicated above).

A factorial design of repeated measurements was used. Each subject was presented with the 18 displays obtained by combining the level of the three independent variables (2x3x3) considered. Each display was presented 20 times to each subject in one experimental session. The distances (in milliradians) between the two nearest noise letters, α , for all cases studied can be easily calculated by simple geometrical principles and are set out in Table 1. The corresponding $\eta(x,i)$ values calculated with our proposed equation (6) are given in Table 2.

Table 1. Distances (in milliradians) between the two nearest noise letters presented to the subjects.

	n	2	4	8
T-N				
0.65°		22.7	16.1	8.7
0.97°		33.9	23.9	13.0
1.30°		45.4	32.1	17.4

Table 2. Strength of the sensory evidence that element x belongs to category i (η values).

	n	2	4	8
T-N				
0.65°		0.92	0.78	0.40
0.97°		0.94	0.85	0.54
1.30°		0.96	0.88	0.63

RESULTS AND DISCUSSION

As we have mentioned above, the response-compatibility effects measured experimentally in Alvarado, Santalla and Santisteban (1999) at a target-letter to noise-letter distance of 0.97° did not agree with the values calculated by CTVA-2D. When using the $\eta(x,i)$ values given in Table 2 however, the discrepancies disappear at this T-N distance.

Therefore, new experimental values have been obtained at 0.65°, 0.97° and 1.30°. The means and standard deviations of the RTs and PRc are shown in Table 3. Theoretical calculations were also carried out to obtain predicted values of RTs and PRc: a) using eta values of 1.0 and 0.01 for compatible and incompatible letters, as it was done until now in the literature (called model CTVA-2D); and b) using the eta values given under Table 2 (model CTVA-2Ds). The values obtained are shown in Table 4.

Table 3. Means and standard deviation of the reaction times (RT in milliseconds) and Proportion of Correct Responses (PRc) observed.

Compatibility of noise letters	T-N DISTANCE					
	0.65°		0.97°		1.30°	
	RT (Sd)	PRc	RT (Sd)	PRc	RT (Sd)	PRc
	n= 2					
Compatible	401 (55)	0.93	400 (41)	0.95	398 (40)	0.98
Incompatible	442 (51)	0.94	429 (61)	0.95	421 (45)	0.92
	n= 4					
Compatible	395 (40)	0.98	386 (44)	0.97	386 (44)	0.97
Incompatible	450 (42)	0.88	426 (53)	0.90	417 (35)	0.88
	n= 8					
Compatible	400 (41)	0.95	392 (47)	0.96	390 (40)	0.98
Incompatible	454 (43)	0.89	427 (41)	0.89	424 (47)	0.85

Table 4. Reaction Times (RT) and Probability of Correct Responses (PRc) predicted by CTVA-2D and CTVA-2Ds models for 2, 4 and 8 distractors.

Compatibility of noise letters	CTVA-2D						CTVA-2Ds					
	0.65°		0.97°		1.30°		0.65°		0.97°		1.30°	
	RT	PRc	RT	PRc	RT	PRc	RT	PRc	RT	PRc	RT	PRc
	n=2											
Compatible	397	0.99	392	0.99	389	0.99	398	0.99	393	0.99	389	0.99
Incompatible	449	0.95	416	0.99	401	0.99	448	0.97	415	0.99	401	0.99
	n=4											
Compatible	389	0.99	387	0.99	386	0.99	392	0.99	388	0.99	386	0.99
Incompatible	472	0.81	429	0.98	408	0.99	466	0.93	426	0.99	406	0.99
	n=8											
Compatible	381	0.99	381	0.99	381	0.99	391	0.99	386	0.99	384	0.99
Incompatible	502	0.53	452	0.91	421	0.99	472	0.98	432	0.99	411	0.99

Note: CTVA-2Ds is the CTVA model extended to two dimensions (CTVA-2D) improved with the $\eta(x,i)$ values obtained using equation (6).

The task carried out in our experiments is very easy and a high accuracy in the proportion of correct responses can be expected, as it is indeed the result obtained for all T-N distances and number of letters used in our experiments (Table 3). The predicted PRc values by CTVA-2D model agree with experimental ones, except at the lower T-N distance and higher number of noise letters used for the incompatible condition, which is underpredicted (see Tables 3 and 4). It is noteworthy that the use of equation (6) for the calculation of the $\eta(x,i)$ values brings the predictions on the proportion of correct responses on agreement with our experimental results (Fig. 2a).

A chi-square test has been used to test the goodness of fit between observed data for the PRc incompatible condition and those predicted by CTVA-2D and CTVA-2Ds models. The calculated values are $\chi^2 = 29.62$ and $\chi^2 = 6.88$ respectively. The critical value for $\alpha=0.05$ and $df = 4$ [(3-1) T-N distances x (3-1) display sizes = 4] is $\chi^2 = 9.49$ and therefore the introduction of the calculated eta values in the CTVA-2D model provides an adequate fit to the observed data.

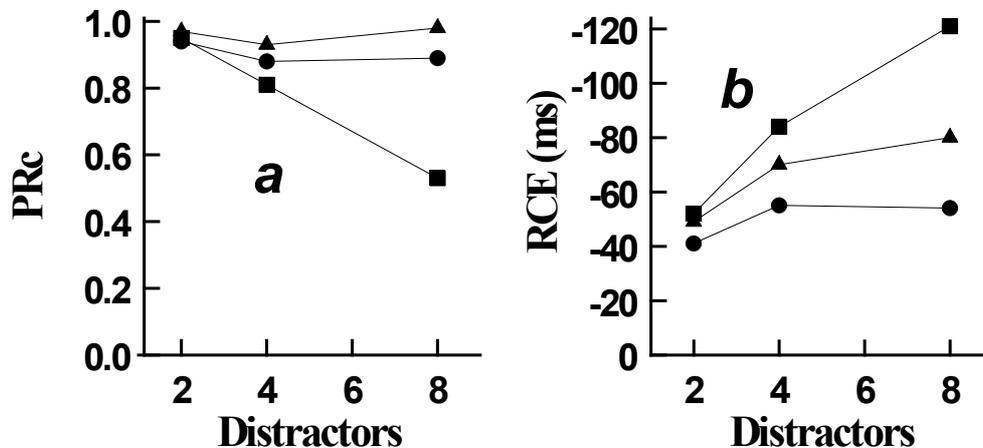


Figure 2. Influence of the number of distractors presented at a T-N distance of 0.65° on: a) the probability or proportion of correct responses, PRc, for the incompatible condition and b) the response-compatibility effect (RCE). Symbols stand for: (●) experimental values; (■) values predicted by the CTVA-2D theory without any correction; (▲) values predicted by the CTVA-2D theory with the $\eta(x,i)$ values calculated using equation (6).

A good agreement between predicted and observed RTs values given in Tables 3 and 4 has been found. Nevertheless, a better test for any model is the comparison between the predicted and observed values for the response-compatibility effect. We have plotted these values for a T-N distance of 0.65° in Figure 2b, where it can be seen that the discrepancies between the theoretical predictions and experimental results are also negligible when only two distractors are present, whatever the η values used (either uncorrected or corrected by equation (6)). With higher numbers of distractors the improvement obtained by applying this correction is also notable. This improvement is clearly reflected in the correlation coefficients calculated between the observed RCE values and those predicted by CTVA-2D and CTVA-2Ds, which were 0.89 and 0.97 respectively.

We should finally like to remark upon the analogy of our equation, which takes into account the correction by grouping of similarity, with that devised by Kubovy and Wagemans (1995), which takes a quantitative account of grouping by proximity in the study of Gestalt phenomena. Kubovy, Holcombe & Wagemans (1998) present later on new experiments which show that the same grouping law explains grouping in different spatial and temporal scales and in more complex patterns. In both cases the attraction function is an exponential of the ratio between the number of dots (noise letters in our case) and the distances between them. This agreement between findings in two fairly unrelated situations strengthens our confidence that the equations used have a firmer foundation than they would have if they answered only to the adjustment of some experimental results.

REFERENCES

- Alvarado, J.M., Santalla, Z. & Santisteban, C. (1998). Efecto de la segregación sobre el procesamiento de la estimulación visual. *Psicologica*, 19, 87-105.
- Alvarado, J.M., Santalla, Z. & Santisteban, C. (1999). An evaluation of the CODE Theory of Visual Attention extended to two dimensions. *Acta Psychologica*, 103, 239-255.
- Alvarado, J.M. & Santisteban, C. (2000). DEVAT: Design of Experiments for Visual Attention Tasks. *Methods of Psychological Research Online*, 5(4), 1-12.
- Baylis, G.C. & Driver, J. (1992). Visual parsing and response competition: The effects of grouping factors. *Perception & Psychophysics*, 51, 145-162
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97, 523-547.
- Eriksen, B.A. & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149.
- Eriksen, C.W. & Hoffman, J.E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, 14, 155-160.
- Eriksen, C.W. & St. James, J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40, 225-240.
- Kramer, A.F. & Jacobson, A., (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception & Psychophysics*, 50, 267-284.
- Kubovy, M. & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, 6, 225-234.

- Kubovy, M., Holcombe, A.O. & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, *35*, 71-98.
- Logan, G.D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603-649.
- Pan, K. & Eriksen, C.W. (1993). Attentional distribution in the visual field during same-different judgments as assessed by response competition. *Perception & Psychophysics*, *53*, 134-144.
- Van Oeffelen, M.P. & Vos, P.G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, *10*, 396-404.
- Van Oeffelen, M.P. & Vos, P.G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition*, *16*, 341-348.
- Yantis, S. & Johnston, J.C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 135-149.

(Manuscript received: 18 February 2004; accepted: 1 September 2004)