

Saliency of road signs in urban areas

R. Brémond, M. Deugnier

Laboratoire Central des Ponts et Chaussées, Paris, France

roland.bremond@lcpc.fr, marion.deugnier@inrets.fr

Abstract: *Traffic control devices are the main information system from the road authorities to the road users. To fulfil this destination, road signs, road markings and traffic lights are designed and set up following governmental regulations, which are stated in terms of location with respect to the roadway, size, colour, contrast, etc.*

These traffic control devices are designed to be seen, but they are embedded among other items of lesser relevance, which may draw away the attention of the drivers. The message born by the signs must be easily captured, despite its complex environment, the visual concurrence between the different visual tasks of driving, and visual distractors in the vision field.

In this paper, we investigate how traffic control devices are dealt with by road users as they collect visual information in an urban environment, in daytime. A laboratory experiment was conducted in order to characterize road signs conspicuity for an urban driving task. We focus on attention saliency, that is to say the ability of items in the vision field to draw the attention of a driver who is not searching it specifically. Hence, the experiment concerns mandatory signs more than information or direction signs.

40 observers were presented with 20 daytime pictures of an urban environment during 100, 300, 600 ms. and 10 s. with general instruction that they are in a driver situation, and the specific instruction to say what they have seen. The statistical analysis of the recorded verbal data shows that "Traffic control devices" is the only semantic category to be more reported as presentation time is longer.

We hope this study may give experimental reference in order to improve saliency algorithms, in a way which would take into account the driving task. It may lead, then, to automatic road diagnostic tools for traffic control devices.

Introduction

Traffic control devices are the main information system from the road authorities to the road users. To fulfil this destination, road signs, road markings and traffic lights must be designed and set up following governmental regulations, which are stated in terms of location with respect to the roadway, size, colour, contrast, etc. Regulations are mostly based on recommendations, such as those published by the Division 4 of the CIE (International Commission on Illumination), which deals with lighting and signalling for transport.

Road signs, as other traffic control devices, are designed to be seen and recognized. One difficulty stands in the fact that they are embedded among other items of lesser relevance, which may draw away the attention of the driver. The message born by the signs must be easily captured by the driver, despite this complex environment and the visual concurrence between the different visual tasks of driving, and between several traffic control devices and visual distractors in the vision field.

In this paper, we investigate how traffic control devices are dealt with by road users as they collect visual information in an urban environment, in daytime. This issue is considered through the visual concurrence between a number of semantic classes (from the driver's point of view), in an experimental task that we want to be close to the driving task. We implemented and validated an experimental protocol which may also be applied to other driving situations, in order to gather robust data in a simple and reproducible way.

An experiment was conducted in order to characterize road signs conspicuity for an urban driving task. We focused on attention saliency, that is to say the ability of items in the vision field to draw the attention

of a driver who is not searching it specifically. Hence, the experiment concerns mandatory signs more than information or direction signs.

Several observers were presented with daytime pictures of an urban environment. The statistical analysis of the recorded verbal data allowed us to understand some aspects of the investigated problem.

State of the art

Road visibility is mainly defined through performance indexes (like visibility level), based on detection tasks [Blackwell 1946 ; Adrian 1989 ; CIE 1981]. Such models are built on experimental data on a large number of subjects, but with no ecological concern. On the other hand, a number of studies do consider the visual task in a driving situation context [Hartmann 1970], or the visual conspicuity of road signs [CIE 2000]. In this paper, we focus on attention saliency rather than research saliency, because it is part of the visual process involving mandatory road signs.

Visual saliency is a component of data processing in the human visual system (HVS), which should be considered in the driving task [Rumar 1979 ; Hills 1980]. Unfortunately, various definitions are available [Engel 1971 ; Cole & Jenkins 1980 ; Hugues & Cole 1988 ; Theeuwes 1991]. Oculometric data was investigated [Cole & Hugues 1990], as scene complexity [Jenkins & Cole 1982 ; Boersema & al. 1984], photometry and colorimetry of the objects [Forbes & al. 1986 ; Hugues & Cole 1986], characteristics of the subjects [Johansson & al. 1970 ; Cole & Hugues 1984]. In this paper, we use verbal data in the experimental protocol.

Vision science and neurosciences have investigated visual saliency mainly with synthesis images, without semantics [Northdurft 1993, 1995, 2000]. Both bottom-up and top-down processes have been investigated, as well as the impact of visual attention [Motter & al. 2000 ; Rao & al. 2002].

Attention models [Niebur & Koch 1998 ; Itti & Koch 1998 ; Treisman & Gelade 1980 ; Navalpakkam & Itti 2005] lead to the concept of saliency map, on the basis on a bottom-up processing of visual information. The central hypothesis is a unique retinotopic map taking into account all the patterns of visual saliency. This model is not based on physiological data, but some works suggest that such physiological ground may be possible [Gottlieb & al 1998].

Material and methods

For this exploratory study, we choose to use experimental psychology methods. The complexity of the driving task and of the visual stimuli in natural driving leads us towards a laboratory experiment, in order to control most parameters and to aggregate individual behaviours through a typology of stimuli and responses.



Fig 1: examples of road scenes (vues of Paris).

We focused on attention saliency, that is to say the ability of items in the vision field to draw the attention of a driver who is not searching it specifically. This allows to use an experimental paradigm based on short

road images presentations: we try to understand what kind of road items are seen during a driving task without visual search.

The experiment is based on short time photographic images presentations (100, 300 and 600 ms.), balanced with a reference, which consists in showing the same images during 10 s. This allows to compare what the subject see with no or very short visual search, and with a quite long search. The scenes come from a database of urban road pictures of different urban roads in France (Rouen, Paris and La Guadeloupe). 40 pictures were selected (see examples Fig 1).

The experiment was performed at the LCPC Paris in a dark room, with a psychophysical software for image presentations and a video-projector. The subjects seat 2 meters from a 2 meters width screen, so that vision field is 50° (see Fig 2).

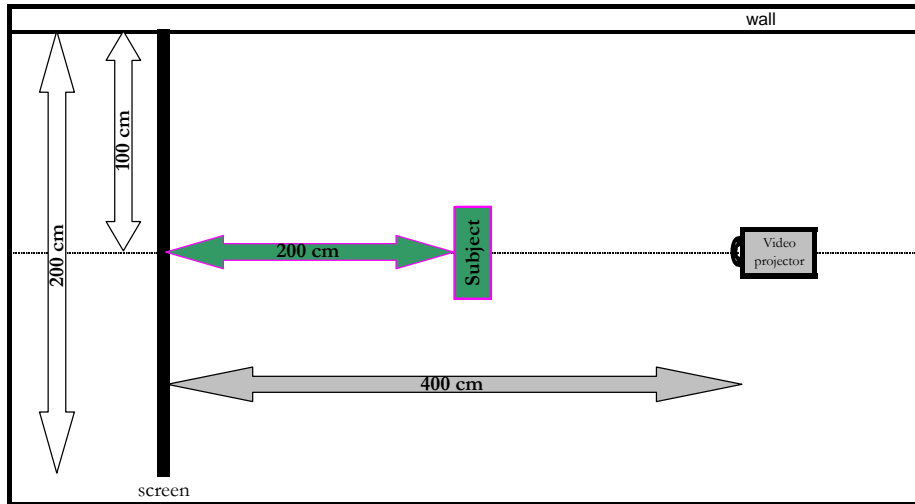


Fig 2 : plan of the experimentation room [width of the visual field about 50°].

44 subjects (27 men and 17 women) participated in this experiment, all having a driving license. 6 are between 20 and 25 year's old, 17 between 26 and 35, 11 between 36 and 45, and 10 over 46. They drive between 0 and 40 000 kilometers a year. 13 subjects have under 10 years of driving experience, 16 between 11 and 20 years, 8 between 21 and 30 and 7 over 30 years.

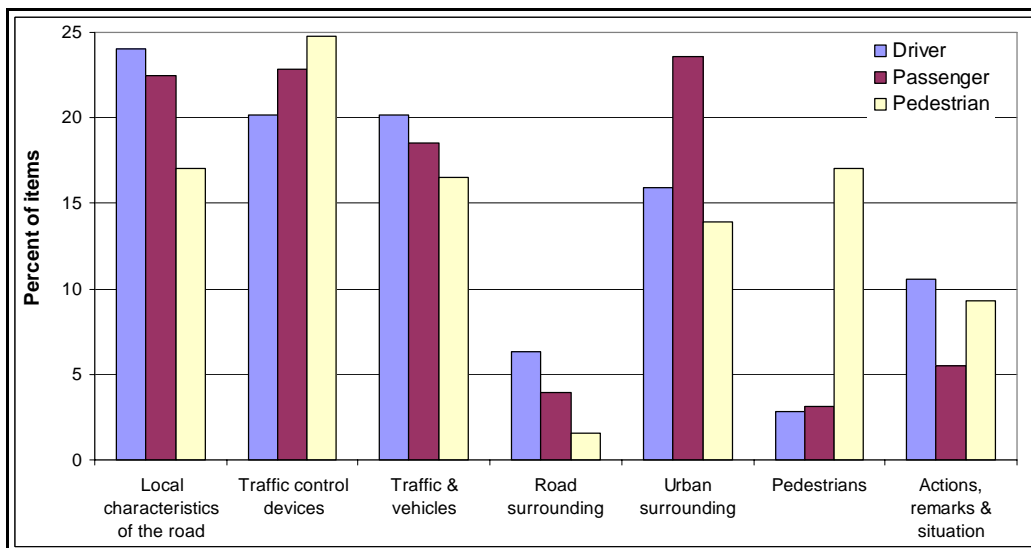


Fig 3: validation of the semantic categories according to the position of the subject.

The subjects are told they are “in a driver situation”, and are asked to tell “what they have seen in the road scenes”, after the presentation of the pictures (short presentation times) or during the presentation (long presentation time). the ecology of the experimental configuration was partially validated in a pre-study (see

Fig. 3) which showed a significant difference between the subjects answers depending on whether they are told to be in a “driver”, “passenger” or “pedestrian” situation.

The image database is splited in two sub-databases and the subjects in two groups. The sub-databases are made of photographic images which are considered closed, which is confirmed by the fact that the differences between the two groups are not significant. This result means that our results aren't specifically linked to the detail of the stimuli's choice.

The images are displayed in four sessions of growing presentation time (100, 300, 600 and 10 000 ms.). A distractive animation appears between each session. Each subject sees 88 stimuli: 20 images presented 4 times, and 8 control stimuli (images displayed 2 times). In each package, images are presented randomly.

The verbal data is recorded with a microphone and a mini-disc recorder. The verbalizations are classified in seven semantic categories, which were confirmed in a pre-study where subjects had to quote the items in a road scene after a 6.4 sec. presentation time. The semantic analysis of the verbal data showed significant difference between the seven categories (*Fig 3*). These differences seem only due to the preliminary indication that the subjects were “in situation of” driver, passenger or pedestrian.

The semantic categories used in this study are: Local characteristics of the road¹, Traffic control devices², Traffic & vehicles³, Road surrounding⁴, Urban surrounding⁵, Pedestrians⁶, Action, remarks & situation⁷.

Results

The distribution of the semantic categories according to the time of presentation of the images (100, 300, 600 ms. and 10 s.) shows that the category "Traffic control devices" becomes more important with the presentation time (a negative result with the student test means a significant increase) (*Fig. 4 and Tab 1*).

No other semantic category shows such an increase with time presentation, and some of them decreases, namely "Road surrounding", "Urban surrounding", "Road characteristics" and "Traffic & vehicles" (*Tab. 1*).

¹ number of ways, size of the street (large, narrow...), if it's one-way, if there is a roundabout, a crossroads, a junction, a curve, a grade separation, a straight way, a slope, secondary roads, a bridge, a median strip, a traffic island, parking spaces, road works..., and the state of the road surface.

² temporary and permanent roadsigns, signposts, markings (except pedestrian crossing), traffic lights and signals.

³ number of vehicles, colour, vehicles which travel or which are parked, trucks, bus, motorcycles, bikes, traffic conditions (dense, moving freely...), and all verbalizations of the kind "road is congested", "road is clear", "there is nobody".

⁴ where the subject is: in a street, road, highway, motorway, urban motorway, out-of-town way, urban way, an entrance or an exit of way, town..., in a town, in the countryside, in an urban area, a residential area, a shopping area...

⁵ If there are trees, open spaces, post, streetlights, bins, buildings, houses, dwellings, publicity, shop sign, store, restaurant, hotel, service station, colour of the sky, weather, and perception of the environment : "it's dark", "it's brightness".

⁶ presence or not of pedestrians, pedestrian crossings, bus shelter, pavement.

⁷ examples of remarks : "I didn't see the colour of the traffic light", "I don't know what contains the road sign"...; examples of actions : "I can't move", "I must stop here"...; examples of feelings : "somebody overtakes me", "no danger", "I must pay attention", "good visibility"...; examples of position in the scene : "I'm on the right side of the way"; what the others do : "a car is starting", "a pedestrian intends to cross the road", "It's not time to cross the road".

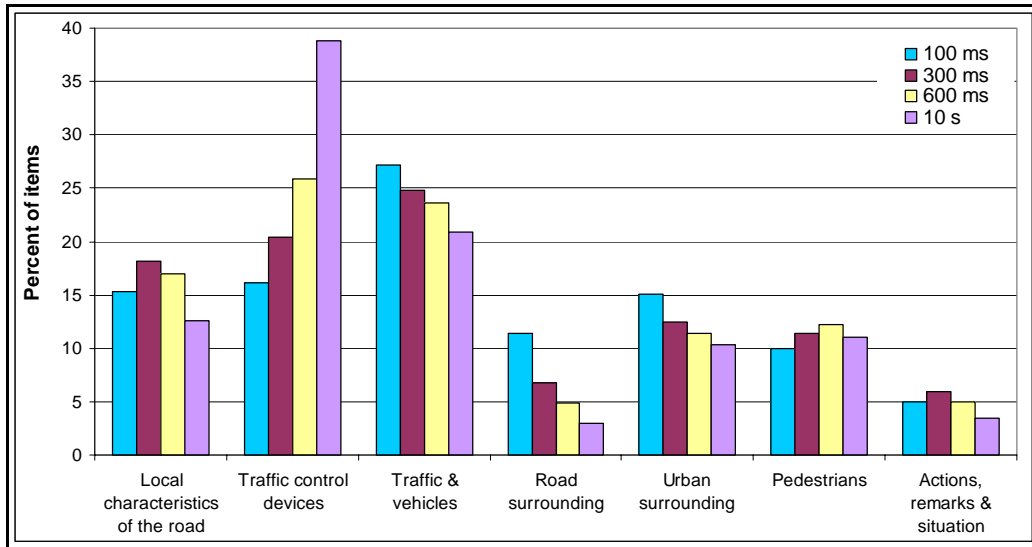


Fig 4: distribution of use of the semantic categories according to the presentation time.

<i>Semantic category</i>	<i>100 vs. 300 ms.</i>	<i>300 vs. 600 ms.</i>	<i>100 vs. 600 ms.</i>	<i>short vs. long</i>
Traffic control devices	-2.36, p .02	-2.08, p .05	-4.48, p .001	-6.19, p .001
Road surrounding	2,27, p.05	NS	3,62, p.001	3,39, p.001
Urban surrounding	NS	NS	2,27, p.05	NS
Road characteristics	NS	NS	NS	2,17, p.05
Traffic & vehicles	NS	NS	NS	2,10, p.05

Tab 1: $t(\infty)$ for various semantic categories and presentation times.

None of the statistics tests shows a correlation between to use of the category "Traffic control devices" and the age or gender of the subjects. Correlations are found with driving experience: the proportion of the "Traffic control devices" among the quoted elements decreases (at 600 ms., $t(\infty) = -2.10$, p.05, and at 10 s., $t(\infty) = -1.96$, p.05) and the proportion of "Road surrounding" increases (at 600 ms., $t(\infty) = 2.78$, p. 01) with the experience of the driver.

Discussion

This study leads to a number of biases with respect to the situation we are interested in, daytime driving in an urban area. As we are interested in the pre-attentive behavior, we give a neutral instruction ("What have you seen?"), short time presentation and use verbalization rather than non verbal data (e.g. car commands in a driving simulator).

A neutral instruction was shown to be useful in order to avoid expectancy biases: in the debriefing of a preliminary study, we found that with more accurate questions (what is the color of the car, etc.), people seemed to build expectations on the basis of the former question, and then diverge from the proper driving behavior, and to the ecology of the experimental situation.

The verbalization of what the subject sees is used as the most salient items for him in his actual visual task, which we have tried to make close to a driving visual task. Thus, we cannot consider the aspects of the scene which are taken into account in the driving task without being consciously perceived. The choice of short presentation times (between 100 and 600 ms.) is due to the hypothesis that we have access to the items stored in the sensorial memory before most of the cognitive screening. This is close to the pre-attentive saliency that we are looking for. The long time observation (10 s.) is aimed to make a reference, to which short time presentation can be compared. We ask to verbalize during the presentation in order

not to overestimate the last perceived objects (it would happen if we had asked to verbalize after the presentation).

We want to know what the subjects look at, and if possible, in which order. From this point of view, the presentation of static images rather than videos makes the comparisons easier. We observe (Tab. 1) that people pay (in proportion) more attention to “Traffic control devices” when they have more time to look at the pictures.

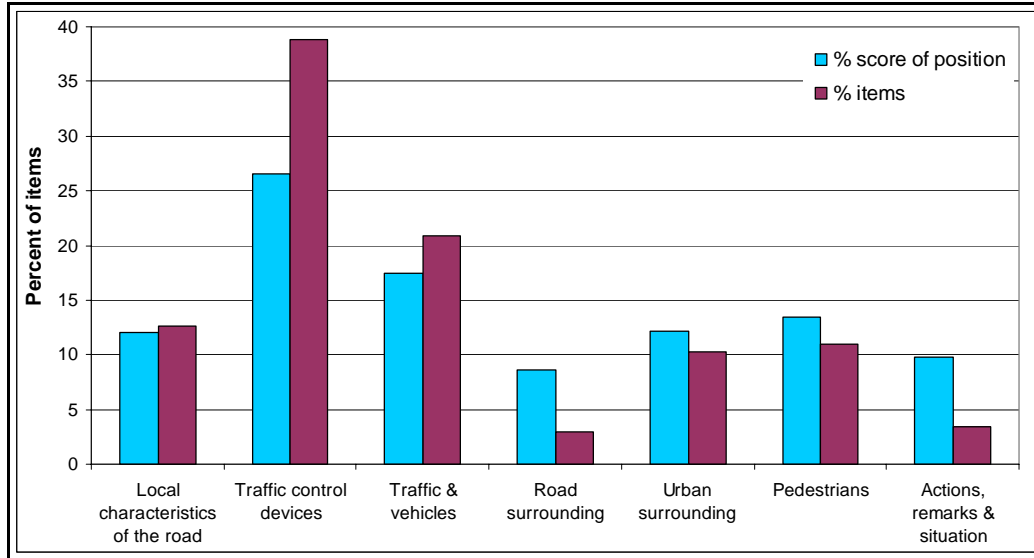


Fig 5: weighted distribution of the semantic categories (visualization of 10 s.) compared with unweighted distribution.

This may lead to the hypothesis that these items are seen as pertinent data, and thus are actively searched (these items are not salient enough, so that people feel the need of searching them). Another hypothesis may be that these items are not in the first range of concern, so that they are not searched at first.

These two hypotheses can be discussed with the help of Fig. 5 and 6, where we take into account the response order and not only the number of response for every semantic category. Fig. 5 shows a weighted distribution of the semantic categories, where an item given in the first place is weighted 1, an item given in second is weighted 1/2, and so on. Only two categories show a significant difference with p.01: $t(\infty) = 2,34$ for “Traffic control devices”, and $t(\infty) = -3,01$ for “Action, remarks & situation”. Fig. 6 shows the proportion of items of each category in the four first places.

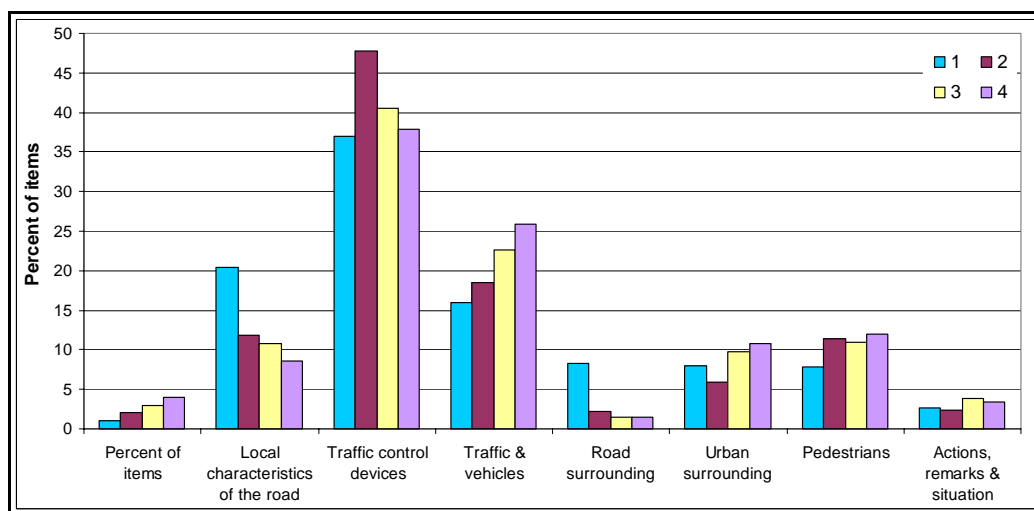


Fig 6: distribution of the semantic categories for verbalization order 1, 2, 3 & 4 (visualization of 10 s.).

Perspectives

This study gives a simple framework for the study of pre-attentive vision while driving, and some results on the importance of traffic control devices. Anyway, a more complete study would use a better classification of urban road pictures, which could allow to make separate studies on different kind of urban environments.

The use of more subjects would allow a more accurate semantic classification, especially inside the “Traffic control devices” category, which could be split into sub-categories (traffic lanes, road signs, etc.).

Such experimental studies can also be used in order to assess the quality of saliency algorithms. Itti, for instance [Itti & Koch 1998], suggests a bottom-up model of visual saliency, based on psycho-physiological data in the first steps of the visual data processing, in order to predict the saliency of any area in an image (that is to say a visual stimulus). This model mainly considers the pre-attentive vision, as we do, but do not consider any task-depending semantic bias (like a Bayesian approach would do).

Fig. 7 shows the most salient areas according to [Itti & Koch 1998] on a specific urban scene. This kind of computation could be classified with the semantic we used and compared with our results. One cannot expect that such a comparison gives the same results with experimental data and with the algorithm, because of the algorithm biases. On the contrary, it may be a useful tool in order to quantify these biases, and to suggest a new and more accurate algorithm for road saliency computation. Such an algorithm could then be applied to road images taken from onboard camera, and be useful for road diagnostic.



Fig 7: most salient areas according to [Itti & al. 2001] on an urban scene.

Another issue for us would be to get closer to the ecological situation of driving, first of all by the use of short videos in order to include dynamic aspects of the vision, even if such an approach raises a number of technical problems in the data interpretation.

Acknowledgments

Thanks to P. Cabon (LAA, Paris V university) for his helpful advice, and to JP Tarel (LCPC) for the computation with Itti & Koch saliency model.

References

- Adrian, W. (1989). Visibility of targets: model for calculation. *Lighting Research Technology*, 21 (4), 181-188.
- Blackwell, H.R. (1946). Contrast threshold of the human eye. *Journal of the Optical Society of America*, 36 (11), 624-643.
- Boersema, T., & Zwaga, H.J. (1984). The influence of advertising signs of the conspicuity of routing information. *Proceeding of the Ergonomic Society Conference*, 2-5th April 1984, sup. n°1 in *Ergonomics*, 27.
- Cole, B.L., & Hughes, P.K. (1984). A field trial of attention and search conspicuity. *Human Factors*, 26 (3), 299-313.
- Cole, B.L., & Hughes, P.K. (1990). Drivers don't search: they just notice. In D., Brogan (Ed.), *Visual Search*. Taylor & Francis, 407-417.
- Cole, B.L., & Jenkins, S.E. (1980). The effect of variability of background elements on the conspicuity of objects. *Vision Research*, 24 (3), 261-270.
- CIE 19.2 (1981). An analytic model for describing the influence of lighting parameters upon visual performance, 2nd ed. *Commission Internationale de l'Eclairage*.
- CIE 137 (2000). The conspicuity of traffic signs in complex backgrounds. *Commission Internationale de l'Eclairage*.
- Engel, F.L. (1971). Visual conspicuity, directed attention and retinal locus. *Vision Research*, 11 (6), 563-576.
- Forbes, T.W., Pain, R.F., Joyce, R.P., & Fry, J.P. (1986). Color and brightness factors in simulated and full scale traffic sign visibility. *Highway Research Record*, 216, 55-65.
- Gottlieb, J.P., Kusunoki, M., & Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391 (6666), 481-484.
- Hartmann, E. (1970). Driver vision requirements. *International Automobile Safety Conference*, New-York SAE.
- Hills, B.L. (1980). Vision, visibility and perception in driving. *Perception*, 9, 183-216.
- Hughes, P.K., & Cole, B.L. (1986a). What attracts attention when driving? *Ergonomics*, 29, 377-391.
- Hughes, P.K., & Cole, B.L. (1988). The effect of attentional demand on eye movement behaviour when driving. In Gale, A.G. & al. (Eds.), *Vision in vehicles II*. Amsterdam: Elsevier Science Publishers B.V.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Review Neuroscience*, 2 (3), 194-203.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Intelligence*, 20 (11), 1254-1259.
- Jenkins, S.E., & Cole, B.L. (1982). The effect of density of background elements on the conspicuity of objects. *Vision Research*, 22 (10), 1241-1252.
- Johansson, G., & Backlund, ?. (1970). Drivers and road signs. *Ergonomics*, 13, 749-759.
- Motter, B.C., & Holsapple, J.W. (2000). Cortical image density determines the probability of target discovery during active search. *Vision Research*, 40 (10-12), 1311-1322.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45 (2), 205-231.
- Niebur, E., & Koch, C. (1998). Computational architecture for attention. R., Parasuraman (Ed.), *The Attentive Brain*. MA : MIT Press, 163-186.
- Nothdurft, H.C. (1993). The role of features in pre-attentive vision: comparison of orientation, motion, and color cues. *Vision Research*, 33 (14), 1937-1958.
- Nothdurft, H.C. (1995). Generalized feature contrast in preattentive vision. *Perception*, 24, S22.
- Nothdurft, H.C. (2000). Saliency from feature contrast: additivity across dimensions. *Vision Research*, 40 (10-12), 1183-1201.
- Rao, R.P.N., Zelinsky, G.J., Hayhoe, M.M., & Ballard, D.H. (2002). Eye movements in iconic visual search. *Vision Research*, 42 (11), 1447-1463.
- Rumar, K. (1979). The visual environment in road safety. In *Proceeding of the 19th Session of the CIE*. CIE 50.
- Theeuwes, J. (1991). Visual selection: exogenous and endogenous control. In A.G., Gale & al (Eds.), *Vision in vehicles III*. Amsterdam: Elsevier Science Publishers B.V.
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12 (1), 97-136.