Design Issues in Gaze Guidance

CHRISTOPH RASCHE*
Laboratorul de Analiza si Prelucrarea Imaginilor
Bucharest Politechnica University, Corp B
B-dul Iuliu Maniu 1-3, Sector 6, B142
Bucuresti 061071, Romania

and
KARL GEGENFURTNER
Abteilung Allgemeine Psychologie, Justus-Liebig-Universität
Otto-Behaghel-Str. 10F, 35394 Giessen, Germany

*Corresponding author: rasche15@gmail.com

The idea of gaze guidance is to lead a viewer's gaze through a visual display in order to facilitate his/her search for specific information. This study elaborates on the process of guiding gaze from one spatial position to another, whereby the goal is to create a guidance process that is as least-obtrusive as possible. A number of guidance aspects is listed and to explore some of those aspects, an experiment is carried out in which observers perform a difficult letter identification task in dynamic noise. To facilitate this recognition task, the viewer is guided by a luminance 'marker'. It is investigated how the observer's visual system reacts to the markers and how the marker's spatio-temporal properties influence the recognition performance. From those results we derive a number of design issues for the process of gaze guidance.

Categories and Subject Descriptors:	Design, Experimentation
General Terms:	H.1.2 User/Machine Systems, H.5.2 User Interfaces
Additional Key Words and Phrases:	Eye-Tracking, Gaze Guidance

---

1. INTRODUCTION

The aim of gaze guidance is to support the viewer during visual inspection of his/her environment by giving suggestions of where to look (Barth et al, 2006a, 2006b). Gaze guidance is potentially applicable in situations where the viewer is confronted with a large visual display (or visual field), which needs to be searched for specific information, e.g. while driving a car, when working at a monitor or when analyzing medical images (McNamara et al 2009; Kim and Varshney, 2008). The (human) viewer itself is undoubtedly the most efficient searcher of visual information, yet a viewer can browse detailed visual information only serially; the viewer may tire; or the viewer may be a novice and lack the experience to find specific information in his/her environment. The aim is therefore to point out potentially interesting spots by means of some visual *marker*, which in turn would draw the gaze toward that position. Thus, there are two parts to a gaze-guiding system. The first one is the computation of visually interesting spots by means of algorithms mimicking human vision or by means of a previously collected set of salient locations obtained from other human viewers (Barth et al, 2006b); the second part is the process of leading gaze through this set of locations in such a way, that the viewer feels least irritated or disrupted by the process. This study is concerned with the second part and to understand the intricacies of this process we start by looking at rudimentary forms of gaze guidance, taking personal-computer displays as an example. Word editors use blinking cursors to signal their present position; operating systems employ blinking icons to signal incoming email, security updates or entry dialogs, which have appeared behind other panels; and banner advertisement on web pages uses blinking or moving objects, or also pop-up windows, to attract a viewer's gaze. Each one of these markers has its advantages and disadvantages. The blinking cursor is effective as long as we stay near it, for instance within the editor. But once we leave the editor window and switch to another window, the memory for the cursor position fades away with increasing duration. A return to the previous cursor position therefore results sometimes in a search. The icons of security updates are sometimes not noticed, because their appearance has become too familiar to us. Finally, advertisement markers can be very irritating (Burke et al 2005).

These examples already hint that there is a range of aspects associated with optimal gaze guidance and that a central issue will be the adjustment of the marker saliency. This quest for a least-obtrusive gaze-guidance system bears similarities to the notion of designing awareness cues, that provide the appropriate attentional 'draw'. For example, in Stanton et al. many guidelines for alarm design are presented – exploiting auditory and visual perceptual processes (Stanton, 1994). In particular, urgency of the information to be presented, has been shown to be directly related to the perception of how distracted a signal may be (Gluck et al., 2007; McCrickard et al., 2001; Obermayer and Nugent, 2000). Gluck et al. discuss in detail how the urgency and utility of information can be tied to the saliency of the presentation of information, and explore a range of possible visual features of signals so as to balance between saliency and disruption (Gluck et al., 2007). McCrickard et al. (2001) explored how animated displays such as tickers and faders can be used to convey information without causing distraction to the ongoing task. The following list of aspects therefore overlaps with many awareness-attracting aspects discussed in the above mentioned literature, but has a bias towards discussing aspects of continuous guidance and eye-movement properties, specifically gaze orienting.

1.1 A List of Aspects

There are many parameters that could influence the guidance process, the parameters of the markers and the parameters of the human visual system reacting to the markers. We have clustered the parameters into a list of aspects. The list is not confined to a particular system, but intends to address the topic in a broad sense.

1. Response Urgency: One may distinguish between different degrees of urgencies to lead the viewer to a conspicuous spot. A high-urgency scenario could be, if a car driver is to be notified about a potentially dangerous situation: then, the marker should act as an alert signal triggering immediate reaction. In this case the marker should be obvious, for instance a large bright marker, combined with an auditory signal to ensure rapid reaction. A low-urgency scenario would be, if an observer browses a large visual display, in which there exist salient spots: in such a case, the marker should act as a suggestion, but does not necessarily require an immediate response. Such a marker should be subtle, ideally subliminal, otherwise it may become irritating and loose its attractiveness and hence its purpose.

2. Marker Frequency: A marker can appear with different frequencies. On the one side, the frequent appearance of a marker may be potentially irritating or tiring leading to its ignorance. The marker frequency can not exceed 3 Hz, because this is the approximate eye movement frequency (3-4 times/sec). On the other side, the occasional appearance of a marker may suffer from potential negligence, thus requiring a stronger saliency. If the frequency varies over time, the saliency of the marker may need to vary accordingly.

3. Marker Occurrence: A marker may occur *sequentially*, meaning only one at a time, or there may be several markers appearing *simultaneously*. In case of the latter, the choice of when to look at which marker may not matter and it would be left to the observer to plan a serial scanning of the spots. If some markers are of higher priority than others, then a serial guidance would be deployed.

If gaze is to be guided at a fast pace, e.g. one or two times a second, the precise marker timing may also be a crucial issue. During some time period before the actual saccade is triggered, ca. 100ms before saccadic onset, visual information does not influence the orienting process anymore (e.g. Nazir et Jacobs 1991, Caspi et al 2004). Thus, the occurrence of a marker during that time has little effect and may not contribute to a fast-paced and smooth guiding process.

4. Marker Range: This aspect addresses the display size and the peripheral decline in visual acuity. Acuity declines with increasing eccentricity from the center of gaze, that is, a signal in the periphery is less detectable than one near the focus. For a grating discrimination task the detectability drops as follows: at 5 degrees eccentricity, which is the perimeter of the parafovea, it has dropped to 32 percent; at 20 degree eccentricity, which is the perimeter of the eye field, the detectability is at 10 percent (Findlay, Gilchrist 2003, p. 15). Hence, in order to render distal markers equally noticeable as close ones, the markers have to be scaled up in size with increasing eccentricity, an issue now called *eccentricity-dependent saliency*. For instance, for a 'grating marker' at 5 degree eccentricity the marker size had to be scaled up by a factor of 3. The decline in grating acuity can be described by an exponential decay, but there exists no general formulation for arbitrary visual structure.

During viewing, the typical saccadic jump distance (= amplitude) reaches up to about 20 degrees, rarely up to 30 (Land et al, 1999; Einhäuser et al 2007). If the display size is limited to this magnitude, it will be browsed to a large extent by eye-movements and to a smaller extent by head movements. For larger display sizes, the proportion of head movements will increase. For markers, which are farther away than 20 degrees of eccentricity, it may require a cueing signals to alert the viewer (aspect 'cue signal').

5. Marker Location: A marker may be *stationary*, e.g. a marker placed on the side mirror of a car, or it may appear at any (*unpredictable*) position in the display. For the former we would expect a viewer to remember its location and make more precise eye movements towards it than in case of the latter. In case of the latter, landing precision may be an issue, depending on the degree of structural detail at the marker's location.

Another potential necessity may be to place the marker slightly beyond its target (with reference to the present gaze position) to account for saccadic undershoot.

6. Attention: An observer may be engaged in another (guidance-independent) action, which is so attention-consuming, that any marker signal may fail to attract the viewer's gaze. In case of a low-urgency situation, this may not matter; in case of a high-urgency situation it may be crucial that the marker appears very salient – possibly coupled with an auditory signal to disengage the viewer from the distracting action. Consequently, the saliency of a marker must correspond to an observer's attentive state, which in turn needs to be tracked. The need to continuously sense the viewers attention has already been suggested by others in studies of human-computer dialogue interfaces (e.g. Qvarfordt and Zhai, 2005, Salvucci et al 2000).

7. Cue Signal: As mentioned repeatedly it may be useful to provide an alert signal for the marker in some situations, that is, a cue signal preceding the actual marker (see aspect 'marker range'. Such cues could be of auditory or visual nature and serve to announce the upcoming appearance or presence of a marker. An example of a visual cue could be for instance a little arrow pointing toward the location of the marker, a cue similar as in Posner's attention experiments (Posner et al 1980). If one knew the viewer's momentary gaze position and if one had control over the visual display, then such a cue could be placed near the viewer's focus to be most effective. This is elaborated in the next 2 points.

8. Eye-tracking: An optimal gaze-guidance system is equipped with an eye-tracker which knows the viewer's gaze position at any given point in time (as already implied in the above discussed points). Such tracking does not need to be overly accurate: recent eye-trackers, geared toward desktop use, may well suffice to operate such a gaze-guiding process. For instance, the eye-tracking solution suggested by Li et al, (2006), provides an accuracy of 1 degree (and costs only 350 dollars), which is sufficient to make use of the idea of eccentricity-dependent saliency (see aspect ‘marker range’) and to sense when the area near a marker has been foveated.

9. Marker Appearance: If the visual field consists of a display of which the guidance system possesses control over each pixel, then there exists the option to place a marker in a *context-dependent* fashion: For instance, a luminance marker can be set by subtly increasing the luminance values at a given salient position. This pixel modulation enables to display markers, which are sufficiently conspicuous but not necessarily irritating. The latter may occur if a fixed-saliency marker is placed into a context from which it pops-out in an irritating manner. The marker's shape may have an influence on guidance; the marker size and location are discussed under aspects no. 4 and 5.

10. Learning: A gaze-guiding system requires time to get acquainted to: the user needs time to learn to respond to the markers without feeling disrupted in his/her regular search behavior. Although it is often stated that human-computer interfaces should require little learning (Jacob 1993), this may not be achievable when learning to interact with a subtle process as pursued here. For the learning process, it may be beneficial to increase the overall level of the marker’s saliency to make the novice aware of the guiding process.

11. Search complexity: A search task may vary in its degree of recognition complexity. For instance, counting the number of occurrences of a visual structure involves merely its detection as it is the case in a study by McNamara (2009). But identifying a structure, as we investigated with a letter-identification task in this study, may require additional processing time and affect search behavior.

12. Reliability: The algorithms computing the salient locations may not always be reliable and therefore generate false marker locations and lowering cue fidelity. If a viewer becomes aware of this unreliability, s/he will likely adjust to it. This has already been investigated in a visual search task by Groenewald et al (under review). They measured that the attentional window changed with marker validity: for valid markers, the attentional window was large in order to capture possible markers, for invalid markers, the attentional window was small.

A central issue is – as mentioned repeatedly - that gaze-guidance feels comfortable. We think that this is particularly necessary for continuous guidance during which a marker is presented frequently. If the viewer's gaze is to be directed to a salient location in a non-irritating manner, then the marker and cue should be subtle. Ideally, a marker would be hardly visible, yet still draw a viewer’s gaze every time it appears (McNamara et al 2008). Toward that goal we carried out experiments, which address the aspects of marker appearance, occurrence and location in a broad manner.

1.2. Experimental Framework

Our primary goal was to create a challenging recognition task, that has a cognitive work load as experienced in a car cockpit or in a PC setting for instance. If no such ‘heavy’ work load existed in our experiments, then the viewer may only passively browse the visual field and react too easily to the markers - in some sense too superficially. Our choice was therefore to combine a noise display with a fast-paced letter identification task. The noise display has been introduced and described previously (Rasche & Gegenfurtner, 2010) and here its qualities are summarized for the purpose of understanding the specific goals in this study:

a) The display is a dynamic (flickering) bar code, or also called noise movie, see figure 1 top for a single frame. The movie is generated from a two-dimensional image, whose power spectrum is correlated in space and time in a 1/f relation of which each row is used as the source for a single frame (stretched to a bar code). The movie thus appears as a mixture of rapid high frequencies and slower low frequencies. We chose this type of noise, because the frequency power spectrum of visual images falls off in a 1/f manner

(Field 1987; Simoncelli, Olshausen, 2001). To ensure that this type of display approximates real-world conditions, a comparison between the statistics of fixation locations and the statistics of non-fixations was made (randomly selected ‘fixations’; see Rasche & Gegenfurtner (2010) for details): the fixation statistics are surprisingly similar to the ones in natural scenes (see also Tatler et al 2005, 2006) and our chosen noise display is therefore a reasonable approximation to a natural stimulus.

b) Visual orienting for a gaze-dependent marker stimulus was tested. The marker consisted of a small increase in luminance for a limited region (see rectangle in bar code for the size of such a marker, figure 1), whereby the luminance increase was dependent on gaze eccentricity using a saturation as mentioned under aspect no. 4 (see above list; to be described again in the method section). During the first few trials of an experiment, observers did not notice the markers, but then learned their appearance. Exhaustive testing showed that the eccentricity-dependent compensation achieved a relatively constant detection rate (ca. 50%) for eccentricities of up to 25 degrees with no deteriorating performance for manual or saccadic reaction times (figure 7 in Rasche & Gegenfurtner, 2010). There was also a substantial amount of saccades toward the markers that were not followed by a button press response (ca. 18%). The eccentricity-dependence adjustment was also successfully implemented in an applied study, in which the size of the mouse cursor depended on gaze eccentricity (Dorr et al 2009).

The saccadic constant error - the amount of saccadic undershoot toward a marker - was ca. 16%, which is approximately twice as large as the one measured in simple displays (8-10%, see Kalesnykas and Hallett, 1994). This indicates that the noisier the display is, the more imprecise is saccadic landing.

But what has not been characterized yet, is the orienting behavior for a recognition task - and not only for a search task. This is certainly of interest when designing a gaze-guidance system in which the recognition of structure is part of the task. To investigate this we carried out a challenging letter identification task in this study (figure 1). The letters appear only transiently and are therefore difficult to detect and to identify, requiring thus full attention. A comparable real-world scenario would be the detection and recognition of road signs while driving in dense fog. To facilitate detection and identification, markers appear at those spatial location where a letter is going to appear. Can a marker compensate for the typical saccadic undershoot? Does the saccadic inaccuracy have an influence on recognition? We also tested a variety of different appearances of the marker and tested the effects of various presentation times.

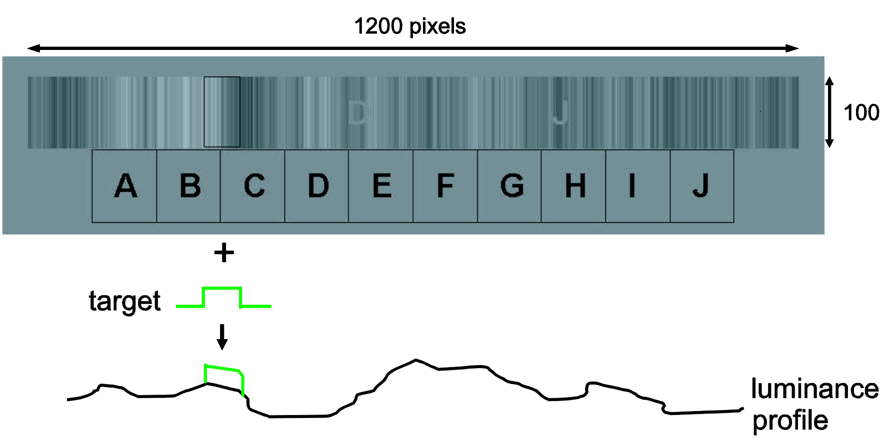

Figure 1. Letter search and identification task. The bar code (1200x100 pixels) represents a still image of a flickering noise movie whose frequency spectrum falls off with 1/f. Two letters are present (with high contrast for purposes of demonstration). Below the bar code the letter menu is displayed, which is used for identification during visual search. A marker was generated by adding a rectangular function to the luminance profile of the bar code (bottom). 10 letters were shown with a frequency of 0.06 Hz each (ca. 6 letters per 10-second trial), for a duration of 500ms at a contrast of 0.1 (not to scale in figure).

2. METHODS

Observers. Male and female students (age 23-30) served as observers and were compensated for their time. All observers had normal or corrected to normal vision. All observers were naive with respect to the aim of the experiment.

Equipment. Observers were seated in a dimly lit room facing a 21-inch CRT monitor (ELO Touchsystems, Fremont, CA, USA) driven by an ASUS V8170 (Geforce 4MX 440) graphics board with a refresh rate of 100 Hz non-interlaced. At a viewing distance of 47 cm, the active screen area subtended 45 by 36 degrees of visual angle on the subject's retina, in the horizontal and vertical direction respectively. With a spatial resolution of 1280 x 1024 pixels this results in 28 pixels/deg. The subject's head was stabilized in place using a chin rest. Eye position signals were recorded with a head-mounted, video-based eye tracker (EyeLink II; SR Research Ltd., Osgoode, Ontario, Canada) and were sampled at 250 Hz. Observers viewed the display binocularly through natural pupils. Stimulus display and data collection were controlled by a PC.

Noise stimulus. The two-dimensional 1/f image *I[x,t]* is generated using a 2D image of normally distributed random pixel-intensity values, whose frequency spectrum was then transformed to describe a 1/f-frequency decline. The image size is 1000*1200 pixels (time and space respectively). Each row is the source for a single frame: the row was stretched vertically to a height of 100 pixels and placed into a gray background presented as 8-bit (40 cd/m^2 luminance). The total intensity ranged from 0 to 1. A frame was shown for 10ms, a movie thus lasted 10ms*1000 (pixels) = 10s and constitutes one trial. Each movie *I(x,t)* was different to avoid potential learning effects.

Marker stimulus. Markers were shown for a duration of *d*=300ms (30 frames) and a spatial width of 1 degree, see figure 1. They were presented spatially and temporally randomly with an average frequency of 0.333Hz. Markers are added as a rectangular function with amplitude a_{mrk} to the luminance profile of the source image. The amplitude depended on eccentricity e by an exponentially saturating function: $a_{mrk}(e) = a_{min} + a_{max} - exp(-e)a_{max}$, whereby a_{min} is a minimal amplitude and a_{max} is a maximal amplitude; the function starts at a_{min} and saturates at $a_{min} + a_{max}$. The parameter values were a_{min} =0.2, and a_{max} =0.5, chosen heuristically after a few initial tests.

Markers appeared 550 ms before onset of a letter with always 100% validity. The markers duration lasted 500ms. The temporal gap between marker offset and letter onset was typically 50 or 100ms to avoid potential masking effects. Markers appeared with varying frequency per condition: 0, 25, 50, 75 and 100%. The conditions with 0% and 100% cueing represent the control conditions for which no supporting cues (markers) appeared at all (0%), or for each letter appearance one (100%).

Marker variations. A number of marker modifications were tested, whereby the above described properties are also called fixed [‘fxd’], meaning that no other modifications were done on the gaze-eccentric marker.

Flickering condition [‘flk’]: The amplitude a_{mrk} alternated between 0 and a_{mrk} with a frequency of 50Hz (every 2nd frame).

Looming condition [‘loom’]: The amplitude gradually increased from 0 to a_{mrk} within a time span of 300ms.

Wiggly condition [‘wig’]: the spatial location of the marker was alternated along the horizontal axis (left/right displacement) around its center point with a frequency of 33Hz.

Letter stimuli. Letters are taken from a 64 x 64 bitmap and appear of size ca. 1x1 degree in the movie (figure 1). A letter was shown with a frequency of 0.06 Hz (ca. 6 letters per 10-second trial), for a duration of 500ms at a contrast of 0.1. The letters in figure 1 are shown with increased contrast for the purpose of illustration. Letter occurred with equal probability.

Procedure. Observers performed blocks of 50 trials, generally 3 blocks per day and 6 blocks per experiment. Each block was preceded a calibration. The letter identification response was performed with the mouse by menu selection (see figure 1). The letters in the menu had the same size as the ones in the noise movie. Each search and recognition condition was carried out by 4 to 5 persons. For the 100% guidance condition the marker appeared 850-900 times (ca. 3 marker presentations per trial). To rule out learning effects, which could possibly occur when performing the conditions in increasing cueing order, observers started with the 50% condition first, followed by other conditions. The 0% condition was carried out last. After the end of a noise sequence, observers were given another 2 seconds to make their identification response for letters that had appeared just before the end of the movie. The instructions were to identify as many letters as possible and to make the best possible judgment. Observers were told that letters appeared of the same size as in the menu, occurred with equal probability and always had 100% validity. Observers performing this experiment had done the search task as described in Rasche & Gegenfurtner (2010) to get acquainted with the type of marker (cue). Observers occasionally saw more letters than they could manually select by the mouse menu. Observers were not given any time constraints when doing the letter-identification. They made the selections quickly but did not feel pressured.

Analysis. To determine whether an observer reacted to the appearance of a target (letter or marker), we chose as a criterion whether gaze shifted (saccadic shift) toward the selection menu during or shortly after the presence of a target. Given the relatively slow mouse-menu selection process – as opposed to just a button press in the search experiment -, no maximal reaction time was defined. Due to the occasional occurrence of multiple letters within a short period, it is difficult to relate individual menu selections with individual identification responses.

3. RESULTS

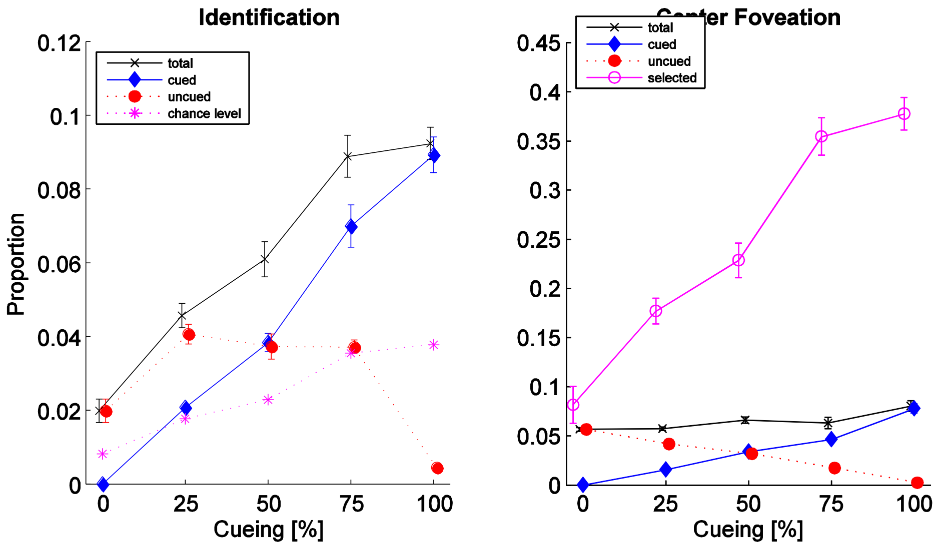

Figure 2. Letter foveation and identification in dependence of the amount of cueing/guidance (0, 25, 50, 75, 100%). Left: Proportion of identified letters. total: cued and uncued (guided/not guided); cued: proportion for cued (marked) letters; uncued: proportion for uncued letters. chance level: proportion of identification responses (see right) divided by the number of letters. Error bars represent standard error of inter-observer performance. Right: Center foveation (1-deg tolerance): selected: proportion of identification responses (letter selections using menu).

To verify that guidance did facilitate the recognition process, the total, cued and uncued performances are plotted separately (figure 2 left). The total identification rate, determined as the proportion correct of the selected letters, increased steadily from 0.02 to 0.09. The absolute identification level was small yet irrelevant to the goal of this experiment. Chance level was calculated as the proportion of manual selections divided by the number of letters (see figure 2, right graph for proportion of manual selections). The performance for cued letters (filled diamonds) increased equally rapid but with a small offset. The uncued identification rate (filled circles) unexpectedly increased slightly from 0% to 25%, but may be explained by an increased propensity to respond when cueing was present. The uncued rate then slightly decreased for 25% to 75% cueing. The results so far clearly prove that guidance facilitated the recognition process.

To obtain first insights into the orienting dynamics we now compare the foveation rate - the proportion of letters to which the gaze was moved to - with the selection rate - the proportion of manually selected letters (identification responses). The comparison was made for a ‘foveation hit’ with a 1-degree tolerance representing the center fovea (figure 2, right graph). For 0% cueing, the center foveation rate was at a value of around 0.06, whereby the selection rate was only slightly higher, revealing that central foveation was almost a requirement to make an identification response. With increasing amounts of cueing, the selection rate increased rapidly (open circles), whereas the cued rate increased slower (the uncued and total rate are shown for control only). This hints that covert attentional shifts must have occurred to obtain ‘certainty’ for the letter identification judgment.

The large increase in identification responses may have several reasons: 1) in the absence of guidance, observers foveated letters too late to identify them properly; 2) in the presence of guidance, observers feel more compelled to make identification responses; 3) in the presence of guidance, the markers have facilitating effects on identification by the transient high-lighting of the letter location.

To obtain further clues about the orienting dynamics, we determined the proportion of letter selections for which 0, 1 and 2 saccades toward the target were carried out, also called no-saccade, one-saccade and two-saccade selections (figure 3). For the majority of selections no saccade was carried out, independent of cueing condition (labeled ‘0’, upper left graph called ‘Total’), hinting that covert attentional shifts are the dominant form of orienting to obtain a letter judgment. These no-saccade selections increased slightly from a value of ca. 0.55 for the 0%-cueing condition to a value of ca. 0.6 for the 100%-cueing condition. The upper right plot (labeled ‘No Saccade’) shows the individual proportions for cued and uncued letters.

A large portion of selections was carried out after one saccade toward the target was made (one-saccade selection), ca. 0.35 for all conditions (labeled ‘1’). A few selections were carried after two saccades toward the target were made, and the proportion of those two-saccade selections was decreasing with increasing amount of cueing (from a value of ca. 0.1 down to 0.05; labeled ‘2’).

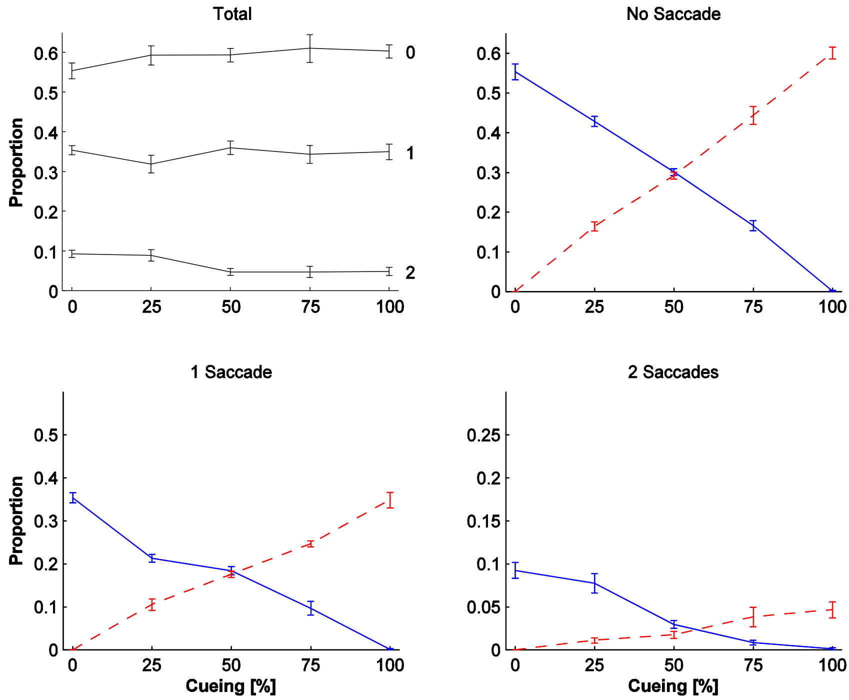

Figure 3. Proportion of saccades - made toward targets before letter-identification selection was carried out - as a function of cueing conditions (0%, 25%, …, 100%). Upper left: Total proportion (letters and cues) for 0 (attentional shift only for identification), 1 and 2 saccades. Upper right: Proportion of identification selections for uncued (solid) and cued (dashed) letters, for which no saccade toward the target was made. Lower left: Selection proportion for one saccade. Lower right: Selection proportion for two saccades. Error bars = standard error of inter-observer performance.

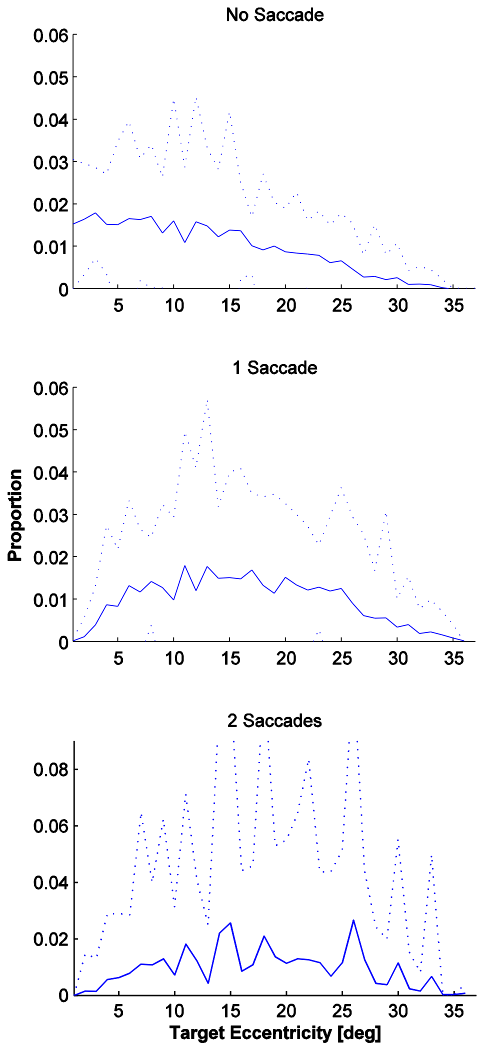

Figure 4. Distribution of target eccentricities for identification selections for which 0, 1 and 2 saccades toward the target were made, averaged across cueing conditions (cued/uncued; 0%, 25%, …, 100%). Top: 0 saccade (no-saccade selections). Middle: 1 Saccade. Bottom: 2 Saccades. Dotted = standard error of inter-observer performance.

We now perform the eccentricity-dependent analysis of the target location and the saccadic landing precision (figure 4 and 5). This was done for each cueing condition (0%,...,100%) and for cues and letters separately, in an effort to find potential differences in orienting behavior. But no statistical differences could be determined. For the distribution of target eccentricities we therefore averaged across all the cueing conditions to obtain a function as smooth as possible (figure 4). For no-saccade selections, the eccentricity distribution started high at the center of gaze (0 degrees) and then gradually declined into the far periphery (more than 30 degrees, see top graph). Thus, attentional shifts were favorably carried out for proximal targets. For one-saccade selections the distribution was even and was centered around 15 to 20 degrees (middle graph). It coarsely matches the one for secondary saccades made in the target search (see figure 8 and 9 in Rasche, Gegenfurtner (2010)). For selections after two saccades the distribution seemed to match the one for one saccade.

For the eccentricity-dependent constant-error (undershoot function) we also did not find any significant differences between conditions. We therefore show the variability for one condition, the 50% condition for one-saccade selections (see figure 5). The function is much steeper than the one for visual search and shows a constant error of ca. 50%, which is about 3 times as much as for a simple visual search task (error of 16%, Rasche & Gegenfurtner, 2010).

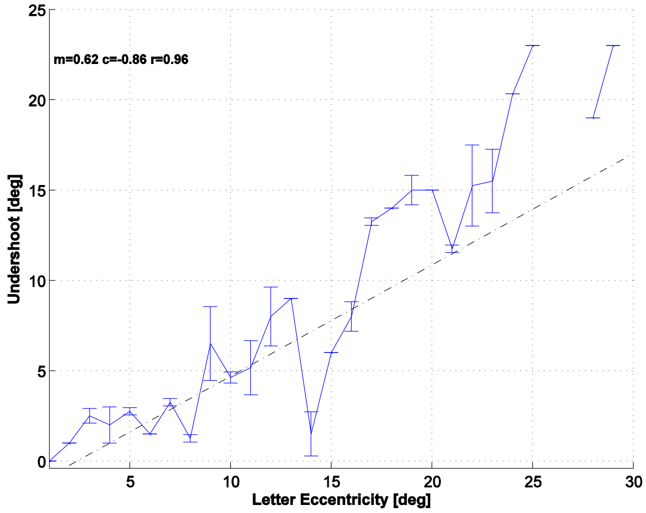

Figure 5. Landing precision (constant error) in dependence of target eccentricity for the letter-identification task, when one saccade before letter selection was carried. Error bars = standard error of inter-observer performance.

Apparently, the purpose of a saccade was not to land precisely on the target, but rather to bring the target letter somewhat closer in order to perform another, spatially shorter attentional shift. Given this potential strategy, it is no surprise to find that the proportion of two-saccade selections decreased with increasing amount of cueing.

Could this lack of increase in saccadic orienting be task specific? For example, observer may have intended to catch as many letters as possible by viewing the noise movie on a global scale and by consequently suppressing additional saccadic shifts toward the letters - although observers were not temporally constrained. Another potential reason for this saccadic orienting ‘inertia’ could have been the choice of temporal stimuli durations. But in principle, they can be considered as sufficiently large: the cue duration was 300ms, followed by a gap of 50ms; the target itself appeared for a duration of 500ms, which even in the uncued condition is long enough to perform a saccade toward it. In summary, even if it is not clear whether the lack of saccadic orienting is task-specific or not, it is still perplexing how robust and far-reaching attentional shifts are.

Now that the principle of gaze guidance is established, we can start testing variations of the marker properties to improve guidance performance. To investigate the timing issue (aspect ‘occurrence’), we varied the temporal gap between marker offset and letter onset (50, 100 and 150ms). This is carried out with the constant (fixed) marker amplitude at a guidance rate of 50% (figure 6). For increasing gap sizes, the total foveation rate steadily increased (triangles); the performance for guided letters and not-guided letters is shown as control. However, for the identification rate, there was a sharp drop for a gap size of 150ms and the performance for a gap size of 100ms seems to be close to the optimum.

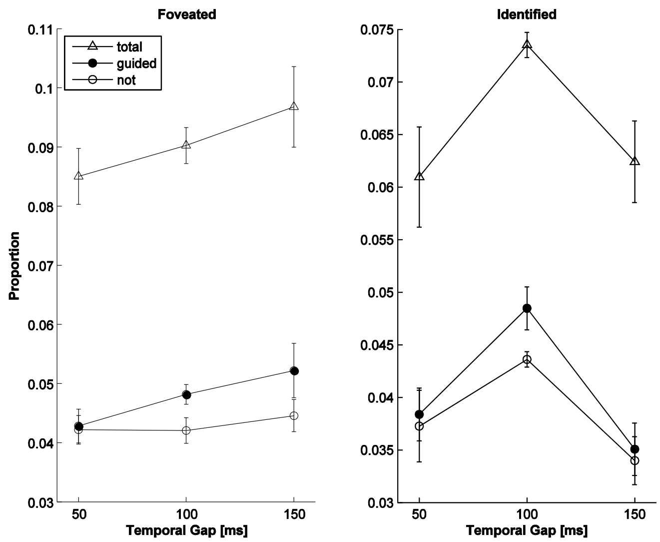

Figure 6: Letter foveation and identification rate for three different temporal gaps between marker offset and letter onset for the fixed-amplitude marker (50% guidance). Left: Foveated letters (total, guided, not-guided). Right: Identified letters.

As a temporal gap size of 100ms seemed the optimum, we used this parameter value when testing 3 other marker variations, a flickering, a looming and a wiggly marker (figure 7). For comparison the performance of the fixed marker used so far, is also plotted (label ‘fxd’). For a flickering marker with alternating amplitude (‘flk’) the foveation performance dropped slightly (left graph in figure 7); for a looming marker (‘loom’) the performance marginally increased; and for a wiggly marker (‘wig’) with an alternating, horizontal displacement along the spatial axis, the performance was highest. Again, the corresponding identification performance looked different (right graph in figure 7). It was lowest for the flickering condition, but highest for the fixed condition. The letter identification performance for guided letters (full circles) was even significantly under the performance for non-guided letters (empty circles). Thus, it seems that guidance even deteriorates recognition performance for this marker type. The implications of these differences for the design of markers are discussed next.

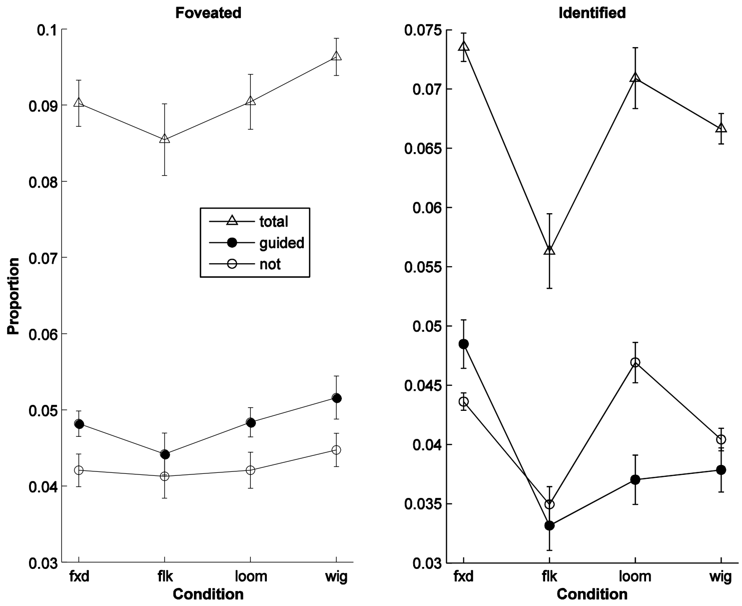

Figure 7: Letter foveation and identification rate for different markers (50% guidance; 100ms gap). Left: Foveated (total, guided, not-guided). fxd: fixed amplitude (eccentricity-dependent marker without further modification); flk: flickering marker (alternating amplitude); loom: looming marker (gradual amplitude increase and decrease); wig: wiggly marker (alternating spatial displacement). Right: Identified letters (total, guided, not-guided).

4. DISCUSSION

4.1 Previous approaches

A specific guidance system was already tested by McNamara et al (2008). In their study, observers were asked to count the number of soap bubbles which were placed into a static, virtual-world-like scene, e.g. 6 fist-sized soap bubbles were placed randomly in a virtual office scene. They used a flickering luminance marker, whose amplitude was set to two distinct levels: a high level represented the obvious marker type; a low level represented the subtle marker. The subtle marker was applied in the periphery only (gaze-contingent), was smaller than the soap-bubble target and was never noted by observers; the obvious marker was simply more salient and was clearly noted by observers. The detection and counting rate was higher for the obvious markers but surprisingly not by much. McNamara's study clearly demonstrates the potential of unobtrusive gaze guidance. Following our list of aspects, the system can for example be classified as a task with low search complexity as it involves only the counting/detection of objects; targets and markers appeared simultaneously and observers were given sufficient time for counting (aspect 'occurrence' and 'response urgency' respectively).

Another gaze-capturing system is the one developed by Kim and Varshney, who designed a method to attract gaze in 3D-graphic displays (Kim and Varshney, 2008). Their markers, called 'persuasive filters', were designed especially for 'meshes' and were created by inverting the center-surround saliency operator. If a higher performance is desired for either system, than our list of aspects provides a systematic approach to address possible sites of improvement.

Both studies were carried out in virtual scenes, which typically contain less visual complexity and noisiness than real-world scenes, in which for instance the luminance of surfaces is already much more inhomogeneous. For guidance in real-world scenes, the markers of the above mentioned studies may not be salient enough to attract gaze as they are generated by very subtle manipulations in a noise-free image. The system that is being developed by Barth's group aims at such real-world scene guidance, e.g. Vig et al, 2009. The goal is to guide the viewer through a brief movie with the purpose to manipulate the viewer's understanding of the movie. In comparison, movies produced by the film industry place the position of the camera such, that a viewer's gaze is placed on the appropriate spot, meaning gaze guidance was already implemented by the director. But for simpler types of movies or scenes, guidance needs to be implemented afterward. To pursue this ambitious goal, Barth et al perform whole-image manipulations which involve the lowering of the saliency of those areas, which are *not* supposed to be focused at a given point in time (Barth et al, 2006a, 2006b). Their marker is therefore not confined to an isolated area but is in some sense the untouched or non-manipulated area.

4.2 Experiences of our studies

We now discuss the guidance experiences made with the dynamic noise movie, starting with the visual search task as characterized in our previous study (Rasche & Gegenfurtner, 2010).

a) For a search task, the landing precision of eye movements toward markers linearly decreased with increasing eccentricity, a 16% error approximately (see Rasche & Gegenfurtner, 2010). For up to ca. 8 degrees eccentricity, undershoot measures only 1 degree and may not be worth correcting for those proximal eccentricities, because the fovea covers an area of 2 degree diameter. But for larger marker eccentricities it may be necessary to consider peripheral compensation, in particular when small targets are to be detected such as the 'Blinking Cursor' in a word editor. This compensation could be done by placing the marker radially beyond its target and turning it off when gaze moves toward it.

We suspect that the landing variability in McNamara's as well as in Kim and Varshney's study is smaller than in our study as they use static scenes only, but it probably is larger than in experiments with typical psychophysical displays, as the observers in McNamara's study carry out a visual search, which likely involves an increase in landing variability.

b) For an attention-intensive recognition task (this study), the letter identification task had revealed that observers did not need to place their gaze upon the letters to make identification judgments, but preferred 'direct' attentional shifts over saccadic shifts. And if a saccade toward the target was carried out, its constant error was 50% (figure 5). Does this mean that the proposed undershoot compensation (item a) can be neglected? Not necessarily. The identification task used in this study was very hectic, a likely scenario in dense traffic situations. But for other applications (see item a above) such attentional shifts may play a smaller role.

c) The exact marker appearance properties influence the performance but only to a small degree (figure 7). They maybe therefore be negligible in certain applications, but could be beneficial in other applications or if an optimization is intended. The marker manipulations we tested were essentially all some form of ‘motion’ stimulus and given that such stimuli are very salient (Franconeri & Simons 2003), one could have expected that they increase performance. It is only the wiggly marker, which showed a slight increase in foveation performance, but for identification performance the motion markers were rather detrimental. The reason may have been that such markers do not combine well with a dynamic noise background. In contrast, the ‘fixed’ marker, which pops out as a constant spot in this restless background, may appear as a ‘calm’ guidance. Thus, the recognition process should not be underestimated: gaze guidance toward a spot is only part of the process, but the perception of structure at that location is another important part.

d) The manipulations with temporal gap sizes aimed at determining the degree of masking (figure 6). Masking is the phenomenon that when two stimuli are presented in rapid succession at the same spatial location, then one stimulus can influence or even prohibit the perception of the other (Coltheart, 1999). Applied to our experiments, this means that a marker can affect the detectability of its guided letter (also called forward-masking). This likely has occurred in case of the 50ms gap, for which the identification rate was smaller than for the 100ms gap. But for larger gap size of 150ms, identification declined again, possibly because of the intrinsic rhythm of the visual system to move on and to rest only a limited period on a fixed spot.

As our experiments were carried out under strict psychophysical conditions, e.g. using a dark room and a chin rest to fix the head, on may wonder whether the results also extend to more natural conditions. Eye-tracking at a PC monitor or in a car cockpit certainly does not provide the same type of accuracy and the eye-position measurements would therefore show a larger degree of variability. Furthermore, under more natural conditions the amount of undershoot or orienting inaccuracy (items a and b) may be even higher. But the more important could be any attempt for compensating this variability with cleverly placed markers.

4.3 Summary of design issues:

We summarize the specific experiences made in this study as a set of issues to be considered when designing or analyzing a gaze guidance system.

1) **Aspect range**: To compensate for the decline in peripheral acuity, the marker's amplitude is increased with eccentricity by an exponentially saturating function: $a_{mrk}(e) = a_{min} + a_{max} - exp(-e)a_{max}$ (a_{min} = minimal amplitude, a_{max} = maximal amplitude).

2) **Aspect location**: If a compensation for undershoot is desired, the marker should be placed radially beyond its target by 18% of target eccentricity. Such compensation is probably required when small, hard-to-detect targets are to be foveated which are embedded in a complex background.

3) **Aspect appearance**: Motion markers are better gaze-capturing events than stationary markers, however they are potentially detrimental to recognition performance at their location.

4) **Aspect occurrence**: In case of guidance toward briefly appearing stimuli, the optimal gap size between marker offset and target onset is ca. 100ms to avoid strong forward-masking effects.

ACKNOWLEDGMENTS

We would like to thank Erik Groenewold for the reliability aspect. Lab support by Nadine Hartig. This work was funded by the Gaze-based Communication Project (contract no. IST-C-033816, European Commission within the Information Society Technologies).

REFERENCES

BARTH, E., DORR, M., BÖHME, M. GEGENFURTNER, K. & MARTINETZ, T. (2006a). Guiding the mind's eye: improving communication and vision by external control of scanpath. In Human Vision and Electronic Imaging XI: Proceedings of SPIE, B.E. Rogowitz, T.N. Pappas, S.J. Daly, eds., 6057, 1-8.

BARTH, E., DORR, M., BÖHME, M., GEGENFURTNER, K. & MARTINETZ, T. (2006b). Guiding eye movements for better communication and augmented vision. *Perception and Interactive Technologies*, volume 4021 of *Lecture Notes in Artificial Intelligence*,1-8.

BURKE, M., HORNOF, A., NILSEN, E., & GORMAN, N. (2005). High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. ACM Transactions on Computer-Human Interaction, 12(4), 423-445.

CASPI, A., BEUTTER, B. R. & ECKSTEIN, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. Proceedings of the National Academy of Sciences of the USA, 101 (35), 13086-13090.

CHAPMAN, P.R. & UNDERWOOD, G. (1998). Visual search of dynamic scenes: event types and the role of experience in driving situations. In: Underwood, G. (ed.), Eye guidance in reading and scene perception, pp. 369-93, Amsterdam: Elsevier.

COLTHEART, V. (ED.) (1999). Fleeting Memories: Cognition of Brief Visual Stimuli. The MIT Press.

DORR M., RASCHE C. AND BARTH E. (2009). A Gaze-Contingent and Acuity-Adjusted Mouse Cursor. Conference on Communication by Gaze Interaction (COGAIN). Lyngby, DK.

FIELD, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4 (12),* 2379-2394

FINDLAY, J.M. AND GILCHRIST, I.D. Active Vision. New York: Oxford University Press, 2003.

FRANCONERI, S.L. & SIMONS, D.J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65,* 999-1010.

GLUCK, J., BUNT, A. AND MCGRENERE, J (2007). Matching attentional draw with utility in interruption Proceedings of the SIGCHI conference on Human factors in Computing Systems ACM Press, San Jose, California, USA 2007 41-50.

GROENEWOLD, E.S., RENKEN, R. AND CORNELISSEN, W. (UNDER REVIEW). Attentional Window Size is Primarily Set by Environmental Relevance.

EINHÄUSER, W., SCHUMANN, F., BARDINS, S., BARTL, K., BÖNING, G., SCHNEIDER, E. & KÖNIG, P. (2007). Human eye-head co-ordination in natural exploration. Network: Computation in Neural Systems, 18, 267-297.

HANSEN, D.W., MACKAY, D., NIELSEN, M. & HANSEN, J.P. (2004). Eye tracking off the shelf. *In: Proceedings of the 2004 symposium on eye tracking research & applications,* 58-58.

JACOB, R.J.K. (1993). Eye movement-based human-computer interaction techniques: toward non-command interfaces. *In* H. R. Hartson & D. Hix, eds, Advances in Human-Computer Interaction, Vol. 4, Ablex Publishing Corporation, Norwood, New Jersey, chapter 6, pp. 151-190.

KALESNYKAS, R.P. & HALLETT, P.E. (1994). Retinal eccentricity and the latency of eye saccades. Vision Research, 34, 517-531.

KIM Y, VARSHNEY A. (2008). Persuading Visual Attention through Geometry. IEEE Trans. Visualization and Computer Graphics. 14(4):772-782.

LAND, M.F., MENNIE, N. & RUSTED, J. (1999). The roles of vision and eye movements in the control of activities of everyday living. Perception, 28, 1311-1328.

LI, D., BABCOCK, J. & PARKHURST, D.J. (2006). openEyes: a low-cost head-mounted eye-tracking solution. *In: Proceedings of the 2006 Symposium on eye tracking research & applications, San Diego,* 95-100.

MCCRICKARD, D.S., CATRAMBONE, R. AND STASKO, J.T., (2001) Evaluating Animation in the Periphery as a Mechanism for Maintaining Awareness. in Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction, (Tokyo, Japan, 2001), 148-156.

MCNAMARA A, BAILEY R AND GRIMM C (2008). Improving search task performances using subtle gaze direction. Applied Perception in Graphics and Visualization. In press.

NAZIR, T.A. & JACOBS, A.M. (1991). The effects of target discriminability and retinal eccentricity on saccade latencies: An analysis in terms of variable-criterion theory. *Psychological Research, 53,* 281-289.

OBERMAYER, R.W. AND NUGENT, W.A. (2000), Human-computer interaction for alert warning and attention allocation systems of the multi-modal watchstation. in Integrated Command Environments: Proceedings of SPIE, Vol. 4126, 14-22.

POSNER, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32A,* 3-25.

QVARFORDT, P. & ZHAI, S. (2005): Conversing with the user based on eye-gaze patterns. *In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '05,* 221-230.

RASCHE, C. AND GEGENFURTNER, K. (2010). Visual Orienting in Dynamic Broadband (1/f) Noise Sequences. Attention, Perception & Psychophysics. 72(1), 96-109.

SALVUCCI, D.D. & ANDERSON, J.R. (2000). Intelligent gaze-added interfaces. *CHI Letters, 2,* 273-280.

SIMONCELLI, E. P. & OLSHAUSEN, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24,* 1193-1216.

STANTON, N. (ed.), Human Factors in Alarm Design. Taylor and Francis, London, 1994.

TATLER, B. W., BADDELEY, R. J. & GILCHRIST, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45,* 643-659.

TATLER, B. W., BADDELEY, R. J. & VINCENT, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46,* 1857-1862.

VIG E., DORR, M. AND BARTH, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. Spatial Vision. In press.