Design Issues in Gaze Guidance

Under review with ACM Transactions on Computer Human Interaction

Christoph Rasche*, Karl Gegenfurtner
Abteilung Allgemeine Psychologie, Justus-Liebig-Universität
Otto-Behaghel-Str. 10F, 35394 Giessen, Germany
Email: rasche15@gmail.com

Abstract

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 list of guidance aspects is discussed and then applied to two specific scenarios, car cockpit and PC monitor. To explore some of those aspects, an experimental framework is introduced in which subjects perform a difficult letter search and identification task in dynamic noise. To facilitate this recognition task, the viewer is guided by a luminance 'marker'. It is investigated how the marker's spatio-temporal properties influence the recognition performance. From those results we derive a number of design recommendations for the process of gaze guidance.

Content indicators:

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

*Corresponding author

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 one 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 intends to argue the following points:

- The complexity of gaze guidance should not be underestimated: trying to comprehend the complexity in its entity will more likely lead to a successful implementation.
- In order to successfully implement a complex guiding system, one should start with a well specified, simple guiding task. A simple task may not satisfy the researcher's hunger for futuristic system, but may be necessary to gain experience and to anticipate potential pitfalls when expanding to more daring systems.
- Today's technology is ready to perform gaze guidance in a PC setting and some examples are given.
- Guidance should be comfortable, e.g., bright flashing dots are unlikely candidates for a smooth guiding process.

Rudimentary forms: Gaze guidance already exists in rudimentary forms, for instance on personal computers: 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.

Traffic signals are another rudimentary form of gaze guidance. Traffic signals are made of different degree of saliency. They exhibit bright, conspicuous colors to notify the driver of potential dangers (e.g., red stop sign), but moderate colors when the road sign contains general information (e.g. signs pointing toward historic sites). The saliency of a road sign is modulated by its context: on a secluded road in a remote area, any traffic signal may be a welcome attraction; in a busy city center, a traffic signal may be drowned in a sea of other signs.

These examples already hint that there is a range of aspects associated with optimal gaze guidance. In case of the blinking cursor it would be optimal if the saliency of the cursor were proportional with eccentricity in order to recapture gaze unerringly. In case of the car cockpit, it would be beneficial, if those road signs were pointed out, which pertain to the current driving task, for instance parking signs when searching for a parking spot. A central issue is therefore the adjustment of the saliency of the visual marker. This point is expanded in the next paragraph.

The process in a nutshell: When humans search a visual field in real-world conditions, they are doing this often in parallel with many other actions, which decrease or even

divert the viewers attention from the actual search task (e.g. leaving temporarily the word editor). Generally expressed, the viewers attention, *A,* fluctuates as a function of time, *A(t)*. During such distractions the marker's saliency, S_m, needs to be increased in order to efficiently recapture gaze. The saliency is therefore dependent on the subjects attention, and is also dependent on the context, *c*, as mentioned above: $S_m(A, c)$. If a viewer is distracted strongly, it may even require a cueing signal to disengage the viewer from the present gaze position. The saliency of the cueing signal may also depend on attention and context, $S_c(A, c)$. Thus the essential cycle of operations is the sensing of the viewer's attention, then adjusting the saliency of the cuing and marker signal, followed by placing an attention- and context-dependent marker at the corresponding spatial locations: these latter two operations are denoted as $P_c(x_c, y_c, S_c)$ and $P_m(x_m, y_m, S_m)$, with (x_c, y_c) and (x_m, y_m) as the corresponding spatial coordinate pairs. The cycle is completed by recognizing the visual information at the marker location, $R(x_m, y_m)$:

$$A(t) \rightarrow S_c(A,c), S_m(A,c) \rightarrow P_c(x_c, y_c, S_c), P_m(x_m, y_m, S_m) \rightarrow R(x_m, y_m) \rightarrow A(t)$$

The formulation is a generic one. Implementations of simple guidance tasks may require only a subset of those outlined operations and functions. The following section elaborates on all of these operations.

A List of Aspects

The list of aspects is not confined to a particular system, but intends to address the topic in a broad sense. One may distinguish between 3 types of aspects: temporal, attentional and spatial (figure 1).

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: As mentioned in the introduction already, 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.

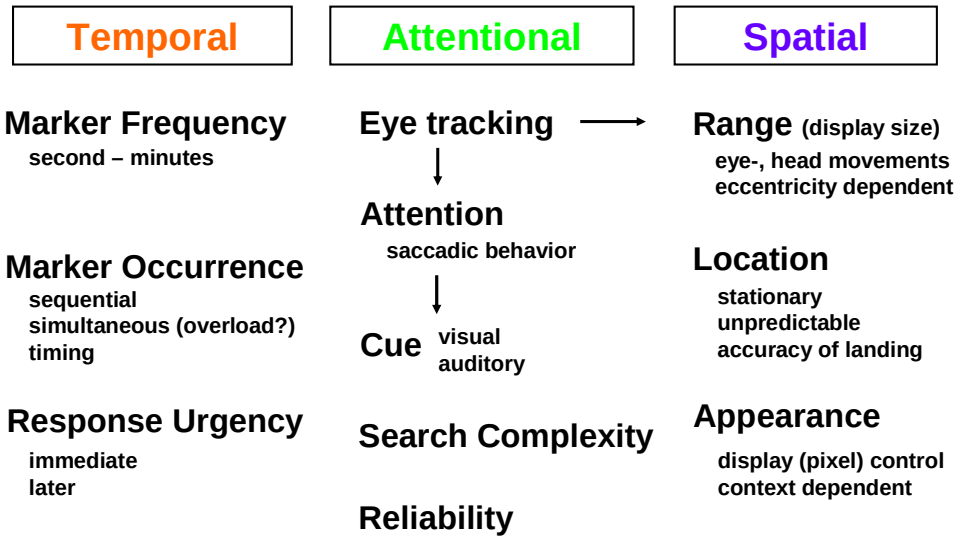

Figure 1. Aspects of gaze guidance, nominally classified into temporal, attentional and spatial aspects.

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 (term $R(x_m, y_m)$ in above equation). For instance, counting the number of occurrences of a visual structure involves merely its detection, but identifying it may require additional processing time. This may affect search behavior and the chase for an optimal search performance may require different markers.

12. Validity: The algorithms computing the salient locations may not always be reliable and therefore generate false marker locations. 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. 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.

With this list of aspects in mind, we can start envisioning specific tasks and elaborate on some of the aspects.

Specific Scenarios

PC monitor:

As pointed out above, the technology for performing gaze guidance on PC monitors exists already. Eye-trackers with an accuracy of 0.5 to 1.0 degrees provide sufficient potential to perform reasonable gaze guidance (Hansen et al 2004, Li et al 2006). Let us start with sketching simple guidance tasks, which do not require that attention is sensed (aspect 'attention'), but only eccentricity E serves as a variable to modulate the marker's saliency $S_m(E)$ (aspects 'marker range' and 'eye-tracking').

1) *Blinking cursor* (of a word processor): Taking the cursor size h as a measure for saliency, $h=S_m$, the size could be made proportional to gaze eccentricity E: When the gaze leaves the word editor, the cursor size would increase with gaze eccentricity. An eccentricity-dependent cursor size could also be beneficial, when the page is scrolled – an action during which one tends to loose the cursor position very rapidly due to the whole-text flow field.

2) *Event notification*: Unless the user awaits an immediate email response by someone, many email notifications – or other notifications - can be rather disrupting. To reduce the likelihood of disruption, email notification could be placed only at points in time when

the user's gaze is outside of the editor or near the icon bar. One could take this a step further and consider a system, which magnifies icons or menu points when gaze is near them. That would be a step toward gaze-contingent displays. Alternatively, one could merely provide cueing signals at the location of the present gaze location.

Users may have also developed strategies to ignore areas of event notification. For instance, a study by Stenfors et al (2003) showed that experienced internet users avoid areas of banner advertisement. Even if a new web site appears, some users do not even fixate the banners once. Stenfors et al even distinguish between different strategies of how to avoid such areas. It is likely that such avoidance or ignorance strategies exist for any kind of notification. Instead, this would be another incentive to implement a clever Gaze Guidance system.

3) *Site anchors*: When users browse web sites, they tend to return to the same locations for faster the purpose of orienting. Stenfors et al call these locations *anchors* (2003). Such anchors can range from the menu bar, representing a static, global anchor, to the mouse pointer representing a dynamic, local anchor. Even the text cursor may be used as an anchor by placing it to specific text locations. The use of such external stimuli is an aid to avoid memory burden (Zhang and Norman, 1994). Clearly, all these forms of anchoring are some sort of gaze guidance. We therefore imagine the situation in which the user has the opportunity to place markers on the screen, which serve as anchors. For instance, when working in a text, these anchors should serve only temporarily and could thus fade

away within a few tens of seconds. Such anchors could be created in an eccentricity-dependent fashion as the blinking cursor.

The marker range on a PC monitor (visual field) varies from about 25 to 40 degrees depending on the viewer's distance to the screen (sitting far and close respectively). Given a maximal saccadic amplitude of 20 degrees, it is likely that visual browsing on the monitor screen involves head movements to some extent, and with the upcoming of wider screens their occurrence is even more likely. For large saccadic amplitudes, there exists undershoot, that is saccadic landings are generally too short. Sometimes, a corrective saccade takes place which brings the target into focus. If a viewer is supposed to be guided sequentially through a display, then repeated undershoot may feel disrupting. It is therefore beneficial to investigate large saccadic amplitudes and consider markers which minimize undershoot.

The great potential of a gaze-guiding system at a PC monitor is the feature of complete display control – in contrast to for instance the visual field as seen from a car.

Car cockpit:

The visual field in a car cockpit is much larger as compared to a PC setting: browsing the road scene and checking the mirrors clearly involves head movements. This makes eye-tracking much more challenging than in a PC setting and accurate eye-tracking may only

be possible when the driver observes the road scene. Still, knowing the approximate gaze position can be crucial: for instance, if a driver is looking into the rear view mirror, and the guidance system attempts to draw the driver's gaze into a side mirror, then the marker needs to be stronger, than if the driver is observing the road. This brings us back to the aspects attention and cuing, which are probably even more crucial during driving due to the presence of dangers.

Research on eye-movement behavior during driving has already addressed the aspect of attention (or intention) in order to develop educational methods to prevent accidents. Originally, it was believed that there exists a difference between the scanning behavior of novice and experienced drivers. But more extensive research questions the presence of any such differences: instead of seeking a single potential cause or mechanism, which could trigger accidents, Chapman and Underwood instead suggest to analyze the moment-to-moment syntax of scanning behavior (1998). Such research is relevant to the construction of a gaze-guiding system, because it can sketch what is required to render a marker's saliency attention-dependent in order to avoid its annoyance. As for the PC setting, algorithms which read a driver's attention or intention are being developed (e.g. Liu 1998). Assuming for the moment the absence of such algorithms, one may at least coarsely adjust the marker's saliency, by sensing where gaze location is approximately. For instance, one could divide the visual field into zones, the road-scene zone and the dash-board zone. During driving, gaze resides mostly in those two zones, but also the

mirrors can be regarded as a zone. In whichever zone gaze is at a given point in time, it may require more effort to lead a viewer's gaze toward another zone, than for instance to lead a viewer's gaze within the road scene zone. Following this zone division of the visual field division, we distinguish two types of guidance:

1) Guidance between the road-scene zone and the mirrors and dashboard zones: The left and right view mirrors appear at ca. 70 and 90 degrees eccentricity respectively, the rear view mirror at ca. 30 degrees. This clearly exceeds typical saccadic amplitudes (ca. 20-30 degs) and involves also head movements. Because the location of those markers is fixed (aspect location), saccadic orienting toward them should be very accurate, but marker saliency remains certainly an issue.

With a guidance toward mirrors, the driver is merely pointed out where potential information lies and s/he as to still interpret it, e.g. the dynamics of a car: whether a car is approaching quickly, is just staying behind, or is in the process of passing. This may bear the pitfall that drivers become too familiar with the markers: the may sense the marker in the periphery, anticipate a situation, but do not really gaze toward the marker anymore – and hence do not actually interpret the scene.

2) Guidance within road-scene zone: This is the most futuristic type of gaze guidance one can think of. Ideally, the observer is notified of immediate dangers, e.g. a car approaching the vehicle from the side. A more realistic goal is that potential dangers are merely pointed out, e.g. a merging car is labeled by a marker, or the car ahead is labeled if one

follows too closely. If eye-tracking within the road-scene zone is of reasonable accuracy, then one may make the size of the markers eccentricity-dependent as we suggested for the word-processor cursor. The technical challenge of guidance in the road-scene is to display those markers. There is less control over the display as in the PC monitor. Markers should of course not obstruct the driver's view, and therefore must be subtle yet still salient.

Experimental Framework

Whatever guidance system is implemented, a central issue is – as mentioned repeatedly - that the gaze-guiding process feels comfortable. 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. No specific scenario is simulated, but the experimental tasks and stimuli should reflect a cognitive load as experienced in a car cockpit or in a PC setting. 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 of display is a dynamic (flickering) bar code, or also called noise movie, see figure 2 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 we determined how the statistics of fixation locations differed from the statistics of non-fixations (randomly selected 'fixations'). Using the method by Kienzle et al. (2007), we computed that fixation and non-fixations differed by ROC area values ranging from 0.54 to 0.62 for different persons. The values are almost as high as ROC area values determined for fixations in natural scenes (ca 0.63, e.g. Tatler et al 2005, 2006). Our chosen noise display therefore evokes similar fixation behavior as in natural scenes.

In the experiments presented here, the task is to detect and identify letters embedded in a background of dynamic noise. 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, markers appear at those spatial location where a letter is going to appear. The appearance of a marker consists of only small manipulations of the

background noise. Although the experiment may appear very simple, the marker's appearance has a number of parameters, which may influence the detection and identification performance.

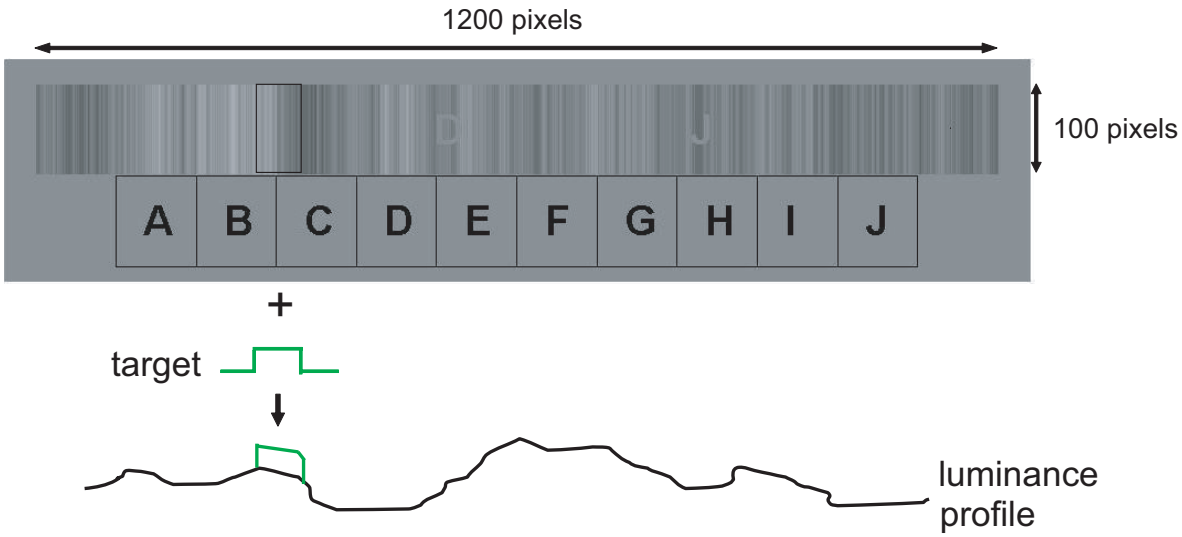

Figure 2. Letter search and identification task. The bar code (1200x100 pixels) represents a still image of a flickering noise movie whose frequency spectrum falls off with 1/f. Two letters are present (with high contrast for purposes of demonstration). Below the bar code the letter menu is displayed, 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).

Methods

Subjects. Male and female students (age 23-30) served as subjects. All subjects had normal or corrected to normal vision. All subjects were naive with respect to the aim of the experiment.

Equipment. Subjects 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. Subjects 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 2. 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: $a_{mrk}(e) = a_{min} + a_{max}\text{-}exp(\text{-}e)a_{max}$, 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%.

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 10). 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 10 are shown with increased contrast for the purpose of illustration. Letter occurred with equal probability.

Procedure. Subjects 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 10). The letters in the menu had the same size as the ones in the noise movie. Each search 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).

Analysis. Subjects were instructed to move their focus toward the targets and make the identification response. Target detection (‘foveation’) is defined as the temporal coincidence of a ‘saccadic hit’ and a button press. A saccadic hit required a saccadic flight toward the target and a spatial landing within 5 degrees of target eccentricity. The temporal tolerance for a saccadic latency was 400ms, respectively. Given the slow

mouse-menu selection process, no maximal reaction time was defined. No attempt was made to correlate mouse selections with foveated letters as this is very difficult to establish.

Results

In a first set of experiments, the search behavior for the markers alone was investigated, omitting the letter recognition task. Subjects were asked to browse the bar code and to react to markers appearing in their periphery by moving their gaze toward them and pressing a button. During the first few trials of an experiment, subjects did not notice the markers, but then learned their appearance. Figure 3 shows the manual reaction times and saccadic latencies in dependence of eccentricity. The manual reaction time was about the same (ca. 400ms) across eccentricities of up to 30 degrees (top curve in upper graph), proving that the exponential compensation for the decline in visual acuity did not deteriorate performance. For a fixed marker, the decline Saccadic latencies decreased slightly with increasing eccentricity from ca. 230 to 180ms, that is, saccades are triggered faster the more peripheral the marker was. For both, the manual reaction time and the saccadic latency, the values were higher by 10 to 15 percent, when the compensation took not place (not shown). Hence, eccentricity dependent compensation did actually improve performance.

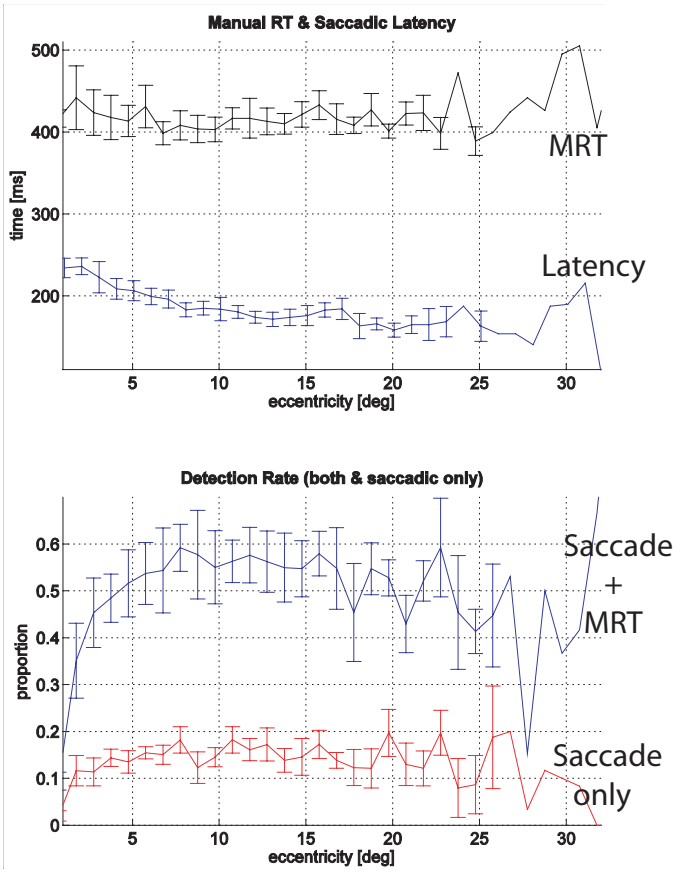

Figure 3: Visual search for markers only (no letter task involved). **Top**: manual reaction times (MRT) and saccadic latencies (Latency) in dependence of marker eccentricity. **Bottom**: detection rate for markers which captured gaze and were signaled by a button press (Saccade+MRT) and detection rate for markers to which only a saccade was made (Saccade only).

The lower graph (figure 3), shows the detection rate for markers (blue), which captured gaze and were also responded by a button press, within 400 and 1200 milliseconds as the

temporal tolerances respectively. Detection rate seems approximately constant across a range of 5 to 20 degrees with values above 50 percent, but then slightly decreased for larger eccentricities. There was also a substantial amount of saccades toward the markers which were not followed by a button press response (shown in red), in a range between 13 and 18 percent.

The luminance level for detected and not detected markers was analyzed as well: markers which were added to a lower luminance profile were less likely detected than markers added to a higher luminance profile (not shown). Some subjects even sensed that they could not properly detect markers of low-luminance level.

Figure 4 shows the landing precision of saccades. A typical saccadic landing is slightly too short of its target, which is called undershoot. The amount of undershoot increases linearly with eccentricity. Because the visual system seems to notice when it has undershot, it sometimes generates a corrective (second) saccade, which however occurred only for 10% of the saccades toward a marker.

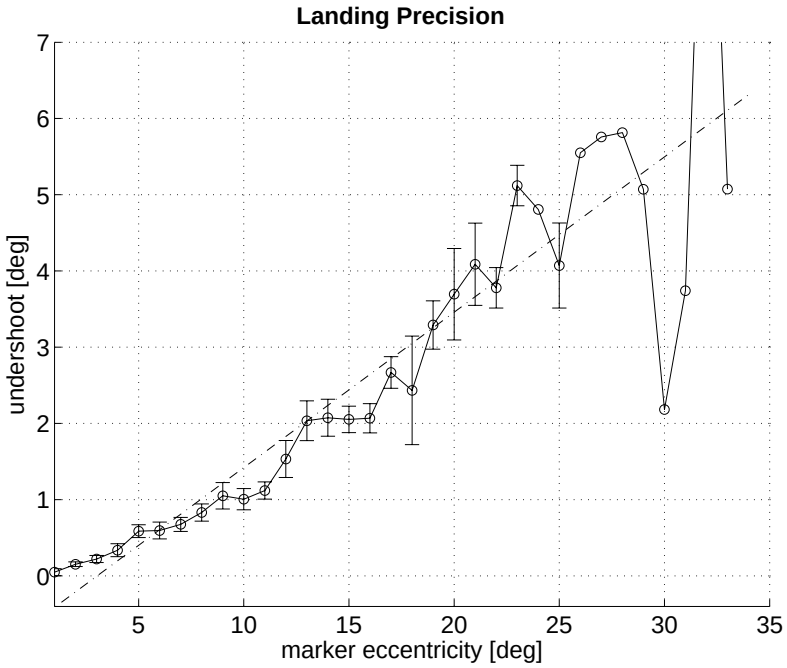

Figure 4: Saccadic landing precision for a visual marker search. Slope=0.19; intercept=-0.51; error at 15 degrees: 16%.

In a second set of experiments, the letter task was carried out with varying degrees of guidance to probe the frequency aspect. Markers appeared 550 ms before onset of a letter and lasted 500ms, leaving a temporal gap of 50ms between marker and letter to avoid potential masking effects – effects which prohibit proper recognition of the stimulus (letter). Markers appeared with different frequency per condition: 0, 25, 50, 75 and 100%. The conditions with 0% and 100% guidance represent the control conditions for which no supporting markers appeared at all (0%), or for each letter appearance one (100%).

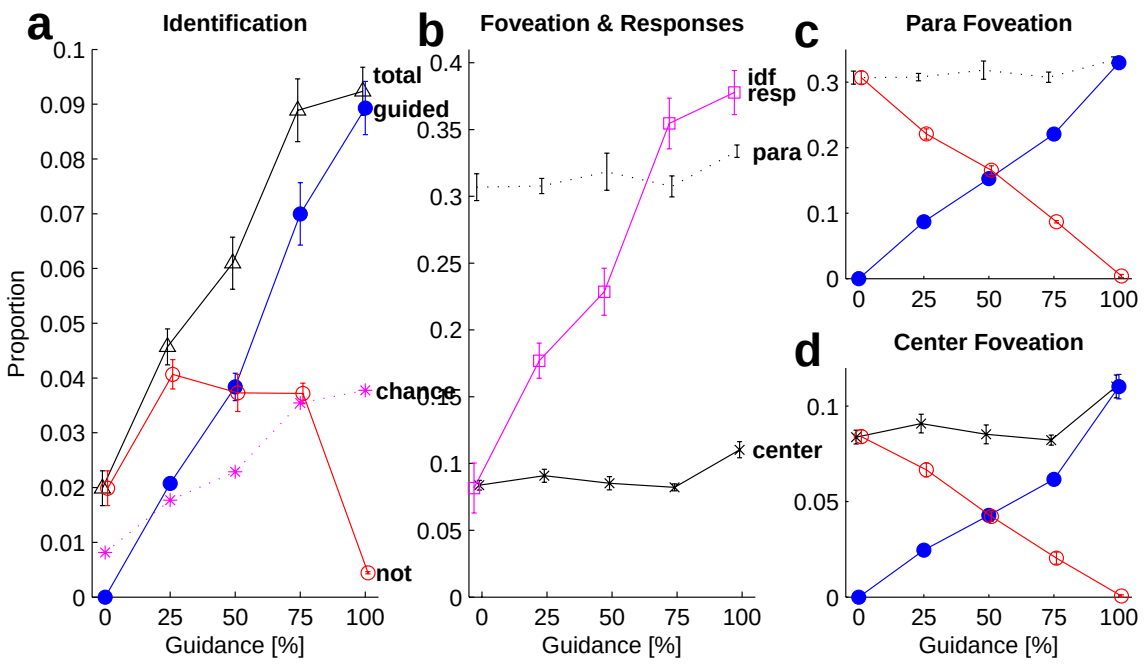

Figure 5: Letter identification and foveation in dependence of guidance (0, 25, 50, 75, 100%). **a**. Proportion of correctly identified letters. total (triangle): total identification rate; guided (filled circles): for guidance with markers only; not (guided - empty circles): for no guidance; chance (dotted): chance level of total identification rate (taken as the proportion of identification responses, see b). Error bars represent standard error of inter-subject performance. **b**. Foveated letters: proportion letters toward which the gaze moved to. identification responses (square): actual manual selections using menu. para: parafoveal foveation (fixed 5-degree radial tolerance). center: exact foveation (1-deg radial tolerance). **c**. Parafovea-foveation rate for guided and not-guided letters. **d**. Center-foveation rate for guided and not guided letters.

The identification rate for correctly selected letters steadily increased from 0.02 to 0.09 with increasing guidance, see curve with triangles labeled ‘total’ in figure 5a. The

absolute identification level is small yet irrelevant to the goal of this study, because we merely intend to prove the principle of gaze guidance for difficult recognition tasks. This increase already proofs that continuous guidance works and we now analyze the responses for guided and not-guided letter stimuli separately, as well as the foveation rates, in order to obtain a more detailed picture of the recognition dynamics during guidance.

The guided responses increase steadily as well (filled circles, figure 5a). The not-guided responses increase slightly, remain steady and then drop (empty circles, labeled ‘not’, figure 5a); the initial increase may be explained by an increased propensity to respond when guidance is present. The chance level for the total identification rate is indicated by the dashed line and is taken as the proportion of manual selections multiplied by 0.1 (the probability of randomly selecting a letter). The proportion of manual selections is the proportion of letter-identification responses using the mouse menu which quadruples from ca. 0.09 to 0.37, see figure 5b (squares). Again, this confirms that guidance facilitated letter identification, but the surprise comes when looking at the foveation rate, which is the proportion of letters to which the gaze was moved to. Two tolerances for a ‘foveation hit’ are used: a 1-degree tolerance representing the central fovea and a 5-degree tolerance representing the parafovea (central and para foveation rate respectively). For 0% guidance, the central and para foveation rate was at 0.09 and 0.32 respectively. Both increase with higher guidance but only by a few percent, which evidences that the

subject does not centrally foveate a letter to make an identification response, but that its parafoveal presence suffices. The subject must therefore perform some covert attentional shifts toward the letter stimulus to obtain an identification judgment. Because the foveation rates hardly increase, it is not clear whether the fovea is ever shifted toward a letter by a saccade (overt attentional shift), or whether the visual system simply awaits the appearance of a letter in its parafovea and uses only covert attentional shifts. To elucidate this, we plot the foveation rate for guided and not-guided letters for both the central fovea and the parafovea (figure 5c and d, respectively). In both cases, the foveation rate for guided letters increases with increasing guidance, whereas the foveation rate for not-guided letters decreases, proving that guidance also involves ouvert attentional shifts (saccades). Summarizing, guidance is indeed exploited for better letter identification, whereby the markers primarily trigger covert attentional shifts and only secondarily overt shifts. The large increase in identification rate may have several reasons: 1) without guidance, subjects foveated letters too late to identify them properly; 2) with guidance, subjects feel more compelled to make identification responses; 3) with guidance, the markers have facilitating effects on identification by the transient high-lighting of the letter location.

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.

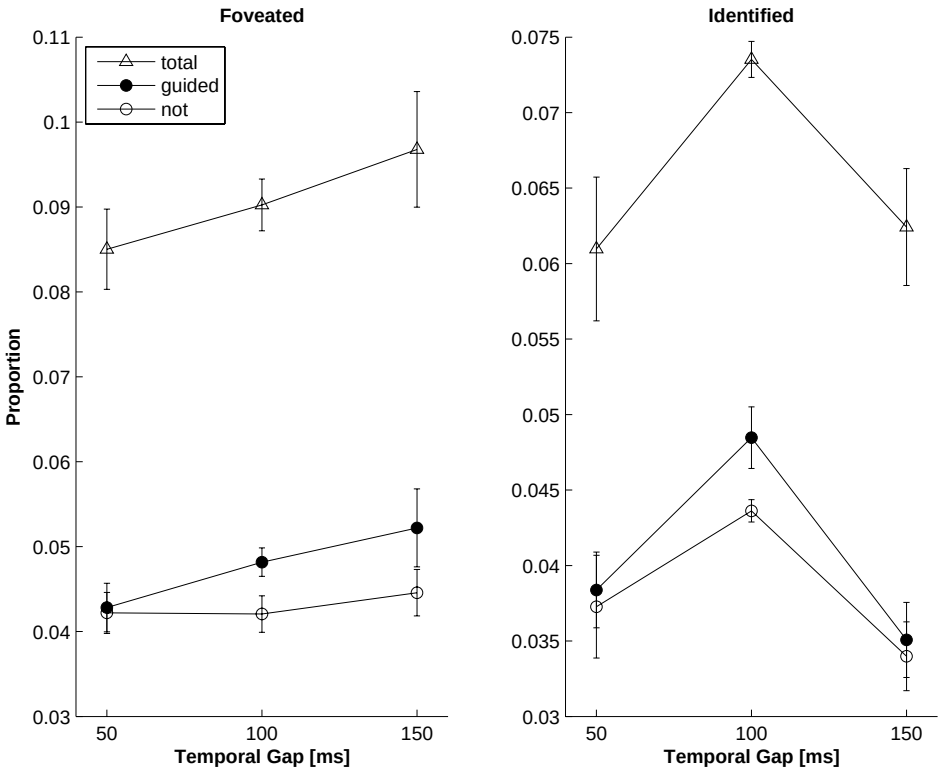

Figure 6: Letter foveation and identification rate for three different temporal gaps between marker offset and letter onset for the fixed-amplitude marker (50% guidance). **Left:** Foveated letters (total, guided, not-guided). **Right**: Identified letters.

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.

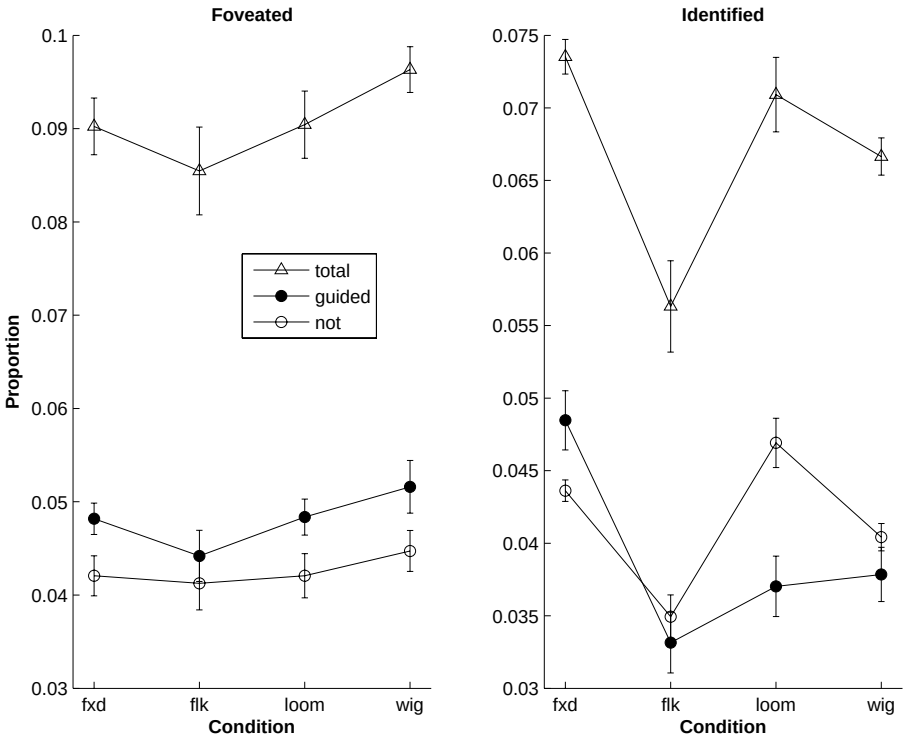

Figure 7: Letter foveation and identification rate for different markers (50% guidance; 100ms gap). **Left**: Foveated (total, guided, not-guided). *fxd*: fixed amplitude (eccentricity-dependent marker without further modification); *flk*: flickering marker (alternating amplitude); *loom*: looming marker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

Discussion

One outlined the complexity of implementing a gaze-capturing process for a gaze-guiding system by discussing a list of aspects, which essentially reflects the complexity

of the human visual recognition process per se. Several simple guidance scenarios were sketched for the PC setting, which may serve as a starting point to gather experiences for larger, more daring approaches.

Previous approaches

A specific guidance system was already tested by McNamara et al (2008). In their study, subjects 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 subjects; the obvious marker was simply more salient and was clearly noted by subjects. 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 subjects 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.

Experiences of this study

The goal of our study was to support Barth et al's quest by investigating the search behavior in dynamic noise displays bare of any structure. The noise movies had a frequency spectrum akin to natural images and provide a high degree of unspecific detail. The noise movies elicited natural eye-scanning behavior and therefore provide a natural distraction (see Rasche and Gegenfurtner). The letter task emulates a difficult recognition task, with the goal to fully engage the viewer's attention. We particularly investigated the aspects location, range, occurrence and appearance. In order to render the markers as subtle as possible (just noticeable or least-obtrusive), markers were merely added to the luminance profile of the background noise – and not just placed into it with fixed amplitude. We made the following experiences with that task:

a) The eccentricity-dependent compensation of the marker amplitude with an exponentially saturating function resulted in detection rates and manual reaction times, which remained approximately constant across eccentricities of up to 20 degrees; saccadic latencies even decreased slightly. For larger eccentricities the detection rate started to slightly decrease, whereas manual reaction times and saccadic latencies remained about the same. We think it is *not* worth improving the detection rate of those

far-eccentric markers, but rather to investigate the issue together with head movements, which are likely to occur when the markers appear in the far periphery.

b) The landing precision of eye movements toward markers linearly decreased with increasing eccentricity, a 16% error approximately, which is plotted as an increase in undershoot in figure 4. 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' or 'Site Anchor' in PC screens. This compensation could be done by placing the marker radially beyond its target and turning it off when gaze moves toward it. In the car cockpit scenario, it may be even essential to have accurate landing, but this should be investigated in combination with head movements. The measured undershoot is actually twice as large as the one measured in simple displays (Kalesnykas and Hallett, 1994; see Rasche and Gegenfurtner for an exact comparison). This indicates that the noisier the display is the more imprecise is saccadic landing. 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 simple displays, as the subjects in McNamara's study carry out a visual search, which likely involves an increase in landing variability.

c) The letter recognition task revealed that subjects did not need to place their gaze upon the letters to make identification judgments, but that parafoveation sufficed – that is, the subjects performed some attentional shifts to obtain an identification judgment (figure 5). Does this mean that the proposed undershoot compensation (item 'b') can be neglected? Possibly. It is only specific implementations which will reveal what is actually required.

d) The exact marker 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 (term $R(x_m, y_m)$ in the above equation) 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.

e) 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. 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 duration on a fixed spot.

f) Subjects did not react well to markers of very low luminance, hinting that subjects seem not to deal well with markers of varying luminance level. This may have also been the case in McNamara's study, but is difficult to analyze in their study as much fewer fixations are collected. A remedy to this may be to introduce a lower limit for the marker luminance level.

Summary of recommendations:

We summarize the specific experiences made in this study as a set of recommendations for gaze-guidance markers:

1) **Aspect range**: To compensate for the decline in peripheral acuity, the marker's amplitude is increased with eccentricity by an exponentially saturating function: $a_{mrk}(e) = a_{min} + a_{max}\text{-}exp(\text{-}e)a_{max}$ (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**: a) Motion markers are better gaze-capturing events than stationary markers, however they are potentially detrimental to recognition performance at their location.

b) If one uses a luminance marker, which is merely added to the luminance profile to make it just-noticeable, it may be necessary to set a lower bound in order to avoid the 'neglect' of very low luminance markers.

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.

Acknowledgements

We would like to thank the following members of Schölkopf department at MPI Tübingen in guidance with some computational methods: Wolf Kienzle with help of the support-vector machines; Jakob Macke with guidance of the reverse correlation technique; Matthias Franz for overview of both techniques. We 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

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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

[6] Findlay, J.M. and Gilchrist, I.D. Active Vision. New York: Oxford University Press, 2003.

[7] Franconeri, S.L. & Simons, D.J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65,* 999-1010.

[8] Groenewold, E.S., Renken, R. And Cornelissen, W. (Under Review). Attentional Window Size is Primarily Set by Environmental Relevance.

[9] 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.

[10] 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.

[11] 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.

[12] Kalesnykas, R.P. & Hallett, P.E. (1994). Retinal eccentricity and the latency of eye saccades. Vision Research, 34, 517-531.

[13] Kienzle, W., Wichmann, F. A., Schölkopf, B. & Franz, M. O. (2006). A nonparametric approach to bottom-up visual saliency. *Neural Information Processing Systems (08/2006).*

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

[15] 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.

[16] 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.

[17] Liu, A. (1998). What the driver's eye tells the car's brain. In: Underwood, G. (ed.), Eye guidance in reading and scene perception, pp. 431-52, Amsterdam: Elsevier.

[18] McNamara A, Bailey R and Grimm C (2008). Improving search task performances using subtle gaze direction. Applied Perception in Graphics and Visualization. In press.

[19] 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.

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

[21] 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.

[22] Rasche, C. and Gegenfurtner, K. (under review). Gaze Control in Dynamic Broadband (1/f) Noise Sequences. Attention, Perception & Psychophysics.

[23] Salvucci, D.D. & Anderson, J.R. (2000). Intelligent gaze-added interfaces. *CHI Letters, 2,* 273-280.

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

[25] Stenfors, I., Morén, J. & Balkenius, C. (2003). Behavioural strategies in web interaction: A view from eye-movement research. In: The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, pp. 633-644. Elsevier Science BV.

[26] 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.

[27] 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.

[28] Vig E., Dorr, M. and Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. Spatial Vision. In press.