# Visibility of temporal blur on a gaze-contingent display

Michael Dorr*
Institute for Neuro- and Bioinformatics,
University of Lübeck

Martin Böhme†
Institute for Neuro- and Bioinformatics,
University of Lübeck

Thomas Martinetz‡
Institute for Neuro- and Bioinformatics,
University of Lübeck

Erhardt Barth§
Institute for Neuro- and Bioinformatics,
University of Lübeck

## Abstract

The spatio-temporal characteristics of the human visual system vary widely across the visual field. Many studies that have investigated these characteristics were limited to the use of artificial stimuli, such as flickering sinusoidal gratings, etc. Here, we present a gaze-contingent system that is capable of modulating the spatio-temporal content of a high-resolution image sequence in real-time. In a first experiment, we measure - as a function of eccentricity - how much the temporal resolution of a natural video can be reduced before this manipulation becomes visible.

**CR Categories:** I.4.9 [Computing Methodologies]: Image Processing—Applications; J.4.1 [Computer Applications]: Social and Behavioral Sciences—Psychology

**Keywords:** gaze-contingent display, temporal sensitivity

## 1 Introduction

The spatio-temporal characteristics of the human visual system vary widely across the visual field. Spatial resolution as a function of eccentricity has been well studied both anatomically and psychophysically ([Duchowski 2000]). The fact that spatial resolution drops sharply towards the periphery has been exploited for image coding ([Itti 2004]) and low-bandwidth video communication ([Geisler and Perry 1998]). [Perry and Geisler 2002] have presented a gaze-contingent display simulating arbitrary visual fields.

The peripheral temporal characteristics of the visual system are less well understood. There are at least two distinct temporal "channels", with properties that vary with eccentricity ([Hess and Fredericksen 2002]). Research has mainly been focused on the use of well-controllable, artificial stimuli, such as flickering sinusoidal gratings, etc. While this approach has yielded many important insights, recent findings have shown that visual performance in free-viewing conditions differs significantly from that in paradigms where a steady fixation has to be held ([Gersch et al. 2004]).

In the following, we will present a gaze-contingent display capable of modulating the spatio-temporal content of a high-resolution image sequence in real-time. This allows us to selectively change the properties of a dynamic and natural visual input.

In a first experiment, we locally and selectively suppress higher temporal frequencies. The locus of the suppression is then varied so as to investigate the visibility of such changes as a function of eccentricity. Note that we do not intend to measure the maximum temporal frequency that can be detected at a given eccentricity; it is the absence of higher temporal frequencies that should go unnoticed.

Our work is motivated by applications involving the guidance of eye movements [Ita 2001]. In previous work ([Böhme et al. 2005]), we were able to predict a small number of candidate locations in a video sequence that were likely saccade targets. To increase the probability of one of these candidate locations to become the actual target for the next saccade, we plan to enhance the visual saliency there and reduce it elsewhere. One possible means of reducing saliency is "temporal blur", i.e. a suppression of high temporal frequencies as described in this paper. In a different experiment using the system presented here, we were indeed able to show that temporal transients in the periphery play an important role in the guidance of saccadic eye movements ([Dorr et al. 2005]).

## 2 System with gaze-contingent temporal resolution

### 2.1 General setup

The processing of high-resolution videos in real-time is still a computationally very challenging task. Therefore, our display system consists of two high-end computer workstations. One computer is connected to a commercially available SensoMotoric Instruments Hi-Speed eye-tracker running at 240 Hz, while the other computer (Pentium 4, 3.2 GHz) performs the image processing, using hand-written assembler code for its SSE2 vector unit.

### 2.2 Varying temporal resolution

To generate the gaze-contingent map of varying temporal resolution, we used techniques similar to those used for the real-time simulation of arbitrary visual fields ([Perry and Geisler 2002]). There, a video sequence is filtered using an arbitrary two-dimensional map that specifies the desired spatial resolution at each pixel location relative to the direction of gaze. This is achieved by blending between levels of a multi-resolution pyramid computed for each frame in the video.

Our approach is similar but uses a temporal pyramid instead of the spatial multi-resolution pyramid. The first level of the pyramid is the original input video, and each successive level is constructed by

*e-mail: dorr@inb.uni-luebeck.de
†e-mail: boehme@inb.uni-luebeck.de
‡e-mail: martinetz@inb.uni-luebeck.de
§e-mail: barth@inb.uni-luebeck.de

applying a temporal low-pass filter to the previous level, then downsampling by a factor of two (i.e., dropping every other frame). We can now obtain a version of the video with a specific temporal resolution at each pixel by interpolating between the two pyramid levels that bracket the desired resolution and upsampling to the frame rate of the original video. The interpolation scheme is described in detail in [Perry and Geisler 2002].

Note that, strictly speaking, interpolating between two pyramid levels with specific resolutions does not yield a result with intermediate resolution. A more accurate technique for approximating the continuous Gaussian scale-space is described in [Köthe 2004]. However, we believe that the simpler approach used here is sufficient for our purposes.

The resolution map we use in our experiment is set to the resolution of the original video $R_\mathrm{o}$ across the whole of the visual field except for a ring-shaped region around the fovea at a certain eccentricity, which is set to a certain reduced resolution $R_\mathrm{r}$ (because of this locally limited loss in resolution, we call this region a "temporal scotoma"). To prevent a hard border between the two regions, the flanks of the "temporal scotoma" have the shape of a Gaussian with standard deviation $\sigma$. Thus, the resolution map, expressed as a function of the eccentricity $\phi$, is

$$R(\phi) = \begin{cases} R_\mathrm{r} & |\phi - \phi_\mathrm{c}| <= \frac{w}{2} \\ R_\mathrm{o} & |\phi - \phi_\mathrm{c}| > \frac{w}{2} + 2\sigma \\ R_\mathrm{r} + (R_\mathrm{o} - R_\mathrm{r})\; G_{\sigma,\mathrm{inv}}(|\phi - \phi_\mathrm{c}| - \frac{w}{2}) & \text{otherwise,} \end{cases}$$

where $\phi_\mathrm{c}$ is the eccentricity of the center of the ring, $w$ is the width of the ring, and $G_{\sigma,\mathrm{inv}}$ is an inverted Gaussian with standard deviation $\sigma$.

## 3 Visibility of temporal blur

We measured the visibility of a reduction in temporal resolution with varying eccentricity. Four image sequences of natural outdoor scenes were taken with a JVC JY-HD10 camera (HDTV standard, 1280x720 pixels spatial and 30 Hz temporal resolution). In order to be able to process the sequences in real-time, they were downsampled to a spatial resolution of 960x540 pixels with the publicly available mencoder ([Mpl ]) software. Sequences were cut to a duration of 20 seconds each; they depicted scenes from a sunny day at the beach, a roundabout, and two shots of pedestrian areas (see Fig. 1). The sequences contained a variety of objects of different size moving at different velocities.

Subjects were seated 55 cm from an Iiyama HM204DT 22" computer screen, running at a spatial resolution of 1280x960 pixels and a refresh rate of 90 Hz. Because the aspect ratios of the screen and the videos differed, black bars were added above and below the video so that the video covered a field of view of 40.0x22.5 degrees (45.9 degrees diagonal). Thus, the maximum horizontal spatial frequency was 12 cycles/degree. Eye movements were measured at 240 Hz and the average latency of the system was 60 ms. This latency may seem quite high compared to the latencies of gaze-contingent displays that alter single images only, but note that building a temporal multiresolution pyramid is by far costlier. Ultimately, every video frame that is displayed is a weighted sum of more than 250 high-resolution frames surrounding it. Latency could be reduced further by sacrificing spatial resolution, but we judged this to be unnecessary as subjects did not report a noticeable lag in the current setup.

Before data collection began, subjects performed a few demo trials of the experiment to familiarize themselves with the visual impression caused by temporal blur. In these trials, the position of the temporal scotoma was controlled by the mouse.

Detection thresholds were then determined with an interleaved staircase procedure. Staircase procedures are commonly used to estimate thresholds in psychophysical 2AFC (two alternatives, forced choice) tasks. After each stimulus presentation, the subject had to respond whether the stimulus was detected, i.e., whether the reduction in temporal resolution led to a visible degradation of image quality or not. If the reduction went undetected, the temporal resolution was decreased even further in the next presentation; similarly, after a detected reduction the resolution was increased again. The step size for these changes was 0.75 Hz. After a predefined number of turns, that is, switches in response from "detected" to "undetected" or vice versa, the staircase procedure ended. The average response at the turns then served as a reliable estimate for the detection threshold.

We measured thresholds for a scotoma width of 2.3 degrees, a flank width of $2\sigma = 0.9$ degrees, and five different eccentricities, namely 1.2 degrees (a foveally presented circular disk), 10.3; 19.6; 28.7; and 37.9 degrees. For each of these eccentricities and each of the four image sequences, a separate staircase was held in memory. Before each trial, one staircase was chosen randomly so that the subject could not know beforehand where in the visual field the "temporal scotoma" would occur.

## 4 Results

Figure 2 shows results for two different subjects. Clearly, the sensitivity to high temporal frequencies drops towards the periphery. The variance across different image sequences can be explained by the different distribution of moving objects on the image plane; i.e., in the roundabout scene, most moving objects are in the center of the image, so that at higher eccentricities, there are almost no high temporal frequencies to suuppress anyway.

It is therefore important to note that we do not claim the function in Figure 2 to be generally applicable. It is more of a heuristic that gives a rough approximation of the sensitivity to temporal blur in natural scenes. We plan to use the domain of "invisible" frequencies above the thresholds in Figure 2 to guide eye movements in a subconscious way.

Of course, more data need to be collected, especially with regard to the width of the "scotoma" that was arbitrarily set to 2.3 degrees here. Also, because of the limited display size, we were limited to eccentricities of up to 38 degrees, while the human visual system has a field of view of about 180 degrees.

## 5 Conclusion

We have presented a gaze-contingent system capable of simulating an arbitrary space-variant temporal resolution map. We believe this system is a useful tool for psychophysical research that will allow us to perform experiments on the spatio-temporal characteristics of the human visual system with natural scenes.

In a first experiment, we have shown that it is possible to invisibly suppress a wide range of temporal frequencies in the visual periphery. We plan to use this property to guide eye movements, in an attempt at more efficient image communication ([Ita 2001]).

Figure 1: Stills from the movies used in the experiment. From top to bottom: “beach”; “roundabout”; “street1”; “street2”.

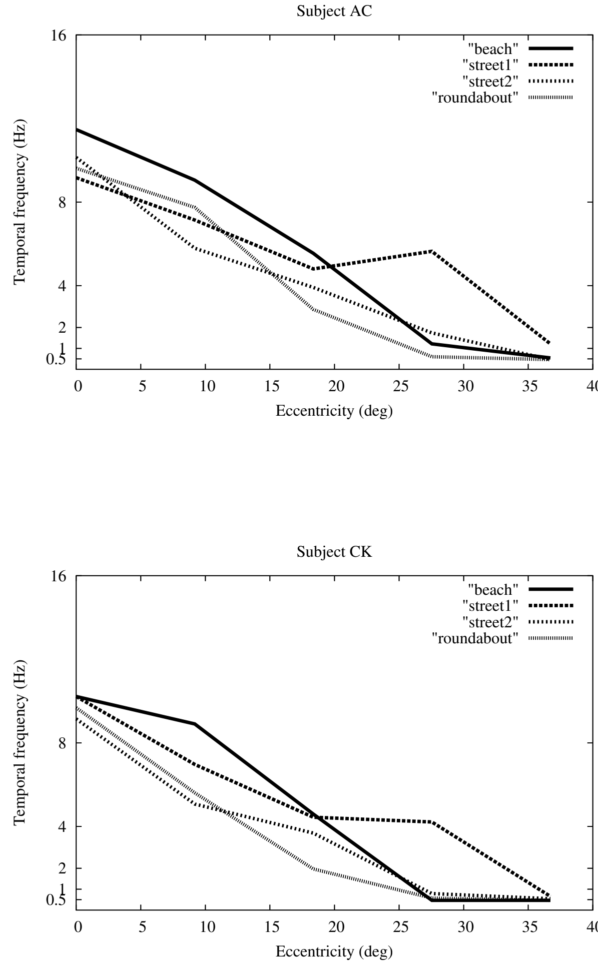


Figure 2: Detection threshold for temporal blur as a function of eccentricity for two different subjects. Each line indicates results for one image sequence.

## 6 Acknowledgements

Research is supported by the German Ministry of Education and Research (BMBF) under grant number 01IBC01B, ModKog.

## References

BÖHME, M., DORR, M., KRAUSE, C., MARTINETZ, T., AND BARTH, E. 2005. Eye Movement Predictions on Natural Videos. Submitted to Journal of Neurocomputing.

DORR, M., BÖHME, M., MARTINETZ, T., GEGENFURTNER, K. R., AND BARTH, E. 2005. Eye movements on a display with gaze-contingent temporal resolution. Submitted to: European Conference on Visual Perception, A Coruña, Spain.

DUCHOWSKI, A. T. 2000. Acuity-matching resolution degradation through wavelet coefficient scaling. *IEEE Transactions on Image Processing 9*, 8, 1437–40.

GEISLER, W. S., AND PERRY, J. S. 1998. A real-time foveated multiresolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging: SPIE Proceedings*, B. Rogowitz and T. Pappas, Eds. 294–305.

GERSCH, T. M., KOWLER, E., AND DOSHER, B. 2004. Dynamic allocation of visual attention during the execution of sequences of saccades. *Vision Research 44*, 12, 1469–83.

HESS, R. F., AND FREDERICKSEN, R. E. 2002. Temporal detection in human vision: dependence on eccentricity. *Ophthal. Physiol. Opt. 22*, 92–102.

2001. Information technology for active perception. http://www.inb.uni-luebeck.de/Itap/.

ITTI, L. 2004. Automatic attention-based prioritization of unconstrained video for compression. In *Proc. SPIE Human Vision and Electronic Imaging IX (HVEI04), San Jose, CA*, B. Rogowitz and T. N. Pappas, Eds., vol. 5292, 272–283.

KÖTHE, U. 2004. Accurate and Efficient Approximation of the Continuous Gaussian Scale-Space. In *26th DAGM-Symposium*, Springer, Heidelberg, C. E. Rasmussen, H. H. Buelthoff, M. Giese, and B. Schoelkopf, Eds., vol. 3175 of *LNCS*, 350–358.

MPlayer suite. http://www.mplayerhq.hu.

PERRY, J. S., AND GEISLER, W. S. 2002. Gaze-contingent realtime simulation of arbitrary visual fields. In *Human Vision and Electronic Imaging: Proceedings of SPIE, San Jose, CA*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 4662, 57–69.