A new masking technique for natural scenes reveals the saliency of an image

Claudia Wilimzig\textsuperscript{1}, Rufin VanRullen\textsuperscript{2}, and Christof Koch\textsuperscript{1}

\textsuperscript{1}California Institute of Technology, Pasadena, CA \textsuperscript{2}CNRS Toulouse, France

Masking is a key experimental tool to precisely control the visibility of visual stimuli. We explore a new technique for masking images, in particular natural scenes: briefly flashing (e.g., 10 ms) a natural scene followed by a brief flash of a negative version of the same scene. The stimulus and its negative version are inversely related like a photo and its negative and can be obtained by subtracting the stimulus from the maximum palette entry in each respective color channel.

This technique is related to experiments showing that for two differently colored lights flashed subsequently, people report a blended version of both stimuli as if the process underlying perception integrates over both stimuli \cite{1}. If the visual system perfectly averaged over stimulus and mask, subjects would report seeing a uniform gray patch. Contrariwise, in some areas of the picture subjects reported deviations from mere gray perception. We show (1.) that these areas are systematically related to predictions of computational approaches to saliency \cite{2}, e.g. such that our masking technique masks everything but the most salient regions of the image; (2.) that this effect cannot be achieved by using standard masking techniques or no masking; (3.) that this masking effect is weakened but not abolished in a dichoptic version of the experiment showing that retinal mechanisms may contribute but are not sufficient to explain this effect.

These results have significant implications for the neuronal coding of saliency and for computational approaches to saliency. In particular, we show that a simple thresholded wavelet transform, computing local contrast intensities with positive and negative values corresponding to two different polarities (ON- and OFF-center cells), corresponds to people’s percepts. Given the similarities between percepts and saliency maps, this may provide an alternative, and less costly, tool for computational approaches to saliency.

Acknowledgments
We thank W. Bair and R. Peters for helpful discussions. This work was supported by the ONR, NGA, and the Alexander von Humboldt foundation.

References