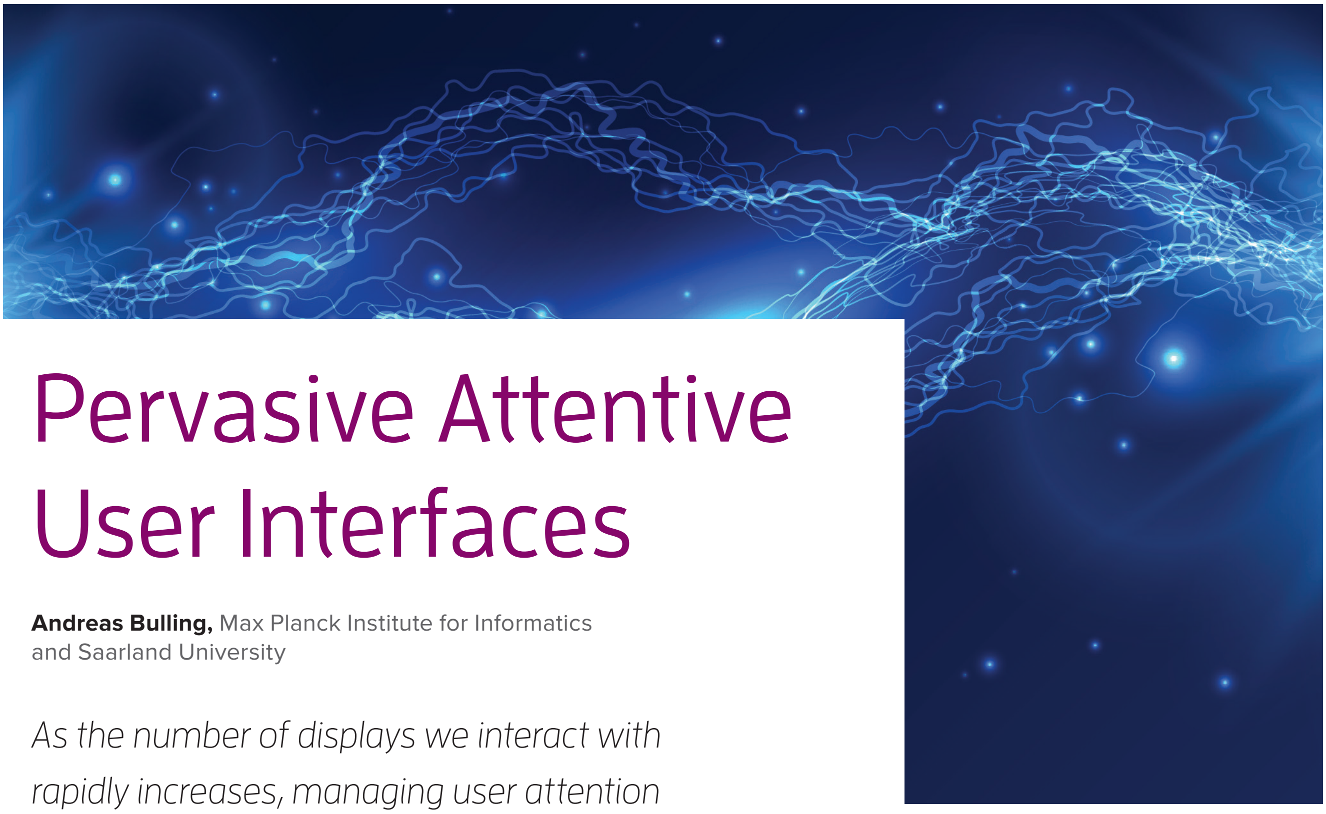


Pervasive Attentive User Interfaces

Andreas Bulling, Max Planck Institute for Informatics and Saarland University

As the number of displays we interact with rapidly increases, managing user attention has emerged as a critical challenge for next-generation human–computer interfaces.

Ten years ago, Thomas Friedman argued that society had left the Information Age and entered the Age of Interruption.[1] At that time, he was referring to interruptions caused by humans, for example through instant messages, emails, or cellphone rings. In recent years, with technology entering every corner of our lives, we've become deluged with interruptions: from notifications on PCs, TVs, and laptops; to push messages on smartphones, tablets, and wearables such as smartwatches and fitness trackers; to ads on public displays.

The frequency of interruptions will only increase as displays become even more ubiquitous. In 2007, people living in cities were already seeing an estimated 5,000 ads per day.[2] By 2020, totaling the projected number of digital signs, wearables, desktop computers, and mobile devices, there could be as many as 9.7 billion displays. This corresponds to an average of eight displays per person, a 25 percent increase from 2015. And this estimate doesn't even include other types of displays such as those in cars and household appliances.

CONTINUOUS PARTIAL ATTENTION

In daily life we constantly shift our attention among various tasks, filtering relevant from irrelevant information as well as processing and acting upon new information. Interruptions by displays further reduce our attention span—that is, the amount of concentrated time we can spend on any single task without getting distracted by other tasks. Consequently, sustained attention is increasingly being replaced by *continuous partial attention*: the act of paying simultaneous attention to multiple sources of information but only at a superficial level. In economics, this well-known phenomenon has led to the "attention economy" theory that acknowledges both the scarcity and superficiality of consumer attention and, consequently, the importance of managing it.[3]

Continuous partial attention fundamentally limits users' ability to efficiently interact with computing systems, as well as these systems' ability to support their users, yet

 0018-9162/16/$33.00 © 2016 IEEE

it remains relatively unexplored in the human–computer interaction (HCI) literature. Here, I argue that managing user attention, and thereby turning continuous partial attention into sustained attention, is one of the most pressing but also difficult HCI challenges. Given that users' sustained attention can be interrupted during all explicit and implicit interactions with computing systems, the HCI community should strive to develop computational methods to estimate and analyze the visual attention of a potentially large number of users, unobtrusively and continuously over long periods of time in their everyday life, as well as user interfaces that leverage attention information.

FROM ATTENTIVE TO PERVASIVE ATTENTIVE USER INTERFACES

As Figure 1 shows, traditional user interfaces deliver information with a minimum of subtlety—they don't consider the amount or type of information being presented or the user's attentional capacity. Because users can only simultaneously process a limited number of competing sources of information, overall information throughput is low.

Current *attentive user interfaces* operate along a spectrum with respect to information throughput and sublety. At one end of this spectrum are displays that subtly present information on the periphery, without demanding that users shift their attention; these peripheral displays require low mental effort to process information but at the cost of minimum throughput. At the other end of the spectrum are gaze-contingent displays that more obtrusively present information at users' focus of attention, achieving maximum information throughput but requiring more mental effort to process that information.

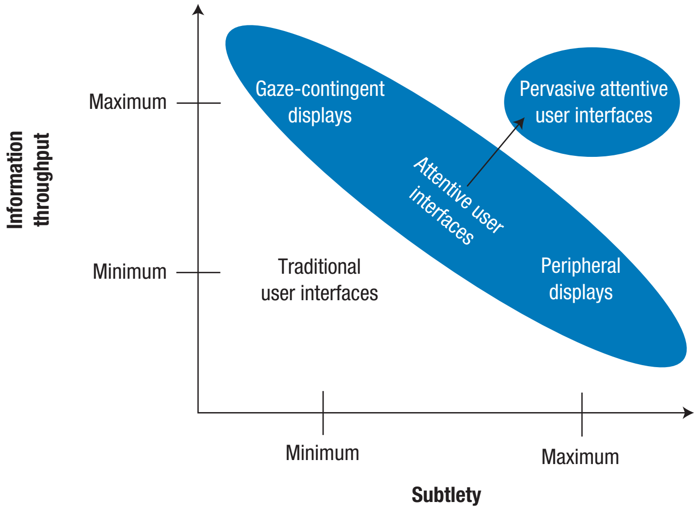


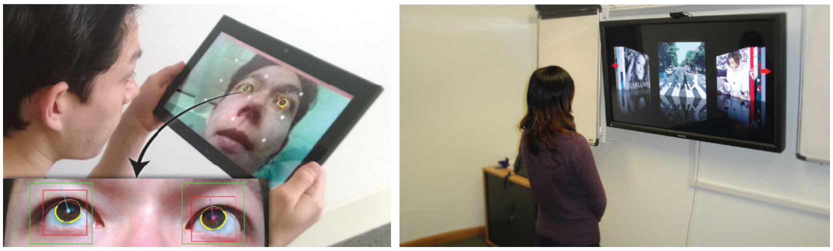
Figure 1. Traditional user interfaces deliver information with a minimum of subtlety and, due to users' limited attentional capacity, must keep throughput low. Current attentive user interfaces—with peripheral displays at one end of the spectrum and gaze-contingent displays at the other end—must trade off information throughput with the subtlety of information delivery. Future pervasive attentive user interfaces will continuously manage users' attention in daily life, simultaneously optimizing for both information throughput and subtlety.

Future *pervasive attentive user interfaces* could manage user attention—perhaps as an "attention account" that, like a bank account, maintains a balance of available attention. Drawing from this account, displays could then dynamically adapt the amount and type of information presented to users based on their current attentional capacity, thereby simultaneously optimizing for information throughput and sublety. In addition, instead of interrupting the user whenever new information becomes available, future interfaces could trade off information importance with users' current interruptibility level and time the delivery of information appropriately—for example for a period of low cognitive load, free attentional capacity, or even boredom.[4]

Attentive user interfaces

What exactly is attention, and how can it be estimated and analyzed to facilitate such optimizations? One definition of attention widely used in psychology, cognitive science, and human vision research describes it as the process of concentrating on a discrete aspect of information while ignoring other perceivable information.

Attention is typically further differentiated into covert and overt as well as top-down and bottom-up.[5] *Covert attention* refers to the act of directing one's mental focus to some information. Estimating covert attention is challenging given that it's encoded in the brain's complex neural dynamics and therefore requires sophisticated technology, such as functional MRI or electroencephalography. In contrast, *overt attention* involves unconsciously turning one's head or shifting one's gaze in the direction of interest and can therefore be observed externally. *Top-down attention* is engaged when executing tasks consciously, whereas



(a) (b)

Figure 2. Advances in computer vision make it possible to robustly estimate user attention on (a) mobile devices such as tablets and laptops using their integrated front-facing cameras and on (b) public displays equipped with wide-angle cameras.

bottom-up attention is a reaction to visual stimuli, such as a blinking light, moving object, or loud sound.[6]

As indicated earlier, we must divide our attention when interacting with multiple computing systems and often struggle to maintain our focus and concentration when faced with competing distractions. This fundamental limitation inspired the development of attentive user interfaces that adapt to users' current attentional focus and capacity. Given the challenges involved in estimating covert attention, most such interfaces focus on estimating overt attention and use eye tracking as the core measurement technique. For example, IBM's prototype SUITOR (Simple User Interest Tracker) system monitored users' gaze and typing behavior during Web browsing and showed potentially relevant information in a subtle ticker at the bottom of the display.[7]

Pervasive attentive user interfaces

While the first generation of attentive user interfaces was limited by major shortcomings in eye-tracking technology, the last couple of years have seen renewed interest in pervasive eye tracking and mobile attention measurement. These activities are partly driven by significant price decreases in commercial stationary and head-mounted eye trackers, which cost only a few hundred euros. Simultaneously, computer vision methods for head pose and gaze estimation using monocular cameras are continuously improving in robustness and accuracy.

These advances point toward a new generation of pervasive attentive user interfaces characterized by six key properties:

- *Unobtrusiveness*. Further miniaturization will make it possible to embed attention-sensing capabilities unobtrusively into eye glasses, head-up displays, handheld devices, everyday objects, as well as ambient systems, making them attentive to the user.
- *Accuracy*. In contrast to existing binary eye contact or coarse on–off screen attention detection, new interfaces will provide accurate visual attention estimates that can, for example, be used to generate fine-grained visual attention maps.
- *Large scale*. While most current interfaces only support attention measurement of individuals, pervasive attentive user interfaces will enable real-time measurement of collective visual attention—that is, attention dynamics of large groups of users.
- *Long-livedness*. Attention measurements won't be limited to dedicated short-term recording sessions but will be conducted over long periods of time in everyday life, thereby forming a holistic, spatiotemporal record of users' visual attention.
- *Seamlessness*. To facilitate large-scale and continuous attention measurements in daily life, next-generation attentive user interfaces will seamlessly switch among multiple sensors to aggregate attention information across numerous users and displays.
- *Context awareness*. Attention measurements will be contextualized by users' current situation and activities—for example, by combining eye tracking with inertial sensors and GPS readily integrated into mobile phones and smartwatches.

These properties are key to addressing the challenges of attention management and continuous partial attention in a multi-billion-display world. However, for pervasive attentive user interfaces to emerge, researchers must still overcome several obstacles, most notably in estimating and modeling attentive behavior.

ATTENTION ESTIMATION

Overt visual attention can be estimated by mapping eyeball rotations, typically inferred from images of users' eyes, to gaze positions in a reference coordinate system such as on a display or in a 3D environment. State-of-the-art gaze-estimation approaches rely on hybrid feature- and model-based methods and require special-purpose hardware, such as infrared lights and stereo cameras, to track users' heads and eyes.[8]

These approaches can achieve high estimation accuracy in controlled settings but are sensitive to changes in lighting conditions, severely limiting their use in mobile everyday settings. In addition, current gaze-estimation systems are typically not interconnected, which prevents seamless measurements across multiple displays or users. While remote systems require the user to stay within a defined tracking box, head-mounted systems require augmenting each individual with cumbersome equipment.

Appearance-based gaze-estimation methods could address some of these

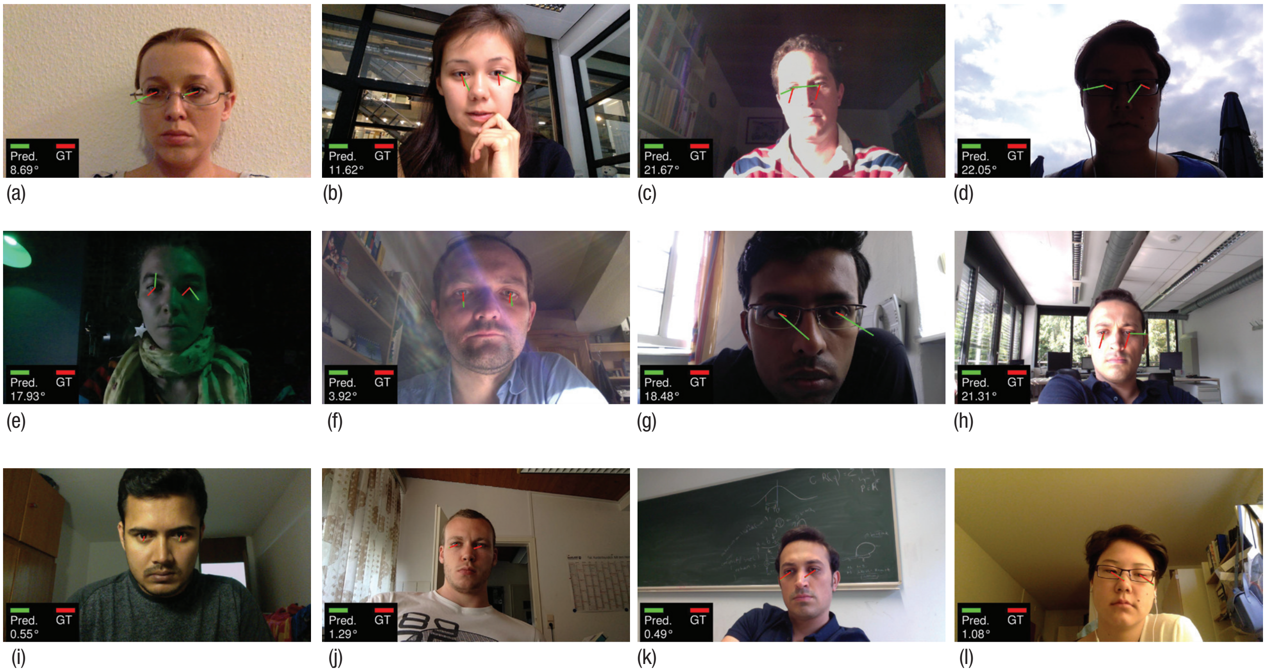


Figure 3. Estimating user attention in unconstrained everyday settings faces several challenges including (a, b) considerable variability in user and background appearance, (c) directional lighting, (d–f) shadows and glare, and (g, h) users' head pose and distance to the camera. Images j–l show best-case examples of state-of-the-art gaze-estimation accuracy in degrees of visual angle as well as predicted (Pred.) and ground-truth (GT) gaze vectors.

limitations. Such methods only require a single off-the-shelf monocular RGB camera and rely on machine learning to directly map eye appearance to gaze directions. As Figure 2 shows, robust attention measurements could be performed on the billions of camera-equipped mobile devices[9] and public displays [10] already in use today.[10]

Current research efforts focus on porting appearance-based gaze-estimation methods from the laboratory to real-world settings.[11] From a computer vision perspective, these settings are characterized by considerable variability in terms of user and background appearance, directional lighting, shadows and glare, and users' head pose and distance to the camera, as Figure 3 shows. State-of-the-art gaze-estimation methods achieve accuracies of about six degrees of visual angle in the practically most useful—but also most challenging—user- and device-independent gaze-estimation task.

However, this accuracy is still rather far from the one degree of visual angle achieved by established model-based gaze-estimation methods. Closing this performance gap will require further advances in computer vision and machine learning methods that are robust to the large variability of real-world settings. Other remaining research challenges are to seamlessly estimate attention from multiple systems, whether placed in the environment or worn on the body; to aggregate this information into a holistic, spatiotemporal record of user attention; and to estimate attention at scale—that is, jointly for several individuals or even large groups of people.

ATTENTION MODELING

Modeling shifts in overt visual attention over time, or visual behavior, directly builds on attention estimation. Visual behavior analysis is extensively studied in experimental psychology and the behavioral sciences—and more recently also in HCI. Previous work has shown that visual behavior is a rich source of information about users, such as their activities or daily routines.[12] Moreover, the link between visual behavior and cognition promises automatic analysis of covert aspects of user state that are closely related to attention, such as cognitive load.[13]

Despite significant advances in analyzing and understanding human visual behavior, most studies have focused on short-term visual behavior. To realize the vision of pervasive attentive user interfaces, computational methods that analyze visual behavior over long periods of time in daily life, such as days or even weeks, are needed.

One of the biggest challenges for long-term attention modeling using head-mounted eye trackers is *calibration drift*: deterioration of gaze-estimation accuracy over time. Such drift can be severe, even for hour-long recordings, and completely unpredictable—for example, if users temporarily remove the tracker or move it on their head.

Another key challenge is to contextualize attentive behavior by considering users' current overall situation, activities, and goals. This requires extending existing user models with models of the environment, available systems, and interactions among multiple users. Although attention is a core aspect of user modeling, it has yet to be examined in unconstrained everyday settings.[14] Researchers have started to explore these areas, but achieving the ultimate aim of robust and multimodal analysis of long-term visual attention will take many years.

As the number of displays we interact with—on our own devices as well as in the environment—rapidly increases, managing user attention has emerged as a critical challenge for next-generation human–computer interfaces. Addressing the core issues of interruptions and continuous partial attention requires new computational methods to unobtrusively estimate and analyze the visual attention of large numbers of users over long periods of time in everyday settings across a multitude of diverse body-worn and ambient displays. If achieved, these pervasive attentive user interfaces will open up exciting new opportunities to not only optimize for user performance, but also interface usability as well as user experience. ■

REFERENCES

1. T. Friedman, "The Age of Interruption," *The New York Times*, 5 July 2006.
2. L. Story, "Anywhere the Eye Can See, It's Likely to See an Ad," *The New York Times*, 15 Jan. 2007.
3. T.H. Davenport and J.C. Beck, *The Attention Economy: Understanding the New Currency of Business*, Harvard Business Rev. Press, 2001.
4. M. Pielot et al., "When Attention Is Not Scarce—Detecting Boredom from Mobile Phone Usage," *Proc. ACM Int'l Joint Conf. Pervasive and Ubiquitous Computing* (UbiComp 15), 2015, pp. 825–836.
5. R. Desimone and J. Duncan, "Neural Mechanisms of Selective Visual Attention," *Ann. Rev. Neuroscience*, vol. 18, 1995, pp. 193–222.
6. E.I. Knudsen, "Fundamental Components of Attention," *Ann. Rev. Neuroscience*, vol. 30, 2007, pp. 57–78.
7. P.P. Maglio et al., "SUITOR: An Attentive Information System," *Proc. 5th Int'l Conf. Intelligent User Interfaces* (IUI 00), 2000, pp. 169–176.
8. D.W. Hansen and Q. Ji, "In the Eye of the Beholder: A Survey of Models for Eyes and Gaze," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, 2010, pp. 478–500.
9. E. Wood and A. Bulling, "EyeTab: Model-Based Gaze Estimation on Unmodified Tablet Computers," *Proc. Symp. Eye Tracking Research and Applications* (ETRA 14), 2014, pp. 207–210.
10. Y. Zhang, A. Bulling, and H. Gellersen, "SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI 13), 2013, pp. 851–860.
11. X. Zhang et al., "Appearance-Based Gaze Estimation in the Wild," *Proc. Conf. Computer Vision and Pattern Recognition* (CVPR 15), 2015, pp. 4511–4520.
12. A. Bulling, C. Weichel, and H. Gellersen, "EyeContext: Recognition of High-Level Contextual Cues from Human Visual Behaviour," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI 13), 2013, pp. 305–308.
13. S.P. Marshall, "The Index of Cognitive Activity: Measuring Cognitive Workload," *Proc. IEEE 7th Conf. Human Factors and Power Plants*, 2002; doi:10.1109/HFPP.2002.1042860.
14. A. Ferscha, "Attention, Please!," *IEEE Pervasive Computing*, vol. 13, no. 1, 2014, pp. 48–54.

ANDREAS BULLING is head of the Perceptual User Interfaces Group at the Max Planck Institute for Informatics as well as the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University, Germany. Contact him at bulling@mpi-inf.mpg.de.