Attention approximation: from the Web to multi-screen television.

Caroline Jay
School of Computer Science
University of Manchester
Kilburn Building, Oxford Road
Manchester, M13 9PL. UK
caroline.jay@manchester.ac.uk

Simon Harper
School of Computer Science
University of Manchester
Kilburn Building, Oxford Road
Manchester, M13 9PL. UK
simon.harper@manchester.ac.uk

Andy Brown
School of Computer Science
University of Manchester
Kilburn Building, Oxford Road
Manchester, M13 9PL. UK
andrew.brown@cs.manchester.ac.uk

Maxine Glancy
BBC R&D
Salford Quays
Manchester, UK
maxine.glancy@bbc.co.uk

Mike Armstrong
BBC R&D
Salford Quays
Manchester, UK
mike.armstrong@bbc.co.uk

Proceedings of TVUX-2013: Workshop on Exploring and Enhancing the User Experience for TV at ACM CHI 2013, 27 April 2013, Paris, France.

Abstract
The move towards the provision of television content over two or more screens represents an enormous opportunity and a considerable challenge. A scientific understanding of how people switch attention between screens during television viewing is key to the development of this technology. We describe how 'attention approximation', a technique we have used to model visual attention and design screen reader presentation of Web content, can be used to investigate viewing behaviour, and ultimately drive the provision of content across multiple screens.

Author Keywords
Eye-tracking, Web, TV, User Interfaces, HCI, Attention Approximation

ACM Classification Keywords
H.5.1 [Information interfaces and presentation (e.g., HCI)]: Multimedia Information Systems.; H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces.

Introduction
Successfully providing television content over two or more screens represents a two-fold technological challenge. Assembling an infrastructure that has the ability to provide secondary content seamlessly is an important first

step, but does not in itself guarantee to add value. To ensure that content provided via additional devices genuinely enhances the viewing experience, we require a scientific understanding of both *what* information viewers want to see, and *when* they want to see it.

Here we focus on the second question, considering how an approach used to model the allocation of visual attention to dynamic Web content can be extended to incorporate further stimuli; in particular television content presented across two or more screens. We encapsulate this approach in the term **attention approximation**, which describes the process of determining a user's locus of attention at varying levels of detail, using a variety of tools.

Attention approximation occurs in two dimensions: space and time. The extent to which we can approximate *where* attention is located at a given moment, or *when* attention is allocated to a particular location, varies in granularity. For example, using an eye-tracker with a standard monitor allows us to specify the location of overt visual attention with a high level of precision: fine-grained approximation is possible in this case. Determining the location of attention on a tablet or television—a situation where high-fidelity eye tracking is not currently possible—would occur at a coarser granularity: for example, is someone looking at the screen or the tablet? Attention approximation involves matching the spatial or temporal resolution with the constraints and demands of the data required, and environment within which it needs to be collected.

Attention Approximation on the Web

The SASWAT project[1] examined how people switch their attention between multiple streams of content on Web pages, from tickers and animations, to hover-over menus, slideshows and suggestion lists. Our motivation was to understand how this dynamic content could be presented to people who access the Web non-visually, in particular blind or visually impaired people who interact with a computer using a screen reader. Screen readers convert content into synthetic speech whilst enabling a level of control over the flow of information, essentially providing linear access to the non-linear, dynamic, resource.

We started by investigating how sighted users interacted with dynamic content. Our aim was to approximate the distribution of attention to Web content as a function of its characteristics. Tracking the eye movements of sighted users whilst they viewed and interacted with a wide range of content gave us a detailed understanding of how people respond to dynamic content in a variety of situations, and a model of the likelihood that someone will view this content as a function of its properties [2]. These studies showed that rather than splitting attention between different sources of updating information, people focused primarily on updates they had initiated, and that the likelihood an update was viewed was affected by predictable characteristics of both the content (its size, duration and the way it was triggered) and the user's activity (determined through the timing of mouse movements and keystrokes).

The rich understanding of user behaviour we obtained from these studies gave real insight into how people used dynamic content, and the advantages it brought them; we were then able to design an audio interface that replicated those benefits in a serial information stream [1]. The translation technique, refined through iterative user testing, was demonstrated in a double-blind evaluation with visually disabled users to offer greatly improved

[1] http://wel.cs.manchester.ac.uk/research/saswat/

screen reader access to dynamic content.

Gathering the data to model fine-grained visual attention on a Web page demanded the precision provided by eye-tracking equipment. Effective audio presentation of the content also requires knowledge of the user's locus of attention, but the detail provided by eye tracking is neither possible nor necessary; in this case it was approximated from the location of the user's focus on the Web page, combined with the predictive model resulting from the eye tracking data. We believe that this fundamental approach — approximating the user's attention at an appropriate level of granularity to guide stimuli presentation — can play a key role in determining how to present television content across multiple screens.

Attention Approximation for Television

How can attention approximation contribute to improving television content provision? Although the human visual system is often considered to be parallel, in reality it is linear, and this is more obvious when attention is shifting between two devices. Understanding which factors influence shifts of attention could allow us to design supplementary content that is more likely to be used, and likely to be used more effectively, while distracting viewers from the main program less.

Monitoring attention during television viewing has been performed before, but not to identify when people switch their attention between content on multiple devices. Determining how to model these shifts first requires us to understand the quality of data (acquired through a range of techniques, including eye-tracking) required and the constraints the experiments must be performed under, both of which will inform the granularity and form of attention approximation that is needed.

Challenges

Although we have demonstrated that rigorously studying and modelling users aids the design of effective user interfaces, it is clear that numerous challenges must be met in order to apply these methods to the domain of television, and particularly television with a second device. To model viewing behaviour through attention approximation we identify the main challenges as follows:

Content While the information to be presented on the second screen may have similarities to Web content, the primary information stream is anticipated to be the television. This differs from the Web in several ways: it is less interactive; it involves video or film rather than text; it uses audio as a key source of information. Attention whilst viewing will thus be governed by very different factors. Audio, for example, may have as large or greater an effect on attention than visuals.

Environment To ensure ecological validity, user studies need to be performed in an environment that closely matches the 'real world'. In the case of the Web, it is not unreasonable to have a usability lab arranged like a typical office — an arrangement that is easy to achieve with the eye-tracking equipment used. For television, however, a more natural setting would mirror the home. This is less easy to achieve, although the BBC usability lab in Salford provides such a setting. As important as the physical environment, however, is the social environment, with anecdotal evidence that much of the current second device usage involves social interaction such as monitoring or posting to online social networks and sending SMS messages. Other people in the physical environment may also influence viewers' attention.

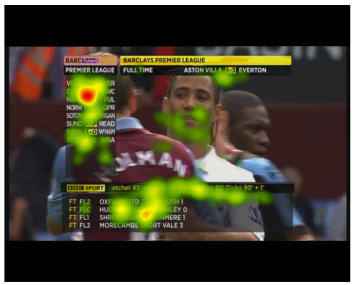

Figure 1: A 'heat map' showing the distribution of a viewer's visual attention whilst watching Final Score.

Methodology Eye-tracking users interacting with a Web browser on a computer monitor lies well within the intended use of the equipment; we are therefore able to collect well-calibrated, high quality data with a good level of precision. Tracking gaze on a television, however, with the less constrained environment of the living room is going to be more difficult, and accurately tracking a person's gaze over two devices harder still. It will also be important to consider attention in other modalities, particularly the auditory domain.

These challenges are not insurmountable, however. The type of information required to approximate attention over multiple devices need not include high-resolution eye-tracking data. Low resolution attention monitoring, such as determining which quadrant of a screen is being attended, or even just recording when attention shifts between devices, can provide empirical evidence that can be used to understand how to coordinate dual information streams. Taking other modalities, particularly the auditory domain, into account, and building on existing knowledge of human perception and cognition, will also help to inform our method. Thus, collection of useful data can run in parallel with development of sophisticated techniques for deeper and more ecologically valid information.

Early studies are underway, exploring the 'middle-ground' of dynamic content on a single television screen (see Figure 1). Examples include news tickers and dynamic sports result summaries. This type of formative study can be performed on standard eye-tracking equipment and give us clues as to the sort of events that trigger attention shift, and point to when users are more or less likely to shift their attention away from the primary content stream. These clues can then be explored in more detail in further studies, and lead into experiments involving multiple devices. Starting with these types of study will also allow us to make relatively direct comparisons between Web and television content, and guide the modelling process.

Summary

If second screens are to be used for presenting television viewers with supplementary content in a way that genuinely enhances the experience, the form and presentation of the content must be scientifically informed. Attention approximation has been used to understand how people interact with dynamic Web content, and we believe it has much to offer this domain. The level of detail at which attention is monitored will need to be adapted to suit the new content and environment, but we believe that the information obtainable by even coarse-grained attention approximation will be extremely valuable and can give real insight into how people use this technology.

Acknowledgements

We would like to thank the EPSRC for its funding contribution to this work (EP/G026238/1).

References

[1] Brown, A., Jay, C., and Harper, S. Tailored presentation of dynamic web content for audio browsers. *International Journal of Human-Computer Studies 70*, 3 (2012), 179 – 196.

[2] Jay, C., Brown, A., and Harper, S. Predicting whether users view dynamic content on the world wide web. *ACM Transactions on Computer-Human Interaction To appear* (2013).