Abstract
Mobile devices are playing an increasingly important role in leisure activities, including TV viewing. Broadcasters see this as an opportunity to enhance a TV programme through the provision of additional information to the 'second screen', but determining how to optimise this experience is one of the grand challenges currently facing content providers. Addressing this issue requires a detailed understanding of interaction with both the TV and the secondary device, but this new form of HCI, as much about passive consumption as active interaction, is not well represented by typical task-based paradigms. This paper describes experiments that use eye tracking to understand one of the key components of this new area of study — determining which device is currently receiving the user’s attention — and discusses the considerable challenge of accurately monitoring attention while maintaining ecological validity.

Author Keywords
Eye-tracking, 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
The near-simultaneous use of multiple devices is becoming increasingly common, and one of the more significant use-cases is using a mobile phone or tablet whilst watching the television. A study by Yahoo! and Nielsen reported that, in 2010, 79% of people occasionally or frequently accessed the Internet over a smartphone while watching TV [1]. Vanattenhoven and Geerts conducted a more in-depth study, capturing the second-screen habits of 12 households, and found that second-screen use was recorded by 11 of them [22], and that second screens were used for activity both related and unrelated to the programme.

This behaviour is emerging naturally as these technologies become an integral part of everyday life, but it is not entirely consumer-led — broadcasters have also been exploring the use of a mobile phone, tablet or laptop computer as a secondary display, and for a number of reasons. Broadly, these encompass content discovery and engagement [17]. The secondary device has been used to provide programme guide information [8] or material to aid accessibility to disabled users [23] in a manner that does not occlude the main television screen. This can be particularly important in situations where the main screen is being shared by a family or other group. Increasingly, by creating and distributing software applications for phones/tablets [13], or placing specifically related material on the web, television broadcasters are using the secondary screen to present synchronised companion content [18] for the primary programme. Retrospectively, this can be seen as a response to the phenomenon of viewers using Internet services on handheld devices whilst simultaneously watching television. Broadcasters’ dual screen companion applications might add auxiliary material to drama productions, coverage of sport, news, documentaries and entertainment [3, 20, 5], with a variety of mechanisms used to achieve synchronisation between the two display devices [15, 9].

Related Work
Despite ‘second screening’ emerging as an important form of content consumption, research in the field is limited. Although some work has been done investigating other aspects of the behaviour (e.g., social), there has been very little research systematically investigating how attention is split during dual-screen human-computer interaction. The most relevant work in this field is by Holmes et al., who used a head-mounted eye-tracker to monitor the attention of people viewing a drama or documentary programme with an ‘app’ running on a second-screen [13]. They reported the proportion of attention given to each device (63% on TV, 30% on tablet) and the lengths of the views on each device (means of just over 2s on the TV, 1s on the tablet). They also noted that attention appeared to shift to the tablet when it updated, and when the TV showed advertising content.

Beyond this, there has been little research: a survey of 1788 papers from the last 5 years of the ACM SIGCHI conference revealed only a handful relating to multiple screens in the general case. Much of this previous work has examined how additional screens affects task performance in a laboratory setting or a digital workspace [24, 10, 16], or focused on examining the use and usability of dual screen systems [12, 2, 6].

There have also been some previous uses of eye tracking to monitor attention, for example over different tasks located on a single monitor [14], or a single task split across two displays [19].
The results from [13] provide a useful starting point for understanding the split of attention whilst watching TV with a second screen. They indicate that people are more likely to view the second screen during ads, and tend to glance for only short periods of time (less than two seconds) at either device. Many questions remain, however: in particular, do these results generalise to other programmes, genres and audiences?; what are the other cues, beyond ad breaks, that affect the distribution of attention?; and does the experimental set up (including the presence of a head-mounted eye tracker) affect the ecological validity of the conclusions?

Methodological Challenges
There are several challenges in developing a method for accurately and replicably determining the split of attention across two or more screens, particularly in the domain of TV. For example:

- TV viewing typically occurs in a relaxed environment, quite different to that of a typical usability lab (‘sit-back’ vs. ‘lean-forward’ [23]). This environment complicates experimental setup compared to an office/desktop computer scenario, with even basic challenges such as viewing distance potentially making data capture difficult (e.g., [11]).
- TV viewing is essentially ‘entertainment’ and it has been argued that the traditional HCI focus on efficiency and task completion is not appropriate for this domain [7, 21].
- While the TV is normally static, the companion device is, by its very nature, portable, and is likely to be moved around considerably during use.

- Accurately monitoring attention over multiple devices, even if they are not moving, is technologically difficult.
- Attention to either device may vary according to social interaction, either directly with people in the room, or virtually with those outside it.
- Running experiments asynchronously is limiting — TV content is often designed for broadcast at a specific time, and second screen interactions of a social nature rely on the shared experience being synchronous.

Building a body of evidence about how people use second screens will require multiple experiments ranging from controlled laboratory-based experiments to more naturalistic data collection in ‘the wild’. We believe that starting this process with relatively high precision formative studies will enable the identification of factors that can be studied further, either in more detail or in more natural environments. In this paper we describe one such study, where eye-tracking, used to monitor when people shifted their attention between a TV and a tablet, highlighted a number of factors that appear to affect the distribution of attention.

Case Study: Companion Content for TV
Our particular interest is identifying suitable times for the presentation of companion content — when should this content be presented so that it most enhances the TV programme, and with minimum disruption to the viewer? In this context, our initial studies have been to develop an experimental technique that will allow us to study when people shift attention between a television and a tablet computer.
In our attempts to address the challenges laid out above, it has been necessary to design an experiment that balances the time spent capturing and analysing data, the data quality, the naturalism of the viewing environment, and the generalisability of the results. In this case, our laboratory is designed to mimic a home living-room, and participants were free to sit in a comfortable chair. It was decided to use two eye-trackers, one for the TV and one for the tablet. This was expected to add to the data quality, allowing us to determine where on each device people were looking, and to speed up the analysis process compared to, for example, video analysis. While a single head-mounted eye-tracker could have been used (as [13]), and may have resulted in higher precision data, it would arguably have resulted in a less natural experience for participants, and more complex data analysis. The requirements of the eye-tracking equipment, however, constrained the position in which they had to sit during the study, and meant that the tablet had to be fixed in a mount. In addition to the eye-trackers, interactions with the tablet were logged remotely.

In parallel with the quantitative data, a considerable quantity of qualitative data was collected. This allowed us to support the eye-tracking data, collect information about participants’ habits in their homes, and hopefully identify problems with the design, but due to space limitations, it is not reported here.

Stimulus
The information presented to the participants was a 15 minute clip from an hour long episode of the programme (“Autumnwatch”) for which companion content had already been created [18]. The episode covered three topics: dolphins, starlings and sea trout. The companion content was designed to be ‘ambient’ in nature, with relatively infrequent changes (synchronised with programme content), and not too much information. Seven different pieces of content were presented on the tablet during this clip as follows:

Swarming and Flocking. On the screen at the start.

Dolphins... did you know? From 0:09. A simple quiz with one question. See Figure 1a.

Where to see a dolphin. From 1:14. A non-interactive map of the British Isles, showing the best places to see bottlenose dolphins.

Starlings... did you know? From 4:49. A simple quiz with one question.

How to spot a starling. From 6:04. 4 slides giving information about starlings (Fig. 1b).

Your starling photographs. From 9:54. 11 slides showing photographs of starlings.

Rivers and streams. From 12:44. Non-interactive information about the habitat of sea trout.

The sea trout lifecycle. From 13:53. An interactive diagram of the sea trout’s lifecycle (Fig. 1c).

Method
Participants watched the TV content on a 47 inch LG television, and viewed the companion content on an Apple iPad tablet computer. Eye-tracking was performed on both devices; the TV using a Tobii X120, and the tablet using a Tobii X2-60. Eye-tracking on the tablet requires the device to be mounted. The Tobii Mobile Device Stand was used for this, fixed to a coffee table in front of the
participant. In addition to the eye-tracking, video feeds were captured from a webcam attached to the tablet eye-tracker, which pointed up at the user’s face, and from 4 cameras embedded in the laboratory. Figure 2 shows how the equipment was set up.

![Figure 2: The experimental setup: the television content was displayed on the main television screen, while the iPad was held on the Tobii mount in front of, and just to the right of the viewer. On the table is the eye-tracking lamp/detector for the TV; the one for the iPad is held on the mount, below the iPad.](image)

12 participants (6 female, 6 male, with ages ranging between 21 and 55, selected to be people who liked nature programmes and were in the habit of using a second device whilst watching television) took part in the experiment. After answering some initial questions (qualitative data were captured, mainly about the dual-screen experience as a whole, but are not reported here), participants were taken to the user laboratory. They were asked to make themselves comfortable, then the experimenter explained the procedure while the two eye-trackers were calibrated. The nature of the companion content was explained, and the participant was allowed to interact with an example (not used in the study). Once ready, the clip was started, and the participant was allowed to watch the TV and interact with the tablet as he/she wished for 15 minutes. Once complete, the participants were thanked and taken to answer some questions about their experience. Technical problems meant that the data from 2 participants were unusable.

**Data Quality**

As far as we are aware, there have been no other studies where two eye-trackers have been used simultaneously to record someone’s gaze over multiple devices. The eye-trackers used are intended to continuously follow one person’s gaze as they sit relatively still and look around a relatively small field of view. It is therefore necessary to examine the data for problems — would there be interference between the eye-trackers; would there be a lag between the shift of attention and the eye-tracker picking up the gaze; would the movement of the user as they look up and down between devices result in inaccurate calibrations?

As a ‘first check’ of data quality, and to see if there were calibration problems, we can look at the heat-maps generated for different segments of TV content and tablet content, and see if they correspond with what would be expected. For example, it is well known that people tend to fixate on faces — do the eye trackers record this in the data from this experiment?

Figure 3 shows illustrative heat maps, indicating where the gaze of 10 participants fell for 2 spells of TV content and 2 of tablet content. The results are as one would expect, indicating that the eye-trackers collected data of
reasonable quality, and that, overall, calibrations were accurate.

Figure 3: Example heatmaps from the two eye-trackers.

Examining problems related to re-capturing the participant’s gaze as it shifts between devices is less simple. In this case, the video footage from the experiments has been examined, noting, for each participant, when he or she shifts attention between the tablet and the TV. Comparing these data (which are themselves subject to some inaccuracy) with the data from the eye-trackers allows us to get a deeper picture of data quality.

For this, the eye-tracking data was divided into half-second slices. For each slice, the device (TV or tablet) was considered to have received attention from a user if at least one saccade or fixation was recorded within that half-second. Comparing these slices with the data obtained from video analysis indicates that the tablet eye-tracker had low rates of false positives (from 0.1 to 3.8%, with a mean 1.1% of slices recorded as having received attention when video analysis indicates there was none), but higher rates of false negative (1.4 to 12.0%, mean 5.6%). This gives mean accuracy and precision rates for the tablet eye-tracker of 93% and 75% respectively.

Investigation revealed that many of the false-negatives were a result of the participant occluding the eye-tracker with his or her arm while interacting with the tablet.

Results

There are many ways to look at the data collected from this experiment. In this section we explore: the overall split of attention — what proportion of time did people spend looking at each device; the characteristics of the glances — did people spend a few long periods looking at each device, or did they skip from one to the other regularly, and; how the attention received by the devices varied over the course of the clip? To answer these questions, it is only necessary to know which device was being looked at at any one time, so the analysis described above was applied, again using 0.5s slices.

The overall split of attention is roughly 5:1 in favour of the TV — the TV eye-tracker detected activity between 57% and 91% of the time (mean 76.5%, SD 10.9%). In contrast, the eye-tracker monitoring tablet use detected activity between 8% and 35% of the time (mean 16.9%, SD 7.8%). Activity was recorded by neither eye-tracker for a mean of 8.0% of the time (SD 5.3%). Some of this will correspond to the occlusion of the eye-tracker that contributed to the false-negative rate, and the video analysis suggests that the level of attention given to the tablet is of the order of 22%.
Figure 4: Distribution of dwell times between 0 and 100s for both devices. The dwell time is the length of time for which attention is detected on the device for all consecutive 500ms slices. Note that there were longer dwells on the TV, but these are not shown in this chart.

Turning to the nature of these attention spans — is this split of attention comprised of many short glances, or of longer sustained spells of attention on each device? Figure 4 shows a histogram of the lengths of glances. It is clear that many short glances are made at both devices, and spells of sustained attention are relatively infrequent.

People dwelt on the TV for longer periods than on the tablet, and dwelt for long periods on the television more often (27% of glances on the TV were 30s or longer, compared to less than 1% for the tablet; Mann-Whitney U test $p < 0.001$). In contrast, a higher proportion of glances at the tablet were very short (51% were 2.5s or less, compared to 30% for the TV; Mann-Whitney U test $p = 0.001$). These dwells are longer than those reported by Holmes et al. [13], although it is not clear whether this is due to the nature of the TV or companion content, or because our data analysis has a lower time resolution.

Figure 5 shows how this attention varies over the course of the clip. It plots the percentage of participants attending each device for each half-second time slice. The TV is plotted on the upper half (0 to 100%), while the tablet is on the lower half (0 to −100%; high negative values indicating more attention on the tablet). The graph is also annotated to show when the second screen content changed automatically, and when there were major changes in scene on the television (i.e., moving between the outdoor studio and location).

The long-term aim of this research is to understand the factors involved in shifts of attention. One of the striking features of Figure 5 is the shift of attention to the tablet each time it automatically updates. Viewers typically attended the tablet within a few seconds of the update: median times between the update and the first view range from 2s to 9s over the 7 segments. Other factors that appear to trigger attention shifts are changes in the TV content: more people view the tablet when the TV is showing presenters speaking from the outdoor studio than when it is showing location footage, and audio cues such as exclamations seem to attract the viewer back from the tablet to the television. Indeed, we hypothesise that shifts from the TV to the tablet are primarily visually driven, while those back to the TV are primarily initiated by audio.

Limitations
While the experiments were technically successful, questions remain about the generalizability of the results. In particular, it is unclear whether people would have used the tablet in the same way if it had not been mounted. It is possible that this limited interaction, as participants might not have felt ownership over the device; its position also meant that automatic updates would have been visible in people’s peripheral vision, which might not have been the case if it were placed flat on a table or chair.
Summary
Two eye-trackers have been used to monitor when viewers shift attention between a TV and a tablet computer; good quality data were obtained that showed not only which device was receiving attention, but where on that device people’s gaze lay. The ‘sit-back’ nature of TV interaction (cf. the ‘lean-forward’ nature of traditional interaction with computers) is a new form of HCI, and brings a number of extra challenges that we have listed in this paper. Arguably the most important among these is the need to keep participants in a relaxed and comfortable state. This experiment was performed in as natural environment as possible, but further experiments are required to confirm whether similar behaviour is seen when participants are free to sit and view the devices without the constraints imposed by the eye-trackers. We plan to perform such an experiment, to determine whether video analysis and/or tablet motion data can be used to detect attention shifts. If a simple binary measure to ‘approximate attention’ [4] (TV or tablet) is sufficient, this would allow experiments to be performed quickly and in a naturalistic manner.

The splitting of attention between two devices is an inherent complication in designing both natively dual-screen content and a second-screen companion for otherwise self-contained television material. The proposed methodology has the potential to inform and improve the emerging practice of user experience design across dual screen television material. The results of the pilot case study have shown the type of insights that might be gathered by larger scale monitoring of attention across two screens. Insights into users’ general frequency of switching attention and durations of focused attention, together with knowledge of the auditory and visual cues which can cause attention to drift or switch across screens, could form the basis of design patterns and other guidance for experience designers in this domain. Also, as companion content is a relatively new and emerging form
of interaction, with a low level of familiarity within the user population, this style of methodology might be usefully applied in a longer term manner, to study the extended effect of user familiarity and aid designers in developing an effective experience language for dual screen television and companion content.

**Acknowledgements**
This work was funded by the EPSRC Impact Acceleration awards programme (EP/K503782/1-P115906).

**Open Data**
The data from this experiment are available at http://wel-data.cs.manchester.ac.uk/, and the data analysis routines at https://bitbucket.org/webergonomicslab/sage-dual-screen-analysis.

**References**


