More is more: Investigating Attention Distribution Between the Television and Second Screen Applications - a Case Study with a Synchronised Second Screen Video Game

Regina Bernhaupt, Raphaël Guénon, François Manciet, Antoine Desnos

MORE IS MORE: INVESTIGATING ATTENTION DISTRIBUTION BETWEEN THE TELEVISION AND SECOND SCREEN APPLICATIONS - A CASE STUDY WITH A SYNCHRONISED SECOND SCREEN VIDEO GAME

R. Bernhaupt, R. Guenon, F. Manciet, A. Desnos

ruwido austria gmbh, Austria & IRIT, France

ABSTRACT

Attention is a key concept for the design of second screen applications that are to be used while watching television (TV). One of the key design goals is to balance the user's attention between the second screen application and the TV content. To investigate the influence of interactivity on attention and overall perceived workload and user experience, we developed video games with varying degrees of interactivity allowing users to play while watching a TV show. The small games were synchronized with the TV show in both the temporal dimension (presenting games as the show progressed) and with the content (enabling users to play games similar to the storyline in the TV show). Results from a laboratory-based user study show that highly interactive video games that are interleaved with the content, draw up to seventy percent of attention on the game (tablet) and are judged with a perceived workload similar to doing real 'work' while watching TV. Nevertheless participants rated the games as fun to play with a high user experience.

INTRODUCTION

Second screen applications are becoming more and more popular and widespread due to the growing popularity of tablets and smart phones. These applications can considerably enrich the TV experience by providing more information or additional features, like background information about actors and movie plots, access to the movie’s web site, or other interactive features like quizzes or voting.

Second screen applications typically allow the user to interact with the content they are consuming such as TV shows, movies or music. The main characteristic is that additional information is displayed on a portable device - usually the smart phone or tablet. The way that the content is enhanced (i.e. added value for the user) seems critical for the user acceptance. Work by Basapur et al. (2012) indicates that the experience is only positive when the additional media are in sync with the TV show and that it provides information that is judged to be relevant. The primary and secondary screens have to give the user the feeling of having a holistic (and synchronized) experience.

As Figure 1 presents, content on the tablet can be classified by degree of interactivity and by degree of synchronization. From the user's point of view, a high level of synchronization means having a holistic experience when using the TV screen together with a second screen application. An example is a game where you play along with the main character on the screen; the game being only available during a specified time or scene during the show. A low level of synchronization would describe an application that provides a TV experience only on the second screen without taking into account the content on the main TV screen. We refer to this category as TV stand-alone apps for second screens.

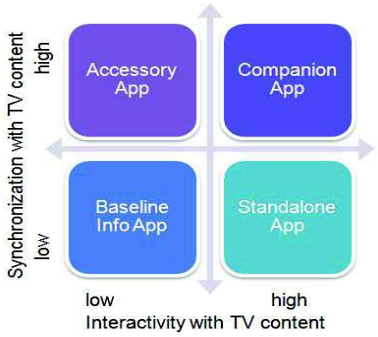

Figure 1 - Classification of second screen TV apps (from Bernhaupt et al., 2013).

In terms of perceived interactivity, a high level of interactivity refers to a user interface design that allows the user to seamlessly interact with content on both screens (e.g. a jump-and-run game that is connected to the main screen content - and ideally would change the course of events in the movie or show). A low or limited level of interactivity would refer to a simple selection of content on the second screen (e.g. an electronic programme guide that could enable parts of the content to be displayed on the TV screen).

What is still unclear when designing for the second screen is what level of interactivity and what degree of synchronization lead to what level of attention. Is the TV becoming superfluous? And finally, what level of attention and overall perceived workload will be just too much and overwhelming for the user? Overall, what we would want to achieve is a positive user experience. But is this still the case when second screen applications are highly demanding?

The goal of this study was to understand how to balance the user's attention in a multi-screen environment by investigating the relationship of games with different degrees of interactivity and synchronization, measuring attention, perceived fun, difficulty of the game and overall perceived user experience in terms of hedonic and pragmatic quality. Our hypothesis was that watching a high-paced TV show while playing small games with varying levels of interactivity and high synchronisation with the content, would increase the perceived workload and therefore affect fun and user experience.

STATE OF THE ART

While watching TV, people use a variety of other devices. During the past decade the devices used changed from standard household appliances and activities (e.g. ironing) (Bernhaupt et al., 2007) to more entertainment-oriented ones: Hess (2011) reports that people perform activities associated with the TV programme currently being watched, like searching the web to obtain information on the show, but also a variety of activities unrelated to the TV programme, including usage of social networks (e.g. Facebook, Twitter) or playing games.

When introducing second screen applications and performing activities on a second device, one key dimension for the design is how to control the user's attention between the two devices. Attention is a concept that refers to how we actively process specific

information present in our environment. It implies that, of the multitude of possibilities to which the user can direct his/her attention, some are ignored in order to deal effectively with others (James, 1880).

Attention *per se*, is measured using behaviour-oriented measurements like the identification of patterns. Visual attention for example is measured with the D2 test (Brickenkamp, 1962). The visual scanning performance is evaluated by asking people to identify in a sequence of the letter “d" all the cases where the letter is accompanied by two small strokes. When measuring attention for interactive systems, researchers have been applying a multitude of methods ranging from standard tests to bio-physiological measurements. Currently eye-tracking is popular. Results from Hawkins et al. (1997) show that the average look at the television is about 7 seconds, with the median length of gaze being 2 seconds. When interacting with a second screen, Holmes et al. (2012) found that 63% of gaze time was on the TV, compared to 30% on the tablet and 7% off-screen. When using a second screen, gazes are becoming shorter than in the traditional settings, averaging under 2 seconds for gazes on the TV and just over one second on the second screen (Holmes, 2012). The main indicators from a user perspective are the overall hedonic quality of the system (aesthetics, identification and stimulation) as well as the emotional state of the user (e.g. fun).

For the design of TV user interfaces and second screen applications it is important to understand how visual attention can affect memory. As Holmes (2012, p.2) has stated: "For example, audio combined with text appears to overwhelm the separate channels, whereas combining audio with a related visual requires less time to process the information with increased performance." Following Paivio (1986) memory can be enhanced if visual and verbal information are paired and available in working memory. To understand how to enhance second screen applications it is thus necessary to evaluate the working memory e.g. via the task-load. A task load is a subjective value that indicates the perceived workload of tasks.

PLAYING ALONG WITH THE CONTENT: AN EXPERIMENTAL APPROACH

Research Question and Method Choice

The goal of this study was to investigate the effect of levels of interactivity for highly synchronized second screen games on attention and perceived fun and overall how synchronized games are perceived in terms of workload and user experience. Our hypothesis was that a high paced TV show along with mini-games with low to high levels of interactivity would be (even for the young generation) too high in workload to be enjoyable.

To investigate this research area we decided to perform a set of experiments starting with a small-scale user study, verifying our initial hypothesis on what the maximum of interactivity in terms of workload is for a young age group. The experimental oriented design of this study thus varied the degrees of interactivity within the game, and measured self-reported levels of attention and perceived workload using the NASA-TLX questionnaire.

To investigate the influence of the varying degrees of interactivity and high synchronisation we decided to perform an experiment combining a high-paced TV show and a range of small games (mini-games) with low to high degrees of interactivity. The level of attention

needed for each activity (watching TV and playing the mini-game) was tuned in terms of activity in the gameplay to investigate how attention gets impacted and so how the overall workload of these games is perceived.

To evaluate attention we used participants' self-judgment as well as observation. The workload was measured using the NASA-TLX scale (Hart & Staveland, 1988). The NASA Task Load Index (NASA-TLX) is a subjective, multidimensional assessment tool that rates perceived workload. A set of different scales (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration) are combined into one measure ranging from 0 to 100, the task load index.

As games that are not fun to play might affect the effort that people are putting in the game, and as systems that are not having the necessary positive user experience can limit overall engagement, we observed perceived fun and overall user experience. Fun was measured after each mini-game by self-evaluation (scale 1-10) and overall user experience was measured using the AttrakDiff questionnaire (Hassenzahl et al., 2004). The AttrakDiff is a questionnaire measuring pragmatic and hedonic quality (in terms of stimulation and identification as well as attractiveness) using a 7-scale semantic differential with word-pairs (e.g. ugly vs. attractive). Results are combined in two measures (pragmatic quality vs hedonic quality) and represented as a diagram showing the attractiveness of the product.

Prototype, Participants and Procedure

To understand the level of attention and perceived workload when playing along a TV show, we developed a set of three mini-games synchronized with a movie currently running on the TV screen. Based on the BBC Sherlock Season 1 Episode 1 TV show, a set of mini-games was developed to be played along. The mini-games were made available on the second screen app depending on the progress in the show (see Figure 2).

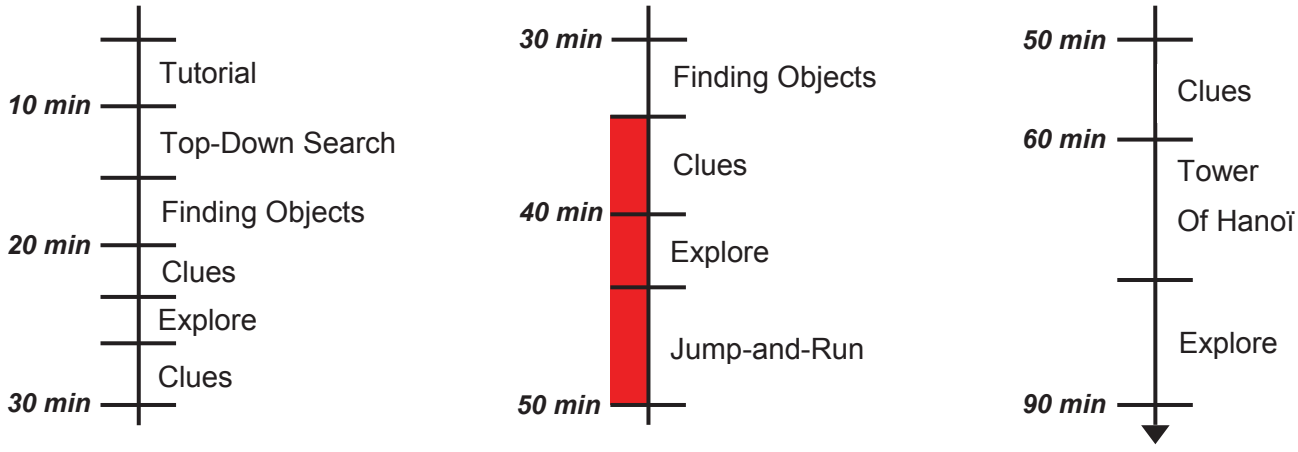

Figure 2 - Timeline of the Sherlock episode with different mini-games associated. In red, section of the episode retained for the prototype.

In the study, three mini-games were played. Game 1 (Clues) was a puzzle game where people had to find clues to solve the problem that was in line with the story (see Figure 3). Synchronization of the game and the content was very high (available time for the game was linked with the show progressing) and interactivity was low. Game 2 (Explore) was an exploration game where people had to find clues in a 3D representation of the streets that

were at the same time used in the show, with a medium level of synchronization and interactivity. Game 3 (Jump-and-Run) was a game where the player had to run along with the main character on the screen to simulate catching a taxi. It had high synchronization and high level of interactivity.

Eight participants, five male, three female, aged 16 to 26 with a mean age of 22 years, took part in the study. The size of the eight households in which the participants lived ranged from one to four members. Participants reported that they played video games regularly on a multitude of devices. Seven participants reported that they had IPTV at home (covering all major operators in the country). Six participants reported watching TV (on a TV screen) at least several times per week, one reported watching once per week, one reported to watching once per month.

Figure 3 - Clues game: The player had to play Puzzle while main character in the TV show found the same clues.

All participants used the Internet, mainly to check e-mails, play games and occasionally to do online shopping. While all participants owned a PC or laptop and a mobile phone, only four participants owned, or had access to, a tablet. None of the participants had used a second screen application while watching TV, before taking part in the study.

Participants were welcomed for the experiment and were seated in a low chair looking at a TV screen; a position that would be typical for watching TV on a couch. Participants signed the participant video allowance, consent form and answered some short questions on their TV watching and gaming behaviour. Participants then were handed over the tablet with the games application already running, and the Sherlock Episode was played on the TV screen. To minimize duration participants were asked to play along with three segments of the 90 min show (red in Figure 2).

After each of the three mini-games, participants had to judge their perceived level of fun, degree of attention to the game, interconnectedness between TV content and game and overall difficulty of the game. At the end of the episode, participants filled out the NASA-TLX and the AttrakDiff questionnaires and answered a set of interview questions related to their experiences with the system. Events in the game (on the tablet) were logged, and participants were video-recorded with a camera. We recorded the user from the front to focus on both the user's face and hands interacting with the tablet.

For the analysis we had two researchers investigating the videos and coding on an agreement basis. All material used, including video and games, were in French, results from interviews were translated by two researchers independently.

RESULTS

All participants playing the game started interacting with the game within the first 10 seconds after the game started, on average after 4.9 seconds (SD=1.9). Each of the three games was played by the eight participants once. From the 24 game sessions played, 22 were played successfully (reaching the game goal within the given time), one player ran out of time while playing the game, while one result was missed due to technical logging problems. While playing the three mini-games (Clue, Exploration and Jump-and-Run) none of the participants reported any problems or difficulties or asked any questions.

Table 1 shows the results for the three different mini-games (G1, G2 and G3) for overall attention and users' self-judged evaluation of fun and difficulty of the game. Results are contrasted with observer-judgements of the attention ratio between TV and tablet.

G1 (Clues)			G2 (Explore)			G3 (Jump-and-Run)		
Attention on TV (%)	Fun (1-10)	Difficulty (1-10)	Attention on TV	Fun	Diff.	Attention on TV	Fun	Diff.
46%	6.1	2.5	37.5%	7.5	4.8	30%	7.9	3.7

Table 1 - Results of observed attention on TV, perceived fun and difficulty of the games.

What can be seen in Table 1 is that the design of the mini-games fitted the purpose: G1 was perceived low in difficulty (low interactivity), and the synchronization with the main TV content was high. Attention to the TV screen was still high with 46%. For G2, the exploration game, with medium synchronization and interactivity, attention to the TV was rated 37.5%, while perceived difficulty was rated medium, and fun was rated rather high (7.5 on average on a scale from1 to 10). For G3, the Jump-and-Run game, synchronization of the content and interactivity were high, leading to low attention for the TV screen, high scorings for fun and a rather low difficulty rating of 3.7. A one-way between subjects ANOVA was conducted to compare the effect of degree of interactivity in the game. There was a significant effect of interactivity on fun at the $p<.01$ level for the three games ($F(2.21)=7.76$, $p<0.01$) and for attention ($F(2.21)=5.29$, $p<0.05$). Difficulty was not significant, but there was a clear trend visible ($F(2.21)=10.71$, $p=0,051$).

Mental Demand	Physical Demand	Temporal Demand	Performance	Efforts	Frustration	NASA-TLX (workload)
62.8	31.8	39.4	56.2	47.5	38.7	59.12
min:35 max:75	min: 5 max 75	min:10 max: 75	min: 10 max: 85	min: 30 max: 65	min 5: max: 65	SD=3.29

Table 2 - Showing results for the six scales of the NASA-TLX and the overall workload score (on a scale from 0 to 100).

The NASA-TLX questionnaire measures perceived workload on a scale from 0 to 100, with results around 60 indicating a medium workload, or an activity similar to a office work (Cinaz et al., 2013). It replicates results from users watching TV and performing interleaved work-tasks indicating workloads around 60 (Du Toit, 2013).

In terms of user experience, measured with the AttrakDiff questionnaire (Hassenzahl, 2004), the second screen application was rated above average regarding the pragmatic and hedonic quality on the 7 point scale, while failing to achieve a rating to be really desired. This can be interpreted as while users perceived the user experience in terms of hedonic and pragmatic quality as acceptable, there is still a lot of room for improvement to reach an overall experience that is really desired by the user.

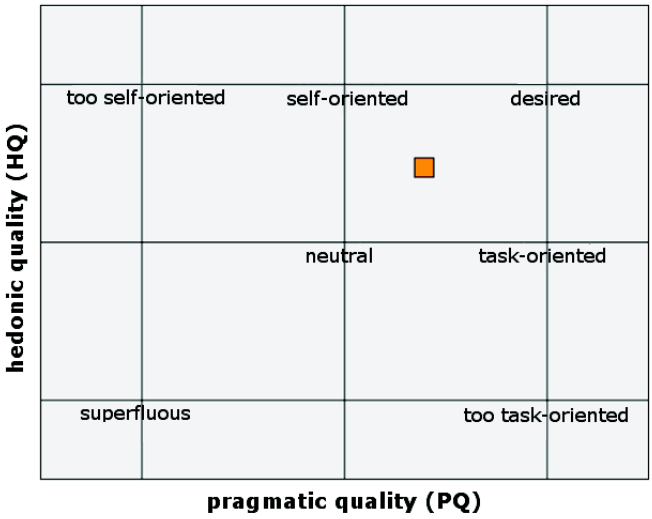

Figure 4 - AttrakDiff results for hedonic and pragmatic quality.

The results show that high attention games are judged positively by users in terms of fun and perceived workload is about the same degree as people would have when performing an interleaved work-task. What we found is that the more the gaming action is synchronized with the movie's storyline, the easier the player can follow the two media streams simultaneously, which is in line with the psychological theory on two-cue coding (Paivio, 1986). On the contrary, highly synchronized games lead to a shift of attention away from the TV to the tablet. What we found in this particular case study (with a rather high-paced TV show, high degree of synchronization and high degree of interactivity) is that the attention of the player is on the tablet and not any longer on the TV show, with a worst case of only 30% of attention on the TV.

SUMMARY AND DISCUSSION

How much attention to the second screen is too much attention? Based on results from this experimental study with young participants (age 16 to 26 years), we found that second screen games that are interleaved with the programmme draw the majority of attention to the second screen (up to 70%). But they are still judged as fun to play and to provide a good overall user experience. What has to be critically discussed is if such a finding can be generalized to other user groups and applications. As indicated by the workload measure, playing an interleaved game is perceived as the same level of workload as performing actual work. This might indicate that for short periods of gameplay a high level of workload might be acceptable, but that it is less likely to be accepted for longer periods. For younger age groups playing such game might be interpreted as a challenge and they really want to get immersed in the game, but for other applications like additional information or social connectedness, this level of workload will very likely not be accepted.

For the design of second screen applications it is advisable to interleave the content as much as possible, with clear indications where the user should look at. This way attention can be controlled and attention for the screen can be better balanced. To limit the workload it can be helpful to interleave the activities on the two screens so the activity of watching and playing feels like doing one task, instead of two separate ones. Separate tasks have been shown to have load ratings up to 80 (Du Toit, 2013), which might not be

advisable for an entertainment application. To summarize, contrary to lots of design advice for second screen activities, the more activity is expended on the second screen (in line with the content) the better the overall user experience.

REFERENCES

1. Basapur, S., Mandalia, H., Chaysinh, S., Lee, Y., Venkitaraman, N., and Metcalf, C. 2012. FANFEEDS: evaluation of socially generated information feed on second screen as a TV show companion. Proceedings of the 10th European conference on Interactive TV and Video. July, 2012. pp. 87 to 96.

2. Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., and Tscheligi, M. 2007. Trends in the living room and beyond. Interactive TV: a Shared Experience. pp. 146 to 155.

3. Bernhaupt, R., Pirker, M., Gatellier, B. 2013. Identification of User Experience and Usability Dimensions for Second Screen Applications: Results from an Expert Evaluation Using Generic Task Models. Proceedings of 2013 International Broadcasting Convention.

4. Cinaz, B., Arnrich, B., La Marca, R., and Tröster, G. 2013. Monitoring of mental workload levels during an everyday life office-work scenario. Personal and ubiquitous computing, 17(2). pp. 229 to 239.

5. Du Toit, H. 2013. Working while watching TV, is it really work?: The impact of media multitasking on stress and performance.

6. Hart, S. G., and Staveland, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology (52). pp. 139 to 183.

7. Hassenzahl, M. 2004. The interplay of beauty, goodness, and usability in interactive products. Human-Computer Interaction, 19(4). pp. 319 to 349.

8. Hawkins, R., Pingree, S., Bruce, L., and Tapper, J. 1997. Strategy and Style in Attention to Television. Journal Of Broadcasting And Electronic Media (41). pp. 245 to 264.

9. Hess, J., Ley, B., Ogonowski, C., Wan, L., and Wulf, V. 2011. Jumping between devices and services: towards an integrated concept for social TV. Proceedings of the 9th international interactive conference on Interactive television. pp. 11 to 20.

10. Holmes, M. E., Josephson, S., and Carney, R. E. 2012. Visual attention to television programs with a second-screen application. Proceedings of the Symposium on Eye Tracking Research and Applications. pp. 397 to 400.

11. James. W. 1950. The Principles of Psychology, Vol 1. Dover Publications, reprint from the Original Article in 1880.

12. Paivio, A. 1986. Mental Representation: A Dual Coding Approach. Oxford University Press.