Improving Interactive TV Experience Using Second Screen Mobile Applications

- -

***Abstract*—The past two decades have seen a shift in the multimedia consumption behaviours from that of collectivism and passivity, to individualism and activity. This paper introduces the architectural design, implementation and user evaluation of a second screen application, which is designed to supersede the traditional user control interface for primary screen interaction. We describe how NSMobile, our second screen application, can be used as a pervasive multimedia platform by integrating user experiences on both the second screen and primary screen. The quantitative and qualitative evaluation of user interactions with interactive TV content also contributes to the future design of second screen applications.**

I. Introduction

The past two decades have seen a shift in the multimedia consumption behaviours of consumers from that of collectivism and passivity, to individualism and activity, both in the home and in mobility situations outside of the home. Concurrently, there has been a secondary shift towards non-linear (i.e., time-shifted) usage patterns as consumers move away from the classical model of linear broadcast TV.

The rise of multi-platform media consumption does not mark the demise of traditional television, however. Platform choice has been shown to be both context and content dependent. For example, televisions are preferred for watching high-quality, long duration, low urgency content, while mobile devices are preferred for content that is of low-quality, short durations, and a perceived high-urgency [4]. The characteristics of the platform itself also affect this decision; predominantly that of form factor [3].

There is also a growing body of research indicating that consumers are beginning to spend a significant amount of time interacting with mobile devices and other technologies whilst watching television, in what has been labelled media multitasking [10]. Although this synchronous viewing behaviour reinforces perceptions of the television as a lean-back medium, that requires low levels of cognitive loading [13], it does not diminish the continuing active role played by viewers. Media multitasking could be perceived as a threat to the traditional television viewing model, but it also presents opportunities to exploit how multiple screens could support and further engage viewers.

Despite the growing number of second screen applications, little is known regarding the impact of mobile second screen applications on interactive TV content as well as the right balance between improved engagement and distraction [6].

In this paper we outline the architectural design and evaluation of a second screen (i.e., mobile phone or tablet-based) application, which is designed to supersede the traditional remote control for primary screen (i.e., predominantly television-based) interaction. We describe how our second screen application, can be used as a pervasive multimedia searching, scheduling, viewing, and socialising platform by integrating user experiences on both the second screen and primary screen. We also outline the design and results of lab-based user evaluations providing both quantitative usability analysis and subjective feedback of user experience.

II. Background and Related work

The features and performance of smartphones have been revolutionised in recent years. These rich features make smartphone an ideal candidate for an effective and truly "universal" remote control that can fully unlock the potential of the next generation of interactive digital TV systems. The major advantages of having smartphone as the next generation remote control include ubiquitous, connected and always on, flexible, personal, and the integrated display. Research shows that many viewers already use online television-related content to complement broadcast TV [10]. If a mobile device is aware of the viewing context it is in a position to intelligently fetch and display relevant online programme metadata or allow viewers to take part in social discussions around the programme.

Meanwhile, the explosion in digital and mobile device ownership has greatly changed the focus of social and interactive TV to enhancing user experiences with tablets and smartphones as second screen devices [5]. We have also seen recent developments around semantic video applications that adapt existing single-screen applications to multi-screen environments based on author or user choices [11] and multi-screen orchestration that connects TV programs with "social sense" using mobile devices [7]. An example is the IllumiRoom project where Microsoft looked into augmenting the area surrounding a television with projected visualizations to enhance traditional gaming experience [8]. The BBC took a similar approach in its Surround Video, an immersive video technology to be integrated in a domestic-scale viewing environment [12]. Kusumoto et al. studied the effects of complementary information and tweets on the media experience indexed by a comprehensive self-report questionnaire [9]. Centieiro et al. designed a second screen betting application for realtime interaction during live sports

TV broadcasts as the means to improve user engagement [2]. There have also been psychological studies on the split of attention, cognitive load, perceived comfort, and the maximum number of screens that could be watched at the same time [14], [1]. In spite of the myriad of mobile applications, most of the existing second screen designs focus on incorporating third-party contextual information as the adjunctive elements to the primary screen. There is currently lack of empirical study how mobile device can directly enable interactive TV and its impact to the overall TV viewing experience.

III. NSMobile Application

A. NSMobile

NSMobile is an integrated second screen mobile application designed to work in conjunction with internet-based IPTV set-top box (STB) or Smart TV (Figure 1). NSMobile provides the following functionality:

- A replacement remote control for the STB.
- An electronic programme guide (EPG) in carousel, grid or list format showing TV and radio schedule for content retrieval. This includes highlighting which recorded items are available to playback immediately (i.e. video-on-demand assets). Moreover, users can: add content to their favourites list for repeat viewing; view their playback history; view online meta-data relating to a particular media item (e.g. a Wikipedia article); follow a conversation about a particular media item on Twitter; rate and comment on media items.
- The ability to utilise roaming screen for multimedia playback. Multiple NSMobile clients can control and stay synchronized to a single STB, and it is easy for a user to move between STBs, transferring their current viewing context with them. The software supports local media playback, allowing seamless session mobility between STBs and NSMobile.
- The support of subtitles, chapters, polls, quizzes, web links, etc., relating to the primary screen content.

NSMobile acts as a wireless control point (client) with the STB acting as the controlled device (server). Signalling is maintained between NSMobile and STB as a means of coordinating interaction (i.e., communicating control instructions and synchronisation information), rather than as a means of transporting media traffic. Once connected, NSMobile is able to send commands (actions) to the STB to simulate remote control key presses, request playback of a given media resource, submit text for an on-screen dialog, etc. NSMobile (one or more) subscribes to events published by the STB, each of which contains all of the information required to synchronise its local representation of the STB's state with the STB. NSMobile updates information and metadata relating to the current media item (e.g., tile, description), playback state (e.g., playing, paused), and the

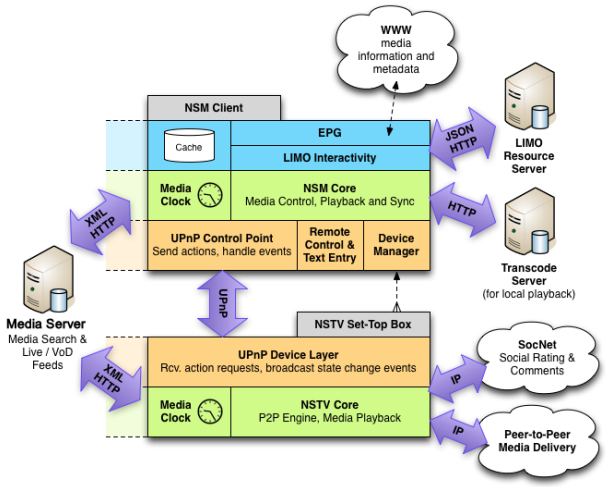


Figure 1. NSMobile Architecture

current playback offset (i.e., bytes and time in seconds from the start of the media). Events are sent to NSMobile clients whenever the state changes and also periodically (every 10 seconds) to ensure their local representation of the STB's clock is synchronised. This is necessary as any skipped frames, or buffering delays on the playback through the primary screen can result in clock drift. The synchronized media orchestration between NSMobile and its paired STBs is implemented using the Lightweight Interactive Media Object (LIMO) framework developed by the BBC. In the LIMO framework, timed metadata can be used to build interactive applications associated with the media content, such as captioning, chapter navigation, comments, or quiz games. It may also be used to attach arbitrary descriptive metadata relating to periods of time within the content, such as location or a list of characters or actors.

B. Content discovery

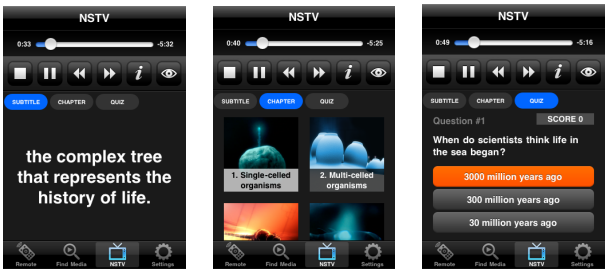
NSMobile provides a virtual remote control, which is rendered as a touch-screen facsimile of the physical remote control (Figure 2). This enables interaction through the STB interfaces on the primary screen. Furthermore, whenever the TV interface requires text input (e.g. for search), the mobile device's touchscreen keyboard can be used to input text rather than the virtual remote control. This process is fully synchronised across the primary and secondary screens. NSMobile also provides a number of supplementary mechanisms for facilitating local content discovery, including EPGs, search, play history, and favourite (Figure 2).

NSMobile is capable of parsing and rendering Atom feeds containing EPG data. These Atom feeds are automatically generated (and stored as MPEG-7) for both live and video-on-demand content within an IPTV experimentation infrastructure, and delivered to the second screen application.

Although there is existing literature outlining the implementation of EPG functionality into second screen appli-

(a) Virtual RC (b) Content discovery

Figure 2. Virtual remote control and content discovery



(a) Subtitles (b) Chapters (c) Quiz

Figure 4. LIMO: Subtitles, Chapters, Quiz

cations (e.g., [13]), there has been minimal research into which interface designs are suited for content discovery on handheld devices. NSMobile implements three manifestations of user interfaces: *Grid*, *Carousel*, and *List* (Figure 3). The usability and user preference of these interfaces will be addressed in the user evaluation section.

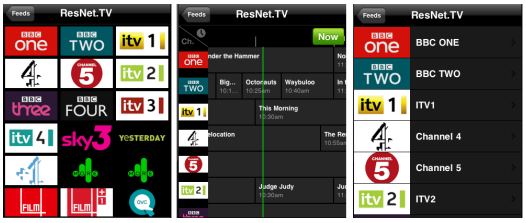


Figure 3. Content Discovery Interfaces: Grid, Carousel, List

Beneath each of these interfaces is the programme and detailed metadata levels presented as a list-based interface of VoD assets, with the option of playing live television.

1) Subtitles: Subtitles can be ported by utilising a set of JSON files with time-encoded values measured in time with the audio in the video stream being displayed on the primary screen. For subtitles in NSMobile (Figure 4), the process of periodic state updates from STB is used to maintain correct timing (e.g., for automatic re-synchronisation after seeking).

Traditionally the subtitles would overlay the video stream. We are used to the convention of having subtitles appear at the bottom on primary screen, as it allows for rapid reading. Moving this to a second screen may have the effect of viewers paying more attention to the second screen and neglecting the primary screen in favour of understanding the speech. The ability to connect and pull speech data could be useful in situations where audio is unavailable, disrupted or cancelled out from background noise. Furthermore, this would allow subtitle information to be displayed in an array of different languages, based on user preferences.

2) Chapters: Chapters provide a shortcut to skip to predefined sections of a TV program. The data is defined in JSON encoded files, with a start and end time, which can be loaded into the second screen engine (Figure 4). This provides the viewer with a fluid interface for skipping through content, rather than having to manually seek through of content. A short preview of the chapter could be played, or even a description could be displayed on screen, detailing what happens within the chosen chapter. Descriptions are valid within the JSON specification for the chapter feature, providing users with more control over what they are skipping to.

3) Quizzing: Quiz functionality within the NSMobile application shares similar underlying principles as the subtitles function. A manifest file has a start and stop time defined to set a period of time in which a question is valid. When this time period is reached, NSMobile triggers a pause of the playback on primary screen. The question is then presented on both the primary and secondary screens. A viewer can then select an answer on the second screen, which is logged by the engine, and then the playback on the primary screen is resumed. After the question times out, a viewer is presented with a message on the second screen, stating the answer is either correct or incorrect. At the end of the multimedia stream the answers are totalled and a viewer is given an overview of their score. One potential use case for this functionality is with educational programs. If the program had a section in which a question is asked and the viewer given a list of answers, a student can practice with a great learning experience, breaking away from the norm of memorization directly from written material. The same feature can also support audience participation and polling of reality TV shows or song contests.

IV. User Evaluation and Discussion

A. Experiment Design

In order to systematically study the benefit of second screen application in improving the user experience of interactive TV content, we conducted a usability test. 12 participants were recruited who had no prior experience with the NSMobile. Of these participants, 9 were male, and 3 female. The age of participants ranged from 25 to 45. The usability test lasted one hour in three sections.

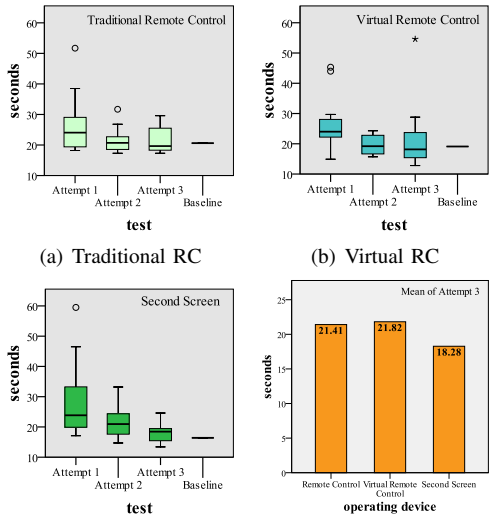
The first section of the test consisted of two quantitative evaluations of the time taken to complete content discovery tasks. The first task involved navigating through a series of menus to the EPG of popular video-on-demand assets. Participants were then asked to play a particular item on this list, and skip 4 minutes into the content. Participants

performed this task with the traditional remote control, the second screen's virtual remote control, and the second screen EPG. The second task involved using the search mechanism to find and play a particular piece of content. Participants performed this task with the traditional remote control, and the second screen. Each test was repeated three times. The order of operation for each task was cycled, in order to mitigate the impact of the learning effect on our results. The second section consisted of a demonstration of various NSMobile features (e.g., roaming screens), and a qualitative interview. The final section was a questionnaire about remote controller preference for particular tasks, their usability, and the desirability of particular second screen features.

The testing environment was designed to facilitate the collection of an array of different metrics (e.g., timing of tasks, calculation of error rates). This involved two overhead cameras, one of which directly looked down over the participant to collect behavioural data. The STB was connected to a HDMI splitter. One output was sent to the TV, and the other to a PC. This PC captured the output, along with that of the overhead cameras.

B. Results and Discussions

For the evaluation of the first section of the usability test, each of the participant's three attempts with each of the three operating devices was timed. Figure 5(a), 5(b), and 5(a) shows the box plots of the collected data across all participants, along with the mean values for the third attempt using each operating device (bottom right). In order to compare the participants' performance with each operating device, we include a measure (baseline) of the performance of an experienced user performing an identical task.



(a) Traditional RC (b) Virtual RC

(c) Second screen (d) Comparison

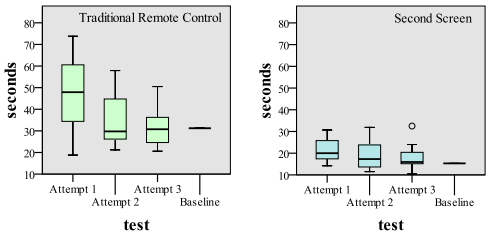
Figure 5. Comparing Operating Devices for Task 1

For the traditional and virtual remote controls, participants perform on average, similar to an experienced user by their second attempt. For the second screen, this gap is larger, but it remains a nominal difference. For all operating devices there are outliers in the first attempt, which can be attributed to certain participants having difficulties in initially understanding the mode of operating each device. The four outliers for the first attempts were spread across three participants. It is observed that by the third attempt for both the traditional remote control ($M : 21.4, SD : 4.1$) and virtual remote control ($M : 21.8, SD : 11.3$), the difference in mean completion time is small, but there is greater variation in the samples.

Both the virtual and traditional remote controls show similar interquartile ranges, and ranges of high-end completion times, however, the difference lies at the lower end. As indicated by the median and short lower tail of the traditional remote control's box plot, this operating device appears to hit a hard lower limit for completion times, but with some consistency in the frequency that participants could do so. In contrast, some participants are able to complete the same task using the virtual remote control at higher speeds. We attribute this to the speed with which discrete button presses can be triggered using the touch screen's multitap mechanism, compared to the delay involved in pressing and releasing a physical button and the infrared processing that must occur. Task completion time with the virtual remote control, however, was observed to be inhibited by an increased error rate. After an initial introduction, participants were able to use the traditional remote control without looking down at the device when pressing buttons. Despite the novelty of the device, this applied not only to repeat presses, but when moving between keys. For the virtual remote control there was a notably higher error rate as participants tried to apply this mode of behavior, with increasing frequency as they gained experience with the device. Participant completion times were hampered by these increasing errors, along with the additional delay from the requirement of looking down and then back to the primary screen after each press, in order to continually reinforce that a correct action is made. NSMobile attempts to mitigate this issue by triggering "clicking" sounds.

For the second screen, the third attempt ($M : 18.2, SD : 3.6$) yielded a lower mean and variability for content discovery than that which was provided by the other operating devices, which used the primary screen interface. Furthermore, unlike the other operating devices that performed similar to the baseline on the second attempt, the results suggest that performance may continue to improve with increased use.

Unlike the first task, the second that involved text input shows a significant difference in the performance of the two devices across all attempts (Figure 6). For the third attempt, completion times using the second screen ($M : 18.04, SD : 5.68$) remained significantly faster than the traditional remote control ($M : 31.96, SD : 8.91$), whilst also showing less variability. This difference can be attributed to



(a) Traditional RC (b) Second screen

Figure 6. Comparing Operating Devices for Task 2

two factors. First, a difference in error rates. The mean error rate across all participants and attempts with the traditional remote control was 1.28, but only 0.11 with the second screen. Second, the greater cognitive load required when typing with the traditional remote control. When typing, participants would often focus on the remote control and key presses, rather than the actual characters appearing on the primary screen. In order to mitigate errors resulting from this behavior, participants would look between the primary screen and remote control after each key press, a process increasingly taxing when multiple presses (on the numeric keypad) are required for a single character. In contrast, the mobile device also allowed the search term to be displayed locally, which expedited any validation process.

In order to evaluate the usability of the second screen application compared to the traditional remote control, a questionnaire was conducted. User ratings are on a scale of 0 to 100, with 100 marking the highest level of usability. The traditional remote control produced a rating of 75.6 (SD : 12.06), whilst the second screen produced a rating of 77.7 (SD : 11.35). Despite the second screen showing an improvement, it is not considered to be statistically significant ($t = 1.1, df = 11, p < 0.294$).

Various supplementary questions were asked to determine device preference for particular tasks (Figure 7 and Table I). The results for device preference for browsing TV programs (Q1) were mixed, but the second screen is seen as highly preferably when searching for video-on-demand content (Q2). The second screen is preferable for fast forwarding through VoD content (Q3), and checking program information during viewing (Q4). Although the majority of participants were neutral towards device preference for checking social media information such as Twitter feeds (Q5), no participants found the traditional remote control to be preferable for this function.

A further series of question was asked to determine functional desirability on the second screen application. The majority of participants expressed an interest in using the second screen to operate a TV, and to browse the TV guide. More specifically, where this interest was expressed, it was that the functionality be integrated into a mobile phone.

"If you've lost your remote, you can just give it a ring."

The participant preference for each of the EPG interfaces is inconclusive with the list at 33.33%, the grid at 25%, and

Table I
USABILITY QUESTIONNAIRE

Which device do you,prefer for the following tasks?	
Q1	Browse TV Programs
Q2	Searching video-on-demand content using the keypads.
Q3	Fast forwarding through content.
Q4	Checking program information whilst viewing the program.
Q5	Checking social media information (e.g., Twitter)

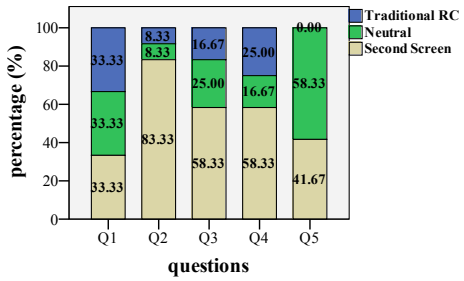


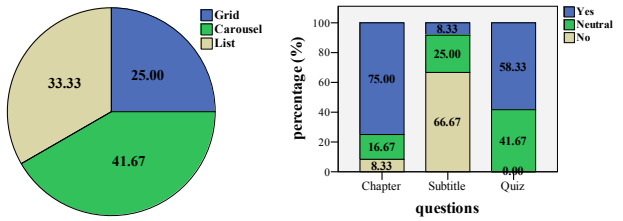
Figure 7. Device preference for designated tasks

the carousel at 41.67% (Figure 8(a)). Although the questionnaire was originally designed to determine EPG preference, it was observed during the trials that preference constituted different things for different tasks. For premeditated, tasks (e.g., playing a video-on-demand asset) the participants preferred the list and grid interfaces. Participants considered the linear format of these interfaces to facilitate the content discovery process, as they were not "cluttered" with other information (e.g., programme listings for other channels). For unpremeditated content discovery the carousel was considered most appealing due to the exploratory nature of the experience. Over half of the participants considered the ability to copy playback across display devices to be desirable. However, the roaming screen variant of moving playback between display devices was perceived to be more preferable. Highlighted during the participant evaluations was the opportunity of this functionality to supersede the traditional model of time-sharing primary multimedia consumption devices within the household.

"I like that idea a lot. If it's my time for having the TV for an hour, instead of it being a scheduled time ... you can take yourself off to the kitchen and pick it up again."

Participants were asked to indicate whether they would use the three second screen functions integrated into NS-Mobile, the results of which are illustrated in Figure 8(b). Chaptering was the most popular of these functions. One participant stated that although they would primarily use this function for "additional material" in conventional contexts (e.g., with DVDs), they were keen to highlight how it could be applied to traditional television.

"if you think about the [news] paper, I read the headlines, and then I always turn to the back to read the sport, and then start again at the beginning ... I don't know whether I'd change my mode, [and start] thinking of TV as modules."



(a) Browsing preference (b) Feature preference

Figure 8. Functionality evaluation

Subtitling attracted little interest amongst the sample. One issue highlighted by participants was the required continual shifting of focus between the primary and second screen.

"the issue with the subtitles is that you're always trying to look at the program."

Participants were open to and proactively suggesting alternative uses of subtitling on second screen. For example, when you wish to store a local transcript of a primary screen's audio, or for public advertisements and how subtitling could eliminate noise pollution.

"on Who Wants to be a Millionaire, you can imagine yourself doing it, can't you?"

The majority of participants (58.33%) found the quizzing functionality desirable, and all consolidated their desires around specific genres: educational, documentaries, and quizzes. Multiple participants stated that they could see a more general application in interactive TV services in order to increase audience engagement. For example, with soaps or talk shows. For all second screen functions, the importance of optionality about their use was stressed.

V. Conclusions

In this paper we present an integrated second screen application NSMobile for interactive TV. NSMobile integrates a framework for timed metadata that enables interactive functions such as session mobility, chapters, subtitles and quizzes. We also presented a user study consisted of both a qualitative and quantitative analysis to investigate the user experience of interactive TV with the help of second screen applications. We found that the perception of second screen varies and it's content specific. Some features do not significantly improve the efficiency of content navigation *per se*, though the novelty of user-TV interactions through a second screen device has an overall positive contributing effect to the user experience. Quizzes and chapters was perceived as desirable functionalities and our future work will look into how such features influence user behaviours in content retrieval and social interactions.

References

[1] A. Brown, M. Evans, C. Jay, M. Glancy, R. Jones, and S. Harper. Hci over multiple screens. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 665–674. ACM, 2014.

[2] P. Centieiro, T. Romão, and A. E. Dias. From the lab to the world: studying real-time second screen interaction with live sports. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, page 14. ACM, 2014.

[3] D. Dearman and J. S. Pierce. It's on my other computer!: computing with multiple devices. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 767–776. ACM, 2008.

[4] D. Geerts, P. Cesar, and D. Bulterman. The implications of program genres for the design of social television systems. In *Proceedings of the 1st international conference on Designing interactive user experiences for TV and video*, pages 71–80. ACM, 2008.

[5] D. Geerts, R. Leenheer, and D. De Grooff. In front of and behind the second screen: Viewer and producer perspectives on a companion app. In *Proceedings of the ACM International Conference on Interactive Experience of Television and Online Video (TVX2014)*, 2014.

[6] D. Geerts, R. Leenheer, D. De Grooff, J. Negenman, and S. Heijstraten. In front of and behind the second screen: viewer and producer perspectives on a companion app. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video*, pages 95–102. ACM, 2014.

[7] H. Hu, J. Huang, H. Zhao, Y. Wen, C. W. Chen, and T.-S. Chua. Social tv analytics: a novel paradigm to transform tv watching experience. In *Proceedings of the 5th ACM Multimedia Systems Conference*, pages 172–175. ACM, 2014.

[8] B. R. Jones, H. Benko, E. Ofek, and A. D. Wilson. Illumiroom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 869–878. ACM, 2013.

[9] K. Kusumoto, T. Kinnunen, J. Kätsyri, H. Lindroos, and P. Oittinen. Media experience of complementary information and tweets on a second screen. In *Proceedings of the ACM International Conference on Multimedia*, pages 437–446. ACM, 2014.

[10] Ofcom. Communications market report 2010. *Ofcom*, 2010.

[11] M. Sarkis, C. Concolato, and J.-C. Dufourd. The virtual splitter: refactoring web applications for themultiscreen environment. In *Proceedings of the 2014 ACM symposium on Document engineering*, pages 139–142. ACM, 2014.

[12] G. Thomas, P. Mills, P. Debenham, and A. Sheikh. Surround Video. *White Paper WHP 208, http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP208.pdf*.

[13] E. Tsekleves, R. Whitham, K. Kondo, and A. Hill. Investigating media use and the television user experience in the home. *Entertainment computing*, 2(3):151–161, 2011.

[14] R.-D. Vatavu and M. Mancas. Visual attention measures for multi-screen tv. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video*. ACM, 2014.