Methods for user studies of interactive TV in a commercial context

Paul Marrow
BT Innovate & Design
Orion 3/7 PP 1
Adastral Park, Ipswich, UK
+44-1473-641456

paul.marrow@bt.com

Sue Hessey
BT Innovate & Design
Orion G PP 9
Adastral Park, Ipswich, UK
+44-1473-606434

sue.hessey@bt.com

Ian Matthews
BT Innovate & Design
Orion G PP 9
Adastral Park, Ipswich, UK
+44-1473-606433

ian.r2.matthews@bt.com

ABSTRACT
Interactive TV services have attracted much attention in a commercial context because they have the potential to generate more revenue through the scheduling of services tailored to viewer's interests. However there is the problem of finding what to watch, which is why we are interested in recommendation for interactive TV. We discuss work on user identification systems, expert evaluation of existing recommender systems, and focus groups, which have provided us with information upon which to base future field trials of recommender systems on an interactive TV service. These trials will use both quantitative (indirect, large-scale) and qualitative (direct, small-scale) methods.

Categories and Subject Descriptors
H1.2 [**Models and Principles**]: User/Machine Systems – *human factors.* H3.5 [**Information Storage and Retrieval**]: Online Information Services – *commercial services.* H5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *video.*

General Terms
Design, Experimentation, Human Factors.

Keywords
Recommendation, Personalisation, Marketing.

1. INTRODUCTION
We conducted our research as part of a project [1] studying recommender systems. Although most recommender systems have been developed for web browsers we are interested in recommendation for an interactive TV service [2]. User studies of the interaction between users and technology have attracted much attention [3], but because we are dealing with an existing commercial service, our research is unusual in that it must not disrupt the existing service, including any marketing, scheduling and business policy. In addition, we aim to ensure this will lead to an improved user experience which in turn provides a commercial benefit to the company.

2. RELATED WORK
Recommender systems are frequently deployed on interactive web browsers, e.g. [4]. Their use in interactive TV is at an earlier stage, but is attracting a great deal of attention because of the demonstrated benefits that recommendation can bring to interactive web technologies. However, at the present time interactive TV is typically a different experience from interacting with a web page, being more a social or household activity rather than an individual one, involving the viewer being seated further away from the screen, and typically not involving the username/password identification that many websites with recommendation use.

3. METHODOLOGY
Prior to carrying out user studies, we have sought to identify information from existing principles of User Interface (UI) design, as well as evaluating the UI designs of existing recommender systems (RSs), and presenting different UI designs to focus groups. While the information that we can gain in this manner is only indirect, it does show some general principles [5].

3.1 Expert Opinion – User Identification
How should users of a TV recommender system identify themselves (or not)? We identified several key points:

a) Logging on with a username and password for individual identification to the TV system is less likely to be successful than for web-based systems, due to the limited functionality of the remote control.

b) TV-viewing is often transient in nature and the system cannot identify who may have left the room when logging in manually.

c) Avatars- virtual representations of users- in the form of customizable characters may partially solve the problem, having been used extensively in games, for example. There is however issues surrounding this approach as avatars may be deemed childish, and privacy would still need to be achieved by inputting a password.

d) Recommender services can be based on "Time/date/device". Specific identification of an individual is not needed, but instead of the delivery device (e.g. family TV, personal TV, mobile device), and where appropriate of the time and date the request was made. This last option avoids additional input by the user and is therefore best suited to TV-based systems.

3.2 Evaluation of existing RSs
We also carried out two evaluations of UI designs of existing RSs [5], in the form of an expert evaluation and a focus group. The 8 recommender systems chosen for both evaluations were all web-based, reflecting their availability in comparison to TV-based

systems. So the results of this evaluation it may be argued apply more strongly to web-based recommender systems than TV-based ones, but we believe that the analysis, based as it was on well-established principles of UI design, gives some insights that are relevant to both online and TV interfaces.[1]

3.2.1 Expert evaluation

Nielsen's principles for UI design [6], a renowned reference for web usability, was used as a template for assessing the usability of the RSs, with a number of RS-specific attributes included to add relevance to the findings. Each RS was evaluated by one highly experienced UI designer.

Qualitative statements and quantitative ratings were applied to each attribute (where available and relevant) and captured in a spreadsheet, so that high and low-rated items could be identified quickly to draw out best- and worst-practice features of each RS. Then three main strengths and three main weaknesses were identified by the evaluators to summarize each RS's UI and listed in a table in qualitative form. Similar statements made about each positive and negative point across different systems were grouped until seven categories that appeared most frequently were identified. Table 1 shows these categories, some of which are general to web site design and some of which are more specific to recommender systems.

Table 1. Categories identified from the expert evaluation

General (relevant to all website design)	Recommender-system specific categories
Ease of use	User information
Appearance	Recommended items
Navigation/browse	Extra functionality
Help + explanation	

Subsequently this table was then analyzed to identify key themes across the UI design of all RS's evaluated.

3.2.2 Focus Groups

Focus groups [7] were identified as the third stage of UI evaluation to add insights from people not directly involved in the research projects to offer their opinions as potential users of the RSs. The focus group participants were company employees but were not participants in the project: due to the size of the company their responses were unlikely to be biased by knowledge of its objectives. The same RSs were presented to the attendees but additionally, one (out of two conducted) focus group was presented a "flat" (i.e. non-interactive) mock-up of the proposed interactive TV RS.

The focus group moderator led the participants through a series of open-ended questions covering general and specific topics relevant to the experience of each RS presented, and opened up the discussion to think freely about how this may be applied on the TV.

3.3 UI design evaluations: conclusion

The blend of methodologies used in this research exercise yielded results and conclusions which are beneficial in anticipating how viewers would react to recommendations delivered via TV when

we carry out user studies on a pre-existing service (see [5] and Table 1).

4. FUTURE WORK

The user experience of interactive technology can be studied at many levels, whether it is on TV, web pages or with other technology. At one level objective measures of user interaction can log user actions and use these to report user behavior to analysts. At another level subjective responses by users can provide detailed information about an individual's own reactions. Other work has shown [8] how these are just two aspects of how reaction to recommender systems can be measured.

In order to carry out an analysis of user behavior of an interactive TV recommender service in a business context, we propose to split our study into two parts. The first will address the business benefits of recommendation in interactive TV, using quantitative information gathered indirectly from large numbers of users. The second will study user experience in more detail, through direct interaction with a smaller group of users across a few households. By combining the results of these studies we expect to gain both academic and commercial insights about recommender services for interactive TV.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] http://www.mymediaproject.org/default.aspx

[2] http://www.btvision.com/

[3] Sundstedt, V., Whitton, M. and Bloj, M. 2009 The whys, how tos, and pitfalls of user studies. ACM Intl. Conf. on Computer Graphics and Interactive Technologies, New Orleans. DOI= http://doi.acm.org/10.1145/1667239.1667264

[4] Dias, M.B., Locha, D., Li, M., El-Deredy, W. And Lisboa, P.L.G. 2008 The value of personalised recommender systems to e-business: a case study. In: ACM Conf. On Recommender Systems 2008, Lausanne. DOI= http://doi.acm.org/10.1145/1454008.1454054

[5] Meesters, L., Marrow, P., Matthews, I., Hessey, S., Bollen, D. and Graus, M.2009. User Design Specification. User Control Test Results Report. MyMedia Project ICT-2008-215006 Deliverable D1.4. Download from http://www.mymediaproject.org/deliverables.aspx

[6] http://www.useit.com/papers/heuristic/heuristic_list.html

[7] Finch, H. and Lewis, J. 2008 Focus Groups. Ch.7 in *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, pp. 170-198. Sage, London.

[8] Meesters, L., Marrow, P., Knijnenburg, B., Bowhuis, D. and Glancy, M. 2009. End-user Recommendation Evaluation Metrics. MyMedia Project ICT-2008-215006 Deliverable D1.5. Download from http://www.mymediaproject.org/deliverables.aspx

---

[1] It should be emphasized that inclusion in, and the results of the study, do not represent an endorsement or criticism of the web site or company concerned.