Evaluating *BluScreen*: Usability for Intelligent Pervasive Displays

Maria Karam[1], Terry Payne[2], Esther David[3]

[1]*Centre for learning technologies, Ryerson University, Canada*
marial.karam@ryerson.ca
[2] *School of Electronics and Computer Science, Southampton University, United Kingdom*
trp@ecs.soton.ac.uk
[3]*Ashkelon College, Israel*
astrdon@ash-college.ac.ik

Abstract

Within a ubiquitous environment, market-based approaches can be used to select the most appropriate material for a public display, depending on factors such as the audience's preferences and diversity of interest. Likewise, strategies used by agents to compete for customer attention should strive to be rational, based on contextual observations of user-preferences within the local environment and include a reward mechanism based on audience responses. But while such systems currently exist, utilizing Bluetooth-enabled mobile phones to uniquely identify and detect the presence of individuals within a localised environment, there is little known about their effectiveness, or even how to assess usability for these systems. In this paper, we present the details a user study that contributed to the development of an interaction model that supports a structured methodology for evaluating intelligent pervasive displays.

Keywords: Pervasive Computing, Context Technology, Mobile Agents, Interaction models, usability studies, intelligent interfaces.

1. Introduction

Public electronic displays are increasingly being used to provide information to users, to entertain (e.g. showing news bulletins), to inform, or to advertise products within public and semi-public environments such as office spaces, airports, city centres, and retail stores. Within these displays, advertisers typically utilise a variety of delivery methods to maximise the number of different adverts displayed, and thus increase their overall exposure to target audiences [1]. However, these methods are typically naïve and fail to take into account details about the current audience, such as their interests or whether or not they have previously seen the advertised content.

In parallel, the prevalence in sensing technology is growing, with both RFID and smartcards being increasingly used as a means of tagging items, or providing access mechanisms to offices or residences. Short-range wireless technologies such as Bluetooth facilitate personal area networks, where devices communicate with peers within a localized space (typically 5-10 meters). This has enabled the development of a number of pervasive and ubiquitous computing projects, such as location detection [2], [3], or identifying co-located users [4]. The synergy of combining such near-field wireless technologies with personal devices is now being realized though a number of *interactive* intelligent displays that support communication with a user through the active use of handheld devices such as PDAs or phones, or to a closed set of known users with pre-defined interests and requirements [5], [6]. Such systems assume prior knowledge about the target audience, and require either that a single user has exclusive access to the display, or that users carry specific *tracking* devices [7], [8] so that their presence can be identified. These approaches fail to work in public spaces, where no prior knowledge exists regarding users who may view the display, and where such displays need to react to the presence of several users simultaneously. However, while this is a growing area of research and development within the domain of pervasive computing, very little research focus has been aimed towards understanding the effectiveness of such systems, and how interaction researchers can effectively evaluate intelligent interfaces. In addition, little is known about what interaction researchers should be evaluating these systems for. In this paper, we present preliminary work towards quantifying a set of parameters for assessing the usability of intelligent pervasive information dissemination systems. This research begins with a user study designed to evaluate the effectiveness of intelligent pervasive displays. The work we present in this paper extends our ongoing research on *BluScreen* [9], [10], an intelligent public display framework that utilises a novel wireless approach to detecting nearby users to improve the selection of adverts for display.
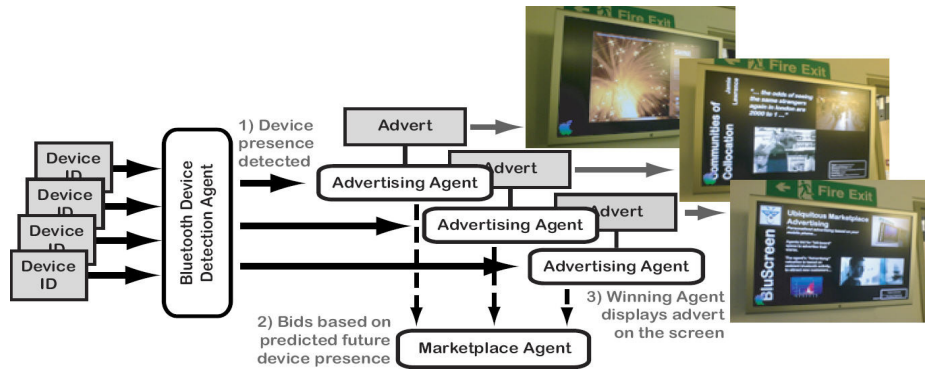
Figure 1. The BluScreen Agent Architecture for a single intelligent display.

1.1 Evaluating pervasive displays

Traditionally, user studies seek to uncover quantitative results to demonstrate ease of use, user satisfaction, or to uncover usability problems with the system. However, these types of studies are not designed to reveal information about the user's perspectives or expectations of these systems. Our research has taken on this challenge by augmenting our user study with a set of qualitative instruments – interviews, questionnaires, and observations – to reveal what the user's role is in this type of interaction. While the quantitative results from this study suggest that the system is effective in meeting its goals of delivering novel information to users, the qualitative results contributed to the development of an interaction model. Interaction models represent an approach advocated by researchers in human computer interaction (HCI) that promotes a more scientific method to gaining an understanding of the factors that effect interactions with a system [11], [12]. It is seen as a valuable tool in supporting the comparison and design of different systems using common parameters [16], while supporting a methodological approach to evaluating and designing interactive systems. Once the interaction model is in place, further usability studies can be conducted that target specific areas of pervasive systems by manipulating the parameters that are presented within the model. This supports a more precise approach to conducting usability studies to identify specific problem areas with the interaction, and ensure that a mechanism is in place to support the development of systems that are improved, not just different. The remainder of this paper is organised as follows: Section 2 describes the deployed system and use of Bluetooth-devices Section 3 presents our user study, which incorporates both quantitative and qualitative methods to evaluate the system. This is followed by a discussion of the process taken to develop the interaction model, and presents an example of its application to the design of usability studies in Section 4. We conclude with a summary of this research, and overview of future work in section 5.

2. *BluScreen*

The current approach we employ for advert selection within *BluScreen* is to maximise the exposure of as many adverts as possible to as wide an audience as possible (i.e. to maximise the number of distinct adverts seen by the population of users). In doing so, the main advantage of our system design is that it achieves this goal without: (i) any prior knowledge on the audience, (ii) the need for any specific action by the user, or (iii) the need for any client-based software. Moreover, unlike interactive public displays, our detection technology facilitates an awareness of several devices simultaneously. Previous publications describing the *BluScreen* system [9], [10], [13] have examined game-theoreic approaches for selecting adverts using mechanisms such as Agent-based Auction Theory [14] or Queueing Theory [15]. Whilst such techniques can be shown to yield optimal (or near-optimal) results through simulation, the evaluations used do not consider how human subjects react to situated displays in a working environment. Thus, within this paper, we aim to establish a methodology for conducting user studies on inference-based pervasive displays.

Three *BluScreen* prototypes have been developed and deployed to evaluate the feasibility of the auction-based approach. In two cases, 60 inch plasma screens were located in the entrance or foyer of different buildings belonging to the School of Electronics and Computer Science, at the University of Southampton. A third deployment consisted of a 23 inch flat-screen display deployed outside an office adjacent to the corner of two corridors and an exit (thus maximising visibility to individuals moving within both corridors). In each case the environment was scanned for Bluetooth devices every 20 seconds\footnote{The choice of a 20 second scanning cycle was determined by evaluating different scanning cycle lengths with varying numbers of nearby devices.}. Twelve adverts were generated, describing a range of topics including research projects, upcoming events, and general information. Several types of agents

have been designed within our *BluScreen* architecture (illustrated in Figure 1), and are described in more detail in [9]

3. Evaluating *BluScreen*

The domain of pervasive computing poses many challenges when attempting to assess the effectiveness of a system. Unlike traditional interfaces, where users explicitly interact with the computer using direct input devices such as the mouse or keyboard, intelligent pervasive displays rely on implicit input; *BluScreen*, through the detection of discoverable devices, accepts input when a user is within range of the system. Further control of the system must then be inferred by the agents, who control the display by presenting adverts based on the information stored in the system about a given user. With this particular class of intelligent pervasive displays, we cannot explicitly evaluate a user's performance with the system, but must instead consider how well the system can achieve its goals. But to do this, we must assess external factors to the system, such as what users expect from the system, and how those expectations are influenced by these factors, and how they can be measured. In the case of *BluScreen*, we approach these user-related factors by augmenting the system evaluation with qualitative methods. This involved conducting an experiment to determine effectiveness of *BluScreen* in achieving its information delivery goals, in addition to conducting interviews with the participants, and asking then to keep notes (self reports) about their experiences, perceptions, and comments about the system. We discuss the details of the study next.

Table 1: The table shows the results from our user trials.

TotalAdverts	% Recall	% Relevant	% Non-relevant	% Extra
13	84.62%	46.15%	30.77%	30.77%
16	62.50%	43.75%	56.25%	0.00%
5	100.00%	60.00%	20.00%	20.00%
7	85.71%	57.14%	28.57%	42.86%
20	80.00%	60.00%	35.00%	0.00%
21	76.19%	33.33%	33.33%	0.00%
20	80.00%	60.00%	30.00%	0.00%
20	40.00%	65.00%	30.00%	5.00%
15.25	76.13%	53.17%	32.99%	12.33%

3.1. The Initial User Trials

The *BluScreen* prototype had been deployed for over one year, with screens in two different locations within the school. To assess the effectiveness of the *BluScreen* architecture, we ran a one-week long study and enlisted 8 male participants between the ages of 23 and 40 who had daily access to both of the displays. Each participant agreed to turn their bluetooth devices onto discoverable for the duration of the study. The system identified participants by their bluetooth devices, and at the end of the week, they were asked to identify the adverts that *BluScreen* recorded displaying to each for both displays. Quantitative measures were recorded as the number of adverts that participants recalled seeing, those that were not recalled, and those that were seen in addition to what the system recorded. Qualitative data was gathered through the interviews and questionnaires administered at the end of the study.

3.2. Quantitative Results

To assess the effectiveness of the *BluScreen* system, we tested the intended goals of the system: to present relevant and novel information to users. These quantitative results are presented in Table 1, and discussed in the following sections.

- **Advert recall:** The average percentage of information that users recalled seeing on one or both of the screens was over 75\%, suggesting that the information presented on the display was acknowledged by the participants. The actual results are shown in table \ref{stats}.
- **Advert relevance:** The percentage of adverts that the participants deemed as interesting or relevant to them was just over 50\%, however at this stage, there are only a limited number of adverts being used in the *BluScreen* prototype.
- **Non-relevant adverts:** The percentage of adverts that participants deemed as un-interesting was just over 32\%, however all participants found a higher number of adverts interesting than not. For the remaining 18\% of adverts, participants said that they could find those adverts interesting however they would need to have more information to make a decision.
- **Extra Adverts:** There were just over 12\% of the adverts that participants noticed in addition to those explicitly presented to them by *BluScreen*. This may have been due to a number of factors including those associated with their bluetooth devices, *BluScreen*, or the user's behaviour.

3.3. Qualitative Results

Responses from the interviews and questionnaires were coded, revealing three main categories, which could potentially influence the effectiveness of the system.

- **User preferences:** Participants said that they would stop and read the adverts that appeared interesting to them. This included new adverts that they hadn't seen or noticed before.
- **Cognitive constraints:** Participants said that the level of attention they gave to the displays was dependent on the amount of time they had to spend around the displays. When in a hurry or focused on

another task, they often ignored the screens, but would spend more time reading the adverts when they were not rushed or while waiting for the lift or sitting in the lobby.

• **Environmental constraints**: Even when participants were within close proximity of the screen, they could not always see the screen. This was due to the configuration of the chairs, or being at the periphery of the screens.

3.4. Discussion

The quantitative results of this study are promising, suggesting the participants noticed the adverts, found most to be relevant and interesting, and felt there was value in the system. However, with the additional factors revealed through the qualitative measure, we were able to capture details about what the users were doing during the interactions, what they were thinking, and what they felt the system should be providing in terms of the interaction. These results are categorised and presented as part of the interaction model in the next section.

4. Evaluating Pervasive Displays

We begin this section with a discussion about interaction models, their role in supporting effective usability studies for novel systems, followed by a discussion about the model, and a proposed methodology for evaluating pervasive system.

4.1. Interaction Models

Michel Beaudouin-Lafon [11] introduces the notion of an interaction model to facilitate the repeatability of experiments across different interactive systems, and to promote a shift away from designing interfaces, towards designing interactions. In previous work, an interaction model was used to describe gesture interaction systems in an experiment that evaluated user tolerance levels for errors in computer-vision recognition systems [12]. That study was situated in the domain of ubiquitous computing for supporting distance interactions with a visual display using device-free gestures. In that research, a similar approach was taken to develop the interaction model for the gesture system, which represented another novel form of interaction. The model focuses on identifying parameters that can be used to evaluate three major influences in designing usable interactions: the user's goals (what users expect from the interaction), system performance (accuracy rates, response speeds, etc.), and the interaction context (environmental, social, or other factors external to the system). We next discuss the methodology we used to develop the interaction model for intelligent pervasive

system interactions, which is followed by an example of how the model can be used to guide future evaluations.

4.2. A proposed methodology

When approaching the evaluation of any novel interaction system, where few, if any systems have received extensive usability tests, we propose an approach to bootstrapping the process of determining what factors, external from the system, influence the system's effectiveness and the users' satisfaction.

Step 1: User trials. The user testing phases of an interactive system explores specific features of the system to ensure that it is working as designed. During this phase, user trials can be employed to determine how effective a system is in meeting its stated goals. This involves having users interact with the system, and gathering quantitative data to determine how well it supports user interactions.

Step 2: Qualitative measures. Qualitative measures are run in parallel to the user trials to gather subjective data about the users experience with the system. These measures can include self reports, interviews, questionnaires, or observations of the participants while they use the system.

Step 3: Extending the interaction model. An analysis of the qualitative data can reveal factors that influence the interaction based on identifying what users expect from the system, and what external factors effect usability.

These factors are then organized into the three categories of the interaction model to reveal additional characteristics about the system, the user, the contexts surrounding the interaction, and how each relates to influence the system's usability. The three categories chosen for this interaction model are presented next, incorporating the results from our qualitative studies as parameters that describe the interaction.

Interaction context refers to factors associated with the environment in which the interaction takes place, and is external to the system. Parameters included are:

• **User Profiles:** A general characterisation of the users based on their expected roles (academic, staff, visitors), their expected exposure to the displays (frequent, seldom), and their areas of interest (events, research, general).

• **Environmental Factors:** Specific factors include the location of the *BluScreen* displays, and the associated contexts that constrain the interaction or restricted access. These include the types of behaviours

that a location supports such as in a walkway (short exposure), an elevator (medium exposure), or a lounge (long exposure). Specific descriptions of the parameters may need to be altered to reflect different contexts, locations, or users.

System performance. Each system seeks to achieve specific performance goals. For *BluScreen*, one goal is to maximize the number of relevant adverts presented to each user however any system requires unique parameters to describe its performance. Those specific to intelligent pervasive displays include:

- **Internal Factors:** For *BluScreen*, information dissemination is one specific system goal. Additional factors include the rate in which the system changes its display, collocation detection, response time once a device is detected, or the duration of time a user must be within range to be detected. These factors can be measured empirically.

- **External Factors**: Different mobile devices may effect collocation detection. For example, devices with stronger signals or longer ranges may improve the systems' performance. Empirical methods are also used to determine external factors, which could be applied as a rating system for different devices.

- **User Goals:** Each system provides specific functionality to its users. For example, *BluScreen* presents novel, interesting, and relevant information on pervasive displays based on detecting a user's presence and preferences.

But as with any interactive device, the goals of the system may be different to the expectations users have in terms of its behaviour or functionality. Several participants in our study expressed this, stating that they expected the system to provide a more explicit form of interaction. This may be due to a misunderstanding of how the system works, or imply based on what the user wishes the system to do. This represents a conflict between the intended goals of the system, and those of the users interacting with it. To address this, we suggest the following parameters be used in evaluations:

- **Expectations**: Identify any of the users' expectations that are not in line with the intended functionality of the system. Discrepancies may lead to lower satisfaction ratings, and can be avoided if these are acknowledged in the results.

- **Cognitive Factors**: Identify any the cognitive states that may occur during interactions. These factors include the user's level of focus, distraction, attention, or interests.

- **Physical Factors:** Identify different behaviours that users will exhibit when interacting with the system. For example, some users tend to walk too quickly to read the text on the displays, while others move more slowly, having more time to view the adverts.

Each of the parameters may be varied, creating different experimental conditions, or used as metrics in designing future systems, or for supporting measurable improvements to the interface, and to the interaction. By addressing each of the categories of the interaction model, evaluations can produce more accurate measures of a system's effectiveness, its usability, and the factors that can be altered to improve subsequent versions. We now demonstrate how the interaction model can be applied towards a methodology for assessing intelligent pervasive displays.

4.3. A methodology for evaluating intelligent pervasive systems

The interaction model provides a structure for organizing and relating the critical parameters that influence system interactions. By using a more standardised approach to assessing usability, researchers and developers can compare different systems along similar metrics, supporting comparisons that are based on specific, measurable parameters [16]. Such approaches are common in the engineering sciences, and given the relative novelty of intelligent pervasive interactions, the interaction model can serve as one method of formalising the evaluation process. Although this interaction model represents preliminary work, there is enough information in place to inform the design of our next study on *BluScreen*, which we introduce next.

4.4. Using the interaction model

We are applying the interaction model to design an experiment to determine the effectiveness of *BluScreen* in a heavily used public space. Each of the categories of the model will be represented in the study (user goals, system performance, and interaction context). Parameters from each category will represent variables, or constants used in the experiment design. The study is in its preliminary design stages, and we are considering the following scenario: participants (user profile suggests they are shoppers, or shop employees) are walking (physical context) to work or to the shops. Along the way, they encounter a series of displays located in a corridor of a shopping mall (location), but have limited time and cognitive resources available to view the displays. Based on these parameters, we want to determine if *BluScreen* can effectively deliver targeted information to users as they walk by the screens. The experiment will require a control condition

(users are carrying detectable devices), and an experimental condition (users do not have discoverable devices). We expect to recruit participants who work at the mall, or the surrounding area to take part in the experiment. While this represents only one set of parameters that define the experiment, different combinations of the parameters will lead to a unique approach to conducting evaluations, enabling the exploration of many factors that define the interactions with intelligent pervasive displays.

5. Future work and conclusions

BluScreen is an intelligent, auction-driven agent framework for delivering targeted material to an audience, observed through the presence of Bluetooth devices. Through the use of an individually rational bidding strategy, agents attempt to determine whether their advertisement will be novel to a given audience, and thus use this to mediate their behaviour. *BluScreen* assumes no prior knowledge of the audience; instead it builds up a profile of user behaviour based on details of device presence observed during different advertisements. But while *BluScreen* is currently deployed as a prototype that seeks to provide intelligent information to users through pervasive displays, future applications of the interaction model can serve as a guide to developing improved versions of the system. The model can also assist in making informed decisions about which new features or approaches can improve the overall functionality of *BluScreen*. Current considerations include enabling explicit interactions, so that user may send and receive messages from *BluScreen*, and exploring the use of facial recognition as an alternative method of identifying users. The interaction model may also be extended across the domain of intelligent pervasive interactions, to support system evaluations, to measure specific interaction effects, and ensure that new versions of a system are better, and not simply different.

Acknowledgements

References

[1] A. Ranganathan and R. H. Campbell, "Advertising in a pervasive computing environment," in WMC '02*: Proc. 2nd Int. Wkshp. on Mobile commerce*. NewYork, NY, USA:ACM Press, 2002, pp. 10–14.

[2] A. Genco, "Three step bluetooth positioning." in LoCA, ser. *Lecture Notes in Computer Science*, T. Strang and C. Linnhoff-Popien, Eds., vol. 3479. Springer, 2005, pp. 52–62.

[3] A. Madhavapeddy and A. Tse, "A study of bluetooth propagation using accurate indoor location mapping." in Ubicomp, ser. *Lecture Notes in Computer Science*, M. Beigl, S. S. Intille, J. Rekimoto, and H.Tokuda, Eds., vol. 3660. Springer, 2005, pp. 105–122.

[4] O.J¨arkvik and B. Ljungquist, "Group detection using bluetooth," in Ubicomp, ser. *Lecture Notes in Computer Science,* M. Beigl, S. S. Intille, J. Rekimoto, and H.Tokuda, Eds.,vol. 3660. Springer, 2005.

[5] K. Cheverst, A. Dix, D. Fitton, C. Kray, M. Rouncefield, C. Sas, G. Saslis-Lagoudakis, and J. G. Sheridan, "Exploring bluetooth based mobile phone interaction with the hermes photo display," in MobileHCI '05*: Proceedings of the 7th international conference on Human computer interaction with mobile devices&services*. NewYork, NY, USA: ACM Press, 2005, pp. 47–54.

[6] J. F. McCarthy, T. J. Costa, and E. S. Liongosari, "Unicast, outcast & groupcast: Three steps toward ubiquitous, peripheral displays," in UbiComp '01*: Proceedings of the 3rd international conference on Ubiquitous Computing. London,* UK: Springer-Verlag, 2001, pp. 332–345.

[7] J. Hightower and G. Borriella, "Location systems for ubiquitous computing," *IEEE Computer,* vol. 34, no. 8, pp. 57–66, 2001. [Online]. Available: citeseer.ist.psu.edu/rey01location.html

[8] R.Want,A. Hopper,V.F. ao, andJ. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.,* vol. 10, no. 1, pp. 91–102, 1992.

[9] T. R. Payne, E. David, M. Sharifi, and N. R. Jennings, "Auction mechanisms for efficient advertisment selection on public display*," in Proceedings of European Conference on Artificial Intelligence (ECAI 2006)*, Trentino, Italy, 2006, pp. 285–289.

[10] A. Rogers, E. David, T. Payne, and N. R. Jennings, "An advanced bidding agent for advertisement selection on public displays," *in Proceedings of the International Joint Conference on Autonomous Agents and Multi-Agent Systems* (in press), 2007.

[11] M. Beaudouin-Lafon, "Designing interaction, not interfaces," in AVI '04: *Proceedings of the working conference on Advanced visual interfaces*. NewYork,NY, USA:ACM Press, 2004, pp. 15–22.

[12] M. Karam and m. c. schraefel, "Investigating user tolerance for errors in vision-enabled gesture-based interactions," in AVI '06: *Proceedings of the working conference on Advanced visual interfaces*. New York, NY, USA: ACM Press, 2006, pp. 225–232.

[13] M. Sharifi, T. Payne, and E. David, "Public display advertising based on bluetooth device presence*," in Proceedings of the Mobile Interaction with the RealWorldWorkshop* (MIRW 2006), 2006.

[14] M. J.Wooldridge, *Introduction to Multiagent Sys*tems. NewYork, NY, USA: JohnWiley& Sons, Inc., 2001.

[15] D. Gross andC. M. Harris, *Fundamentals of Queueing Theory.* Wiley, 1998.

[16] W. M. Newman, "Better or just different? on the benefits of designing interactive systems in terms of critical parameters," in DIS '97: *Proceedings of the conference on Designing interactive systems.* New York, NY, USA: ACM Press, 1997, pp. 239–245.