FINAL AUTHOR VERSION – INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES – ACCEPTED JUNE 23rd, 2009
DOI 10.1016/j.ijhcs.2009.06.002

Considerate Home Notification Systems:
A User Study of Acceptability of Notifications in a Living Room Laboratory

Martijn H. Vastenburg, David V. Keyson, Huib de Ridder

Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628 CE, Delft, The Netherlands

{m.h.vastenburg, d.keyson, h.deridder}@tudelft.nl

Corresponding author:

Martijn Vastenburg, m.h.vastenburg@tudelft.nl, tel. +31-15-2784960, fax +31-15-2787316

Abstract
Recent field experiments on acceptability of notifications in the home showed that people generally want to be informed of urgent messages as soon as possible, whereas non-urgent messages should not be presented at all. A possible way to improve the acceptability of a notification might be to adjust the presentation mode and the timing of notifications to the message content and to the state of the user. For example, acceptability might be improved by considering user activities when selecting the best time to present the message. The relation between acceptability, presentation mode and timing has not been formally studied in a controlled home setting before. This paper presents the results of a user study, in which ten participant couples were asked to engage in everyday home activities, and to subjectively rate factors that were expected to influence acceptability. The study was situated in a living room laboratory in which the user activities and the timing of notifications were controlled. Questionnaire data was evaluated using cluster analysis in order to construct a semantic model that describes the relationship between user, system and environment. The key findings in the present study are: (1) acceptability could be improved by adjusting the level of intrusiveness of the presentation to message urgency: urgent messages should be presented intrusively, medium-urgent messages unobtrusively, and (2) non-urgent messages should be postponed until the message urgency has increased, or skipped if the message urgency never exceeds the predefined presentation threshold.. Surprisingly, the user activities at the time of notification were not found to influence acceptability. These findings have resulted in a model of acceptability of notifications for the design of future home notification systems.

Keywords: notification systems, considerate home environments, ubiquitous computing, user attention, ambient displays, user engagement

1 Introduction

An increasing number of products in the home are competing for the user's attention (Den Hartog et al., 2004). Email notifiers, medicine reminders, washing machines, mobile phones, instant messengers and many other notification providers push information to their users, even though the user-perceived value of some of the messages can be questioned. Considering the growing number of information providers, and the increasing number of messages across products, users might soon be overwhelmed with notifications. As early visionaries of ubiquitous computing, Weiser and Brown (1997) recognized the need for calmness; when computers are all around, these systems need to be designed "so that the people being shared by the computers remain serene and in control". The shift from functional use and performance to meaningful presence of technology has also been recognized by Hallnäs and

DOI 10.1016/j.ijhcs.2009.06.002

Redström (2002); they emphasize the need to design products that co-exist with users and with other products in their life-world.

A large body of knowledge exists on technology for building intelligent products that are considerate of the user as well as the context of use (e.g., Horvitz et al., 1998; Garlan et al., 2002; Altosaar et al., 2006). Notably, several projects were recently conducted in the area of context-aware telephony (e.g., Khalil and Connelly, 2006; Avrahami et al., 2007; De Guzman et al., 2007). The focus in these projects tends to be on providing the caller with contextual information about the receiver in order to decrease the frequency of mismatch between the callers' decisions and the receivers' desires. Even though people might appreciate context-aware notification systems similar to context-aware telephones, the relation between the user state, context, and the acceptability of notifications in the home has not been studied before. One needs to know how people experience notifications in the home, before one can design a context-aware system to improve acceptability of notifications.

1.1 Considerate home notification systems

A recent field study showed that acceptability of notifications in the home can be represented using a cost-benefit model (Vastenburg et al., 2007). Acceptability can be considered as a tradeoff between the cost of interrupting the user's activities and the value of receiving the notification message. Attention level was found to be a predictor of the cost of interrupting the user's activities, and message urgency was found to be a predictor of message value. Therefore, the cost-benefit model is based on two primary factors of acceptability: attention level and message urgency (Fig. 1). Cost-benefit models could be made part of considerate home notification systems, in order to to improve acceptability of notifications. Similar cost-benefit mechanisms are used in the *attentive user interface* paradigm (Horvitz 1999; Horvitz et al., 2003; McCrickard et al., 2003; McCrickard and Chewar, 2003; Vertegaal, 2003) in which system actions are optimized in terms of minimizing attention cost and maximizing utility. The attentive user interface paradigm is based on the observation that attention is a limited resource. Subtle cues about attention are available; for example, speech can be used as a cue for people being engaged in a conversation. At the time of conversation, the attention resource might be fully used for the conversation activity; a notification at the time of a conversation might therefore interrupt the conversation. An automated system might be able to sense these cues, but many cues tend to be ambiguous. Systems therefore need to be able to reason about uncertain evidence, in order to take appropriate actions.

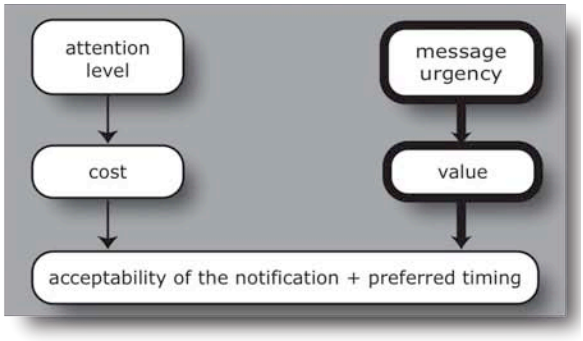

Fig. 1. Cost-benefit model of acceptability of notifications resulting from prior field studies. The subjective acceptability and preferred timing are linked to the attention level and the perceived message urgency. The bold arrow indicates message urgency to be the primary indicator of acceptability and preferred timing

As early pioneers in the area of attentive user interfaces, Horvitz et al. (2003) demonstrated the potential use of Bayesian networks for computing the cost and value of notifications.

DOI 10.1016/j.ijhcs.2009.06.002

Their *Notification Platform*, a cross-device alerting system based on attention-sensitive mechanisms, collects notifications from multiple sources, and automatically selects the best time for presentation. The platform collects attentional cues by using perceptual sensors (e.g., a microphone), device interaction monitoring (e.g., tracking keyboard events), application monitoring (e.g., calendar events) and monitoring time of day and date. The attentional cues are then used to derive the attentional focus and workload, which in turn are used to compute the cost and value of actions. The system not only computes the actual cost, but also the expected cost of delayed presentation. The expected value of messages was predicted using a classifier, which was trained with sample messages.

The approach taken by Horvitz et al. in the office setting, might also be applicable in the home environment. Whereas much is known about the factors underlying acceptability of notifications in an office setting, not much is known about the home environment. Do people experience notifications in the home in a way similar to notifications in the office? And how do acceptability, presentation mode and timing relate in a home environment? The factors underlying acceptability in the home need to be known, in order to be able to train a computational model to capture this tacit knowledge. The present study therefore serves as a first step towards creating a "Notification Platform" for the home environment.

Key factors underlying acceptability of notifications in the home include *message urgency* and *engagement in activities*. In our earlier field studies, urgent messages –such as medicine reminders- were found to be valuable to the users, whereas non-urgent messages -such as a reminder to water the plants- were perceived as being of low value. Engagement in activities, defined in terms of concentration, social interaction, urgency of activities, and interruptibility, was also found to affect acceptability. When participants were highly engaged in their activities, for example in a telephone conversation, the cost of interruption was shown to be high. When user engagement in activities was low, for example when participants had just returned from shopping, the cost of interruption was shown to be low.

Cost and benefit of notifications could potentially be manipulated by intelligent systems. To increase acceptability, the cost of notifications could be lowered. Perceived cost primarily depends on the effect of the interruption on the ongoing user activities and the intrusiveness of the presentation of notifications. Cost can therefore be affected by changing the mode and timing of the presentation. Presentation mode can move from the center to the periphery of the user's attention (Weiser and Brown, 1997; Ishii et al., 1998), thereby reducing the intrusion level. To avoid high-cost interruptions of activities, messages can also be postponed until a moment in time that is less disruptive for the users.

Acceptability could also be improved by increasing the benefit of notifications. Perceived benefit of notification messages depends not only on the message content. It seems reasonable to expect that perceived message value and urgency depend on many factors including message structure, style, phraseology, and the relationships between messages. Perceived benefit might also vary between users and between situations. Therefore, a single notification message might result in entirely distinct user ratings of perceived benefit. Notification systems could increase the benefit of notifications in many ways, e.g., by postponing messages until the message urgency has increased, by aggregating messages, by formulating and presenting messages in a more urgent style, or by showing a message to a different user.

1.2 Level of intrusiveness

Ubiquitous computer systems might be embedded in the actual activities of everyday life, resulting in "calm" technology that moves back and forth between the center and the periphery of human attention (Weiser and Brown, 1997). Towards creating systems that adapt their level of intrusiveness to the context of use, researchers and designers face a design

DOI 10.1016/j.ijhcs.2009.06.002

challenge in terms of creating displays that support multiple levels of intrusiveness in conveying information (Ishii et al., 1998; Abowd and Mynatt, 2000).

An example of a display that supports multiple levels of intrusiveness is AuraOrb (Altosaar et al., 2006), an ambient notification system that uses eye contact as a social awareness cue. The orb conveys messages in the least-disruptive way. It starts by showing an ambient light notification, i.e., a visual cue in the periphery of the user's attention, to preserve user attention. A summarized version of the message is presented on the orb, i.e., in the center of the user's attention, only when the user shows interest by looking at the orb as indicated by eye contact. When the user touches the orb, the full message is shown on a computer screen. Although the AuraOrb is an elegant example of how products could move between unobtrusive and obtrusive presentation in a socially acceptable manner, the orb is unaware of the value of the messages itself, resulting in potentially unwanted interruptions in case of non-urgent messages. To be able to link presentation style to message urgency, AuraOrb would need to be linked to an urgency prediction mechanism.

In designing a display that supports multiple levels of intrusiveness, the level of intrusiveness can be directly related to the perceived contrast between the display and the environment. In a "calm" living room, a highly animated display would immediately draw the user's attention. On the other hand, in a lively environment, users might not even be aware of an animated display. Furthermore, transitions between presentation states are crucial in terms of intrusiveness. Abrupt changes in the display attract the user's attention (Matthews et al., 2005). In short, messages could be presented non-intrusively using slow transitions, static presentation, and low color contrasts. A high level of intrusion could be created using fast transitions, animation (McCrickard et al., 2001), high color contrasts and audio.

In the case of notification systems, system messages are expected to be less disruptive to user activities when presented in the periphery of the user's attention (Maglio and Campbell, 2000; McCrickard and Chewar, 2003). Non-urgent messages could be presented non-intrusively, whereas urgent messages might need to be presented in the center of the user's attention, since immediate action is needed. The effect of level of intrusiveness on acceptability will be studied in the present study.

1.3 Everyday activities

In work situations, the primary goal for notification systems is to communicate information effectively and efficiently. Prior studies in the area of human interruptibility and notification systems generally focus on tasks, goals and attention resources, and measure the cost of interruptions in terms of objective task performance scores (McFarlane, 1998, 1999; McFarlane and Latorella, 2002; McCrickard et al., 2003; Gievska and Sibert, 2005). Even though experiential effects of notifications such as annoyance and anxiety have been studied before (Bailey and Konstan, 2006; Iqbal and Baley, 2008), the main body of earlier work focused on task performance rather than user experience.

In order to embed notifications into everyday life, an understanding is needed on how people experience notifications during everyday activities. Whereas professional activities tend to be well-structured and goal oriented, user activities in domestic environments are typically informal and unstructured (Abowd and Mynatt, 2000; Hughes et al., 2000; Eggen et al., 2003). It therefore may be more appropriate for a notification system in the home to use subjective, non performance-based rather than objective, performance-based measures of message value and cost of interruption. These subjective measures should eventually be automatically predicted by an aware system, thereby enabling the calculation of the acceptability of notifications without disturbing the users.

Present studies that do consider acceptability of notifications and interruptibility in the home setting are generally limited to explorative studies. An interesting example is a study by

DOI 10.1016/j.ijhcs.2009.06.002

Nagel et al. (2004), which relates interruptibility and user activities. Nagel et al. used experience sampling to measure activities and self-reported availability to interruption in the home setting. Subjects were asked to fill out a survey, which was presented on a PDA approximately twice per hour. People were asked to describe their social setting (being alone or being together with other people), location (a room in a house, a location at work, etc.), their activities (watching TV, sleeping or napping, etc.) and their availability to interruptions ("Would this be a good time for an adult family member to catch up on today's events with you?"). Interestingly, some leisure activities (e.g., watching TV) turned out to be predictors of availability, whereas other leisure activities (e.g., reading a book) had only marginal significance as a predictor. These differences may have been caused by individual differences. Face-to-face conversation was found to be the most prominent activity (46% of the surveys). Conversations were found to be significantly and negatively correlated to availability; engaging in a conversation apparently made people less available to interruption. Leisure activities accounted for 29% of all activities reported, including watching TV, game playing and listening to music. Interviews suggested that transitions between activities would link better to being unavailable than the particular activities themselves, because supposedly transitional activities provide mental bridges between social roles. The study by Nagel et al. shows the significance of user activities in relation to interruptibility, and suggests that people at home are less receptive to interruptions during activity transitions.

Nagel's finding of people not being receptive to interruptions during activity transitions is not in line with the findings by Miyata and Norman (1986) and Ho and Intille (2005). According to their studies, attentional costs of interruptions between tasks might are lower, since the attentional demands caused by the previous task have ended, and new tasks have not started yet. In a study on notification systems for mobile devices, notifications that were delivered during activity transitions were generally found to be more easily accepted by the participants (Ho and Intille, 2005). In the case of everyday activities, user engagement in activities is expected to be lower when transitioning between activities, resulting in a high acceptability of notifications. An intelligent notification system in the home could eventually link notifications to activities as well as activity transitions, similar to linking notifications to a workflow system in a work setting (Maglio and Campbell, 2000; Carroll et al., 2003). In studying notifications, one should thus not only measure the acceptability during activities, but also in between activities.

Our prior field study showed that concentration level, social interaction level and urgency of the current activities are useful attributes towards determining the extent of user engagement in activities in relation to interruptibility. These attributes will be taken into account in the present study while measuring how people experience notifications in relation to their current activities.

1.4 Focus

Today's notification systems, such as mobile phones and PDA's, are generally not capable of adapting the presentation mode and timing of notifications to the ongoing user activities and the state of the environment. To create a considerate mechanism for scheduling and presenting notifications in the home, a better understanding is needed of how acceptability of notifications is influenced by contextual factors, presentation factors, as well as (subjective scores of) message urgency. Our prior field study focused on two factors of acceptability: engagement in activities and message urgency. The present study concentrates on a potentially relevant factor that was not included in the earlier study: level of intrusiveness. Furthermore, the relation between engagement in activities and acceptability will be studied again, this time in a more controlled setting.

DOI 10.1016/j.ijhcs.2009.06.002

The remainder of the paper is organized as follows. Section 2 describes the expected results. The present study incorporates research methodologies for controlled registration of user experiences in a living room laboratory. The resulting study design, in which ten participating couples spent an evening in the lab, is described in section 3. Section 4 describes the key findings from the study, while the remainder of the article is used to discuss the results and future steps.

2 Expected results

Fig. 2 shows the expected acceptability of notifications presented in the center (left-hand panel) and the periphery of the user's attention (right-hand panel). The left-hand figure presents the results of our prior field study, in which all notifications were presented with a high level of intrusiveness. The right-hand figure shows the expected acceptability when using a non-intrusive presentation mode.

Acceptability of low-urgency and medium-urgency messages is expected to improve when presented non-intrusively, since non-intrusive notifications are expected to be less disruptive of ongoing activities as compared to intrusive notifications. Acceptability of high-urgency messages is expected to be low when presented non-intrusively; immediate user response is needed, so participants would like to see urgent messages in an immediate and intrusive way. Acceptability is expected to drop when users are highly engaged in their activities; therefore acceptability of low-urgency messages would be low even when messages are presented non-intrusively. Based on this model of acceptability, a considerate notification system could select the optimal presentation mode and timing based on the message urgency and the degree of user engagement in activities.

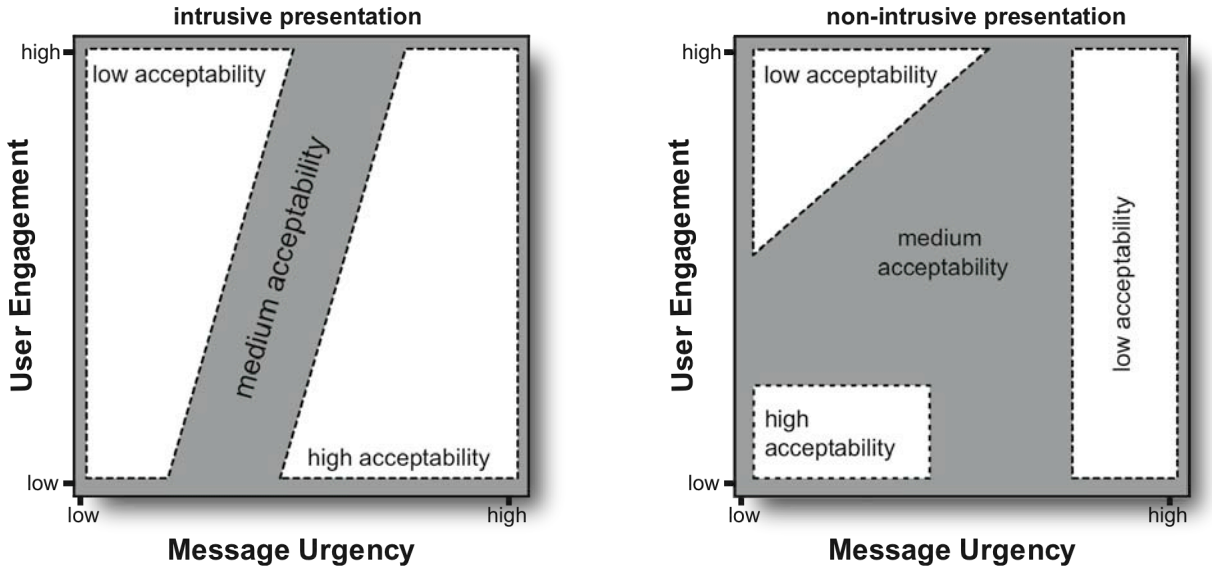

Fig. 2. Expected acceptability of notifications presented in the center (left-hand panel) and the periphery of the user's attention (right-hand panel). The horizontal axis shows the message urgency; highly urgent messages are expected to be acceptable when presented intrusively, whereas low-urgency messages would only be acceptable when presented non-intrusively. The vertical axis shows user engagement; acceptability is expected to be negatively related to user engagement, since people supposedly do not want to be interrupted when highly engaged in activities.

3 User study

A user study with twenty participants was conducted in a living room laboratory. Notifications were varied along three dimensions (Fig. 3): *user activity* (A1: watch TV, A2: read a book, A3: drink tea together, A4: play a game), *message urgency* (low, medium, and

DOI 10.1016/j.ijhcs.2009.06.002

high), and *presentation mode* (NI: non-intrusive, I: intrusive). Timing in relation to user activities has been studied by adding activity A5: activity transition; notifications were presented during the four main activities (A1-A4), as well as between activities (A5). Through a questionnaire, the user experiences on the notifications, the user activities, and the dependent variables *general acceptability*, *preferred level of intrusiveness*, and *preferred timing* were collected.

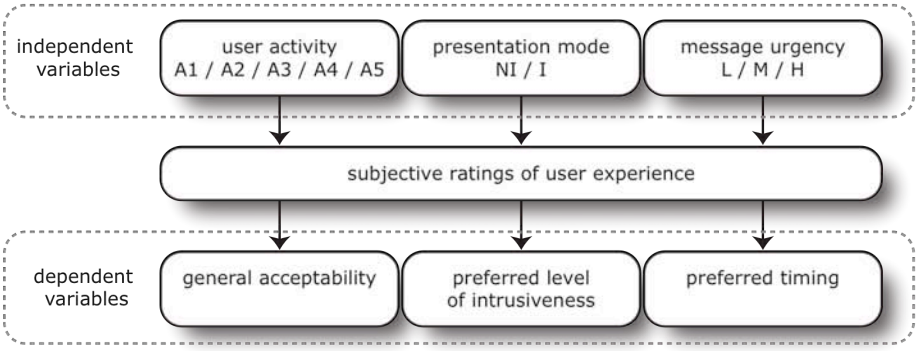

Fig. 3. Notifications were varied along three dimensions: user activity, level of intrusiveness, and message urgency. Using a questionnaire, the user experiences on the notifications and the dependent variables were collected.

3.1 Notification messages

A set of 15 diverse informational and alerting messages were selected from an existing set of user-rated messages from our prior field study. The original set of messages was defined on the basis of plausibility by a panel of three product designers, such that participants could relate to the messages in terms of their living situation. User ratings on message urgency from this earlier study were used to select messages for the present study: 5 messages that were rated *low-urgency*, 5 *medium-urgency* and 5 *high-urgency* (Table 1). An implicit assumption underlying the present user study is that message urgency can be predicted by future considerate home notification systems based on for example *message content*, *sender*, or *time of day*.

Classification	ID	Notification message
low-urgency (L)	L1	Don't forget to double-lock the front door.
	L2	A good program is about to start on TV.
	L3	To save energy, the thermostat should be set lower.
	L4	The plants in the garden need water.
	L5	Coffee is ready.
medium-urgency (M)	M1	Garbage will be collected tonight.
	M2	The washing machine has finished, please put the laundry in the drier.
	M3	Don't forget to get some bread out of the freezer for breakfast.
	M4	You need to be at work 30 minutes earlier tomorrow.
	M5	The videotape needs to be returned to the video rental shop tonight.
high-urgency (H)	H1	The lady next door has fallen and is in need of care.
	H2	Someone is touching your car.
	H3	Smoke has been detected in the shed.
	H4	A burglar might have entered the study.
	H5	The roof is leaking.

Table 1. Notification messages, originally in Dutch, and classifications.

DOI 10.1016/j.ijhcs.2009.06.002

3.2 Notification display

With the help of a product designer, an ambient notification display was developed for the experiment. The display enabled both non-intrusive and intrusive presentation of messages, while harmonizing with the living room environment. To this end, the prototype display projected notification messages on a wall in the form of yellow post-it notes (Fig. 4), roughly 1 meter in diameter. Post-it notes were chosen for visualizing the messages, since post-it notes are generally used to convey reminders. Two ambient color lamps, positioned in two opposite sides of the living room, were linked to the display, and could be used to attract the user's attention. Although audio signals could have been used to attract user attention, the notification display was based on visual presentation only; the product designer did not consider audio to add to the overall quality of the display. As will be shown below, according to a panel of product designers, the resulting display was able to immediately grab user attention without using audio signals.

When a message was presented non-intrusively, a static post-it gradually faded in, from transparent to visible in 40 seconds. The message was written in black, and the ambient light color changed slowly from orange to yellow. When presented intrusively, a post-it popped up within a second, and kept "wiggling". The top-left corner of the post-it was fixed, the post-it moved similar to a clock swing. The color of the ambient light "wiggled" accordingly, from orange to yellow, and the message was written in red.

The prototype display was evaluated by a panel of product designers. The evaluation was based on the *ambient heuristics*, a set of heuristics for evaluation of ambient displays as proposed by Mankoff et al. (2003). The ambient heuristics consist of 12 items including *visibility of state*, "*peripherality" of display*, and *match between design of ambient display and real world*. Five designers were asked to give a score on a scale from 1 to 5 for each item of the ambient heuristics.

The panel was told that the "ambient notification display" could be used to present notifications in a living room using a combination of video projection and colored lights to attract user attention. Furthermore, they were told that messages could be shown in two modes; messages shown in the foreground mode were supposed to require immediate user attention, messages shown in the background mode were supposed to require user attention without time pressure.

Fig. 4. The ambient notification display, which was developed specifically for the experiment, enabled both non-intrusive and intrusive presentation of messages.

DOI 10.1016/j.ijhcs.2009.06.002

The display scored very high on *visibility of state* (4.6 out of 5), which means that the states of the system (i.e., no message, non-urgent message, urgent message), and the transitions between states were evident. The display scored high on "*peripherality" of display* (4.0 out of 5), which means that the display was unobtrusive and remained so unless user's attention was required. Finally, according to the designers, the *match between design of ambient display and real world* was good (3.8 out of 5). Based on these results, the display was considered suitable for the experiment.

3.3 Procedure

Participants. Twenty subjects (11 male, 9 female) participated in the study, ages ranging from 22 to 47 (M=32; SD=5.9). Subject pairs ("couples") were selected based on their home situation, being couples living together. Eighteen participants were employed, two participants were students.

Setting. As the timing of notifications, the display conditions as well as the user activities had to be strictly controlled, the study was situated in a living room laboratory including furniture, a television, a CD player, a coffee maker and reading material. Each couple participated in the lab study which took approximately three hours.

Instruction. Participants were told that a prototype living-room notification system was being studied. The supervisor indicated that the aim was to explore how people experience notifications; feedback from the participants would be used to improve the timing and presentation of messages of the prototype. The supervisor explained the study would take place in a living room lab, because the prototype could not yet be deployed in the field. Participants were instructed to conduct the assigned activities, and to experience these activities as if they were at home. Furthermore, participants were encouraged to imagine that the notifications were real (e.g., the coffee is ready). Participants were asked to suppress feelings of frustration that might result from the high number of notifications that was presented in the three-hour experiment sessions. Both participants of each couple were instructed to complete a questionnaire on paper, as shown in Table 4, immediately after observing a notification. A questionnaire took on average approximately 30 seconds to complete.

Observation and control. During each experiment session, the two participants were left alone in the living room. Two observation cameras and a microphone were used by the supervisor to monitor their activities.

Activity-assignments. Four activity-assignments were used in the experiment, with varying levels of concentration and social interaction (Table 2). Each assignment was conveyed verbally to both participants, for example, "At home, you sometimes relax in front of the television. Please put yourself in this situation, and watch a relaxed program." Participants were given on average 15 minutes per assignment, thereby allowing them to enter fully into the assigned activities.

The activity-assignments were selected from our prior field study, in which concentration level and social interaction level were found to affect the acceptability of notifications. No significant relation between physical activity level and acceptability was found, therefore physical activity level was not used as a selection criterion when selecting activity-assignments. Urgency of activities was found to be closely linked to concentration level; therefore, urgency of activities has not been used as an additional selection criterion. Each activity-assignment was used twice in each session, resulting in eight activities per participating couple. The order of activities was randomized for the first five couples, and reversed for the remaining five couples.

DOI 10.1016/j.ijhcs.2009.06.002

	social-low	social-high
concentration-low	watch a relaxed program on television	drink tea together
concentration-high	read a book or magazine	play a game together

Table 2. Overview of the activity-assignments used in the experiment, varying in mental load and social interaction level.

Notifications. Notification scenarios were created beforehand, one scenario was used per session. The scenarios stated the order of activity-assignments, the presentation mode, and the order of the notification messages (Table 3). During each activity, three notifications were presented, with regular intervals of approximately five minutes. Six notifications were scheduled between assigned activities (during activity transitions), resulting in a total number of 30 notifications per session. Each notification message was used twice during each session, once non-intrusively and once intrusively. The order of the notification messages was randomized for the first five couples, and reversed for the last five couples.

user activity		presentation mode		message urgency
watch TV (A1) read a book (A2) drink tea together (A3) play a game (A4) activity transition (A5)	X	non-intrusive (NI) intrusive (I)	X	low-urgency (L) medium-urgency (M) high-urgency (H)

Table 3. Overview of notification treatments. Notifications were presented for each combination of the independent variables *user activity, presentation mode,* and *message urgency*, resulting in 30 notifications per session.

Exit interviews. At the end of each evening session, participants were asked to provide feedback on the notification system used in the experiment and on the experiment itself.

3.4 Questionnaire

Table 4 shows the questionnaire on paper with seven-point rating scales, which was used to collect subjective data on the notification acceptability and presentation (4 questions), on the notification message (1 question) and on the user activities (4 questions). To reduce the time spent on the questionnaires, the questions on user activities (Q6-Q9) were asked only once per activity, since activity ratings were not expected to vary within one user activity. In analyzing the questionnaires, all user ratings on the seven-point rating scales were coded as ordinal scales from 0 (low) to 6 (high).

DOI 10.1016/j.ijhcs.2009.06.002

Part 1: Notification acceptability and presentation			
Considering the message and your activities at the time of notification:			
Q1. General acceptability	not acceptable	ooooooo	very acceptable
Q2. Presentation	not intrusive	ooooooo	very intrusive
Q3. Preferred presentation for this notification	not intrusive	ooooooo	very intrusive
Q4. Preferred timing for this notification	now	ooooooo	much later
Part 2: Notification message			
Considering the message, without considering your activities at the time of notification:			
Q5. Message urgency	not urgent	ooooooo	very urgent
Part 3: Activities			
Q6. My concentration level	not concentrated	ooooooo	very concentrated
Q7. My interaction with others	no interaction	ooooooo	much interaction
Q8. Urgency of my activities	not urgent	ooooooo	very urgent
Q9. Appropriateness of this moment for interruption	not appropriate	ooooooo	very appropriate

Table 4. A questionnaire on paper was used to collect user ratings on the acceptability of the notification, the urgency of the message, and the user activities.

4 Results

4.1 General acceptability

Each of the twenty individuals evaluated thirty notifications, resulting in 600 completed questionnaires. To understand the categorical dependency of general acceptability of notifications on the independent variables *user activity*, *presentation mode* and *message urgency* (Fig. 3), a classification tree (Fig. 5) was constructed using chi-squared automatic interaction detection (Exhaustive CHAID) using SPSS 14.01™. The classification tree suggests general acceptability to be primarily dependent on message urgency (χ^2=114.67, p<0.001). Acceptability tended to be higher (U=19985, p<0.001, r=-0.41) for high-urgency messages (Mdn=5) than for low- and medium-urgency messages (Mdn=4).

Furthermore, acceptability of low- and medium-urgency messages was found to be associated with presentation mode (χ^2=15.12, p<0.05); acceptability tended to be higher (U=16243, p<0.005, r=-0.16) when presented non-intrusively (Mdn=4) as compared to intrusively (Mdn=3). Lastly, acceptability of high-urgency messages was also found to be associated with presentation mode (χ^2=13.63, p<0.05); acceptability tended to be slightly higher (U=4341, p<0.05, r=-0.12) when presented intrusively (Mdn=5) as compared to non-intrusively (Mdn=5). No significant association between acceptability and the user activity was found (χ^2=17.86, p=0.81).

DOI 10.1016/j.ijhcs.2009.06.002

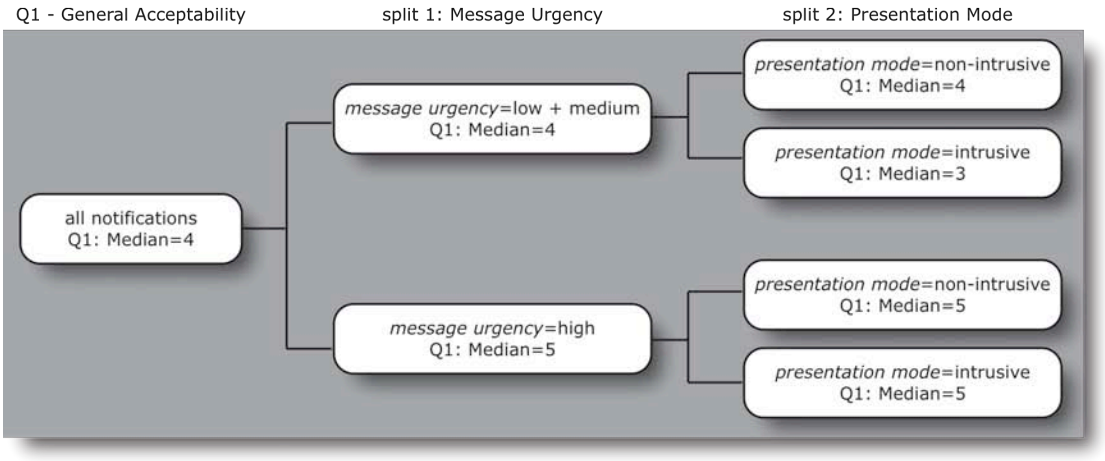

Fig. 5. Classification of acceptability based on the independent variables *user activity, presentation mode* and *message urgency* using Exhaustive CHAID, significance level 0.05. No significant relation between acceptability and user activities was found.

4.2 Message urgency

The actual user ratings on message urgency for low- and medium-urgency messages deviated from the anticipated scores based on the pre-classification of messages (Fig. 6). The perceived message urgency for the high-urgency messages (Mdn=6) was significantly higher ($U=4387$, $p<0.001$, $r=-0.74$) than for the low-urgency (Mdn=1) and medium-urgency (Mdn=1) messages. No significant differences were found between low-urgency and medium-urgency messages ($U=19132$, $p=0.439$, $r=-0.04$). Interestingly, the user ratings on message urgency by the female participants (Mdn=4) were significantly higher ($U=37751$, $p<0.005$, $r=-0.13$) as compared to male participants (Mdn=2).

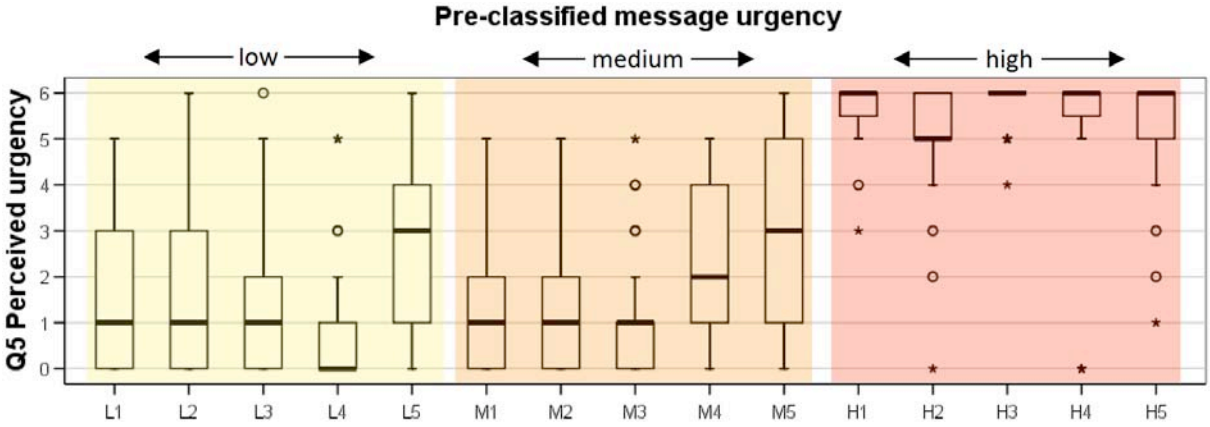

Fig. 6. Box-plot of the perceived urgency of the notification messages. The horizontal axis shows the 15 messages used in the experiment. The vertical axis shows the message urgency ratings, ranging from low (0) to high (6). Circles denote outliers, stars denote extreme values.

In box-plots, values which are more than 1.5 box lengths from either end of the box are labeled outliers; extremes are more than three box lengths from the box. In examining the outliers and extremes, it was found that these values could not be attributed to only few participants; the 31 outliers and extremes were caused by 14 participants. Notably, message L4 ("The plants in the garden need water.") was rated as low-urgent by all participants, except for three female participants, who considered the message to be highly-urgent.

The pre-classification procedure used in the experiment was useful in terms of creating a diverse set of messages that covered the urgency spectrum from low to high, but the

DOI 10.1016/j.ijhcs.2009.06.002

procedure was not good at predicting the message urgency of low- and medium-urgency messages for individual users. Therefore, the remainder of the analysis is based on perceived rather than the pre-classified message urgency scores.

4.3 Presentation mode

The user ratings show that the independent variable *presentation mode* performed as expected. The perceived level of intrusiveness (Q2) was significantly higher (U=5639, $p<0.001$, r=-0.77) when presented intrusively (Mdn=5), as compared to non-intrusive presentation (Mdn=1).

There was a positive relation (r=0.74, df=598, $p<0.001$) between message urgency (Q5) and the preferred level of intrusiveness (Q3). Participants explicitly indicated that non-urgent messages should be presented non-intrusively; the more urgent the message was considered to be, the higher the level of intrusiveness should be (Fig. 7).

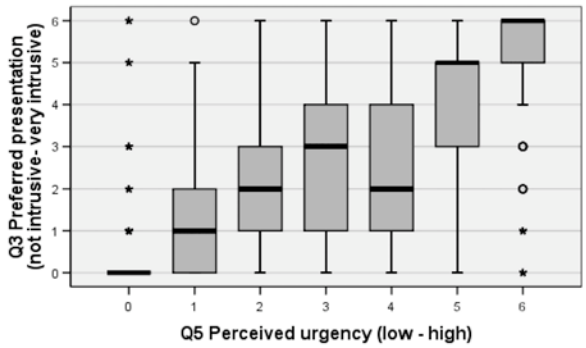

Fig. 7. Box-plot of preferred level of intrusiveness (Q3) for each of the perceived urgency levels (Q5). Non-urgent messages should be presented non-intrusively; the more urgent the message was perceived, the higher the level of intrusiveness should be. Circles denote outliers, stars denote extreme values.

4.4 Engagement in activities

Activity assignments were used in the experiment to study the relation between user activities and acceptability of notifications. Six out of ten couples indicated in their exit interviews that they could empathize with the activity assignments. The subjective scores for concentration (Q6) and social interaction (Q7) for each of the activity assignments (Fig. 8) corresponded to the anticipated scores. User concentration was significantly higher (U=15883, $p<0.001$, r=-0.40) when reading a book and when playing a game (Mdn=4), as compared to watching TV and drinking tea (Mdn=3). Social interaction was significantly higher (U=6087, $p<0.001$, -0.69) when drinking tea and playing a game (Mdn=5), as compared to watching TV and reading a book (Mdn=1). For activity transitions, users scored low on concentration (Mdn=2) and moderate on social interaction (Mdn=3).

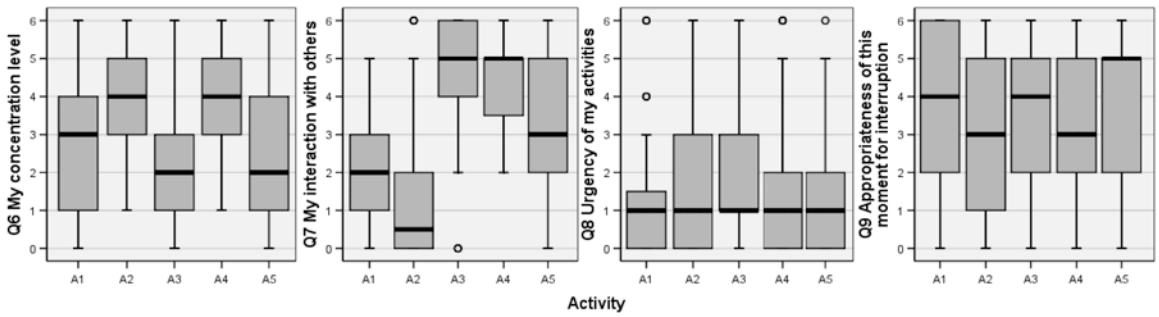

Fig. 8. Box-plots of the subjective scores for concentration (Q6), social interaction (Q7), urgency of activities (Q8) and interruptibility (Q9) for each of the activities used in the experiment: watch TV (A1), read a book (A2), drink tea together (A3), play a game (A4), and activity transition (A5). The vertical axes range from 0 (low) to 6 (high). Circles denote outliers, stars denote extreme values.

DOI 10.1016/j.ijhcs.2009.06.002

Perceived urgency of activities (Q8) was found to be low throughout the experiment. Similar ratings on interruptibility were found for watching TV, drinking tea, playing a game, and activity transitions.

User-rated interruptibility scores (Q9) show minor differences between activities; participants would rather be interrupted ($U=11646$, $p<0.005$, $r=-0.16$) when in between activities (Mdn=5) as compared to when reading a book (Mdn=3) or playing a game (Mdn=3).

4.5 Timing

A significant negative relation ($r=-0.66$, $df=598$, $p<0.001$) was found between message urgency (Q5) and preferred timing (Q4). Participants indicated that they wanted to see all high-urgency messages immediately; low- and medium-urgency messages were to be postponed (Fig. 9).

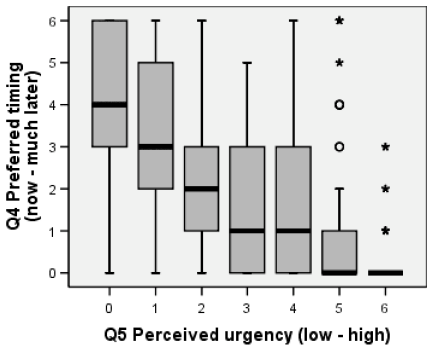

Fig. 9. Box-plot of timing (Q4) for each of the perceived urgency levels (Q5). Circles denote outliers, stars denote extreme values.

5 Methodological issues

In the present study, the acceptability of notifications was examined in a living room laboratory using prescribed activities. Several effects, caused by the nature of the study, might have influenced the results:

- The laboratory setting and the forced user activities might have resulted in user experiences that did not completely correspond to regular activities in a natural setting. Since the assigned activities had no sense of urgency attached, the cost of interruptions might have been lower than in a home setting.
- The user ratings of engagement in activities might have been influenced by the relevance of the notification message. When a highly urgent message ("Smoke has been detected in the shed.") was shown, participants might have rated their current activities as less urgent.
- Although asked to treat all messages as authentic messages, participants knew the notifications were artificial. Some participants indicated they found it hard to empathize with the messages. The lack of authenticity in the user feedback could have lead to lower ratings of acceptability.
- Although participants were instructed to rate each notification apart from earlier notifications, frustration caused by the high number of notifications might have resulted in lower acceptability ratings. A realistic notification system would probably aggregate messages based on their availability and urgency, which would lead to fewer notifications.
- The high number of notifications may have washed out the effect of user activity on acceptability. Since participants knew they were to be interrupted within minutes, they might have lowered their engagement in activities.

DOI 10.1016/j.ijhcs.2009.06.002

In view of these methodological issues, both ecological and external validity of the experiment need to be considered with care. Eventually, no matter how much effort is put into imitating the home setting in the laboratory, natural user experiences can never be experienced in an artificial environment. Even though the home would have been the ideal setting for the experiment, the requirements in terms of user control and timing made a field study unfeasible.

In the exit interviews, the majority of the participants indicated that the experiment session resembled an evening at home, and none of the participants mentioned the high number of notifications as a possible cause for frustration. We do therefore believe that the experiences as perceived by the participants, although restricted by the nature of the experiment, are close to natural experiences, and the results could therefore serve as a good first step towards testing in the field.

In the case of the present study, the artifacts as listed above could be solved by using a realistic system with real messages for a longer period of time in a real home setting. A simplified feedback mechanism, using for example thumbs up/thumbs down rather than questionnaires, could be used to minimize the interruption of activities caused by the experiment.

6 Discussion and conclusions

To find out how acceptability of notifications is influenced by the level of intrusiveness of the presentation, and how acceptability depends on timing in relation to user activities, a user study was conducted in a living room laboratory. Acceptability of notifications in the home has not been studied in a realistic and controlled setting before. The study showed that acceptability of low- and medium-urgency messages could be improved by adapting the level of intrusiveness; acceptability of low- and medium-urgency messages was higher when presented non-intrusively as compared to intrusive presentation (Fig. 5). Furthermore, the acceptability of highly urgent messages was slightly higher when presented intrusively as compared to non-intrusive presentation.

The results support the findings of both our prior field study as described in (Vastenburg et al., 2007) and related studies by other researchers (e.g., McCrickard and Chewar, 2003): message urgency was found to be the primary indicator of acceptability. Participants wanted to see highly urgent messages immediately (Fig. 9) and intrusively (Fig. 7).

Contrary to our expectations as shown in Fig. 2, user activities at the time of notification were not found to be related to acceptability (Fig. 5). Earlier work on interruptibility does suggest a relation between user activities and receptiveness to notifications. For example, Nagel et al. (2004) hypothesize that people are less receptive to notifications when engaged in a conversation, and Miyata and Norman (1986) put forward activity transitions as a good time for interruptions. We therefore expected acceptability to be low when people were engaged in highly demanding or highly social activities, whereas high acceptability was expected in between activities. No significant relation between user activity, activity transitions and acceptability was found. A possible explanation relates to individual differences; the non-existent relation between activities and acceptability might have been caused by differences between participants. Whereas for example activity transitions might be a good time for notifications for some of the users, no general significant relation was found for all participants. Nagel et al. (2004) encountered similar problems when studying the relation between user activities and interruptibility.

Based on the findings of the present study, one might adopt a simple strategy when managing notifications. First, highly urgent messages would be presented immediately and

DOI 10.1016/j.ijhcs.2009.06.002

intrusively. Medium-urgency messages would be presented in the periphery of the user's attention. Low-urgency messages would be either postponed until the message urgency has increased, or skipped if the message urgency never exceeds the predefined presentation threshold.

Towards creating considerate home notification systems that adapt the timing and presentation mode of messages based on contextual factors and message content, it seems important to focus on anticipating perceived message urgency. Whereas all highly urgent messages were perceived as highly urgent by all participants, a higher degree of inter-subject variation was observed for messages judged as low- and medium-urgency. No significant differences in perceived urgency were found between messages that were pre-classified low-urgent and medium-urgent (Fig. 6). Therefore, an automated classification mechanism might need to consider individual differences between users in message urgency ratings towards predicting message urgency for low- and medium urgent messages.

Based on the results of the present study, the initial model of acceptability of notifications (Fig. 1) needs to be updated. First of all, acceptability was found to be related to the presentation mode. Furthermore, although perceived urgency clearly is the primary indicator of acceptability, the initial model did not show the need to be able to predict perceived urgency. Perceived message urgency is expected to be related to the message itself (message structure, phraseology, relationship between messages), the context (user activities, state of the environment) and the user (user values, user state). These factors have been included in the updated model in Fig. 10; further studies are needed to understand how perceived message urgency is related to the message itself, the context, and the user.

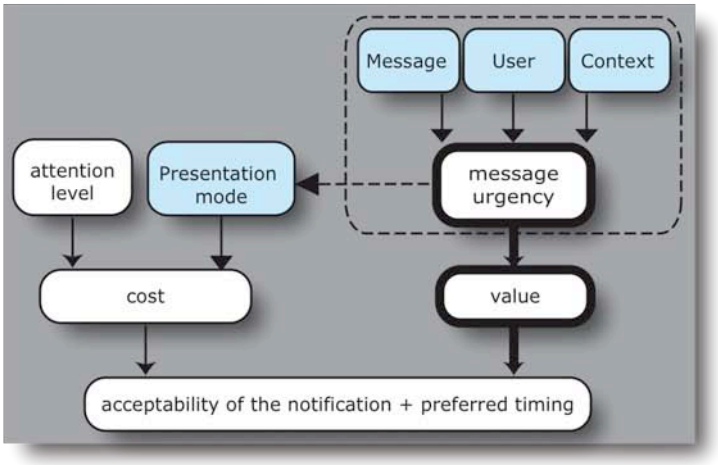

Fig. 10. Updated model of acceptability of notifications. Perceived cost of notifications is linked to attention level and presentation mode. Perceived value of messages is linked to perceived message urgency. Prediction of message urgency remains a major challenge in the development of considerate home notification systems. The bold arrows indicate message urgency to be the primary indicator of acceptability and preferred timing.

In comparison to existing examples of attentive user interfaces, such as the Notification Platform, user attention is merely a minor factor in our model of acceptability. Based on our studies, the primary challenge towards increasing acceptability of notifications in the home is to predict message urgency. When developing a mechanism for predicting message urgency, attention level might be found a predictor of message urgency. New studies are needed to better understand these factors underlying message urgency.

DOI 10.1016/j.ijhcs.2009.06.002

7 Future work

Prediction of the perceived urgency of messages remains a major challenge in the development of future considerate home notification systems. The predefined classification of message urgency used in the present study, which was based on actual user ratings from our prior field study, turned out to be a good predictor of high urgent messages. Predicting message urgency of low- and medium-urgency messages was found to be more difficult; personalized models might be needed to consider perceived message urgency of individual users, either based on the message alone, or in respect to the context (user activities, state of the environment) and the user (user values, user state). Additional user studies are needed to measure perceived message urgency and to create personalized prediction models. User profiles could be used as a start; a learning mechanism would be needed to improve the models in time.

The studies in the field and in the lab have provided practical cues for building a considerate system. Rather than focusing on additional controlled experiments, as a next step, a working notification system could be used to collect real usage data and to improve the general understanding of acceptability in the home. Based on the straightforward prediction mechanism that resulted from the present study, notification systems might eventually be able to adapt the presentation mode and timing of messages to the needs of the users and the context.

Acknowledgements. The authors thank the participants for their input, and the members of ID-StudioLab for their advice and support. They also thank the anonymous reviewers of IJHCS for providing helpful comments on earlier drafts of the paper. Particular thanks to Miguel Bruns for his support with the design of the display, Marc de Hoogh for statistical support, and Michel Varkevisser for his feedback on the early drafts of the manuscript.

8 References

Abowd, G.D., Mynatt, E.D., 2000. Charting past, present, and future research in ubiquitous computing. ACM Trans. Comput.-Hum. Interact. 7(1), 29-58.

Altosaar, M., Vertegaal, R., Sohn, C., Cheng, D., 2006. AuraOrb: Social Notification Appliance. Proceedings of CHI 2006, Montréal, Canada, 381-386.

Avrahami, D., Gergle, D., Hudson, S.E., Kiesler S., 2007. Improving the match between callers and receivers: A study on the effect of contextual information on cell phone interruptions. BIT 26(3), 247-259.

Bailey, B.P., Konstan. J.A., 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. Computers in Human Behavior 22 (4), 685-708.

Carroll, J.M., Neale, D.C., Isenhour, P.L., Rosson, M.B., McCrickard, D.S., 2003. Notification and awareness: Synchronizing task-oriented collaborative activity. Int. J. of Human-Computer Studies 58(5), 605-632.

Den Hartog, F.T.H., Baken, N.H.G., Keyson, D.V., Kwaaitaal, J.J.B., Snijders, W.A.M., 2004. Tackling the complexity of residential gateways in an unbundling value chain. Proceedings of 15th IEE International Symposium on Services and Local Access (ISSLS 2004), London, 1-10.

De Guzman, E.S., Sharmin, M., Bailey, B.P., 2007. Should I call now? Understanding what context is considered when deciding whether to initiate remote communication via mobile devices. Proceedings of Graphics Interface 2007, Montreal, Canada, 143-150.

Eggen, J.H., Hollemans, G., Van de Sluis, R., 2003. Exploring and Enhancing the Home Experience. Journal on Cognition, Technology and Work 5, 44-54.

Garlan, D., Siewiorek, D.P., Smailagic, A., Steenkiste, P., 2002. Project Aura: Toward Distraction-Free Pervasive Computing. IEEE Pervasive Computing 1(2), 21-31.

Gievska, S., Sibert, J., 2005. Using task context variables for selecting the best timing for interrupting users. Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence (sOc-EUSAI '05), New York, 171-176.

Hallnäs, L., Redström, J., 2002. From Use to Presence: On the Expressions and Aesthetics of Everyday Computational Things. ACM Transactions on Computer-Human Interaction 9(2), 106-124.

Ho, J., Intille, S.S., 2005. Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. Proceedings of CHI 2005, Portland, Oregon, USA, 909-918.

Horvitz, E., Breese, J., Heckerman, D., Hovel, D., Rommelse, K., 1998. The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 256-265.

Horvitz, E., 1999. Principles of Mixed-Initiative User Interfaces. Proceedings of CHI'99, 159-166.

Horvitz, E., Kadie, C., Paek, T., Hovel, D., 2003. Models of Attention in Computing and Communication: From Principles to Applications. Communication of the ACM 46(3), 52-59.

Hughes, J., O'Brien, J., Rodden, T., Rouncefield, M., Viller, S., 2000. Patterns of Home Life: Informing Design for Domestic Environments. Personal Technologies 4, 25-38.

Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B., Yarin, P., 1998. ambientROOM: Integrating Ambient Media with Architectural Space. Conference Summary of CHI'98, 173-174.

Iqbal, S.T., Bailey, B.P., 2008. Effects of intelligent notification management on users and their tasks. In: CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. ACM, New York, NY, USA, 93-102.

Khalil, A., Connelly, K., 2006. Context-aware Telephony: Privacy Preferences and Sharing Patterns. Proceedings of CSCW'06, 469-478.

Maglio, P.P., Campbell, C.S., 2000. Tradeoffs in Displaying Peripheral Information. CHI Letters 2(1), 241-248.

Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J., Lederer, S., Ames, M., 2003. Heuristic Evaluation of Ambient Displays. Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA, 169-176.

Matthews, T., Dey, A.K., Mankoff, J., Carter, S., Rattenbury, T., 2005. A Toolkit for Managing User Attention in Peripheral Displays. Proceedings of the 17th ACM Symposium on User Interface Software and Technology (UIST'04), 247-256.

McCrickard, D.S., Catrambone, R., Chewara, C.M., Stasko, J.T., 2003. Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. Int. J. of Human-Computer Studies 58, 547-582.

McCrickard, D.S., Catrambone, R., Stasko, J.T., 2001. Evaluating Animation in the Periphery as a Mechanism for Maintaining Awareness. Proceedings of Interact 2001, 148-156.

McCrickard, D.S., Chewar, C.M., 2003. Attuning notification design to user goals and attention costs. Communications of ACM 46(3), 67-72.

McFarlane, D.C., 1998. Interruption of People in Human-Computer Interaction. Dissertation, The School of Engineering and Applied Science, The George Washington University.

McFarlane, D.C., 1999. Coordinating the Interruption of People in Human-Computer Interaction. Proceedings of INTERACT'99, 295-303.

McFarlane, D.C., Latorella, K.A., 2002. The Scope and Importance of Human Interruption in Human-Computer Interaction Design. Human-Computer Interaction 17, 1-61.

Miyata, Y., Norman, D.A. (1986). Psychological Issues in Support of Multiple Activities, in: Norman, D.A., Draper, S.W. (Eds.), User-Centered System Design. Hillsdale: Lawrence Erlbaum Associates: 265-284.

Nagel, K.S., Hudson, J.M., Abowd, G.D., 2004. Predictors of Availability in Home Life Context-Mediated Communication. Proceedings of the 2004 ACM conference on Computer supported cooperative work, 497-506.

Vastenburg, M.H., Keyson, D.V., de Ridder, H., 2007. Considerate Home Notification Systems: A Field Study of Acceptability of Notifications in the Home. Personal and Ubiquitous Computing 12(8), 555-566.

Vertegaal, R., 2003. Attentive user interfaces: Introduction. Communications of ACM 46(3), 30-33.

Weiser, M., Brown, J.S. (1997). The coming age of calm technology, in: Denning, P., Metcalfe, B. (Eds.), Beyond calculation: The next fifty years of computing. New York: Springer-Verlag, 75-86.