

Effects of Content and Time of Delivery on Receptivity to Mobile Interruptions

Joel E. Fischer^{*}, Nick Yee⁺, Victoria Bellotti⁺, Nathan Good⁺, Steve Benford^{*}, Chris Greenhalgh^{*}

^{*}The Mixed Reality Laboratory
University of Nottingham
Nottingham, NG8 1BB, UK

{jef, sdb, cmg}@cs.nott.ac.uk

⁺ Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

{yee, bellotti, good}@parc.com

ABSTRACT

In this paper we investigate effects of the content of interruptions and of the time of interruption delivery on mobile phones. We review related work and report on a naturalistic quasi-experiment using experience-sampling that showed that the receptivity to an interruption is influenced by its content rather than by its time of delivery in the employed modality of delivery – SMS. We also examined the underlying variables that increase the perceived quality of content and found that the factors *interest*, *entertainment*, *relevance* and *actionability* influence people’s receptivity significantly. Our findings inform system design that seeks to provide context-sensitive information or to predict interruptibility and suggest the consideration of receptivity as an extension to the way we think and reason about interruptibility.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors.

Keywords

Receptivity, content, interruption, experience-sampling, push vs. pull, context, mobile, SMS, empirical study, quasi-experiment

1. INTRODUCTION

The spread of ubiquitous computing goes hand in hand with a rise of ubiquitous access to information. A recent study found that the number of people accessing news and information daily on their mobile devices in the U.S. more than doubled in the last year [5]. As applications mature away from the desktop, so does the potential of unintended side effects such as ‘information overload’ and the ‘attention economy’ [14], whereby information resources compete for the individual’s attention and it becomes crucial to filter out irrelevant information.

Proactive information push is a controversial instance of resources competing for attention. On the one hand, it is clearly becoming a core challenge in context-aware computing to provide the user with the right type of information at the right time [13]. On the other hand, it is especially information pushed through such channels as email, instant messengers and the mobile phone that causes distractive interruptions and that fragments interactions

across media [22]. The work presented here is relevant and related to both the positivistic field of context-aware computing and the problem-driven work that investigates effects of interruption, interruptibility and interruption management.

Research in context-aware systems seeks to model the user’s context in terms of “location, identity and state of people, groups and computational and physical objects” [10]; for example, in order to provide the user with information that is relevant to her context [32]. If this information is pushed to the user, it may be perceived as a disruptive interruption, depending on the obtrusiveness of the delivery method. Effects of interruption during tasks [1, 8, 9, 16, 31] and the prediction of interruptibility, for instance with sensors in office settings [3, 15] have been studied widely.

Whereas the prediction of *interruptibility* is more interesting from a technical systems-oriented perspective, e.g. as a trigger to an action, *receptivity* is a related concept that anticipates a user’s subjective overall reaction to an interruption, which may encompass both interruptibility *and* experience of the interruption. Indeed, even though a message may be interruptive, people can be receptive to it if the content justifies the interruption, e.g. “It’s a boy!”. For the purpose of this study, the more inclusive and user-centered notion of receptivity is preferred to the more system-centered notion of interruptibility.

2. INTERRUPTIONS IN CONTEXT

We begin by reviewing key concepts in interruption studies. We follow the idea that interruptions are an inevitable constituent of the ways we communicate [28], therefore we posit that the most promising intervention for context-aware systems for interruption management is to optimize the *timing* of the interruption according to the predicted interruptibility or receptivity, rather than to let systems make the decision to hold back interruptions entirely. The interest in timing then becomes twofold: first, what is the *impact* of interruptions at different times? Second, how can we *predict* opportune moments for an interruption? The impact of timing has been studied in laboratory studies in order to inform systems design by operationalising timing in terms of the individual’s position in the primary task [1, 8, 9, 16, 31]. Due to research in cognitive psychology the phases of task processing are fairly well understood [20].

However, the finding that opportune moments for interruptions lie at the breakpoints between phases of cognitive task processing [1] can only be turned into design recommendation for systems that

Copyright is held by the author/owner(s).

MobileHCI 2010 September 7 - 10, 2010, Lisboa, Portugal.

ACM 978-1-60558-835-3.

can monitor the user's current task. Whereas this may be feasible in relatively controlled settings such as offices by installing both software sensors (e.g. to log desktop interaction) and hardware sensors (e.g. cameras and microphones to monitor social situations or first person activity away from the computer), it would appear to become difficult in mobile settings and would require the users to wear sensors to detect physical activity [13].

In this paper, we attempt to explore the impact of interruptions at different times on receptivity in a mobile setting.

2.1 Local and relational contextual factors

What are the factors that influence a person's receptivity to an interruption? It depends on the situated interplay of a whole host of psychological and environmental factors. For example, Ho and Intille [13] identify at least 11 factors from related work that influence a person's interruptibility such as "activity of the user", "utility of the message", "emotional state of the user", "modality and frequency of interruption", "task efficiency rate", "previous and future activities", "social engagement of the user", and "history and likelihood of response".

We go on to contrast the characteristics of local contextual factors that are traditionally looked at in studies of interruptions with contextual factors that accommodate a *relational perspective* [12].

The contextual factors that have been studied for their influence on receptivity or interruptibility mostly account for the local context of the person *independently* of the interruption itself.

Local contextual factors have been roughly distinguished into two categories: cognitive and social factors [12] (even though other local contextual factors related to physical activity or body position and environmental factors such as physical affordances of the environment are missing from these categories). Social factors have been studied to discern the impact of the social surrounding on the interruptibility of the interrupted person, e.g. the presence of other people [2], or the impact of specific organizational or cultural norms of the place where the person is interrupted on how interruptions are managed [14, 28]. Cognitive factors that have been studied include the interrupted person's mental workload when interrupted [1], or the person's current activity [2].

Whereas local contextual factors describe the individual's context at the moment of the interruption, relational factors take into account properties of the interruption itself. We suggest that relational factors should be taken into account in addition to local contextual factors when trying to estimate a person's receptivity. For instance, out of the 11 factors identified by Ho and Intille [13], only one takes into account the relation between properties of the content of the interruption and the user ("importance of the message to the user").

As it is the process of transmitting information that fundamentally causes an interruption, we take on a perspective inspired by communication theory. In order to gauge a person's relational context for each interruption, we could analyse the answer/s to Lasswell's famous formula of communication [18], which summarises concisely what we mean by relational factors:

Who says what to whom in what channel with what effect? [18]

At first sight, this question focuses solely on the process of the transmission of a message and seems to lack psychological and environmental factors. However, inspecting its parts closely, each of them inherit more factors and most importantly, they always set the individual receiver (*to whom*) in relation to the factors.

First the receiver's social relation to the sender (*who*) is important [12]. Trust, credibility, authority, and familiarity of the sender probably will influence how receptive the receiver will be to the message. For instance, a study has found the identity of the sender to be the main factor in call handling decisions [12]. The receiver's receptivity may also be influenced by the sender's known or anticipated local context [12].

Next, the content (*what*) of the interruption may play a significant role in how receptive the receiver will be. Studies that take properties of the *content* of an interruption into account have manipulated the interruption content to be relevant to the current task, and the content of others to be irrelevant [9, 17, 31]; and some found significant effects for relevance [9, 17]. A study of interruption in the home also discovered that the "urgency" of a message is a stronger predictor of the acceptance of an interruption than the people's current engagement in activities [30]. Studies of acceptance of mobile advertising in China and Taiwan found "informativeness", "entertainment" and "credibility" to be the most important factors of the advertisement message that influence the consumers' acceptance; and "irritation" to have a negative effect [29, 25]. Whereas the former study [30] used fictional messages with varying implied urgency for the recipient (e.g. "the fire alarm in the shed has detected smoke"), the latter studies [29, 25] were based on questionnaires.

Furthermore the characteristics of the channel the interruption is delivered in (*in what channel*) may impact receptivity or the management of the interruption. For example, whereas a ringing phone may act as a summons [23], chat in instant messengers may be ignored until a convenient later moment [21].

Lastly, *what effect* the interruption may have is not only interesting to observe, for instance how interrupted people manage their social accountabilities [28], anticipating the effect could also be an important factor in predicting receptivity.

2.2 Motivation

The absence of a "relational perspective" such as the consideration of the *who* (sender/source) and the *what* (content) of interruptions in most of the interruption research has previously been noted [12]. This is reflected in experimental methods: instead of sending genuine content most of the studies that investigate interruptibility interrupt their participants only with a request to confirm their interruptibility either on a scale [15]; or they ask how much time the interrupted would have for an interruption [14]; or they use generic types such as "how receptive are you to a 'phone call' or a 'reminder'?" [13].

Despite that the content and the timing of interruptions have been acknowledged as cognitive characteristics of the interruption [31, 17], these authors operationalise content somewhat narrowly as the relevance of the interruption content to the *current task* and the timing of an interruption is operationalised in terms of the temporal position in the *current task*. However, we argue that one can be receptive to an interruption even though the content is not relevant to the current task. Also, these operationalisations inherit a requirement to control the current task; thus calling for a controlled laboratory study, which makes it difficult to transfer the methods to a more naturalistic mobile setting. This also holds for the finding that opportune moments for interruptions lie at the breakpoints between phases of cognitive task processing [1].

Mobility of work and life is arguably a definitive characteristic of our culture today, and thus it may be desirable to study the

accompanying characteristics of interruption in a mobile setting, rather than, for example in a setting where the occurrence of the interruption is limited to one room, such as an office setting.

No study was found that interrupted participants with *real* messages containing *genuine* content in mobile settings. Therefore, in line with the relational perspective, this study employs a method where participants were interrupted with real messages with genuine content on their mobile phones in order to examine how manipulation of interruption content and time of delivery affects receptivity *in situ*.

3. RESEARCH QUESTIONS AND HYPOTHESES

The mobile phone seems to be an ideal vehicle to study mobile interruptions. It is probably the single technological instance we are most used to get interrupted by anywhere, anytime. With *timing* of interruptions being the key intervention by systems for interruption management we also wanted to explore the impact of a preliminary operationalisation of timing outlined below. To evaluate the impact of a relational factor on receptivity we chose to look at *content*.

- RQ: How do the interruption content and the time of delivery impact receptivity to pushed information on mobile devices?

Furthermore, what properties of the content make people more receptive to information pushed to their personal mobiles? In line with the relational perspective, rather than properties of the content per se, these are investigated as relations people have to the content; namely, how people relate to *what* is said, to *how* it is said, and to the *effect* of the message. The factors we chose to study as probable underlying factors of receptivity are *interest* in the content of a message (*what*) and *actionability* of a message (*effect*) in addition to the factors *relevance* and *entertainment* (*how*) identified by previous research as outlined above [9, 25, 29, 30, 31, 17]. *Actionability* is understood to be the likelihood to influence the recipient's future actions. We hypothesize the factors to be strong indicators of receptivity. The study therefore builds on the following two major hypotheses:

H1: Interest in content is a predictor of receptivity: People are more receptive to "good content" than to "bad content", based on their individual interest ratings.

H2: Interest in content, perceived entertainment value, relevance and actionability of a message's content are all valid factors that influence people's receptivity.

The perception of the timing of an interruption may depend on the nature of the *delivery method* (asynchronous vs. synchronous). Some delivery methods are inherently more disruptive than others and the interruption is more immediate. A ringing phone, for example, is inherently immediate and can be understood as a summons [23]. In contrast, asynchronous modes of delivery such as email, SMS or instant messengers do not demand attention in such an immediate way. They even afford *plausible deniability* [21], in that it is socially acceptable not to respond immediately. Presumably, the cost of ignoring an incoming interruption that was delivered in an asynchronous way is less as its content can be tended to at a convenient later moment whereas the content of a missed phone call may be lost indefinitely. Hence, the timing of interruptions in an asynchronous channel may be evaluated more liberally than in a synchronous channel. The asynchronous nature of SMS effectively enables the individual to tend to the

interruption at a convenient moment – even if it is just seconds after the interruption was noticed, it effectively becomes a self-selected attention shift rather than a forced attention draw. As the modality of delivery in this study is SMS, we expect that the *perceived timeliness* of the interruptions do not differ significantly between times individuals predict they are especially receptive to interruptions and other times.

We decided to study receptivity in a naturalistic mobile setting where participants do not have to carry additional equipment and wanted to test if we could simply operationalise timing as *time of day of delivery* (henceforth called *time of delivery*). As a control measure, we compare the participants' self-reports of timeliness at participant-selected *good times* for interruption with random *other times*, where times are specific clock-times, e.g. 11:15am.

4. EXPERIMENT DESIGN

In a 2x2 within-subjects design, we manipulated *type of content* (good and bad content) and *time of delivery* (good and other time). We employed both the experience-sampling method (ESM) [7] and a diary method over a period of 10 working days. Six text messages were delivered each day to the participant's mobile phone. Participants were asked to rate their receptivity to the content, utilizing an experience-sampling approach whereby participants assess experience *in situ* [6]; i.e. with the instruction to reply with a rating ASAP.

As the levels of the variables "good" and "bad" are subjective, it is important to stress that the messages were different for each participant. The messages had in common only that their respective content categories were rated highly (for "good" content) or lowly (for "bad" content) by the participants in the setup interview.

The daily survey was set up to unpack what factors of the content make people more receptive, i.e. it looked at *relational qualities* of the content. We stated in our initial hypothesis H2 that we believe that *interest*, perceived *entertainment*, *relevance* and *actionability* all influence the receptivity. Are the receptivity ratings correlated with how interesting and entertaining they find the content, with how relevant and actionable a message is perceived?

The fact that people's daily information consumption routines differ individually and the fact that we conducted repeated measures calls for a within-subjects analysis approach of the data.

4.1 Everyday information consumption

The goal for the content categories we presented to the participants was that all of the participants were interested in some and not interested in other categories. To arrive at the content categories we conducted interviews to study the everyday information consumption behavior of the sample.

In a preliminary set-up study, we carried out informal interviews with 10 co-workers to find out about their information consumption habits. We asked them to itemize what categories of information they consumed frequently, if it was on a regular or on an irregular basis and if they consumed the information out of general interest or out of a situational need. As information categories consumed out of general interest, they named news, sports, hobby/leisure-related and work-related information ("*to stay on top of what's happening*"¹). Information tied to the co-

¹ All quotes by participants unless stated otherwise.

workers' situation was named more often and the category terms seemed more specific, e.g. information on traffic, weather, movies, products and places to go. Even though the categories were diverse, what they shared was that they had an immediate impact on the interviewee's situation and may have influenced subsequent actions. For example, weather reports were followed to inform what to pack for a future travel; movie reviews were read to see if there was anything worth watching and eventually leisure plans for the weekend may have been influenced by this; product reviews and prices were compared to inform a buying decision.

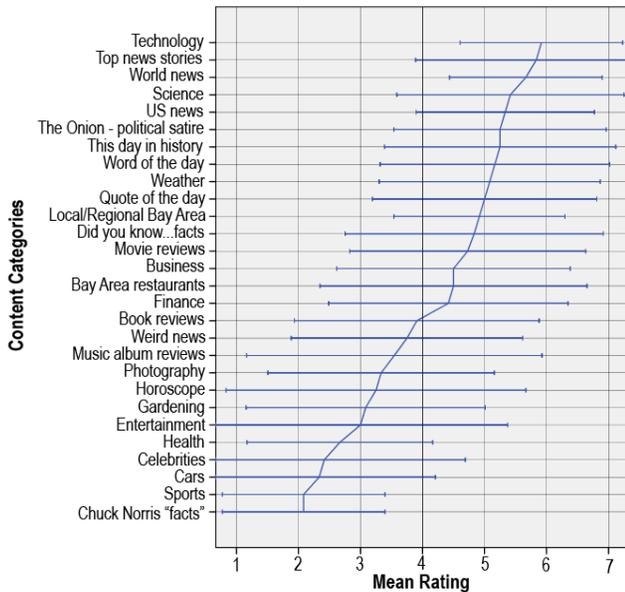


Figure 1: The sample's content preferences (mean + SD) as collected in the setup procedure.

On a broad level, our set-up study identified two modes of information consumption. The first is general interest or serendipity in information consumption, having no immediate impact on activities. The second is situationally directed information use that explicitly informs future actions. For a more detailed explication of information needs see [26].

Since it would be an ambitious technical project in its own right to predict and cater to situational information needs, our limited systems development resources restricted us to choosing general interest information categories which were easier to serve to volunteers for our study.

4.2 Participants and individual preferences

Individual diversity in the ways people relate to information made us realize that we could not send the same types of information to every participant. In order to compare the impact of the factor *content* we needed a constant across participants' relation to the content. As we selected information presumed to be of general interest, a straightforward way to introduce a constant was to ask them how interested they were in the particular content category, and thereby inform what type of information they would get as *good* and as *bad content* throughout the study.

Based on the findings from the set-up study interviews augmented with what we found to be feasible to provide from RSS feeds on the web, we developed a fairly extensive list of 28 general content

categories. The list was offered to our main study participants to allow them to express both interest and disinterest in different categories (see content categories in figure 1).

We recruited 11 co-workers as participants (3 female and 8 male). None of them were involved in or knew details about the study. As part of the setup process we asked the participants to rate their interest in the 28 content categories provided on a scale from 1 (not at all interested) to 7 (very interested) (see fig. 1). According to their ratings, we then chose 3 categories they rated highly (6-7) to provide the "good content". Another 3 categories they rated lowly (1-2) were chosen to provide the "bad content". Figure 1 shows the mean interest ratings of the 11 participants, with the content categories ordered according to descending interest rating.

In addition, we asked them to give us a general time window of 10hrs during the day where they would be willing to receive messages on their mobile phones. Within this window, we asked them specifically to provide three times where they thought they would be receptive to messages. These 3 times would serve as "good times". In addition, we chose three "other times", whereby we attempted to create an even distribution of messages throughout the specified 10hr window.

4.3 Framing receptivity: gut reaction and perceived timeliness

In our study, we sent text messages and then asked for an immediate reply with a rating of their receptivity (via a text message reply). The challenge was to instruct the participants in a way that the ratings reflected their perception of receptivity. We chose to ask the participants to rate their *gut reaction* to the message on a scale from 1 (low) to 7 (high), because the concept it refers to is both subjective and situated in the particular time it occurs and is tied to the circumstances of the situation. We explicitly instructed the participants to include such guidelines in the assessment of their *gut reaction* as the initial feeling towards the content of the message at that particular time; how the interruption of the incoming SMS felt; the particular circumstances of the environment such as appropriateness of location or who else is present.

Because of the aforementioned anticipated limitations of the validity of the factor *time of delivery*, i.e. an inherent uncertainty if the participants' chosen good times really represented good times, we also asked them to provide a control rating of the *perceived timeliness* of the time of delivery on the same scale of 1 (low) to 7 (high).

4.4 Procedure

An SMS server was set up to deliver the text messages according to a daily schedule for each individual participant. When a time to send a message was detected in the daily schedule, a task was performed that looked up the RSS feed that provided the content according to the individual preferences for each participant. For each participant, three content categories provided the *good content* and three categories provided the *bad content*. They were randomly rotated in their assignment to the three good times they said they would be receptive and three other times we chose for them, so that participants would not always get the same type of message at the same time of day. This rationale was informed by the goal to get a sufficient number of ratings for each of the possible 2x2 factor combinations for a within-subjects analysis. With six messages per day sent for 10 days and 4 different

combinations of the factors, each participant would get an average of 15 messages per combination.

Table 1: distribution of the independent variable across messages that received a response.

	Good time	Other time	Total
Good content	98 (23.4%)	107 (25.6%)	205 (49%)
Bad content	108 (25.8%)	105 (25.2%)	213 (51%)
Total	206 (49.2%)	212 (50.8%)	418 (100%)

4.5 Daily survey

We did not want to overload participants by asking them to send a whole host of ratings by text, so a web-based survey was set up to reveal the role other factors play in receptivity. At the end of each day, an online survey presented each of the six messages as a reminder. For each message, the participants were asked to provide a rating of their interest and entertainment value of the content on a scale of 1 (low) to 7 (high). Furthermore, they were asked to pick categories that best describe the relevance and actionability of each message.

4.6 Exit interviews

Finally, participants were interviewed to learn about their experience during the study. Semi-closed or open ended questions were asked to get an overall picture of the experience, looking at change over time, specific moments, e.g. particularly good or bad timing of messages. Also, we asked if and how the messages ever impacted the recipients’ activity or were relevant to it, and investigated their general receptivity to information pushed to their mobile phones, and possible alternatives, e.g. information “pull”.

5. EXPERIENCE-SAMPLING RESULTS

The 11 participants received a total of 641 text messages (56 pp on average) over a period of 10 working days. The intended 60 messages pp were not quite achieved due to technical difficulties with message delivery. We received 418 (38 pp on average) responses in total (i.e. to 65.2% of the messages we sent). Table 1 shows the distribution of the independent variables content and time of delivery among the messages that received a response.

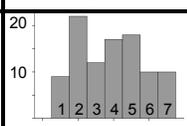
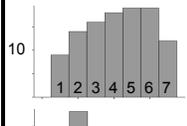
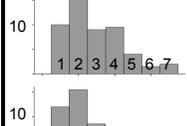
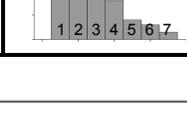
5.1 Gut reaction

The mean *gut reaction* to all of the 418 messages sent was slightly below neutral with a mean rating of 3.42 on a 7-point Likert scale, the median was 3, the SD was 1.8.

To perform a within-subjects ANOVA we aggregated the gut reaction scores on the individual level by averaging per condition (see table 2); a common approach for dealing with unequal amounts of repeated measures per individual [13, 27]. The dataset that was derived in this manner comprised one averaged score of gut reaction for each of the four conditions per participant. Inspection of Q-Q plots showed that deviation from normality was tolerable and skewness (.253) showed that the distribution was still approximately symmetric, according to a widely used rule of thumb [4].

A within-subjects ANOVA showed that *content* had a significant effect on *gut reaction*, with $F(1, 10) = 30.95, p < .001$. Partial eta squared = .76, which is a large effect.

Table 2: Gut reaction descriptives between subjects by independent variable conditions.

IV conditions	Frequencies	Mean	Std. Deviation	N
Good content, good time		3.85	1.84	98
Good content, other time		4.21	1.82	107
Bad content, good time		2.85	1.55	108
Bad content, other time		2.84	1.62	105

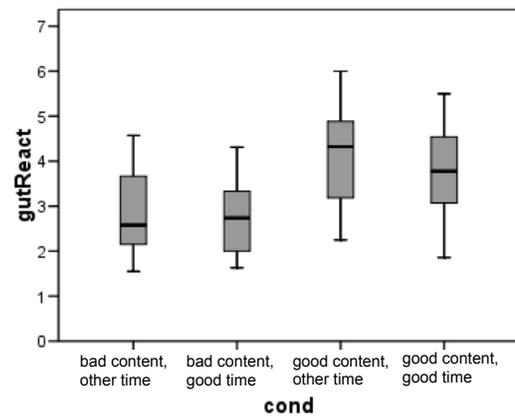


Figure 2: Boxplots of gut reaction by 2x2 factors content and time of delivery.

The mean *gut reaction* for *good content* (3.96) was significantly higher than for *bad content* (2.86) (see fig. 2). The factor *time of delivery* (good vs. other time) did not lead to significant differences in the gut reaction ($F(1, 10) = 2.8, p = .125$); and there was no significant interaction of *content* and *time of delivery* ($F(1, 10) = 1.93, p = .195$). Table 2 and fig. 2 show the means and SDs of *gut reaction* across the conditions.

The finding that *content* has a significant effect on people’s gut reaction supports our hypothesis that people are more receptive to content they have expressed higher interest in (H1). Furthermore, the findings suggest that we can predict receptivity by studying individual information consumption behavior, as the manipulation of the factor *content* was informed by the previously collected interest ratings in the content categories.

5.2 Perceived timeliness

The control measure *perceived timeliness* revealed that the perceived timeliness for good times was not significantly different than for other times. A within-subjects ANOVA done in a similar manner as for *gut reaction* above showed that neither *content* ($F(1, 10) = .92, p = .36$) nor *time of delivery* ($F(1, 10) = 2.82, p = .12$) had a significant effect on *perceived timeliness*. The interaction was also not significant, with $F(1, 10) = .65, p = .44$.

The fact that the *perceived timeliness* does not differ between predicted good times and randomly selected other times may mean that time of delivery in an asynchronous medium such as SMS is less important as messages can be tended to later. It is equally plausible that people are simply bad at predicting the times when they are receptive and thus the validity of the factor *time of delivery* is limited in that the levels “good” vs. “other” are not different from each other.

5.3 Discussion with qualitative findings

To review the statistical findings above, we found that when pushing content to people’s cell phones, their gut reaction to content they had previously expressed interest in was significantly higher than to bad content. The fact that participants’ gut reaction to *good content* was significantly higher than to *bad content* shows that content matters.

Furthermore, the fact that significant effects were found for *content* and not for *time of delivery* indicates that interest in content is a more stable predictor of the gut reaction to a message than the time of delivery. As one participant so poignantly put it:

“Content really trumps the timing.”

This is widely echoed throughout the interviews. When asked directly, 8 participants said that content was more important to the gut reaction than time of delivery; one said that both were important and two said that timing was more important – however one of these last two immediately qualified his statement in an interesting way:

*“Timing first, content second, I mean, well, it’s interesting...I never got a message while I was holding my phone, it was always while I was doing something else. So the first problem is, that I’m doing something else, that’s why timing is the first part, and then I look at the message, and then **if the content is not important to me... it was not a good moment to interrupt me with that. But I’m likely to be a little bit more forgiving if the content is interesting to me.**”*

On the one hand, the participant suggests that the timing is more important than content; on the other hand, he mitigates his assertion by what he says in the highlighted area about his receptivity. This statement points to a notable quality: the assessment of the timing of an interruption is related to the content of the interruption and presumably other factors. Even if the timing is bad, to use the participant’s words, we are likely to be more forgiving if the content adheres to certain qualities; in this case, interest. The notion that the appropriateness of an interruption is evaluated by its content rather than by its timing is even more clearly illustrated by the following statement from another participant, which nicely emphasizes how timeliness of an interruption is assessed through its content.

*“Sometimes it was OK to be interrupted, but not with that particular kind of snippet...**depending on what the interruption is, it is a good time or it’s not a good time.**”*

Through statements such as these, we see that the impact of an interruption is assessed *ad hoc* via its content. This echoes Dourish’s insight that contextuality cannot be determined a priori; “It is an emergent feature of the interaction, determined in the moment and in the doing” [11]. Research on interruption management has put a lot of effort into predicting the user’s interruptibility by sensing their local context [15, 13]. The implication of the finding that the impact of an interruption is

assessed by its content rather than its mere time of delivery is that the prediction of interruptibility has to take into account the content of the interruption.

Since all participants provided us with supposed *good times* to interrupt, the nonexistent statistical effect of *time of delivery* on the participant’s *perceived timeliness* suggests that people are bad at predicting the times at which they will be receptive to messages and shows that *time of delivery* is an incomplete operationalisation of timing – not that timing doesn’t matter. In the exit interview, when asked about the times he had said he would be receptive, one participant replied:

“[...] Those are basically times where I thought I wouldn’t be too engaged in things – it turns out that most of the times I actually was engaged with things.”

The participants picked these times without complaining about the difficulty of this task and with some care, which can be seen by the fact that, at the end of the study, only one participant didn’t remember the times he said he would be receptive to messages. Only 5 participants said in the interviews that the majority of messages were delivered at a bad time (as opposed to 9 that disliked the content of the majority of the messages). This further supports the finding that in the participants’ perception content trumps timing:

“When the content was bad, I felt annoyed, when the timing was bad, I just let it go.”

This “letting it go” in this participant’s statement points to a key quality of interruption specific to the SMS medium. This is that, perhaps unlike some modes of communication, receipt of a message on one’s mobile phone can be ignored. It can be equally well tended to later – at a convenient time.

That interruptions can be ignored when we focus on the task at hand has been acknowledged as a property of being engaged in “task-driven processing” [20]. And with SMS content delivery, the majority of our participants appeared to be able to stay in that mode with little disruption from an incoming item as was echoed in numerous quotes:

*“Timing...I mean, if it’s a little bit off, it’s OK, because **I can always push it off ‘til later.** The thing buzzing or beeping once doesn’t bother me too much.”*

*“When I wasn’t available I could **just not look at it.** And that’s fine.”*

*“There were plenty of times, where the timing was not good, but **I just ignored it,** it’s not like I really disliked it. The thing came in...the timing was not good, so I didn’t look at it.”*

*“Timing wise... **you don’t have to look at it immediately,** you can go back to it later. Phone allows that, so timing is probably less important than interest. I like the flexibility of that, text messages come in, they are lightweight, you don’t have to respond. If you do have to respond, the person will probably phone you.”*

The quality of SMS communication this person describes as “flexible” and “lightweight” in that “you don’t have to respond” alludes to the existence of *plausible deniability* [21] in SMS communication.

6. LOOKING CLOSER AT CONTENT

We now present and discuss the results from the web-based daily survey. In order to reveal possible underlying factors of receptivity, participants were asked to rate their interest in the

content, how entertaining they thought the content was, and the relevance and actionability of the content for each message.

Participants completed the daily survey for only 268 messages they responded to (out of 418 in total). A reason may have been that the web-based survey could only be accessed for the current and the previous day. Despite the fact that a reminder was sent out every evening surveys were often forgotten, and each omission meant that 6 messages at once were without survey responses.

6.1 Content factors

6.1.1 Interest

The median *interest* rating was 3 (mean: 3.24) on a 7-point Likert scale. A within-subjects ANOVA showed that participants were, unsurprisingly, significantly more interested in *good content* than in *bad content*, with $F(1, 10) = 21.09, p = .001$. Interest ratings for *good content* ($M=3.97$) were significantly higher than for *bad content* ($M=2.53$). This finding supports our hypothesis that interest is a predictor of receptivity (H1). Also, this strengthens the validity of the factor *content*: content in the *good content* categories really was significantly more interesting than content in the *bad content* categories. Again, *time of delivery* did not have a significant effect ($F(1, 10) = .083, p = .779$), nor did the interaction of *content* and *time of delivery*, with $F = .115, p = .742$.

6.1.2 Entertainment

The median for *entertainment* was 2 ($M=2.41$) on a 7-point Likert scale. Even though *good content* ($M=2.74$) was rated slightly more entertaining than *bad content* ($M=2.09$), a within-subjects ANOVA showed that the differences were not significant, with $F(1, 10) = 3.39, p = .095$. *Time of delivery* was not a significant factor, nor was the two factors' interaction.

6.1.3 Relevance

The majority of messages rated (66.4%) was not relevant to the participants, 29.5% of the messages were rated as being generally relevant regardless of the time read, and 4.1% of the messages were rated as being relevant at the time read or soon after.

6.1.4 Actionability

Actionability referred to the likelihood that the receiver of the message would subsequently take action or alter his or her actions in some way informed by the content of the message. *Actionability* is an ordinal variable. In the daily survey, the participants were asked: *To what extent did/will the content of the message change/influence your actions?*

Participants rated an overwhelming majority of 75% of the messages as unlikely to influence their future actions, and just 19.4% to be somewhat likely to influence future actions. Only 4.5% of the messages were rated as being very likely to influence future actions, while only 3 messages (1.1%) were rated to have already influenced their actions.

6.2 Underlying factors of receptivity

We have already shown that individually perceived good content fosters significantly higher receptivity than does bad content. But what makes *good content* good? To shed light on this, the interrelation of the factors measured in the daily survey and the participant's *gut reaction* was studied. How much do people's *interest* in the content, its perceived *entertainment* value, the *relevance* and *actionability* of the content influence people's receptivity?

A word of caution is probably appropriate here. The reader should keep in mind that the content factor ratings were collected in a different context (web-based, at the end of the day) than the gut reaction ratings that form the main part of the study (mobile, throughout the day). Furthermore, the chosen factors do probably not conclusively explain the latent factors that underlie receptivity.

Table 3: Mean gut reaction scores and within-subject effects by factors. Values before averaging shown in brackets.

Factor	Levels	Mean Gut react	Std. Deviation	N	$\chi^2(2)$	Exact p	Kendall's W
Interest	Low (1-2)	2.51 ^{2,3}	1.09 (1.2)	11 (138)	18.7	.000	.851
	Medium (3-5)	3.87 ^{1,3}	.82 (1.1)	11 (113)			
	High (6-7)	6.04 ^{1,2}	.57 (.78)	11 (48)			
Entertainment	Low (1-2)	2.97 ^{2,3}	1.09 (1.6)	11 (190)	16.2	.000	.738
	Medium (3-5)	3.93 ^{1,3}	1.44 (1.4)	11 (87)			
	High (6-7)	6.28 ^{1,2}	.27 (.86)	11 (22)			
Relevance	Not relevant	2.30 ^{2,3}	1.05 (1.6)	11 (199)	17.6	.000	.802
	Currently rel.	4.17 ¹	.34 (.81)	11 (12)			
	Generally rel.	4.83 ¹	1.54 (1.7)	11 (88)			
Actionability	Unlikely	2.87 ^{2,3}	1.02 (1.6)	11 (228)	19.6	.000	.890
	Somewhat likely	4.44 ^{1,3}	1.69 (1.8)	11 (56)			
	Very likely	6.04 ^{1,2}	.64 (1.5)	11 (12)			

¹ The mean difference to the first level is significant at the .05 level.

² The mean difference to the second level is significant at the .05 level.

³ The mean difference to the third level is significant at the .05 level.

To unpack the interrelations, we averaged the gut reaction for each participant per each level of the content factors. In order to be able to compare the effect size of each factor, we collapsed *interest* and *entertainment* from 7 into 3 levels. The binning introduces unequal variance between the levels (see the differences in the SD-column in table 3), so we conducted nonparametric Friedman tests for ordinal repeated measures (*k*-related samples) suitable for non-normally distributed populations for the four factors. To analyze the differences among the factor categories, pair-wise comparisons were conducted by means of a Wilcoxon test for 2 related samples. We report Kendall's W as a measure of effect size.

The Friedman tests showed that differences in *gut reaction* across different levels of the four factors were all significant at the .01 level. Furthermore, pair-wise comparison indicates that the mean *gut reaction* increased along with rising factor levels. For the ordinal variables interest, entertainment and actionability, the pair wise comparison shows that the mean gut reaction differs significantly for each level at the .05 level. Table 3 depicts the mean *gut reaction* per level of factor, including pair-wise comparisons and effect size (Kendall's W).

6.2.1 Discussion

The analysis supports our hypothesis (H2) that *interest*, *entertainment*, *relevance* and *actionability* of the content are all valid underlying factors of receptivity. This finding implies that qualities of the content influence receptivity in a substantial way and need consideration when studying interruptibility.

Interest was found to be a strong factor to influence receptivity; it proves to be the valid predictor of receptivity that we hypothesized it to be. This is also supported by the many statements of “interesting” content leading to a high rating of *gut reaction* apparent in the interviews, as seen in the discussion of the experience-sampling results above.

The fact that *content* did not have a significant effect on the participants’ rating of *entertainment* showed that the content we used was not well distributed on the scale of entertainment, i.e. just not that entertaining. However, it proved to have a significant effect on the participant’s *gut reaction*, indicating that the entertainment factor of content does play a significant role in the receptivity to the content, as suggested by earlier work [29, 25].

Only 79 out of 268 messages (29.5%) were rated as relevant (of which only 55 were classified as good content). Note that the average *gut reaction* for messages that people considered being generally relevant (4.83) is higher than their average *gut reaction* to *good content* (4.02). Therefore, the provision of content from categories people said they would be interested in does not necessarily imply that the *gut reaction* is always good, but if you manage to say something *relevant*, chances are higher that people will be more receptive, i.e. have a better *gut reaction*. In the interviews, some views about receiving irrelevant messages emerged: “I found it annoying to get messages that were irrelevant”, or “...content that was irrelevant to me I have no interest in getting pushed to me.” This is complementary to the finding that relevant interruptions are less disruptive [8].

Regarding the *actionability* of content, the fact that the majority of messages were classified as unlikely to influence future actions indicates a low signal-to-noise ratio of our content. So, while the data tentatively suggests that achieving high *actionability* may be a good predictor of receptivity (it had the largest effect), chances to deliver a message with likelihood to influence the recipient’s actions are low, at least with general content categories such as we used in this study.

Anecdotes from the interviews exemplify situations where the participants’ activities actually were influenced by a message. Messages that influenced the person’s activities often coincided with high *gut reaction* ratings of the messages. As a result of a message, one participant shared the content with his friends and thus used it as a “*conversation starter*”. Two mentioned the weather report in this context: “I think the weather thing let me form my plans differently a couple of times” and “Influenced in a way that it confirmed that I could do something, the weather ones”. Most commonly, five participants reported to have looked up the content of the news story, a movie or a music review online as a result of the message.

7. LIMITATIONS OF THE METHOD

The experience-sampling method is challenging in that the researcher has to deal with an incomplete description of the experience. Almost never can a response rate of 100% be achieved. The problem is intensified by the fact that we are investigating timing. A critique of the design may be that the factor *time of delivery* is biased, as it may well be the case that non-responses are simply a result of bad times of delivery. Thus, we looked at the distribution of the independent variables among the 223 messages that were not responded upon. An almost equal amount of non-responded messages was sent at good times (112) and at other times (111). A chi-square goodness-of-fit test showed that the distribution of the frequencies was not biased towards

good or other times, with $\chi^2(1) = .004$, exact $p = 1.0$. This shows that non-responded messages were not systematically biased toward bad times of delivery.

Likewise, the distribution of non-responded messages between good content (117) and bad content (106) is not significant; A chi-square goodness-of-fit test showed no significant bias towards *good or bad content*, with $\chi^2(1) = .543$, exact $p = .503$.

7.1 Response time

The ESM is particularly designed to be an ad hoc method by which the quality of an experience is assessed in the moment. We looked at the time participants took to respond to the prompts to elicit if the reaction of the participants actually happened *in situ*.

The distribution of *response time* was encouraging in terms of timeliness of response: 50 % of the messages were responded to within 373 seconds (6.21 mins), 76.1% were answered within 40 minutes after reception, and still 90% of the messages got responded to within 2 hrs. A within-subjects ANOVA did not yield any significant differences in average log-normalised *response time* for *perceived timeliness* ($F(6, 12) = .31, p = .92$) and *gut reaction* ($F(6, 6) = 2.56, p = .16$).

Different practices with data exceeding a certain response time are common; a lag of 30 minutes has been suggested as a boundary for valid data [24]. If we filter out all data points that exceed 30mins response time, our results do not change in a significant way. A within-subjects analysis on the basis of the 298 messages that were responded to in less than 30mins does not change the fact that *content* is a significant factor for *gut reaction* ($F(1, 10) = 21.02, p = .001$) and *time of delivery* is not ($F = .01, p = .915$).

8. DISCUSSION

8.1 Push vs. Pull

Even though a comparison of modes of delivery as in “push vs. pull” was not the focus of this study, questions about the general receptivity to pushed information in the interviews inevitably led to the discussion of alternatives.

Also, the purpose of this study was not to evaluate an “information push system”, as we deliberately delivered uninteresting content to the participants – obviously a property strongly discouraged in the design of a “real” system. Therefore, participants’ negative reactions towards getting content pushed to their mobile phones did not come as a surprise.

Two participants developed strong aversion against getting content pushed to them; one even said that she “...grew to really dislike hearing my phone tell me a text came in”. The other one said that he realized within three messages he disliked it: “the phone is only for keeping in touch with my own, close personal network”. This supports the notion that the delivery of content to mobiles might potentially be perceived as very intrusive, breaching the owner’s privacy [19]. Interestingly, the other person with strong negative feelings also mentions her close social network when describing her disappointment: “I was just not excited about receiving the texts. I get excited from texts from friends and family, but not these.” It seems that the expectation that a device is for the sole purpose of communication with a close personal circle influences the acceptance of it as a provider of other services. Only three participants mentioned that they prefer to get content pushed, as opposed to pull it themselves on demand. Six out of our 11 participants prefer to pull content, and

frequently use explanations such as that they favour being in control of what kind of content they consume at what time.

However, they often qualified under which conditions information push would be acceptable. Information push seems to have its merits too, one participant said: “*I like it pushed, because it frees myself from having to remember to go pull for it*”. Example content that participants see fit to be pushed to them includes reminders, event-driven news, information with high priority or urgency, and alerts of unpredictable events such as a rain. A sense of control seems important here as well, examples given are either on the side of explicit user feedback, such as mechanisms for the user to subscribe to certain content or signal unavailability, or on the implicit system side, such as algorithms that learn the user’s content preferences. Asked if he prefers information push or pull one participant said:

“*Pull it, absolutely. Although, if it says, here is something new for you, pick it up when you’re ready, that would be something different than getting a text message with an alert.*”

This participant’s statement points out that information push is acceptable under the condition that its delivery is not interruptive and that the content can be accessed on demand. This is certainly not a new paradigm of information delivery and has for instance been realized in subscribe and consume-on-demand RSS technology. However, it is interesting that many of our participants describe this kind of system as the ideal solution after taking part in the study.

All in all, the experience with an information push system from the study did not put our participants off the idea of information push in general, quite to our surprise. Instead, it seemed to lead them to conclude that an ideal system would be a hybrid between push and pull, a system that proactively provides smart and individually pre-filtered content unobtrusively and lets users consume content on demand at their own pace.

8.2 Receptivity vs. Interruptibility

The concepts of receptivity and interruptibility have often been treated synonymously in related work. However, we argue for a subtle but important difference. Interruptibility is the more prevalent concept that seems to be mostly informed by the rationale to predict an opportune moment when to interrupt someone [e.g. 2, 13, 15]. Hence, it is a concept that focuses on helping the *sender* of a message by reasoning about the *receiver’s* local contextual factors, such as current activity. Receptivity on the other hand has been defined as “one’s willingness to be interrupted” [3]. We extend this notion by saying that receptivity adds an important quality to the study of the interruption process: it places the anticipation of the *receiver’s* experience of the interruption into the focus of attention, i.e. it is inherently a concept related to the relational contextual factors as outlined above. Note that the goals of the two approaches remain the same: to ease some of the burden of inevitable interruptions. We prefer the concept of receptivity, as it is more user-centered in that it includes and anticipates the subjective and inherently uncertain and hard to predict experience of and reaction to the interruption.

9. CONCLUSIONS

The results from our study show that the content of a message plays an important role in influencing people’s receptivity. The participants’ gut reaction to good content was significantly better than to bad content. Also, our results show that interest, entertainment, relevance and actionability of the content all have a

significant effect on the gut reaction, giving some evidence that they are important informational qualities that influence people’s receptivity.

9.1 Content trumps time of delivery

The main finding from the study is that in the context of the asynchronous delivery method SMS, content trumps the time of delivery. We arrive at this finding mainly by contextualizing the results from the statistical analysis with the participant’s qualitative responses from the interviews. Furthermore, people’s comments from the exit interviews suggest that content can also dominate other operationalisations of timing more generally than our *time of day of delivery* operationalisation. Not only does the content influence people’s gut reaction, it appears that the content of an interruption also seems to outweigh its timing when people assess its appropriateness. This finding may inform any endeavour of understanding interruption management and building context-aware systems. Interruptibility is not a fixed property of time. The timing of an interruption is assessed by evaluating local and relational contextual factors including, but not limited to, the content of the interruption.

This finding implies that we must attend to properties of the content of an interruption when trying to assess someone’s interruptibility in addition to other local and relational context factors. For example, we have to consider the following kinds of questions: Is the content interesting to the interrupted? Is it relevant or actionable to the interrupted? And is it urgent or does it have a high priority for the interrupted?

9.2 Dealing with timing

However, by no means does the finding that content trumps time of delivery imply that the timing of an interruption can be neglected. The fact that *time of delivery* was not a significant factor shows that people are bad at predicting the times they are receptive, not that timing in general doesn’t matter. The control measure *perceived timeliness* also showed that our attempt to operationalise timing did not validly represent good and other times for people, they were just other times. In future work we need to address the problem of finding appropriate ways to manipulate the timing of interruptions in mobile settings validly, in order to accommodate for the situated uncertainty of timing.

Another reason why in our study we did not find the time of delivery to have an impact on the gut reaction may be that we found most of our participants to be well-trained in ignoring the interruption caused by the incoming message. People are accustomed to dealing with interruptions in their everyday lives; they have developed methods to cope with them [28]. An incoming message is not necessarily perceived as an interruption; the disruptiveness of an interruption is determined not only by the timing of a signal, the pressure to respond seemed to be more interruptive. Also, the medium of delivery plays an important role. Whereas the disruptiveness of a synchronous medium such as an incoming phone call is undoubted, SMS is an asynchronous medium, where messages can be tended to later with little or no decrement in convenience.

Nevertheless, the qualitative responses suggest that the timing of an interruption does influence people’s receptivity. Good timing is more variable and complex to predict as it is mediated by the local context of the recipients, such as their situated engagement in ongoing activities and their social surroundings. In order to make any sense of this immediate context, it would be necessary for a

system to do a great deal of detailed sensing and advanced inferencing about the recipient's surroundings and current state of engagement. Conversely, properties of the content as discussed in this paper are far more stable and straightforward predictors of a person's receptivity.

10. REFERENCES

- [1] Adamczyk, P. D. and B. P. Bailey. If not now, when?: the effects of interruption at different moments within task execution. *Proc. CHI 2004*. ACM Press (2004).
- [2] Avrahami, D. Fogarty, Hudson, S. E. Biases in human estimation of interruptibility: effects and implications for practice. *Proc. CHI 2007*. ACM Press (2007).
- [3] Begole, J., Matsakis, N. E., Tang, J. C. Lilsys: Sensing Unavailability. *Proc. CSCW 2004*. ACM Press (2004).
- [4] Bulmer, M. G. Principles of Statistics, Dover Publications (1979).
- [5] comScore: Mobile Internet Becoming a Daily Activity For Many. <http://bit.ly/aYOHkx>. Accessed: 23.02.2010
- [6] Csikszentmihalyi, M., Larson, R., Prescott, S. The ecology of adolescent activity and experience. *Journal of Youth and Adolescence* 6, 3 (1977), 281-294.
- [7] Csikszentmihalyi, M., LeFevre, J. Optimal experience in work and leisure. *Journal of Personality and Social Psychology* 56, 5 (1989), 815-822.
- [8] Cutrell, E. B., Czerwinski, M., Horvitz E. Effects of instant messaging interruptions on computing tasks. *Ext. abstracts CHI 2000*, ACM Press (2000).
- [9] Czerwinski, M., Cutrell, E., Horvitz, E. Instant messaging: Effects of relevance and time. *People and computers XIV: Proc. HCI 2000*, British Computer Society (2000).
- [10] Dey, A. K., Salber, D., Abowd G. D., A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction* 16, 2 (2001), 97 - 166.
- [11] Dourish, P. What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8, 1 (2004), 19-30.
- [12] Grandhi, S. A., Jones, Q. Conceptualizing interpersonal interruption management: a theoretical framework and research program. *Proc. Hawaii International Conference on System Sciences 2009*.
- [13] Ho, J., Intille, S.S. Using context-aware computing to reduce the burden of interruptions from mobile devices. *Proc. CHI 2005*, ACM Press (2005).
- [14] Hudson, J. M., Christensen, J., Kellogg, W. A., Erickson, T. "I'd be overwhelmed, but it's just one more thing to do": availability and interruption in research management. *Proc. CHI 2002*, ACM Press (2002).
- [15] Hudson, S., Fogarty, J., Akesson, C., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J., Yang, J. Predicting human interruptibility with sensors: a Wizard of Oz feasibility study. *Proc. CHI 2003*, ACM Press (2003).
- [16] Iqbal, S. T., Bailey, B. P. Leveraging characteristics of task structure to predict the cost of interruption. *Proc. CHI 2006*, ACM Press (2006).
- [17] Kalyanaramen, S., Ivory J., Maschmeyer, L. Interruptions and online information processing: The role of interruption type, interruption content, and interruption frequency. *Proc. Annual Meeting of Int. Communication Association 2005*.
- [18] Lasswell, H. D. The structure and function of communication in society. In Bryson, L. (ed.) *The Communication of Ideas*. New York, Harper (1948).
- [19] Lei, S., Zhang, K., Sha, E. Applying Situation Awareness to Mobile Proactive Information Delivery. In *Emerging Directions in Embedded and Ubiquitous Computing*. Berlin, Heidelberg, Springer (2007), 592-603.
- [20] Miyata, Y., Norman, D. A. Psychological issues in support of multiple activities. In Norman, D. A. (ed.) *User-Centered System Design*. Hillsdale, NJ, Lawrence Erlbaum Associates (1986), 265-284.
- [21] Nardi, B. A., Whittaker, S., Bradner, E. Interaction and outeraction: instant messaging in action. *Proc. CSCW 2000*, ACM Press (2000).
- [22] Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. *Proc. CHI 2005*, ACM (2005).
- [23] Schegloff, E. A. Sequencing in Conversational Openings. *American Anthropologist* 70, 6 (1968), 1075-1095.
- [24] Scollon, C. N., Kim-Prieto, C. Diner, E. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies* 4, 1 (2003), 5-34.
- [25] Shen, X., Chen H. An Empirical Study of What Drives Consumers to Use Mobile Advertising in China. *Proc. Grid and Pervasive Computing - Workshops, 2008*.
- [26] Sohn, T., Li, K. A., Griswold, W. G., Hollan, J. D. A diary study of mobile information needs. *Proc. CHI 2008*. ACM Press (2008).
- [27] Stone, A., Shiffman, S. Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine* 24, 3 (2002), 236-243.
- [28] Tolmie, P., Crabtree, A., Rodden, T., Benford, S. "Are you watching this film or what?" Interruptions and the juggling of cohorts. *Proc. CSCW 2008*. ACM Press (2008).
- [29] Tsang, M. M., Ho, S.-C., Liang, T.-P. Consumer Attitudes Toward Mobile Advertising: An Empirical Study. *International Journal of Electronic Commerce* 8, 3 (2004), 65-78.
- [30] Vastenburger, M. H., Keyson, D. V., de Ridder, H. Interrupting people at home. *Proc. 2004 IEEE Systems, Man and Cybernetics*.
- [31] Xia, L. and D. Sudharshan. Effects of Interruptions on Consumer Online Decision Processes. *Journal of Consumer Psychology* 12, 3 (2002), 265-280.
- [32] Xu, K., Zhu, M., Zhang, D., Gu, T. Context-aware content filtering & presentation for pervasive & mobile information systems. *Proc. Ambient media and systems*. Quebec, Canada, ICST (2008).