

TOWARDS SOCIAL AWARENESS IN UBIQUITOUS COMPUTING: A TURNTAKING APPROACH

Esko Kurvinen and Antti Oulasvirta

July 2, 2003

HIIT
TECHNICAL
REPORT
2003-1

TOWARDS SOCIAL AWARENESS IN UBIQUITOUS COMPUTING: A TURNTAKING APPROACH

Esko Kurvinen and Antti Oulasvirta

Helsinki Institute for Information Technology HIIT

Tammasaarencatu 3, Helsinki, Finland

PO BOX 9800

FIN-02015 HUT, Finland

<http://www.hiit.fi>

HIIT Technical Reports 2003-1

ISSN 1458-9451

Copyright © 2003 held by the authors.

Notice: The HIIT Technical Reports series is intended for rapid dissemination of articles and papers by HIIT authors. Some of them will be published also elsewhere.

TOWARDS SOCIAL AWARENESS IN UBIQUITOUS COMPUTING: A TURNTAKING APPROACH

Esko Kurvinen and Antti Oulasvirta

Helsinki Institute for Information Technology, Helsinki, Finland

Abstract

Context-awareness of social interaction is difficult for computers that are denied the social learning process of humans. Consequently, there has been scepticism about the enterprise. Basing on a turntaking analysis of group invitations, a complex social phenomenon, this pessimistic view is refuted. Turntaking provides an operationalizable account of social interaction and adds to the context debate by highlighting the dynamic and constructive aspects of context. Turns (actions) are recognizable by present-day technologies, but their sequential structures must be “hard-wired” by social scientists, as must be the meanings of static factors (e.g., location). Six challenges for attempts to implement turntaking as a model of social awareness are discussed.

Keywords: Context-Awareness, Turntaking, Ubiquitous Computing

1 Introduction

Mark Weiser envisioned ubiquitous computing devices “weaved into the fabric of everyday life until they are indistinguishable from it” [16]; a view that ran counter to the then prominent view of workstation-based human-computer interaction (HCI). One of the main unsolved challenges for Weiser’s manifesto is context-awareness of social activities. Without social sensitivity, provided services may be needless, proactive actions wrongly timed, and interaction styles inappropriate, eventually causing disturbance to social lives. This paper investigates one approach to solve this problem.

A basic problem for all computer models of the social is caused by the fact that computers can never learn social interaction as humans do because they are not social beings. Therefore, this knowledge has to be “hard-wired”, at least to some extent, into the context recognition apparatus. Because social phenomena are complex, and thus hard to model, some researchers have pessimistic views on the enterprise [e.g., 5].

How, then, could social activities be modelled? The prevailing approach has been to collect and interpret sensory data, such as presence of people [12], audio level in a room as indication of activity [12], activity of communication channels such as email [7], calendar entries and attendance history [9], body temperature, heart rate, galvanic skin response [11], presence of individual objects in vicinity and digital information attached to them [12], room temperature or lighting [14], and use these to infer the context—the Who, What, Where, When, and Why for example [1]. These static factors are, no doubt, useful in adapting to some simple and specific social activity. However, something more is needed when addressing complex social activities.

Context, though a relatively new concept for technology developers, has long been studied in social sciences, anthropology, ethnography, pragmatics, and linguistics [3, 4]. These have highlighted the dynamic and constructive aspects of context. Particularly, the turntaking approach investigated here emphasizes that events have sequential structure that unfolds in time [13]. The context of the participants is actively interpreted and constructed in the actions of an individual, to be interpreted and renewed again in the subsequent actions of the other participants [13, 15]. We here refer to these actions as turns. Context-renewing turns often consist of speech, but may also include various kinds of nonverbal acts, as will be shown later. Temporal dynamics make an important addition to static factors: For example, closeness of two persons (a *static factor*) in a cafeteria may not be enough to make inferences whether they are engage in the same activity, but we need to combine this information with bodily orientations and speech distributions, which requires monitoring actions over time.

Four important points can be discerned. First, turntaking considers context as being *tightly coupled* with the activity. Similar concerns have made by HCI researchers who approve that no general description of context can be given, but context depends on the particular application (see Issue 1 of Vol 5 of Personal and Ubiquitous Computing, [10]). Second, contexts are *multiply determined*—any situation can invoke several contexts. For example, when talking to a colleague, we may orient to her profession, but as well to her gender [15]. Third, people are not seen as passive parts of contexts but actively *constructing and transforming* them by taking turns in accord with their goals. Fourth, some amount of shared *preunderstanding* is necessary. Being able to take part in turntaking constitutes what it

means to be a part of that social group and deviations from the expected patterns cause disruption to social engagement. Conversely, appropriate actions may build group coherence.

In what follows, we aim to show, through three concrete examples, that this conceptions of context can be reasonably applied in an analysis of a complex social phenomenon, group invitations. We then discuss theoretical and practical challenges for implementing turntaking-inspired working models in ubiquitous computing devices.

2 Turntaking in Group Invitations

In the following, we present three cases of group invitations, ranging from a relatively simple case to one of the most complex cases in our data. The three cases analyzed here are chosen to illustrate how the participants of the invitation actively transform their social and bodily contexts in turns. In all of the invitations described here, the a person or a group of people (the *inviters*) propose another person or a group of people (the *invitees*) to join them to a *meeting* at some time and place in the future. Our interest is not so much in conversational dialogues [3] or direct and explicit invitations such as invitations to events like parties, weddings, or meetings at a workplace. Instead, we analyze invitations that are embedded in action and remain partly or completely implicit and yet recipients are able to recognize them as invitations and act accordingly. We wish to show how the context change is a sequential, *turn-by-turn* process between several participants. The turns in our examples consist of speech, SMS messages, gestures, movements, gazes, and bodily orientations.

The data is collected in a user-centered product concept design project for ubiquitous computing. The data gathering methods combine ethnography, diaries, and focus groups with such data analysis techniques as contextual inquiry [2]. The data presented here are extracted from observations conducted in fall 2001 in Helsinki.

2.1 Invitation to Café

The first case comes from observations of a group of four 22 to 24 year-old amateur actors (and close friends), three women and one man. The theater group spends lots of time together in a cafeteria called Kafka, located in the Helsinki city center in the lobby of a theater sponsoring their play. The café has become an important meeting place for the group—it is usually visited 3–5 times per week. Visits to Kafka are made alone, in pairs, or together with a larger group of friends. Even if just passing by, group members often take few extra steps to drop in to see if their friends are there.

When one of the group members wishes to meet another at Kafka, he/she often sends an SMS invitation containing plainly “Kafka”. This is usually enough; with the condensed message, the sender makes visible the plan of going to Kafka in the next hours or so. The message “Kafka” is not an explicitly agreed codeword, but it has developed or streamlined into an invitation as a result of practices of the group. Because of SMS, there is no need to fix a time beforehand. As it often happens, those invited just drop in later because they know that the inviters are likely to spend few hours at the cafeteria anyway. There is no need to specify a list or number of participants beforehand either. Because the meetings often lack a specific motive, nobody is likely to miss anything important if they do not show up. Thus, the meaning of the message, while being narrowed down to an invitation, is capable of projecting

several alternative responses. Details of the meeting are issues that can be dealt with in the possible replies or conversations via SMS or mobile phone.

To summarize, the inviter takes the initiating turn in creating a context by sending an SMS "Kafka". For the invitee, this marks a change in the inviter's context that could be, but does not have to be, reacted upon. This invitation, consisting of one nodal turn, while often failing to realize as a meeting, succeeds in producing group awareness and coherence [8].

2.2 Invitation to Sing-and-Play

Our second case is extracted from observations of three women, Irene, Jane, and Beth, at a playground located in eastern Helsinki (hereafter *the Park*). Every Thursday 9 am, Irene and Jane arrange a sing-and-play for small children and their parents at the Park. Schedules are posted for visitors to read at the front door of the main building.

In the following we describe one Thursday morning and the invitation to attend the sing-and-play:

Jane walks towards the workshop, a shed on the perimeter of the Park, where she gives Beth her wireless phone so that Beth can be on duty if someone calls while she and Irene are singing with the children. On her way to the workshop Jane nods to and greets some of the parents that she passes (see Figure 1A). "Feel like singing?", she asks from one of them. Jane arrives to the workshop and gives the phone to Beth. She tells her that they are going to the "lawn over there" and asks if she "wants to join". Beth says she has "got things to do".

Jane then walks towards the place where the singing is going to take place. Instead of taking the shortest route there, she walks first towards the center of the Park, closer to the main building where most of the parents and children are. There she turns towards the singing place. One of the adults asks: "Where is it?" Jane says that if the cote is not in order, they have to go the lawn next to the cote. "So are we starting now?" someone asks. "Yes", replies Jane.

Jane then passes a woman wearing an unusual blue dress. "Someone is so cute today" Jane comments. The woman tells she is dressed as a fairy. Jane proceeds towards the cote, where Irene joins her as she is checking the condition of the floor (see Figure 1B). Jane and Irene notice that it is not ok, because someone has had a fire in the center of the cote. They discuss and complain that these things happen quite often. They decide to move to the lawn next to the cote.

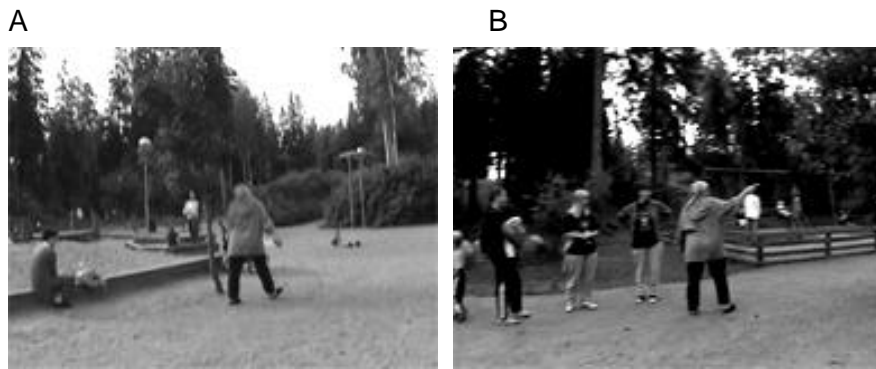


Figure 1. A, Jane walking through the Park, making casual remarks to parents on the way. She selects her route to make herself visible to those parents likely to attend sing-and-play, thus making indirect invitations. B, Jane on her way to the cote with parents willing to join the Sing-and-Play

Here, the inviter's turns take the form of casual remarks to present people and the presence of the inviter at the time and place of sing-and-play. The invitations are embedded in the small details of Jane's interactions with the potential participants for the play. Jane greets some, but not all people in the park. While doing so, she makes sure that those who have participated in the game before will notice that she has arrived and the play is about to begin. Her selection of the walking route and greetings and casual remarks to the people are not just compliments but they also function as invitations to participate.

Similar to our first case, the invitations to attend the sing-and-play need not to be presented as direct verbal questions or requests. Even though the sing-and-play is a scheduled event, it requires further specification of time and place, notifications to the potential participants, and gathering of those who eventually attend. That there is a sing-and-play every Thursday morning and that there is a poster on the wall build preconceptions that help other participants to reason Jane's intentions from her behavior. Furthermore, invitations are left implicit not only because the shared preconceptions make it possible, but also because it gives the invited people possibility to not to participate without having to give an excuse. This is contrast to the explicit invitation to Beth who, possibly in order to avoid implications to her preferences (not liking to attend?), gives an excuse that she's "got things to do".

2.3 Invitation to Lunch

Our third example comes from a set of observations conducted at an editorial office of a weekly IT-magazine located in the center of Helsinki (hereafter *the Office*). The Office is a somewhat typical open space where each worker has his/her own table either arranged next to someone else's or separated by low cubicle walls (see Figure 2). From his/her desk each worker can see most of the Office and recognize whereabouts and activities of co-workers at a glance. This is functional since editing a magazine is teamwork that calls for flexibility and ad-hoc management of tasks and resources. The work thus requires "context-sensitivity" from the part of the employees.

The journalists of the Office often have lunch in small groups. Though some journalists almost always have lunch together, the composition of the overall group may vary from day to day. The activity of organizing a lunch requires some negotiation before who goes to lunch

with whom, when, where and who may also join, is agreed upon. The invitation was usually achieved in somewhat in the following order:

1. Typically, prior to the invitation to a larger group, two or more people have already discussed on what to eat and where. This is carried out either in face-to-face conversations or via email. Most often a rough estimate of time, “noonish” or “after I’ve checked my email” for example, has been agreed upon at this time.
2. When the agreed time for lunch was due, first participants start to gather next to the coat rack close by the lifts. Typically, they put their coats on unhurriedly, just hanging around next to the lift for a while. Only rarely they shout out loud or ask directly if someone would like to join for lunch. Rather, the invitation to attend lunch was presented by making one’s loitering next to the exit visible to others.
3. At this point, when the invitation is on, other people in the office can, and often did, self-invite themselves and joined the group of lunch-goers. Sometimes the co-workers projected a receipt of invitation (“Are we going now?”), a request of verification that the invitation was understood correctly (“Going for lunch?”) or a re-quest for more information (“Kebab-place?”). If an orientation to an invitation was made clear, either an acceptance (joining the group) or negative reply, often in a form of an excuse, was presented (“I have to finish this article first”).

Here, a group of people invites others to lunch by loitering with their coats on in front of an elevator. All of the actions presented above, while not always containing an explicit formulation in respect to the invitation, are parts of the turn-by-turn production of the local context of going-for-lunch embedded in a larger context, the work setting of the Office. The participants know that at a given time, other journalist may have his/her work in a phase not allowing for interruptions. They therefore have to balance between the convenience and relaxation of having lunch together and possible disturbance caused by the invitation. That is why the invitation to attend lunch has evolved into an embedded, yet easily recognizable, routine of hanging about at the exit with overcoats on before stepping out of the Office.

In contrast to our second example, there was no explicit prior agreement behind the invitation to lunch. Still, going for lunch, as a result of it being a daily routine, is easily recognized by the participants and offers similar resources for interpretation of the behavior of the people in the Office. Whereas in our first and second example SMS messages and greetings were directed at specific individuals, the implicit invitation here does not have any recipients at all. Moreover, in contrast to the first two examples, the invitation was achieved in co-operation with others. A single person standing next to the exit does not make an invitation. The other journalists, still sitting at their desks, while recognizing that an invitation has been presented may select themselves as being invited, even when the invitation did not specify any recipients.

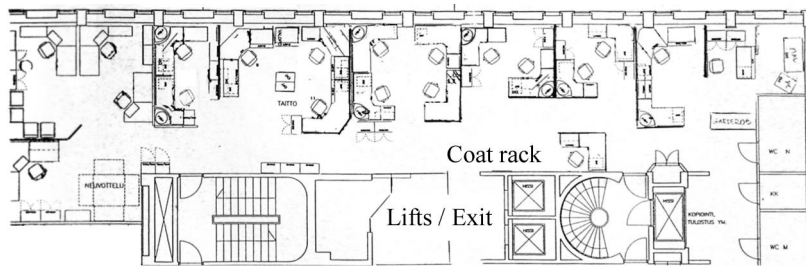


Figure 2. Layout of the Office.

3 Discussion

Through the three cases, we wish to have showed that a focus on dynamic and constructive aspects of context is necessary for understanding complex social activities such as invitations. Our analysis highlights that static factors of context are only a starting point in which people define what is relevant for them at a given time. Con-texts are not based on explicit agreements; rather, they are interactionally achieved through turns. We end the paper by discussing six challenges to this approach.

1. *Recognizing static factors.* Static factors form the preconditions or starting points for turns. For example, the sing-and-play invitation takes place only in one particular place and time of day (see 2.2). How could the device recognize them? In our opinion, because most of the present research has focused on static factors, we are likely to have some ideas on how to recognize them. As reviewed in Introduction, great many useful techniques have been developed for these purposes.

2. *Recognizing turns.* How to recognize turns? These “atoms” of interaction, seem to mostly detectable by present-day technologies, but maybe less than static factors:

– *Speech.* A great deal of interactions in our cases was organized around speech (2.2 & 2.3). Whereas speech recognition may not be possible for a while, it might be possible and useful to recognize where and when it occurs and how it overlaps and is paused [7] and how talk is distributed between participants (2.3).

– *Movement.* Recognizing not only peoples’ presence in a setting, but their direction of moving, could be used as an index of turn, as illustrated in 2.2 where the presence and direction of the inviter at the expected time indicated an invitation.

– *Bodily orientations and eye-gaze.* In human-to-human conduct, bodily orientations and eye-gaze play an important role. Knowing the direction of gaze and posture may thus help to interpret the focus of attention and turns of the person. For example, in the invitation to lunch case (2.3), people who oriented to the invitation by looking at the inviters had to come up with excuses why not to join them.

– *Messaging.* In contrast to technological development that generally moves toward contents of higher fidelity, the invitation to the cafe case (2.1) shows how very short and uninformative contents (“Kafka”) may be meaningful in context of locally developed and group specific practices [see also 6].

3. *Recognizing sequences of turns.* Turntaking is orderly and adheres to a sequential structure. For example, the three-step description of invitation to lunch (2.3) would do as a loose temporal description. A consequent problem for “hard-wired” model of turntaking is posed by the fact that there may be many possible turntaking activities that match any individual turn and vice versa (multiply determined turns).

4. *Interpreting meanings of turns.* How can be known which goal-pursuit, which turntaking activity is going on? Characteristic to our cases is that invitations are often presented trusting that the features of intertextual elements, existing collaborate practices and bodily orientations, “fill in” and make the invitation understandable. From this perspective, the problem lies not so much in recognizing turns but in what they mean: learning that the message “Kafka”, shape and timing of Jane’s walking route and her casual remarks to parents, or journalists’ loitering around the coat rack are all interpreted as invitations by the participants. It seems unlikely that any context-interpreting system could learn this purely by following the co-occurrences of people in the same location/time. This implies that some “semantics” of invitations or context changes in general must be preprogrammed. Whether this is really needed and to what extent it is feasible are yet unknown.

5. *Generality of models.* A related challenge is that the model must be general enough but still leave room for idiosyncrasies inherent in all behavior. This calls for data on 1) how turntaking activities vary and 2) how frequent they are.

6. *Sensor selection.* For obvious cost-efficiency reasons, not all sensors can be included in a device. The problem is that qualitative data such as ours, however, can only give slight hints on what sensors are useful in many situations. This knowledge must be therefore supplemented by quantitative estimations of how frequently activities occur and how well sensors can predict them. These estimations can be obtained, for example, from log analyses, surveys, or codings of videotaped activities.

To summarize, although turntaking seems to be based on sound conceptions of con-text and proved to be a useful analytical tool for understanding complex everyday social interactions, the above six challenges must be resolved in any practical application in context-aware computing. The second author is involved in such undertaking in the domain of mobile instant messaging.

Acknowledgements

We thank Michael Beigl, Jan Blom, and Hannu Kuoppala for helpful comments.

References

1. Abowd, G. D., Mynatt, E. D.: Charting past, present and Future Research in Ubiquitous computing. In: ACM Transactions on Human-Computer Interaction, Vol 7 (2000) 9-58.
2. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann Publishers, San Francisco (1998)
3. Goffman, E.: Replies and Responses. In: Goffman, E.: Forms of Talk. University of Pennsylvania Press (1981) 5-77
4. Goodwin, C., Duranti, A.: Rethinking Context: an Introduction. In: Duranti, A., Goodwin, C. (eds.): Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press (1992) 1-42
5. Greenberg, S.: Context as Dynamic Construct. Human-Computer Interaction, 16 (2001)
6. Koskinen, I., Kurvinen, E, Lehtonen, T-K.: Mobile Image. IT Press, Helsinki (2002)
7. Mynatt, E.D., Back, M. Want, R., Baer, M., Ellis, J.B.: Designing Audio Aura. In: Procs. of the Human Factors in Computing Systems (1998) 566-573
8. Mynatt, E.D., Rowan, J., Jacobs, A., Craighill, S.: Digital Family Portraits: Supporting Peace of Mind for Extended Family Members. In: Procs. of the Conference on Computer Human Interaction. ACM Press, New York (2001) 333-340
9. Mynatt, E.D., Tullio, J.: Inferring Calendar Event Attendance. In: Procs. of the 2001 International Conference on Intelligent User Interfaces (2001) 121-128
10. Nardi, B.: Context and Consciousness: Activity Theory and Human-Computer Interaction. MIT Press, Cambridge, MA (1997)
11. Picard, R.W.: Affective Computing. MIT Press, Cambridge, MA (1997)
12. Rekimoto, J, Ayatsuka, Y., Hayashi, K.: Augment-able Reality: Situated Communication through Physical and Digital Spaces. In: Procs. of the 2nd International Symposium on Wearable Computers. IEEE Comp. Soc. (1998) 68-75
13. Sacks, H., Schegloff E. A., Jefferson, G. A.: Simplest Systematics for the Organization of Turn Taking in Conversation. Language 50 (1974) 696-735

14. Salber, D., Anind, K., Dey, G., Abowd, A.D.: The Context Toolkit: Aiding the Development of Context-Enabled Applications. In: Proceedings of CHI'99, Pittsburgh, PA (1999)
15. Schegloff, E. A.: In Another Context. In: Duranti, A., Goodwin, C. (eds.): Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press (1992) 191-228
16. Weiser, M.: The Computer of the Twenty-First Century. Scientific American (1991) 94-141