How to Evaluate the Social Effects and User Experience of Systems Enhancing Collocated Interactions?

Thomas Olsson
Pradthana Jarusriboonchai
Susanna Paasovaara
Jarno Ojala
Ekaterina Olshannikova
Aris Malapaschas
Kaisa Väänänen
Tampere University of Technology
Korkeakoulunkatu 1, P.O.Box 553,
33101 Tampere, Finland
{firstname.lastname}@tut.fi

Workshop of Proxemic Mobile Collocated Interaction, in conjunction with *CHI'16*, May 7–12, 2016, San Jose, CA, USA.

Abstract
Enhancing collocated interactions with interactive technology has quickly gained plenty of interest in the HCI community. Our approach within this research domain is to design and study mobile and wearable technology that encourages and motivates social interaction between collocated people. This paper sheds light on the considerations related to evaluating systems that address the complex and delicate issues related to social interaction. How do we evaluate the effectiveness of enhancing social interaction? How do we know if a system or a prototype is of high quality? What are the quality attributes? We highlight various evaluation challenges we have encountered in our studies and provide considerations for future research.

Author Keywords
Collocated interaction; proxemic interaction; face-to-face; wearable technology; proactive systems; interaction design; user experience; evaluation.

Introduction and Related Work
While social computing has traditionally focused on enabling remote computer-mediated communication, recent research has shown that interactive technology can play a role also in collocated social interactions.

Face-to-face (F2F) interaction is a fundamental source of happiness and a powerful vehicle to represent oneself and build trust between strangers (see e.g. [3]). Interactive technology can support F2F interaction by, e.g., enabling content transfer between devices, increasing awareness of nearby people, and supporting multi-user interactions with large display areas (e.g. [2,8,11]). For example, Range [7] is an interactive public whiteboard supporting co-located, ad-hoc meetings. AutoSpeaker ID, Ticket2Talk and Neighborhood Windows [12] enable people to reveal information about their background and interests in different situations with proactive technology. CueSense [4] and MemeTag [1] are wearable devices for similar purposes. Also games and playful applications have been used particularly to encourage people getting together [9]. Overall, the design space for enhancing collocated social interaction is extensive, covering, e.g., systems that support, allow, enable or facilitate interaction.

Despite the active research and innovative designs, we know relatively little about the effects and user experience (UX) of such systems, especially in long-term. Like in many other areas within HCI we could ask: How can we evaluate the quality, user experience and effectiveness of the systems we design? And what actually is quality and UX in this application area?

Furthermore, there are various social issues that have generally not been in focus in previous research but that are topical and could benefit from ICT-based solutions. Loneliness, social ignorance, exclusion, and discrimination are vast societal issues that call for research activities also from the HCI community. For example, in modern societies people often feel unconnected to the others around them [13]. One might not know who the others are, what to talk about, or if they could be interrupted in the first place. Interestingly, it has been argued that some of such issues are partly caused by the use of technology. For example, people can get very engaged in the activities with their mobile devices and tend to create a private mobile bubble around themselves, becoming absent-minded to surrounding people [10].

We claim that particularly such complex social phenomena and challenges require new approaches in designing and evaluating interactive technology based solutions. For example, when aiming to make an impact on social interaction one could benefit from proactive and socially aware applications that not only *support* but also *encourage* or even *initiate* social interaction. This leads us to new challenges related to design and evaluation: which of the – often sensitive – social issues can be solved with interactive technology? What are the desirable design goals for such systems from both individual and societal perspectives? How do we know if the goals have been met?

Evaluation Challenges in Our Research

To dive deeper in the evaluation challenges, in the following we give an overview the various challenges and considerations that we have identified in three of our own prior and ongoing studies.

Social Display [6] provides light-weight visual cues about a user's current activity with a mobile device. The cues are displayed on a secondary screen attached to the backside of a mobile device (Figure 1). This aims to break the mobile bubble, increase awareness of nearby people about oneself, and thus encourage

Figure 1. The Social Display revealing the user's currently used app to surrounding people.

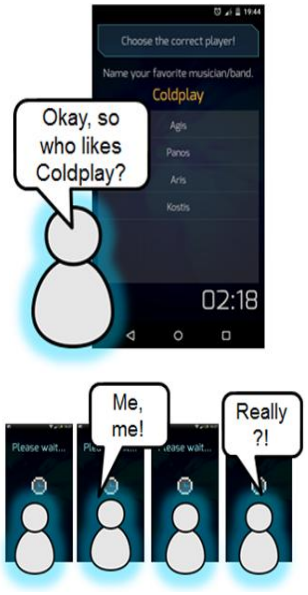

Figure 2. Screen shots and use case example of Who's Next.

serendipitous social interactions. We conducted a ten-day field trial with 13 participants using the prototype. The study was methodologically extensive and relatively long-term and the results indicate that the prototype did increase awareness of users' mobile activities and occasionally triggered interactions with others. Nevertheless, we identified various challenges regarding evaluating the system.

First, measuring the very subjective experiences of felt privacy and self-presentation require a long-term trial study in authentic situations, which often becomes too costly and time consuming. Even a 10-day period is too short for the users to understand all the risks and possibilities or to establish routines in using the application. Second, measuring the secondary users' (i.e. spectators') experiences and increased awareness is very challenging. Only a part of the consequences could be recognized and logged: Only the actually initiated interactions could be noticed by the primary user who was our information source. The intentions and thinking of others around them could not be captured because not every passer-by can be involved in a study. Third, it is hard to generalize the results beyond the specific culture where the study was conducted and the participants' contexts of use. Social approachability is a delicate construct that is affected by numerous cultural, social and situational aspects.

Who's Next [5] is a tool for icebreaking in newly-formed groups that meet face-to-face for the first time. With a turn-based quiz game about personal information it aims to relieve tension and create a delightful atmosphere (Figure 2). In a study with five group sessions it was found to create a playful and relaxed atmosphere; a simple playful application managed to encourage participants to interact with each other. However, it is hard to assess – or especially prove – its true effectiveness or compare it with other icebreaking tools.

First, the first meeting of a group can only happen once. Therefore, comparisons to other possible solutions would have to happen between groups. This is not a sound option either as the variance in personalities within the groups can be a significant factor affecting the results more than the tool itself. Second, any measurements during the activity can interrupt the natural interaction and affect the fragile social atmosphere. As social experiences are strongly subjective, we need to ask people's opinions rather than indirectly measure the interaction. However, the act of inquiry makes people aware of the studied aspects, which can bias their experiences and thinking. Third, recruiting groups of people that would consist of strangers is laborious and prone to judgement errors. Also, due to recruitment challenges, this study suffered from biased sampling, which is common in HCI. This meant including relatively many participants who are socially rather active and open-minded, rather than people with social inhibitions who could actually benefit from the system through increased and technology-facilitated social interaction.

Finally, we are currently working on Next2You, a playful social application for getting to know who the strangers in the vicinity actually are. The application automatically exchanges interesting user-created pieces of information in the users' profiles whenever the users are within the Bluetooth range from each other. Various playful features like receiving achievements by meeting face-to-face aim to

encourage – or even enforce – sharing detailed information and social interaction. For this system we have so far only conducted focus groups for early concept evaluation but we expect to encounter various evaluation challenges with the functional system.

First, there has to be enough of users in a specific geographical area (i.e. critical mass) to keep the application active enough in order to maintain the users' interest. This is particularly relevant in this kind of a system that is based on user-generated content and has no meaningful use without other users. Second, as the application is based on mutual sharing (reciprocity) and unplanned opportunistic encounters in everyday life, we need to somehow ensure that the user-created content is interesting enough without, however, affecting the motivations of use. Especially after a "cold start" where only few people are actively using the application the users can easily stop using it and the true potential of the concept remains as a mystery and impossible to evaluate. At the same time, external incentives like prizes for most active users must be used carefully and the application needs to be rewarding and motivating enough in itself. Third, evaluating a new concept with a lo-fi proto like a scenario or video storyboard can give misleading information as the participants might lack first-hand experience in long-term and real situations. Therefore, the upcoming evaluations should not be too much based on the previous findings but explore the ground with an open mind. Moreover, strictly predefined measures like Likert statements can narrow the scope unnecessarily while open-ended questions can mislead the participants to focus on irrelevant aspects (e.g. the usability of a mock-up UI instead of the appropriateness of the concept).

Considerations for Evaluation Practices

Overall, we have encountered and anticipated various relevant challenges to tackle in the future research. Some of them are general challenges that are present also in other areas of HCI, as well as in psychology and other disciplines that use people as informants. Particularly the challenges in sampling and achieving sufficient ecological validity in the study setting are general research challenges. However, we claim that many of the challenges are even more significant when the target of research is social interaction and utilizing technology to enhance it. The quality and experience of social interaction are affected by, e.g., match of the personalities, current moods, the topic of discussion and the type of joint action. Social phenomena are generally so sensitive to various contextual factors that a lab test or even a quasi-experiment in the field might not provide realistic results.

The challenges bring us new interesting research questions to address in the future. *What are the best indicators of good quality* in this area? For example, is success defined by number of social interactions or rather the quality of them? Qualitative data about subjective experiences can be hard to use as credible proof. But if we use objective measures to investigate the resulted interactions (i.e. estimating the experience from the use of application) it might not tell why it was experienced good or bad.

Furthermore, *what evidence of effectiveness is actually enough?* What is enough for convincing other researchers for the publication-worthiness of a study or selling the ideas to the industry to encourage them try out the ideas in larger scale? In fact, a practical consequence of these evaluation challenges is the

practice that we often see in companies today: Digital services are launched very quickly and constantly iterated according to feedback and log data. While this kind of trial & error –based approach is generally recommendable and very common also in science, it can lead to incremental innovations that just mimic the earlier best practices and successful ideas. However, as science aims to explore novel and unorthodox ideas and approaches, thus also identifying new ideas for business, we still need ways to evaluate which of the various ideas is actually worth developing further. While the industry might opt out of evaluation before having something concrete to show, in science we need to be able to evaluate already early, non-functional ideas (e.g. because the idea is not feasible with current technology) and validate the assumptions we have made. In other words, we need to ask *what kind of prototypes are enough to be able to identify the benefits, drawbacks, opportunities and risks comprehensively and reliably enough to inform design activities and strategic choices?*

Finally, the role of interactive technology in social processes needs to be considered as a part of a larger solution, e.g. supporting non-technical solutions like societal discussion and setting norms and regulations. A technical tool – or intervention – aiming to somehow enhance collocated interaction can in fact only play a small role in what the outcome is. *How can we best assess the effects of the designed systems when there are other elements in the overall solution?*

Conclusions

To conclude, we want to outline a few activities to consider in future research. First, *we should define what are the desirable qualities of user experience (and other quality attributes) in systems for collocated interaction*. What particular social experiences is it desirable to design for? The existing user experience models could be enriched with more detailed social experiences that collocated interactions can engender.

Second, after defining the phenomena what to evaluate, we need to *develop the quality criteria and measurements for them*. When studying complex and sensitive things like social interaction we need to develop mixed methods approaches for analyzing and evaluating the systems and their use more comprehensively than today. This means evaluating both the subjective user experiences and the objective effectiveness, utilizing both predefined measures and open-ended questions, looking into both short-term & long-term behavioral impacts, and understanding and measuring the tradeoffs between positive and negative experiences (e.g. the lost privacy vs. increased opportunities to socialize because of the Social Display). Also, we need to understand both the individuals' viewpoints and the groups': A group opinion is not necessarily the average of the individuals'.

Finally, in long-term we would ideally also *identify what are the indirect society-level indicators of quality and success* of such systems. Can we see the long-term impact by looking into the user reviews in mobile application markets or from the number of people's contacts in social media, or should we optimistically look into larger traces of increased wellbeing and happiness in, e.g., national surveys? All in all, we believe that the suggested research activities could allow the community to make even more valid evaluations and well-informed system designs in the future.

Acknowledgements

The research was funded by Academy of Finland (grants 264422 and 283110).

References

1. Richard Borovoy, Fred Martin, Mitchei Resnick, Brian Sflverman, and Chris Hancock. 1998. Meme Tags and Community Mirrors : Moving from Conferences to Collaboration. *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, ACM Press, 159–168. http://doi.org/10.1145/289444.289490

2. Xiang Cao, Michael Massimi, and Ravin Balakrishnan. 2008. Flashlight jigsaw: an exploratory study of an ad-hoc multi-player game on public displays. *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, ACM Press, 77. http://doi.org/10.1145/1460563.1460577

3. Erving Goffman. 1959. Presentation of Self in Everyday Life. 55, 17–25.

4. Pradthana Jarusriboonchai, Thomas Olsson, Vikas Prabhu, and Kaisa Väänänen-Vainio-Mattila. 2015. CueSense: A Wearable Proximity-Aware Display Enhancing Encounters. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*, ACM Press, 2127–2132. http://doi.org/10.1145/2702613.2732833

5. Pradthana Jarusriboonchai, Aris Malapaschas, Thomas Olsson. 2016. Design and Evaluation of a Multi-Player Mobile Game for Icebreaking Activity. *To appear in proceedings of CHI'16*. ACM Press.

6. Pradthana Jarusriboonchai, Thomas Olsson, Aris Malapaschas, Kaisa Väänänen. 2016. Increasing Collocated People's Awareness of Mobile User's Activities: Field Trial of Social Display. *To appear in proceedings of CSCW'16*, San Francisco, USA. ACM Press.

7. Wendy Ju, Brian A Lee, and Scott R Klemmer. 2008. Range : Exploring Implicit Interaction through Electronic Whiteboard Design. 17–26.

8. Reuben Kirkham, Thomas Ploetz, Sebastian Mellor, et al. 2013. The Break-Time Barometer – An Exploratory System for Workplace Break-time Social Awareness. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*, ACM Press, 73–82. http://doi.org/10.1145/2493432.2493468

9. Christian Licoppe and Yoriko Inada. 2012. 'Timid encounters': a case study in the use of proximity-based mobile technologies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2759-2768. DOI=http://dx.doi.org/10.1145/2207676.2208675

10. Minsam Ko, Chayanin Wong, Sunmin Son, et al. 2015. Lock n' LoL: Mitigating Smartphone Disturbance in Co-located Social Interactions. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*, ACM Press, 1561–1566. http://doi.org/10.1145/2702613.2732819

11. Andrés Lucero, Jussi Holopainen, and Tero Jokela. 2011. Pass-them-around: collaborative use of mobile phones for photo sharing. *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, ACM Press, 1787–1796. http://doi.org/10.1145/1978942.1979201

12. David W. McDonald, Joseph F. McCarthy, Suzanne Soroczak, David H. Nguyen, and Al M. Rashid. 2008. Proactive displays: Supporting Awareness in Fluid Social Environment. *ACM Transactions on Computer-Human Interaction* 14, 4, 1–31. http://doi.org/10.1145/1314683.1314684

13. Sherry Turkle. 2011. Alone Together: Why We Expect More from Technology and Less from Each Other. Retrieved from http://dl.acm.org/citation.cfm?id=1972496