User-Centred Design and Evaluation of Ubiquitous Services

Rahat Iqbal, Janienke Sturm, Olga Kulyk, Jimmy Wang and Jacques Terken
Department of Industrial Design, Eindhoven University of Technology
P.O.Box 513, 5600 MB Eindhoven, The Netherlands
{R.Iqbal, J.Sturm, O.A.Kulyk, C.Wang, J.M.B.Terken}@tue.nl

ABSTRACT
Theoretical and technological progress has revived the interest in the design of services for the support of co-located human-human communication and collaboration, witnessing the start of several large-scale projects over the last few years. Most of these projects focus on meetings and/or lecture situations. However, user-centred design and evaluation frameworks for co-located communication and collaboration are a major concern. In this paper, we summarise the prevalent approaches towards user-centred design and evaluation, and we develop two different services. In one service, participants in a small-group meeting receive real-time feedback about observable properties of the meeting that are directly related to the social dynamics, such as individual amount of speaking time or eye-gaze patterns. In the other service, teachers in a classroom receive real-time feedback about the activities and attention level of participants in the lecture. We also propose ways to address the different dimensions that are relevant to the design and evaluation of these services (the individual, the social and the organisational dimension), bringing together methods from different disciplines.

Categories and Subject Descriptors:
H.5.3 [**Information interfaces**]: Group and organization interfaces – Collaborative computing - Computer-supported collaborative work; J.4. [**Computer applications**]: Social and behavioural sciences – Sociology.

General Terms Design, Documentation, Human Factors.

Keywords Ubiquitous/Pervasive computing, Evaluation, Human-human interaction, Peripheral feedback, Social dynamics, Meetings, Lectures

1. INTRODUCTION
Whereas early work on supporting collaboration focused on co-located (same-time same-place) collaboration, in the nineties the attention shifted towards supporting distributed collaboration. Much of this work focused on providing *computer tools* to support humans in their collaborative work, such as co-authoring features, shared displays, automatic summarisation, etc. Over the last decade, new visions and developments in technology have put co-located collaboration in the focus of attention again. With respect to visions, Weiser put forth a vision of ubiquitous computing, according to which people and environments are augmented with computational resources [1]. These resources provide information and services unobtrusively whenever and wherever required.

The ubiquitous computing paradigm, with computers being everywhere, calls for new technology that prevents humans from feeling overwhelmed by information. 'Calm technology' implements this idea by putting computers in the periphery of our attention until needed [2]. In line with this vision of ubiquitous computing and calm technology, several recent projects (such as AMI, Interactive Workspaces and CHIL) envisage systems that - rather than being used as a tool - support human-human communication in an implicit and unobtrusive way, by constantly monitoring humans, their activities and their intentions. These systems provide services that, acting as pro-active butlers, free the conversational participants from having to operate and attend computers, enabling them to focus on human interaction and communication. In order to understand the context and intentions, perceptual components are required that can recognise facial expressions, speech, gestures, attention etc. to monitor, for example, who is present, who is speaking and about what. In addition, rather than only supporting the work process, another important aspect of these services is to monitor and support the *social dynamics* (the rhythm of the conversation), which entails taking a facilitator role. Although this aspect has not been addressed by many projects, considering and supporting the social dynamics may solve frequently occurring problems during group collaboration and thereby improve the quality of the work.

As has been amply documented in the proceedings, methodologies for the user-centred design of systems and services supporting co-located communication and collaboration need to go far beyond traditional human-computer interaction (HCI) approaches/methodologies. In designing such systems, we should take into account the experiences with group support systems, which have been shown to fail on several aspects related to the social dynamics, such as the fact that existing groupware applications often require additional work without any clear benefit for the users or that using groupware can lead to disruption of social processes [3]. For the evaluation of these systems, techniques from traditional HCI should be augmented with techniques used in collaborative work.

In line with this vision of ubiquitous/pervasive computing and calm technology, we initiated our work in the context of an EU

SIGDOC'05, September 21-23, 2005, Coventry, United Kingdom.

funded Integrated Project called CHIL (Computers in the Human Interaction Loop) [4]. CHIL envisages systems that support human-human communication in an unobtrusive way, by constantly monitoring humans, their activities and their intentions. These systems provide services that, acting as pro-active butlers, free conversational participants from having to attend and operate computers, enabling them to focus on human interaction and communication. In order for the system to be able to understand the context and intentions, it needs perceptive components that can recognise facial expressions, speech, gestures, attention etc. to monitor, for example, who is present, who is speaking and about what. In addition, rather than only supporting the work process, another important aspect of these services is to monitor and support the social dynamics (the rhythm of the conversation), which entails taking a facilitator role.

The remainder of the paper is organised as follows: section 2 provides an overview of related projects aimed at supporting face-to-face meeting or lecture scenarios. Section 3 discusses prevalent approaches to requirements engineering and design. Section 4 is directed towards studying evaluation issues and requirements for new evaluation metrics. Section 5 focuses on small-group meetings and the lecture scenario and describes the design considerations. Section 6 discusses different aspects of this survey highlighting user-centred design techniques and evaluation measures. Section 7 concludes this paper by providing an outline for future research.

2. SUPPORTING MEETINGS AND LECTURES

Several large-scale projects in Europe and the United States that focus on supporting human-human communication have addressed issues related to supporting, capturing and understanding co-located, multi-party interaction in the context of meetings or lectures. The current section gives an overview of these projects (without aiming to be exhaustive).

Many projects in this field adopt a task-oriented approach, providing tools that facilitate the collaboration process. For example, the Stanford Interactive Workspaces project [5] explores new possibilities for people to collaborate in technology-rich spaces, using computing and interaction devices on many different scales [6]. The idea is that user's attention should remain focused on the work being done, rather than on the mechanics of interaction. One of the guiding principles of this project has been the reliance on social conventions: Users and social conventions take responsibility for actions and the system infrastructure is responsible for providing a fluid means to execute those actions. An augmented meeting space has been built (the iRoom) with technology such as large displays, wireless and multimodal I/O devices, and seamless integration of mobile and wireless "appliances" including handheld PC's. Research areas that are addressed in this project include: multi-device, multi-user applications; multimodal and fluid interaction; integration of large (wall-sized) displays with advanced visualisation capabilities; integration of computing "appliances" including PDA's, scanners, digital cameras, etc.

The MIT MeetingManager is a multi-user multimodal collaboration tool for planning, facilitating, and browsing structured meetings [7]. The MeetingManager is a combination of the meeting process-focused work of the CSCW community, and the free-form meeting support that is the focus of the HCI community. The MeetingManager has four components that work together: a planner, a facilitator, a summariser, and a browser. Before the meeting takes place, the participants use the planner to make and store the agenda. During the meeting, the facilitator takes the stored agenda for real-time meeting assistance. After the meeting, a brief summary of the recorded multimedia is e-mailed to the participants. Finally, the meeting is stored in a database available for convenient browsing. Planned research activities include adding components for more natural human-computer interaction, such as hand gesture recognition, speaker detection and tracking, and non-verbal discourse cue recognition.

The InterSpace project at the Fraunhofer Institute [8] investigates novel group interaction techniques for collaboration with multiple and heterogeneous devices. Due to the heterogeneity of the devices, software must be designed in a way that enables collaboration even when people use different devices. Some devices, such as small personal digital assistants (PDA's), are designed for personal use, whereas a large interactive display fits well for group work situations, but is also suited for personal information access. The context of a device (such as its location or the presence of other people) may also influence the kind of activities it can be used for. An environment can only be of benefit for its inhabitants, if it is designed in a user-centred and integrated way. Context has a significant influence on the design. Therefore, interaction styles and user interface concepts have to be developed that are - on the one hand - abstracted and independent of actually used devices, and - on the other hand - can be tailored for different devices (implying different interaction modalities) while maintaining the overall consistent look and feel.

Several projects focus mainly on recording meeting data and facilitating access to the recorded data. The European M4 project (Multimodal Meeting Manager), for instance, is concerned with the construction of a system to enable structuring, browsing and querying of an archive of meetings that have been analysed automatically by means of multimodal sensors [9]. As the true information of meetings is created from interactions between participants, true understanding of meetings can only emerge from considering their group nature. The project focuses on enabling analysis of human interaction in small group meetings. The smart room of the project is envisaged to automatically identify its inhabitants, transcribe what they say, infer emotional states and facilitate the exchange of information, using the following technologies: audiovisual speaker tracking, speech segmentation and enhancement, audiovisual person identification, and group action recognition.

The European project AMI [10] (Augmented Multiparty Interaction) targets computer enhanced human interaction in the context of smart meeting rooms and remote meeting assistants [11] The project aims to enhance the value of multimodal meeting recordings and to make human interaction more effective in real time. These goals are being achieved by developing new tools for computer-supported cooperative work and by designing new ways to search and browse meetings as part of an integrated multimodal group communication, captured from a wide range of devices. The underpinning technologies that are used are human-human communication modelling, speech recognition, computer vision and multimedia indexing and retrieval.

The ICSI Meeting Recorder project [12] also focuses on recording, recognising and understanding meetings. Natural multi-party interaction presents significant challenges not just to speech recognition, but also to speaker technologies, to discourse modelling, to spoken language understanding, and to audio retrieval. Research efforts range from low-level processing of the speech signal, including higher-level analyses of meeting structure, content, and interactions.

Many projects over the years have focused on enhancing the note taking process and facilitating access to meeting and lecture contents afterwards, such as FiloChat (HP), Dynomite (Xerox), eFacilitator (MIT), etc. [13,14,15]. The large-scale eClass project (formerly known as Classroom2000) at Georgia Tech studies the impact of ubiquitous computing on education [16]. The ultimate goal of the project was to revolutionise the classroom experience through the evolutionary introduction of a natural and useful capture, integration and access service. A prototype classroom environment was built with the necessary software infrastructure to seamlessly capture much of the rich interaction that occurs in a typical university lecture. By capturing the different streams of activity in the classroom (electronic annotations, audio, video, and Web browser activity) and presenting an easily accessible interface that integrates those streams together, the need for mundane note-taking is reduced, which allows the student to engage in and better understand the classroom discussion.

Few projects explicitly address the social dimensions of group collaboration. DiMicco et al. examined how a shared display, showing the level of participation of each participant affects the behaviour of the group during a collaboration task [17]. The results indicate that the presence of such a display influences the behaviour of group participants in the extremes of over and under participation. The ATR Media Information Science Labs applies ubiquitous sensor technology to capture and analyse the dynamics of multi-party human-to-human conversational interactions [18]. They try to capture the dynamics of the conversational roles that participants take by analysing their verbal and non-verbal behaviour. Their aims are to empirically examine conversational participation processes and to investigate the possibilities of using information on fine-grained verbal and non-verbal exchanges in conversations for sharing memories and experiences. Their system provides conversational partners with real-time information feedback on the status of ongoing conversations.

The CHIL project (Computers in the Human Interaction Loop) aims at combining the task-oriented and social dynamics approaches. The objective of the CHIL project is to create environments in which computers serve humans who focus on interacting with other humans as opposed to having to attend to and being preoccupied with the machines themselves. The project aims to design computer services that model humans and the state of their activities and intentions.

3. APPROACHES TO REQUIREMENTS ENGINEERING AND DESIGN

In this section, we will look at approaches towards requirements engineering and design that have been taken by the large-scale projects that were summarised in the previous section. Additionally, we pay attention to the social aspects of the system design and investigate how these techniques deal with social issues. We believe that the prime importance of ubiquitous computing is the understanding of the social characteristics of work, people, and real context of work. This is also what makes the realisation of ubiquitous computing difficult.

Research shows how and why many large-scale projects in the past have failed [19]. One of the key reasons of their failure is inadequate analysis of user requirements. Most importantly, social, political and cultural factors have not been brought into question for the development of these systems. For example, the failure of ‘office automation’ systems to support a group of individuals performing their usual tasks was based on inadequacy of incorporating social aspects of groups and merely focusing on functional requirements of individuals. It is in this respect that traditional analytic approaches are found wanting, representing an intrusion of the ‘engineering mentality’ into areas where it is inappropriate. To get a better idea, we provide an overview of how researchers are dealing with the issue of user requirements[1].

Dix and his colleagues raised the point that requirements for an interactive system may not be thoroughly specified from the beginning of the life cycle [20]. The best practice is therefore to build some features of the potential design and test them out on real users. Such an approach, also known as rapid prototyping, has been used in several projects supporting multimodal human interactions. For example, the Fraunhofer Institute has applied the rapid prototyping approach to some of the projects being developed in the institute in order to capture user requirements and to uncover the mistakes or misinterpretations of information in previous requirements. The advantage of a rapid prototyping approach is that it shortens the life cycle overcoming the problems of incomplete requirements by means of several design iterations.

Another user-centred design approach is based on story or scenario development. Scenarios are particularly useful where the user requirements are not so clear [21]. The usage of scenarios for system design is not a new approach; researchers have been deploying such approaches for the last two decades. However, previous approaches focused on single user applications. Recently researchers introduced scenarios to meet the needs of a team or a collaborative work [22]. Researchers at Xerox PARC, for example, employed scenarios of interaction in order to develop ‘Audio Aura’ that explores how peripheral awareness of relevant office activities could be enhanced using ambient sound in a mobile setting [23]. They developed scenarios based on available information about how people work together - for instance, gathering at the coffee bistro, dropping by people’s offices - in order to understand issues related to interaction within the PARC environment. The deployment of this approach provided insight and thus helped designers to improve the design.

Some of the projects, particularly those aiming at developing an interactive system that involves data capturing, apply an approach known as interaction analysis at the early stage of the project. The basic assumption of this approach is that researchers start developing video-based observation and analysis methods to elicit post-hoc reflections from the users [24]. A similar approach has been applied in the Multimodal Meeting Manager (M4) project, focusing on annotation and analysis of meeting room data by

[1] A more extensive overview of methods and approaches to support requirements engineering and design is available from the COMIC Project (Computer-based Mechanisms of Interaction in Cooperative work) (COMIC, 1993).

researchers in conjunction with the meeting participants in the early stages of the project. At a later stage, the project focused on user studies, by enquiring users about the use that people already made of recorded meeting data [25]. This helped the researchers to find out the events of interest and provide effective means of browsing these events.

In other projects, the point of departure for system development was provided by a strong foundation of user studies. For example, extensive user studies have taken place to inform the design of Dynomite [26]. In order to understand how people use notebooks, an artefact walkthrough was conducted on their note-taking behaviour. Following the walkthrough interviews were conducted. Everything was videotaped for later analysis. The investigators asked the interviewee about the type of notebooks they use; the ways in which they take notes; how they use notes; the potential of audio as an enhancement to the handwritten notes and their frustration with their current notebook and their ideas for an ideal notebook.

As the development moves beyond the individual user, we should recognise the socially organised character of work and include this aspect in the requirements engineering process. To acknowledge the fact that work has a social dimension, researchers move from the laboratory to the field. Applying ethnographic methods of investigation can unfold the social aspects of work practices in the 'real world'. An ethnographic approach was taken, for example, in the Flatland project, to inform the design of the system [27]. Together with questionnaires and interviews, this study has been able to unfold and examine the details of social organisation of daily practices in office environment. For the development of Filochat the requirements were captured through observation and user studies [13]. The observation study provided a good basis to conduct a series of interviews afterwards. Whittaker and his colleagues conducted two group studies: firstly, they interviewed people who had experience of using audio recording in the offices in order to identify the pros and cons of the usage of audio; secondly, they focus on non users of audio in order to investigate their note-taking activities in the meetings. This has provided a significant amount of data for comparison purpose. Further examples of the ethnographic approach are provided by Poltrock et al. (2003) [28] and Crabtree [29].

Many researchers argue that integration of different methods may produce better results and thus helps to achieve a system meeting the needs of the users [30]. For example, Van der Veer et al. developed a conceptual model for groupware task analysis based on a combination of traditional HCI approaches and ethnography [31]. The value of the eClass project also lies in the fact that a variety of methods was deployed for data collection [32,33]: web-log analysis with session tracking, questionnaires, controlled experiments and classroom observations. The CHIL project also uses a combination of different methods in order to get an understanding of communicative processes and to identify typical behaviours in different kinds of meetings. Some of these methods were taken from user-centred design methodology including focus groups, and others from social sciences such as ethnography. The purpose has been to get a detailed and fine-grained impression of the activities of the team and of the function of the artefacts that are used by the team, in order to propose new services. In addition to this, a detailed survey was conducted in order to collect data on different aspects of the behaviour of participants during lectures.

4. EVALUATION AND SOCIAL IMPLICATIONS

In the previous section, we have discussed several approaches that have been deployed to capture user requirements for the development of ubiquitous computing. Regardless of the quality and the richness of the requirements engineering techniques, the developed system still has to be tested to ensure that it behaves in accordance with expectation and meets the user requirements. In the subsequent discussion, we will focus on how researchers and developers have deployed different styles of evaluation in order to assess their systems. We will also argue that we need to divert our attention to other disciplines and import models and theories that can be an asset for the development of ubiquitous computing or collaborative technology.

Usually a distinction is made between *formative* and s*ummative* evaluation. Formative evaluation is meant to inform designers and developers designing the service or application and getting user feedback about preliminary versions. Summative evaluation is meant to inform the client or the external world about the performance of the service or application in comparison to a situation where there is no such service available, or to a previous version, or to competing services; in brief, to demonstrate the usefulness of the system. Since most systems are still in a preliminary stage of development, so far most evaluation has been of a formative nature.

In the context of ubiquitous computing, formative evaluation is primarily aimed at answering questions concerning bottlenecks in the interaction of the users with the services and at revealing shortcomings of the perceptive components, the ingredients of the context model and the reasoning and decision algorithms. Concerning the first, questions are for example: what are the user needs concerning interaction technologies for recording personal notes; how can users provide the system with the information it needs to be able to adapt to the personal needs and context of the user. Concerning the second, questions are for example: what information does the system need to capture in order to drive the services; what information needs to be provided by the context model in order for the system to decide to take a particular action; what is the performance of the algorithms by which the system anticipates a particular user need.

For many years, researchers have been using traditional HCI methods of assessment, but with the introduction of CSCW, researchers strived to incorporate other evaluation methods as the traditional methods mostly focus on single user applications and do not adequately address collaboration issues. Importantly, the methods should rely on some understanding of the context in which the system will be operated. One example of such a method is Collaboration Usability Analysis (CUA) [22], which allows the researchers to model the task that people will perform. The focus of CUA is on developing a scenario incorporating tasks, individual task instantiations, collaborative task instantiations and actions. However, many researchers believe that the only way to get a true picture of the system is to evaluate it with real users in real situations. Studying a system in a real context with real users - although it helps to contextualise the evaluation - is expensive and time-consuming, and it may be of little use or even impossible if the system is not fully developed [34].

Only a few projects have addressed questions that are more related to summative evaluation. For instance, in the context of the eClass project, extensive evaluation has been conducted to assess the contribution of the service to students' performance [33]. Putting the prototype system of the eClass project into operation in a real context provided valuable results focusing on both technical and social aspects. For the evaluation of eClass, a mixture of evaluation methods was applied. Observation studies were carried out to investigate the real use of the system in real context. Additionally, an extensive amount of questionnaires was collected from a wide number of students as well as lecturers. On-line logging was used to validate the results obtained from the survey. In addition, the evaluation employed controlled experiments. Two groups of students were selected, one with eClass support and one without. Finally, a survey was performed to get to know how students feel about privacy while using eClass. The Filochat system [13] has also been subjected to summative evaluation: the actual use of the system has been evaluated by the meeting participants in Hewlett Packard Research Laboratories. The participants were observed during the meeting while they were using the system, and afterwards a set of questions were forwarded to them to inquire their satisfaction. A series of evaluations have provided promising results. Summative evaluation was also conducted in the Dynomite [14]: the designers evaluated the usefulness of properties and keywords for low-overhead indexing, examined the usage of the indices to organise and display information and determine whether audio highlighting is usable and useful. In addition, the designers performed a pilot study to test the usability and usefulness of audio highlighting Finally, in a similar fashion, DiMicco et al. investigated the effect of displaying the social dynamics on group performance, albeit not extensively [17].

In sum, it appears that, with the exception of a few projects such as eClass, Filochat and Dynomite, many projects focus primarily on formative evaluation, and substantial evaluation of the usefulness of services has not taken place. In particular, the evaluation of the usefulness of services in terms of their effects on the social dimension is a relatively unexplored area.

5. OUR DEVELOPMENTS

In the EU-funded Integrated Project "Computers in the Human Interaction Loop" (CHIL), perceptive technologies are developed that enable us to conceive systems that perceive and interpret the ongoing course of events in situations of co-located communication and collaboration, such as lectures and small-group meetings. In the framework of the CHIL project, we conceive services that can take action autonomously or provide feedback to the participants in such meetings about the state of affairs in an unobtrusive manner. In this section, we present two different services. In one service, participants in a small-group meeting receive real-time feedback about observable properties of the meeting that are directly related to the social dynamics, such as individual amount of speaking time or eye-gaze patterns. In the other service, teachers in a classroom receive real-time feedback about the activities and attention level of participants in the lecture. We hypothesize that providing this kind of feedback pervasively may help meeting participants and lecturers to adjust their behavior to the demands of the meeting or lecture

5.1 Supporting Lectures

Computers increasingly find their way in the educational system. For example, many teachers nowadays make use of notebooks to deliver their lectures. Electronic devices can be used for educational purposes in many different ways, inside the classroom, for example for electronic exercises, as well as outside the classroom, for example by granting access to lecture content after the lecture. In this way, laptops are becoming essential items for students and teachers, both inside and outside the classroom.

As a first step in the design process a focus group was conducted that was aimed at collecting teachers' opinions concerning issues relating to the use of electronic devices during their lectures. The main questions were whether they make use of electronic devices during their lectures, either for students or for themselves, and what types of services they would consider to be useful. The focus group consisted of five experienced teachers and professors at Eindhoven University of Technology. It took place in a friendly environment and was led by a facilitator. The main conclusions of the focus group substantiate the general merits and demerits of the services that are reported in the literature as reviewed above.

Following the main findings from the focus group, we develop a service that supports the teacher by providing information about students' attention and interest levels during the lecture. As an indication of the level of attention, the service shows how many students are actively taking notes, as opposed to playing games, chatting, or browsing the web. Interest level represents the interest of the students who attend to the lecture (either taking notes or not).

5.1.1 Peripheral Display

In order to support the lecturer, we have developed a peripheral display as shown in Figure 1. It provides two kinds of information: the attention level of the student and the interest level of the student. The attention level is depicted in the form of a pie chart, showing the percentage of students that are busy taking notes in red ("BUSY") and the percentage of students that have finished taking notes in green ("READY"). Green in this case means that the lecturer may move on, whereas red denotes that students are not yet ready to move on. The grey area (or "OTHER") indicates students who are doing other things on their notebooks, such as chatting or browsing the web and whose activity is irrelevant to the teacher's decision whether to move on or not. The interest level for BUSY and READY students will be detected automatically from nonverbal cues (for the current simulation, interest level is indicated by means of a slider bar. This information is presented to the teacher in the form of a scale, with red indicating a low interest level and green indicating a high interest level. The horizontal line indicates the average interest level of the students across the whole lecture. The peripheral display is updated dynamically in real-time. The update rate of the information will be optimised by experimentation, so as to make sure that the information accurately and meaningfully reflects the current situation, but is not too distracting for the lecturer.

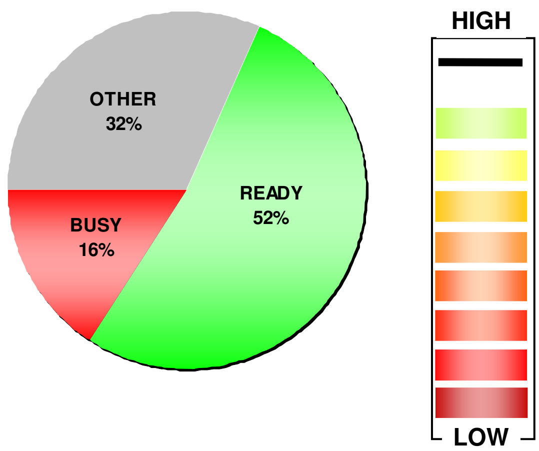

Figure 1: Visualization of attention level (left) and interest level (right)

5.2 Supporting Small Group Meetings

Current technology supports mainly content and information exchange during meetings, whereas social aspects have been addressed only recently. Focusing on improving suboptimal group performance, a recent study reported that feedback on speaking behaviour that is presented on a shared display during the meeting can influence the participation level [17]. In the framework of the EU-funded CHIL project, we develop a service that conducts real-time processing of audio-visual cues and generates an unobtrusive representation of this information during the meeting. After a user requirements study, we designed a visualization service, which presents information on current and cumulated speaking activity in combination with information about the visual focus of attention of speakers and listeners. Our goal is to make the members aware of the current meeting flow, and in this way influence the group's social dynamics, which we define as the way verbal and nonverbal communicative signals of the participants in a meeting regulate the flow of the conversation [2][3].

5.2.1 Design concept

The visualization is projected in the centre of a table, as shown in Figure 2 for a four meeting participants setting. Each table side represents a participant (P1, P2, P3, P4). The visualization contains the following components: (1) The right-hand circle (coded **Sa**) represents how much attention a participant received while speaking from the other participants since the beginning of the meeting. (2) For the current speaker, this circle is surrounded by an outer, lighter-coloured ring representing how much visual attention s/he receives from the other participants. (3) The middle circle (coded **S**) represents the participant's cumulative speaking time since the beginning of the meeting. (4) Again, for the current speaker, this circle is surrounded by an outer, lighter-coloured ring, the size of which represents the duration of the ongoing turn. (5) The left-most circle (coded **A**) indicates how much visual attention the participant – as a listener - has received from the other participants while they were speaking (added up across all other participants). The different circles are distinguished by different colours. The information is updated dynamically in real-time. Visual attention will be computed from head orientation data.

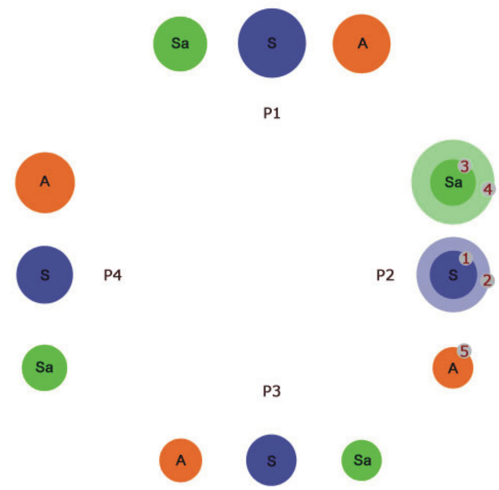

Figure 2: Time-dependent visualization of current and cumulative speaking activity and attention for each participant

6. DISCUSSION

While we have seen a successful move towards supporting human-human communication and collaboration by means of unobtrusive services using perceptive technologies, substantial work remains to be done. With the emergence of ubiquitous/pervasive computing, researchers have urged the importance of extending the attention beyond the basic usability issues and investigate the sociological, organisational and cultural impact of computing. For instance, Ramage has provided substantial arguments in his thesis in favour of deploying multiple methods of evaluation in order to adequately address all the aspects of collaborative work [30]. As we also observe, the focus of the evolving large-scale projects is primarily on the *social* and *organisational* effects and therefore, only usability measures at the level of *individual* user-system interaction will not be sufficient.

The usefulness of services for co-located human-human communication and collaboration needs to be conducted at multiple levels: the individual level, the social level and the organisational level [cf. 35]. With respect to the individual level, questions arise such as to what extent the services enable lecture attendants or meeting participants to reach their individual goals. For instance, a service helping students to make personal notes during lectures, in combination with a browsable record, may help them perform better in the courses (see the evaluations conducted in the eClass project). Similarly, a combination of a personal notes service and a browsable record may help meeting participants to retain a more accurate record of the meeting outcomes for later use, while resolving the conflict between having to write down what was discussed before and attending to what is being discussed right now. Clearly, methodologies from educational and cognitive psychology can be applied here, and the eClass project already did quite a good job, although more detailed investigations will be needed. (Resolving the conflict between attending to the present and summarising what went before may also have social benefits, as a service relieving meeting participants of the burden for maintaining an accurate record during the meeting will enable them to contribute to the meeting more effectively.)

A second issue at the individual level concerns situations where the system acts on its own account, when it perceives certain needs or desires of the user. In this case, the system needs to decide whether and when to perform a certain action. When evaluating such services, the adequacy of system actions needs to be assessed against the cognitive load imposed by the system actions: system actions will potentially interfere with the current activities of the user. In order to investigate these questions, methodologies from cognitive psychology can be applied. Recently, a number of studies on the effects of interruptions have addressed this issue [e.g. 36].

The *social level* concerns two aspects. In the first place the social dynamics have an *instrumental function*, in the sense that they constitute an important determinant of the quality of the process and the outcome of a meeting (the organisational goals). The extent to which different participants have the opportunity to contribute to the conversation will directly affect the added value of collaboration over individual work, as it provides the opportunity to profit from multiple views and complementary expertise. Also, the feeling of having the opportunity to contribute increases the degree of commitment (although this may be dependent on the cultural background). In the second place, the social dynamics have an *intrinsic function*, as they directly affect the satisfaction of the participants during and after the meeting and play an important role in team building.

With respect to evaluation, these different aspects bring different requirements. The social dynamics in terms of their instrumental function will be included in the evaluation from the organisational perspective, as it concerns the effectiveness and efficiency of the group process. With respect to the intrinsic function, evaluation might include both subjective and objective measures. Subjective measures involve questionnaires as mentioned below or interviews that address the satisfaction of individual participants. Although such methods provide indispensable information, a disadvantage is that they only provide information *after the fact*, and are often quite global, whereas one would like to relate subjective judgments directly to particular events and episodes during a meeting.

Both with respect to the *social* and *organisational level*, there are basically two ways to proceed. In one approach, investigators use fictitious problems that require people to optimise their communicative processes in order to arrive at the solution of the problem. One such example is the hidden-profile game, which has been developed within business schools to investigate group performance. A hidden-profile task requires that subjects role-play as members of a committee that must decide how to address several problems. Each subject is required to assume one of four different roles, where each role possesses unique, yet complimentary, information. Hidden-profile tasks are often used to investigate how the composition of a group and the communication technology influence information sharing.

In the other approach [cf. 37] groups are observed and group performance is assessed as they are engaged in their normal activities. Paris and colleagues [37] distinguish three primary types of measures: descriptive measures - what is happening at any given time and documenting individual and team behaviour; evaluative measures - which judge performance against identifiable standards; and diagnostic measures - identify the causes of behaviour and contribute inputs to the feedback process necessary to improve subsequent performance. Diagnosing problems is best done through process measurement. The second measure has been a central point of focus while the first and the third have not been fully addressed. Towards this direction, the work of DiMicco et al. [17] and Katagiri et al [18] makes a first step.

One aspect that Salas and co-workers have addressed is the cognitive aspect of communication and collaboration, emphasising the importance of shared mental models in expert team working. Obviously, certain cognitive conditions should be met for successful communication to take place. This perspective paves the way for evaluation methodologies that take inspiration from the domain of cognitive psychology.

In addition, evaluation should link more directly to the social dimension as such, focusing on the social dynamics of human-human communication and co-located collaboration. Here, inspiration can be taken from ongoing research in business administration departments, although further work is needed. Primary evaluative measurement techniques still consist of observation of the process and of outcome evaluation by experts, often employing checklists. Also, team members are asked to fill questionnaires themselves. Questionnaires such as ITAP's Global Team Process Questionnaire (GTPQ) require team members to fill in a questionnaire that mainly addresses enabling conditions for effective and efficient team performance [38]. The GTPQ has been widely applied in many fields. It has been subjected to various statistical tests, which showed that the questionnaire is reliable and has construct validity within the categories that have been established for summarising the results of the questionnaire. Therefore, the strengths and weaknesses of teams within five dimensions can reliably be identified using the GTPQ, they can be compared to others and they can be monitored over time.

Concerning the *organisational* goals, we may look at measures that relate to quality of the output, to throughput and resource consumption. With respect to quality of the output, again expert judgements are the primary source of information. Measures related to throughput and resource consumption need to be applied with caution, as time saving may not always be the primary target of the organisation. Instead, using resources effectively and strengthening the team spirit in the long run will pay off to the organisation more than simple time saving would.

7. CONCLUSIONS

In this paper we have described state-of-the-art techniques directed towards the development and evaluation of ubiquitous computing in the context of meeting and lecture scenarios. We have also enlightened the long acknowledged issue in the field of collaborative computing to address the social and organisational aspects of services that support co-located human-human interaction and collaboration. We believe that little attention has been given to these issues in the past, while the rapid proliferation of ubiquitous computing in recent years has encouraged researchers to move beyond the basic usability issues and to focus on the social and organisational impact of computing.

We argued that the development and evaluation of ubiquitous computing requires a wider mix of disciplines such as HCI traditional methods, cognitive science, sociology, anthropology, social psychology and user-centred design methods.

In this paper we have also outlined two services that provide a basis for further investigations into the field of ubiquitous computing and calm technology. These services will be further investigated, taking into account important issues of information awareness, cognitive load associated with switching from primary to secondary task and specific requirements for evaluation of this type of feedback.

8. REFERENCES

[1] Weiser, M. (1993): 'Some Computer Science Issues in Ubiquitous Computing', *Communications of the ACM*, 36, 75-84.

[2] Weiser M. and Seely Brown, J. (1996): 'The Coming Age of Calm Technology', in P. J. Denning and R. M. Metcalfe "*Beyond Calculation - The Next Fifty Years of Computing"*, Copernicus/An Imprint of Springer-Verlag.

[3] Grudin, J. (1994):'Groupware and social dynamics: Eight challenges for developers', *Communications of the ACM*, 37(1), 92-105.

[4] http://chil.server.de

[5] http://www.ipsi.fraunhofer.de

[6] Johanson, B., Fox, A., and Winograd, T. (2002): 'The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms', *IEEE Pervasive Computing Magazine*, 1(2).

[7] Oh, A., Tuchinda, R. and Wu, L. (2001): 'Meeting Manager: A Collaborative Tool in the Intelligent room', *Proceedings of Student Oxygen Workshop*, 2001.

[8] http://www.m4project.org

[9] Renals, S. and Ellis, D. (2003): 'Audio Information Access from Meeting Rooms', *Proceedings of ICASSP 2003*, Hong Kong.

[10] http://www.amiproject.org

[11] McCowan, I., Gatica-Perez, D., Bengio, D., Moore, D., and Bourlard, H. (2003): 'Towards Computer Understanding of Human Interactions', *Proceedings European Symposium on Ambient Intelligence (EUSAI)* (invited keynote paper), Eindhoven.

[12] http://www.icsi.berkeley.edu

[13] Whittaker, S., Hyland, P., and Wiley, M. (1994): 'Filochat: Handwritten Notes Provide Access To Recorded Conversations', *Proceedings of CHI '94*, Boston, Massachusetts, USA.

[14] Wilcox, L.D., Schilit, B.N., and Sawhney, N. (1997): 'Dynomite: A Dynamically Organized Ink and Audio Notebook', *Proceedings of CHI'97*, 186-193.

[15] Fox, H. (2003): 'The eFacilitator: A Meeting Capture Application and Infrastructure for Intelligent Environments', *Proceedings of Student Oxygen Workshop*, 2003.

[16] Abowd, G.D. (1999): 'Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment', *IBM Systems Journal, Special issue on Pervasive computing*, 38(4), 508–530.

[17] DiMicco, J.M., Pandolfo, A., and Bender, W. (2004): 'Influencing Group Participation with a Shared Display', *Computer Supported Cooperative Work*, ACM Press, USA. pp. 614-623.

[18] Katagiri, Y., Bono, M., and Suzuki, N. (2004): 'Capture Conversational Participation in a Ubiquitous Sensor Environment', *Pervasive Workshop on Memory and Sharing of Experiences*,.

[19] Goguen, J.A., and Linde, C. (1993): 'Techniques for Requirements Elicitation, Requirements Engineering', *IEEE Computer*, pp. 152-164.

[20] Dix, A., Finlay J., Abowd, G., Beale, R., *Human Computer Interaction,* Prentice Hall, 2004.

[21] Bardra, K.E. (2000): 'Scenario-Based Design of Cooperative Systems Re-designing an Hospital Information System in Denmark', *Group Decision and Negotiation*, 9, 237-250.

[22] Pinelle, D., and Gutwin, C. (2003): 'Task Analysis for Groupware Usability Evaluation: Modeling Shared-Workspace Tasks with the Mechanics of Collaboration', *ACM transactions on Computer-Human Interaction*, 10(4), 281-311.

[23] Abowd, G.D. and Mynatt, E.D. (2000): 'Charting Past, Present, and Future Research in Ubiquitous Computing', *ACM Transactions on Computer Human Interaction*, *Special Issue on HCI.*.

[24] Ehrlich, K. (1999): 'Designing Groupware Applications: A Work-Centered Design Approach', in Beaudouin-Lafon (ed.): *Computer Supported Cooperative Work*, 1999, John Wiley & Sons Ltd.

[25] Wellner, P., Flynn, M., and Guillemot, M. (2005): 'Browsing Recorded Meetings with Ferret', in Bengio, S. and Bourlard, H. (Eds.) *MLMI*, pp.12-21, Springer-Verlag Berlin Heidelberg.

[26] Wilcox, L.D., Schilit, B.N., and Sawhney, N. (1997): 'Dynomite: A Dynamically Organized Ink and Audio Notebook', *Proceedings of CHI'97*, 186-193.

[27] Mynatt, E.D., Igarshi, T., Edward, W.K., and LaMarca, A. (1999): 'Flatland: New dimensions in office whiteboards', *ACM conference on Human Factors in Computing Systems*.

[28] Poltrock, S., Grudin, J., Dumais, S., Fidel, R., Bruce, H., and Pejtersen, A.M. (2003) 'Information seeking and sharing in design teams', *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, November 09-12, 2003.

[29] Crabtree, A. (2003): *Designing Collaborative Systems: A Practical Guide to Ethnography*, London: Springer-Verlag.

[30] Ramage, M. (1999): 'Evaluation of Cooperative Systems', *Ph.D. thesis*, Cooperative System Engineering Group, Lancaster University.

[31] Van der Veer, G.C., Welie, M.V, and Thorborg, D. (1999): 'Modeling Complex Processes in GTA'*,* *Ergonomics*, 42(11).

[32] Brotherton, J.A. (2001): 'Enriching Everyday Experiences through the Automated Capture and Access of Live Experiences: eClass: Building, Observing and Understanding the Impact of Capture and Access in an Educational Domain', *Ph.D. Thesis*, Georgia Tech.

[33] Brotherton, J.A. and Abowd, G.D. (2004): 'Lessons Learned from eClass: Assessing Automated Capture and Access in the Classroom', *ACM Transactions on Computer-Human Interaction*, Vol. 11, no.2, June 2004, pp. 121-155

[34] Steves, P.M., Morse, E., Gutwin, C., and Greenberg, S. (2001): 'A Comparison of Usage Evaluation and Inspection Methods for Assessing Groupware Usability', *Proceedings of Group 2001*.

[35] Neale, D.C., Carroll, J.M. and Rosson, M.B. (2004): 'Evaluating Computer-Supported Cooperative Work: Models and Frameworks', ACM Press: *Proceedings of CSCW'04*, Chicago, pp. 112-121

[36] Horvitz, E., Kadie, C., Paek, T., and Hovel, D. (2003): 'Models of Attention in Computing and Communication: From Principles to Applications', *Communications of the ACM*, 46 (3), 52-59.

[37] Paris, C.P., Sala, E., and Cannon-Bowers, J.A. (2000): 'Teamwork in Multi-person System: A Review and Analysis', *Ergonomics*, 43(8), 1052-1075.

[38] Bing, J.W. (2001): 'Developing a Consulting Tool to Measure Process Change on Global Teams: The Global Team Process Questionnaire™', paper presented at the *2001 Academy of Human Resource Development Conference*, Tulsa, OK.