

Multimodal support for social dynamics in co-located meetings

Jacques Terken · Janienke Sturm

Received: 10 June 2009 / Accepted: 22 October 2009 / Published online: 10 March 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In this paper, we present a system that employs perceptual technologies (i.e. technologies that perceive the context through sensors such as cameras and microphone) to provide feedback about people's behaviour in small group meetings. The system measures aspects of behaviour that are relevant to the social dynamics of the meeting, speaking time and gaze behaviour, and provides visual feedback about these aspects to the meeting participants through a peripheral display. We describe the system properties and the perceptual components. Also, we present a study aimed at evaluating the effect of such a system on meeting behaviour. Groups of participants, amounting to 82 participants in all, discussed topics of general interest. Analysis of the data of 58 participants showed that feedback influenced the behaviour of the participants in such a way that it made over-participants speak less and under-participants speak more. Analysis of the micro-patterns of six participants indicated that feedback on gaze behaviour had little effect on the interaction dynamics. We conclude that perceptual technologies can be used to build services that may help people to improve their meeting skills and we consider some ways in which such systems may be deployed in meetings.

Keywords Co-located collaboration · Social dynamics · Meeting support · Multimodal interaction · Eye gaze

1 Introduction

Meetings are more and more important in structuring daily work in organizations. For example, according to a survey [9], executives spend on average 40–50% of their working hours in meetings. At the same time, meeting participants feel that 50% of that time is unproductive. This situation is determined not only by task-related factors (e.g. a difficulty of choosing the right items for the agenda, and/or of focusing the attention on relevant issues), but also by the complexity of the social dynamics in small groups, which may hinder the performance of teams. Some participants talk too long, parts of the discussion may actually involve only a subset of the participants, and the social and task roles that participants play are not supportive of conducting an effective and satisfactory meeting.

In order to improve the social dynamics, external interventions such as facilitators and training experiences are commonly employed. Facilitators participate in the meetings as external elements of the group, and their role is to help participants to maintain a fair and focused behaviour, as well as to direct and set the pace of the discussion. Facilitators are expensive, however, and are not available to all teams. The recent advancements of perceptual technologies provide opportunities for developing automated services that can provide participants with feedback about the social dynamics, both during and after the meeting, with the ultimate goal to enable them to display more effective behaviours and enjoy a more satisfactory group experience.

The use of technology to support group meetings has appeared as early as 1971 [17]. Most available technologies are directed at supporting task-related activities such as the creation and preservation of content and exchange of information. Tools such as an electronic whiteboard, a

J. Terken (✉) · J. Sturm
Technische Universiteit Eindhoven,
Den Dolech 2, 5612 AZ Eindhoven, The Netherlands
e-mail: j.m.b.terken@tue.nl

projector, video and audio recorders and electronic minutes have been used for brainstorming, idea organizing and voting; the associated methods for working with these tools have been refined over the last two decades. Support for organizational and social aspects of meetings, such as time-keeping and agenda-tracking, group effectiveness and satisfaction of meeting members, has received relatively little attention [20], although the interest on those topics has risen recently. The availability of technologies that are able to perceive and process rich multimodal information makes it possible to explore the possibility of providing some of these services (semi-)automatically. In this paper, we present a multimodal system that monitors groups and provides real-time feedback about the participants' speaking time and eye gaze. Also, through an evaluation study, we show that such feedback affects the social dynamics of the meeting.

2 Related work

In the field of CSCW, the focus is often on distributed meetings. The social relationships between meeting participants have been recognized as a fundamental aspect of the meetings' efficacy since the seminal work of Tang [32]. Many different attempts have been made to bring the social dynamics at a "visible" level. For example, Dourish and Bly [8] investigated the effects on groups of providing information about the distributed meeting context without using a full video-conferencing system. They designed a system, called Portholes, consisting of a simple chat-based system augmented with a shared database of regularly updated visual information available at all sites. Their findings suggest that across-distance awareness can provide more effective communication and improved interaction and can contribute to a shared sense of community. Another example in this respect is the work of Erikson et al. [10], who proposed the idea of "social translucence", that is, graphical widgets that signal cues that are socially salient, for example, the presence and activity of those involved in the current conversation. The claim is that such functionality makes it easier for people to carry on coherent discussions, observe and imitate others' actions, create, notice and conform to social conventions and engage in other forms of collective interaction.

In our work, we deal with face-to-face (co-located) meetings. Again, most of the research in this area is aimed at providing easy access to computerized services for individuals or groups to efficiently accomplish their tasks, and support for organizational and social aspects of co-located meetings has received relatively little attention. For example, in the CHIL project [34], most of the services provided were aimed at offering better ways of connecting

people (the Connector service) and supporting human memory (the Memory Jog). Recently, there has been some interest in the automatic analysis of group interaction, but this has been focused mostly on technological challenges, e.g. McCowan et al. [22], Jayagopi et al. [12]. McCowan et al. [22] developed a statistical framework based on Hidden Markov Models to detect actions that belong to the group as a whole, using multimodal features extracted from individuals' actions. For example, "discussion" is a group action which can be recognized from the verbal activity of individuals. Brdiczka et al. [4] proposed a fusion algorithm that detects subgroup activities in a meeting. Our approach takes a different perspective, aiming at improving team cohesion and individual relational skills. An example of work closer to ours in this respect is DiMicco et al. [6], which investigates the effects of providing the team members with feedback about their own speaking activity during a face-to-face meeting. Our approach, though similar in spirit, is different, because we address eye gaze behaviour in addition to speech activity to bear on the automatic analysis of relational behaviour.

3 Monitoring social dynamics

In this section, we focus on social dynamics and we are interested in investigating whether we can influence the social dynamics of a meeting by providing feedback to the meeting participants. In this context, we define social dynamics as the way verbal and nonverbal communicative signals of the participants in a meeting regulate the flow of a conversation (who has the floor). Analyses of conversations in meetings have shown that the flow of conversation is governed by two mechanisms [25]. First, the current speaker may select the next speaker; this may be done through a combination of verbal and nonverbal signals, e.g. by addressing a participant explicitly and/or by gaze behaviour and additional cues. Secondly, if the current speaker did not select the next speaker, the next speaker may select him/herself: if the current speaker has finished, one of the other participants may take the turn (possibly after a brief transition phase where several participants try to get the floor simultaneously).

From these observations, it follows that both verbal and nonverbal aspects of the behaviour of the participants influence the social dynamics of a meeting. Here, we explain three relevant determinants of the flow of conversation.

3.1 Plain speaking time

Since interrupting the speaker is bound to social conventions, the current speaker determines how long she/he will

speak that is to say, within certain limits. Speaking means having the opportunity to control the flow of conversation and influence the other participants.

3.2 Speaker eye gaze

The current speaker controls the flow of conversation by having the privilege of selecting the next speaker. This may be done through verbal means, such as when the speaker names another participant and asks for his opinion, but often it is done in a more subtle way, by nonverbal means such as eye gaze [2, 14, 15, 30]. In addition, when addressing all participants, the speaker should take care to look at all participants in due time in order to avoid giving the impression that she/he is neglecting particular participants.

3.3 Listener eye gaze

The participant who is speaking is being gazed at by the other participants, indicating that she/he is in the focus of attention [29, 33]. When the speaker is speaking for a long time, other participants may lose interest, which is signalled by gazing elsewhere.

Recently, researchers have taken inspiration from the observation that socially inappropriate behaviour may result in suboptimal group performance and they have developed systems that monitor and give feedback on social dynamics [3, 6, 20, 24]. The systems capture observable properties of the meeting participants, such as speaking time, posture and gestures, analyse the interaction of people and give feedback through offering visualizations of the social data. In Madan et al. [20], for instance, a wide range of vocal features, aspects of body language and physiological signals are measured to calculate a behaviour-based index of group interest, which is then shown to the participants on either a private or a public display. In DiMicco et al. [6], feedback is provided about the speaking time of different participants, visualized through a histogram presented on a public display. Evaluations showed that real-time feedback on speaking activity results in more equal participation of all meeting members.

These findings and observations lead us to believe that automatic feedback on audio-visual behaviour of meeting participants may help to improve the social dynamics of the meeting and increase the satisfaction of the group members with the discussion process. In the framework of the CHIL project (<http://chil.server.de> [34]), we designed a service that generates unobtrusive feedback to participants in a meeting about the social dynamics, presented in real time on the basis of captured audio-visual cues. Our goal is to make the members aware of their behaviour and in this way, influence the group's social dynamics.

We formulate the following hypotheses concerning the influence of feedback on social dynamics:

(H1) Speaking time will be distributed more equally in sessions with feedback than in sessions without feedback. Concretely, participants who under-participate without feedback will participate more when receiving feedback, and participants who over-participate without feedback will participate less when receiving feedback.

(H2) Speakers' visual attention will be distributed more equally among listeners when feedback is present than without feedback.

(H3) Visual attention from listeners for the speaker will be higher in sessions with feedback.

(H4) Participants' satisfaction about group communication and performance will be higher in the presence of feedback.

We focus on meetings with a protocol that invests participants with equal rights and responsibilities to contribute to the meeting, as for instance in a case where a committee needs to take a joint decision and every participant has information relevant to the decision or the members of a team need to reach agreement about a further course of action. In such collaborative meetings, everyone should be able to contribute to the meeting, regardless of the quality of the individual contributions and their impact on the final decision. This means that speakers who try to monopolize the discussion impede the progress of the meeting; the risk is that not all arguments relevant to the topic of discussion come to the surface, which may ultimately lead to a "groupthink" situation, when members of the group conform their opinion to what they believe to be the consensus of the group [13]. It has been shown that not sharing the available information has an adverse effect on the outcome of such meetings as it results in inferior decisions [19, 26, 28]. We should add, though, that there certainly are types of meetings where balancing the participation is less favourable, for example, instructive meetings or presentations, but these are beyond the scope of this paper.

In a previous study [18], the concept was evaluated through a Wizard of Oz approach, in which the behaviour of meeting participants (speaking activity and eye gaze/head orientation) was monitored in real time by human observers. While promising results were obtained, post-hoc analyses showed that the reliability of the monitoring task was below accepted standards, in particular for eye gaze/head orientation. It was therefore decided to build implementations of the required perceptual technologies and redo the evaluation.

In the remainder of the paper, we first describe the prototype of a service, which involves capturing and presenting information on speaking activity and eye gaze of speakers and listeners during the meeting. We then present a study evaluating the effects of the prototype on

participants' behaviour in meetings. We conclude with a discussion of our findings.

4 Prototype

4.1 Visualization

The prototype visualizes three types of information contributing to the social dynamics of the ongoing meeting (see Figs. 1, 2): (1) speaking time of each participant; (2) eye gaze of speaker; (3) eye gaze of listeners. The information is updated dynamically in real time. For a meeting with four participants, the visualization consists of four sets of three adjacent circles projected on the meeting table, as shown in Fig. 1. The individual sets of circles are projected in front of the individual participants, as shown in Fig. 2. At the beginning of the meeting, all circles are small. During the meeting, their size increases depending on the speaking activity and eye gaze behaviour of the participants, as follows.

1. Speaking activity: the size of the middle circle (coded S, for speaking activity) represents the participant's cumulative speaking time since the beginning of the meeting. For the current speaker, this circle is surrounded by a lighter-coloured ring, the size of which represents the duration of the ongoing turn.
2. The left-most circle (coded AS, for attention from speaker) indicates how much visual attention the participant—as a listener—has received since the beginning of the meeting from the other participants while they were speaking (added up across all other participants). The rationale for displaying the cumulative eye gaze each participant received as a listener is that eye gaze from the speaker acts as an inclusion and turn-taking cue: by gazing at a particular participant, the speaker draws that participant into the meeting



Fig. 1 Visualization of social dynamics during a meeting

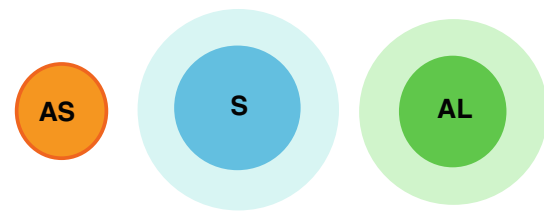


Fig. 2 Visualization of current and cumulative speaking activity and eye gaze. *S*, speaking activity. The size of the *inner circle* represents the cumulative speaking activity of participant since beginning of meeting. The size of the *outer ring* represents the duration of the current turn. *AS*, attention from speaker. For each participant, the size of the *circle* represents how much visual attention she/he received from the other participants when speaking, summed since the beginning of the meeting. *AL*, attention from listener: The *inner circle* represents the cumulative attention from the listeners since the beginning of meeting. The size of the *outer ring* represents the number of listeners currently looking at the speaker

- (provided the participant is also gazing at the speaker), and by gazing at a particular participant at the end of the turn, the speaker invites that participant to take over the turn. As a consequence, a relatively low AS score for a participant indicates that the other participants did not gaze at him/her a lot and did not invite him/her to take turns.
3. The right-hand circle (coded AL for Attention from Listeners) represents how much attention the participant has received since the beginning of the meeting from the other participants while she/he was speaking. For the current speaker, this circle is surrounded by a lighter-coloured ring, the size of which indicates how many participants are gazing at him/her currently. Thus, a small outer ring reveals to the current speaker that the other participants are losing interest, and a small inner circle reflects a lack of interest from the other participants for the participant while she/he was speaking in the previous part of the meeting. The rationale is that lack of interest may be a direct consequence of a participant's tendency to consume excessive speaking time.

The different circles are distinguished by different colours (the codes are not shown in the actual visualization). In order to facilitate users' understanding of the meaning of the different circles, a short mnemonic is displayed underneath each circle.

The information about speaking time and attention from listener and attention from speaker was shown in a rather abstract way in front of the participants on the meeting table for two reasons. In the first place, we wanted to enable participants to derive the relevant information at a quick glance, encouraging them to focus at major trends instead of giving attention to small changes and differences. In the second place, projection on the table made the

visualization appear in the periphery of the visual field of the participants. Peripheral displays present information in an unobtrusive way so that the information is present for inspection on demand but does not monopolize attention at inappropriate moments [1, 21].

4.2 Technology

The visualization is generated on the basis of combined audio (speech) and visual (focus of attention) cues, captured in real time during the meeting. In order to determine speaking time for individual participants, each participant is equipped with a close-talking microphone. The microphones are connected to a Terratec 8 channel audio-controller, which sends the microphone signals to a server that continuously detects if participants are speaking or silent. The server then determines voice onset and offset for each individual microphone signal, on the basis of which speaker diarization is performed.

Eye gaze both of speakers and listeners is estimated on the basis of head orientation. Head orientation may be considered a reliable indicator of gaze direction: average accuracy of eye gaze estimation was 88.7% in a meeting scenario with four participants [29]. To detect head orientation, participants are wearing headbands with two pieces of reflective tape. The two pieces of tape reflect IR light from IR emitters, which is registered by infrared sensitive cameras mounted to the ceiling of the meeting room. The two pieces of tape enable the cameras to pick up two separate coordinates for each headband, which are sent to a server. On the basis of these two coordinates, the server estimates the angle of the headband of each participant relative to its perpendicular axis (looking straight ahead) in a two-dimensional horizontal plane. This is the basis for determining the eye gaze direction of the participant, as shown in Fig. 3 for one participant. If the orientation of the headband is between lines A and B, eye gaze is towards participant II, if the orientation of the headband is to the left of line A, eye gaze is towards participant I, and if the orientation of the headband is to the right of line B, eye gaze is towards participant III. The angle of 35° between lines A and B was determined empirically during several pilot tests.

Combined audio and visual data are sent to a server that controls the visualization that is shown on the meeting table.

Although it is obvious that for real meetings the set-up with headbands and close-talking microphones is far too invasive, we considered the technological equipment suitable for evaluating our concept in a laboratory setting. Less invasive technology that is suitable for real meetings is under development, for example, speaker localization and diarization on the basis of input from microphone arrays

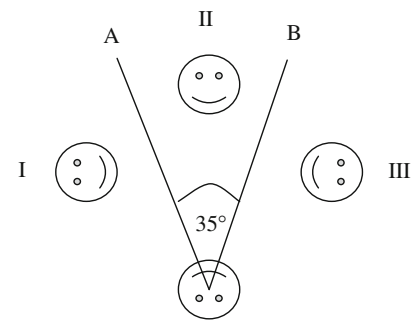


Fig. 3 Schematic diagram showing the relation between the measured orientation of the headband and the visual focus of attention of the participant

and camera-based head pose estimation (see among others the CHIL project [34]).

5 Evaluation

We conducted an empirical test to evaluate whether the service influenced the social dynamics of meetings. In this section, we describe the set-up of the test, the performance of the technology and the results of the test.

5.1 Methods

5.1.1 Participants

Eighty-two participants participated in the experiment, divided across 21 groups, nineteen groups consisting of four people and two groups of three. Twenty-four participants with various educational and social backgrounds were recruited from a database listing volunteers for experiments. The other 58 participants were students of the faculty of Industrial Design of the Eindhoven University of Technology. All participants were native speakers of Dutch and were paid a small fee for participation. In some of the groups, some of the members already knew each other. One group was an existing student team.

5.1.2 Design

The experiment applied a within-subjects (or rather “within-groups”) design. Each group participated in two discussion sessions. In each discussion, the members had to reach agreement on a particular topic. For each discussion, a discussion topic was provided. In one discussion session, participants were presented with feedback about speaking time and eye gaze in the form of the visualization shown in Fig. 2 (the Feedback condition); in the other condition, no feedback was provided (the No Feedback condition). To avoid order effects, the order of Feedback and No

Feedback conditions was balanced across groups. The same was done for the two discussion topics.

5.1.3 Experimental task

Two adjusted hidden-profile decision tasks were given to the participants. Hidden-profile tasks are discussion tasks where “the superiority of one decision alternative over others is masked because each member is aware of only one part of its supporting information, but the group, by pooling its information, can reveal to all the superior option” [27]. The specific tasks we used were adapted from the hidden-profile tasks used in DiMicco et al. [6]. One of the tasks was to select the best student out of three candidates to admit into a university programme; the other task was to choose the best location out of three for a new 24-h supermarket. These hidden-profile tasks comprise quite a large amount of information for each group member, and it is likely that the participants need to address their paperwork now and then in order to recall the facts required to come to a decision. In order to prevent participants from looking at their papers rather than at each other, we reduced the amount of information people had to memorize and took away all paperwork during the discussion. In our adapted hidden-profile tasks, all group members received the same facts but each group member had to defend a different position, representing a particular set of beliefs and values (a profile). For example, for the student selection task, one group member received the assignment of prioritizing financial bonuses associated with admission of particular students, whereas another member received the assignment of prioritizing intellectual ability as a criterion for admission. The goal for the group was to reach consensus about the optimal rank-ordering of candidates. The adaptations of the hidden-profile tasks gave satisfactory results in previous tests [18].

5.1.4 Procedure

At the start of the experiment, participants signed a consent form in which the rights of participants were given and which asked for permission for audio and video recording. After this, in both conditions, the group first had a 5-min warm-up discussion about a topic that they could select from a list provided by the experimenter. This short discussion served to familiarize the group members with each other and the environment and with the feedback. To keep both conditions as similar as possible, a warm-up discussion was included in the No Feedback condition as well.

After the warm-up discussion, the experimenter handed out the instruction for the first task. Participants then had 10 min to study their profile and the information about candidates individually, to make a preliminary choice and

to memorize their arguments. During the next 20 min, the participants discussed the three candidates and tried to reach agreement.

After each discussion, participants were asked to fill in a questionnaire. In the No Feedback condition, the questionnaire addressed group-related issues, whereas in the Feedback condition, the questionnaire addressed both group and service-related issues (see Sect. 5.2 for further details). After the participants completed the questionnaire, a group interview followed, addressing questions about how the discussion went; in the Feedback condition, the interview was extended with questions concerning the visualization.

In order to get an impression of the participants’ intention to use the system in future meetings, a fake third task was introduced and the participants were asked to indicate individually whether they would like to use the system for this third and final task. After this, participants were told that the third task did not exist and the experiment had finished.

5.1.5 Evaluation metrics

Measures for *speaking time*, *speaker’s attention* and *attention from listeners* were obtained from log files of the speech activity and head orientation trackers. For speaking time, each participant’s speaking time was expressed as the percentage of time that the participant had been speaking of the total speaking time for that session. In addition, we calculated to what extent the amount of speaking time for the participants was equally distributed, applying the Gini coefficient as a measure of equality, see (1) for the definition of the Gini coefficient for groups of four participants.

$$\text{Gini} = \frac{2}{3} \times \sum_i |\text{participation}_i - 25\%| \quad (1)$$

The Gini coefficient sums, over all the group members, the deviations of each person from equal participation (25% for a group of four), normalized by the maximum possible value of this deviation [7, 35]. Its values range from 0 for very high equality to 1 for low equality.

For speaker’s eye gaze, we calculated to what extent the speaker’s eye gaze is distributed equally over the three other participants (the listeners) during the whole meeting, using the Gini coefficient.

For attention from listeners, for each individual speaker, we calculated the average number of listeners gazing at the speaker throughout the meeting. The average number of listeners is expressed as a percentage of the maximum number of listeners.

Subjective judgments about participants’ attitudes towards the system and towards the group were collected by means of a questionnaire containing Likert-type scales and group interviews. The group satisfaction questionnaire (83 items) combined existing questionnaires about team

member satisfaction, task cohesion and perceived viability (capability of the group to continue working as a team in the future). The service-related questionnaire (28 items) combined existing questionnaires addressing issues of control, privacy, ease of use, usefulness, intrusiveness, enjoyment, trust, attitude and intention to use. The questionnaires were taken from Graziola et al. [11]. The group interviews addressed several specific topics in more detail, such as positive and negative aspects of the system, the influence of the visualization on the discussion and the perceived reliability of the information that is shown.

5.2 Results

5.2.1 Reliability analysis

To assess the reliability of the speech activity and head orientation trackers, the automatic loggings of the trackers were compared with manual annotations. Three meeting fragments of 2 min were randomly selected. For these fragments, an expert coder manually annotated speech activity as well as head orientation for each participant, using the ANVIL video annotation tool [16]. The resulting annotation is referred to as the reference annotation.

Speech diarization can be considered a segmentation task, i.e. detecting when a person speaks. Detection of head orientation on the other hand is a combination of a segmentation task and a classification task: besides deciding when a participant changes his/her gaze direction, it should be detected to whom the gaze of the participant is directed. For *segmentations* (identifying onset and offset of speech activity and changes in gaze direction), we used segmentation accuracy as a measure of reliability. Segmentation accuracy is defined as $100 - \text{SbER}$ (Segment boundary Error Rate), where SbER is the sum of segment boundary insertions, deletions and misplacements divided by the total number of segment boundaries in the reference annotation. To calculate the SbER, we set a level of tolerance, indicating the time window within which the segment boundaries can still be considered to match the boundaries in the reference annotation. The tolerance window was set to 1 s. For speech diarization, a segmentation accuracy was obtained of 57.3%. For head orientation, a segmentation accuracy was obtained of 40.7%. For *classifications* of gaze direction, we used Cohen's Kappa as a measure of reliability. Kappa measures pairwise agreement among a set of coders making category judgments, while correcting for chance agreement [5]. It ranges from 0 to 1 with large values indicating better reliability. Kappa was calculated using those segments for which there was agreement on both segment boundaries (37.5% of the segment boundaries for head orientation). The Kappa for classification of head orientation was 0.81, meaning that head orientation

was reliably coded in the segments which were correctly segmented.

We consider the figures acceptable, although we admit that increasing the accuracy of the perceptual components is advisable. Further analysis of the automatic and manual segmentations showed that the total distribution of speaking time for individual speakers was quite well preserved. That is, while the size of the outer ring for component S in Fig. 2, which reflects the start and end of a turn, may not be fully accurate, the size of the inner ring appears to give a satisfactory representation of the cumulative speaking time for individual speakers. The same applies with respect to head orientation: while the display may not always accurately reflect the shift and direction of the current visual attention, the distribution of visual attention across meeting participants and the proportional durations appear to be satisfactory. For that reason, we consider it justified to use the automatically obtained data for further analysis.

5.2.2 Social dynamics

Speaking time: due to technical problems in some sessions, we had complete speech and head orientation data from only 15 groups (13 groups of 4, 2 groups of 3). The total speaking time of individual participants was relatively well correlated between the No Feedback and the Feedback condition (Pearson correlation coefficient $r = 0.44$, $N = 58$, $p = 0.001$), meaning that participants who speak relatively little in the No Feedback condition also speak relatively little in the Feedback condition and participants who speak relatively much in the No Feedback condition also speak relatively much in the Feedback condition. Speaking time was divided fairly equally over the participants in both conditions: we found Gini coefficients of 0.14 in the No Feedback condition and 0.11 in the Feedback condition. The difference in equality between the two conditions was not statistically significant ($t(14) = 0.942$, $p = 0.362$).

In order to test the hypothesis that participants who speak less than average (the under-participants) or more than average (the over-participants) will adapt their behaviour as a result of the feedback, we categorized the participants into three categories. This was done only for the groups with four participants. The participants whose total speaking time was more than one standard deviation below average in the No Feedback condition were categorized as under-participants (seven speakers, 13.5%), those whose speaking time was more than one standard deviation above average were categorized as over-participants (seven speakers, 13.5%); the rest was categorized as middle participants (38 speakers, 73%). Table 1 shows the average percentages speaking time in both conditions for participants in each of the three categories.

Table 1 Average percentages speaking time in No Feedback and Feedback conditions for all participants and separately for under-participants, middle participants and over-participants

| | All (groups of 4) | Under-participants | Middle participants | Over-participants |
|-------------|-------------------|--------------------|---------------------|-------------------|
| <i>N</i> | 52 | 7 | 38 | 7 |
| No Feedback | 25 | 12.4 | 25.3 | 36.4 |
| Feedback | 25 | 22.7 | 24.6 | 29.2 |

Table 2 Number of participants in different percentage bins for No Feedback and Feedback conditions (*N* = 52)

| Percentage participation | <12.5 | 12.5–17.5 | 17.5–22.5 | 22.5–27.5 | 27.5–32.5 | 32.5–37.5 | >37.5 |
|--------------------------|-------|-----------|-----------|-----------|-----------|-----------|-------|
| No Feedback | 2 | 5 | 10 | 16 | 11 | 5 | 3 |
| Feedback | 0 | 4 | 12 | 21 | 9 | 6 | 0 |

T-tests were performed on the speaking time data of each category. We found that the under-participants significantly increased their speaking time in the Feedback condition when compared to the No Feedback condition ($t(6) = -3.3$, $p = 0.02$). The decrease in speaking time for over-participants in the Feedback condition compared to the No Feedback condition was nearly significant ($t(6) = 2.32$, $p = 0.06$). No significant difference between the two conditions was found for the middle participants ($t(37) = 0.75$, $p = 0.46$). The results thus indicate that participants who are at the extremes of the speaking time range tend to change their behaviour so as to become less extreme. One might argue that this finding could be explained simply in terms of a regression towards the mean, the phenomenon that measures which have extreme values at one point in time are likely to be less extreme when measured on a different occasion, for statistical reasons. However, closer inspection of the results renders this explanation unlikely. Table 2 gives the distribution of participants over different percentage bins in the No Feedback and Feedback conditions.

Under an explanation in terms of a regression to the mean, the shape of the distribution should remain approximately the same. As can be seen, the overall distribution becomes narrower in the Feedback condition, with participants being centred more closely around the mean. Furthermore, related research has also indicated that people tend to change their behaviour on the basis of visual feedback, while an explanation in terms of a regression to the mean was ruled out [7]. Therefore, we consider it safe to assume that regression to the mean is not a conclusive explanation for our findings.

Attention from speaker: the distribution of the speaker's attention over listening participants throughout the meeting was rather unequal in both conditions (Gini coefficients are 0.54 in the No Feedback condition and 0.55 in the Feedback condition). The difference between the Gini coefficients in the two conditions is not statistically significant ($t(57) = -0.686$, $p = 0.495$), indicating that feedback about the way

Table 3 Average scores for perceived viability, task cohesion and team member satisfaction (7-point scale, 1 is low appreciation, 7 is high appreciation) in No Feedback and Feedback conditions

| | Perceived viability | Task cohesion | Team member satisfaction |
|-------------|---------------------|---------------|--------------------------|
| No feedback | 5.1 | 5.1 | 5.6 |
| Feedback | 5.0 | 5.0 | 5.6 |

speakers divided their attention across listeners did not lead them to divide their attention more equally. Closer inspection of the data showed that for most speakers (73% in the No Feedback condition and 83% in the Feedback condition), the participant seated opposite was the main visual focus of attention. This may be due to using head orientation instead of eye gaze to estimate visual attention or to the arrangement of participants around the table.

Attention from listeners: the average attention level (i.e. the average % of listeners looking at the speaker) was 41% in the No Feedback condition and 42% in the Feedback condition. The difference in attention level between the two conditions is not statistically significant ($t(57) = -1.25$, $p = 0.22$), indicating that listeners did not pay more visual attention to the speaker as a result of the feedback.

5.2.3 Questionnaire and interview results

The questionnaire and interview data are based on all 82 participants. The questionnaire data concerning group satisfaction showed only minor differences between the Feedback and the No Feedback condition. Participants' attitudes towards the system were moderately positive: the average scores on different subscales were between 4 and 5 on a 7-point scale. Lower scores were obtained for usefulness (average 3.5) and control (average 3.9). The questionnaire results are summarized in Table 3 (group-related dimensions) and Table 4 (service-related dimensions).

Table 4 Average scores for service-related dimensions (7-point scale, 1 is low appreciation, 7 is high appreciation), Feedback condition only

| Trust | Usefulness | Privacy | Intrusiveness | Ease of use | Enjoyment | Control | Attitude | Intention to use |
|-------|------------|---------|---------------|-------------|-----------|---------|----------|------------------|
| 4.3 | 3.5 | 4.5 | 4.5 | 4.5 | 4.4 | 3.9 | 4.4 | 4.0 |

As can be seen, the presence of feedback did not influence perceived viability, task cohesion and team member satisfaction. With respect to the appreciation of the service, the average scores are around the mid point of the scale (neutral), tending slightly towards the positive end of the scale for trust, privacy, intrusiveness, ease of use, enjoyment and attitude towards the service and slightly towards the negative end of the scale for usefulness. The neutral score for Intention to Use and the slightly negative score for Usefulness are corroborated by the answers to the question whether the participants would like to use the system in future meetings. After the second task, participants were asked for their preference for using the system or not for the third and final task (which actually did not exist). Fifty-one per cent of the participants indicated that they wanted to use the system for the third task, for various reasons, such as ‘the system shows interesting information about my behaviour’ and ‘it is fun to use the system’. Thirty-one per cent of the participants preferred not to use the system again. The most prominent reason for not wanting to use the system again was the distraction from the meeting task that it causes.

During the interviews, several participants indicated that the meaning of the circles was not immediately clear to them (which is in line with the relatively low questionnaire scores for usefulness). For most participants, the speaking time circle was the most intuitive one and therefore this information was most used. Several participants mentioned that the circles enabled them to better divide their attention, while other participants found that the circles introduced some kind of competition. Some participants indicated that measuring head orientation or eye gaze was not the most reliable way to measure attention, because it captures only visual attention and they may pay attention to the speaker even when they are not looking at him or her. Most people, however, found that the circles adequately reflected speaking activity and focus of attention during the meeting.

5.2.4 Analysis of micro-patterns

Having shown that providing feedback on speaking behaviour and eye gaze affects participants’ social behaviour in meetings, a legitimate question is whether the effect arises from feedback on speaking behaviour or eye gaze or both. As the effectiveness of feedback on speaking behaviour was already shown in DiMicco et al. [6], we focused in particular on the effectiveness of feedback on

Table 5 Average number of invitations by speaker to under-participants and over-participants in No Feedback and Feedback conditions

| | Under-participants (N = 6) | Over-participants (N = 6) |
|-------------|-------------------------------|------------------------------|
| No feedback | 17.2 | 31.0 |
| Feedback | 20.7 | 28.3 |

eye gaze behaviour. As explained earlier, one type of feedback indicated to what extent participants were gazed at by the current speaker. The rationale behind providing this type of feedback was that turn-taking conventions decree that the speaker may invite a participant to take the turn by looking at him/her at the end of the utterance. A potential cause of unequal participation in meetings is therefore that speakers give a preferential treatment to some participants by inviting them through eye gaze to take the turn while neglecting other participants. Feedback on gaze behaviour might make the speakers aware of this asymmetry and lead them to adjust their gazing behaviour such that they also invite the other participants. In order to evaluate this reasoning, the number of invitations was calculated in No Feedback (NFB) and Feedback (FB) conditions for under-participants (UP) and over-participants (OP). Invitations were defined as turns where the speaker gazed at the participant under consideration at the end of the turn. Data of UP and OP were analysed only for those participants who showed a change in participation rate (measured by speaking rate) larger than 5% from NFB to FB (more participation for UP and less participation for OP). For both UP and OP, six participants were selected on the basis of this criterion. Table 5 shows the average number of invitations to UP and OP in FB and NFB conditions.

Although the results are in the predicted direction, an analysis of variance with FB/NFB as a within-subjects variable and UP/OP as a between-subjects variable shows that only the difference between UP and OP is significant ($F_{1,10} = 5.13, p = 0.047$) but the difference between FB versus NFB is not significant ($F_{1,10} = 0.01$) and the interaction is also not significant ($F_{1,10} = 0.52$). In other words, over-participants receive more invitations, both in FB and NFB. From these results, we conclude that feedback on eye gaze does not affect the gazing behaviour of the speakers.

Table 6 Average percentages of accepted invitations by under-participants and over-participants in No Feedback and Feedback conditions

| | Under-participants (<i>N</i> = 6) | Over-participants (<i>N</i> = 6) |
|-----------------|---------------------------------------|--------------------------------------|
| No feedback (%) | 32 | 52 |
| Feedback (%) | 51 | 51 |

Secondly, we investigated whether there was a difference in the percentage of accepted invitations. To this end, we calculated the number of cases where an invitation indeed resulted in a response from the invited participant, relative to the overall number of invitations to the participant. The results are shown in Table 6.

An analysis of variance with FB/NFB as a within-subjects variable and UP/OP as a between-subjects variable shows that the difference between UP and OP is not significant ($F_{1,10} = 1.62$), but the difference between FB versus NFB is marginally significant ($F_{1,10} = 4.79$, $p = 0.053$) and the interaction is significant ($F_{1,10} = 5.93$, $p = 0.35$). The marginal significance of the main effect of Feedback must be attributed to the low score for under-participants in the No Feedback Condition. In other words, the results indicate that under-participants tend to accept more invitations in the Feedback condition than in the No Feedback condition. A plausible interpretation is that the tendency of under-participants to respond more often to invitations in the Feedback conditions is caused by their awareness that their speaking activity is lagging behind.

6 Conclusion and discussion

We have presented a prototype service providing real-time feedback on the social dynamics of meetings to participants in small collaborative group meetings. The prototype captures and visualizes speaking time and gaze behaviour. The feedback is displayed to meeting participants during the meeting through a dynamic peripheral visual display. Analyses of the reliability of the speech and head orientation trackers showed that the prototype is able to detect speech activity and eye gaze direction (as estimated from head orientation) at a satisfactory level of reliability for experimentation purposes.

The system was evaluated in a within-subjects evaluation with 58 participants. The results of the evaluation showed that the visualization of speaking behaviour and eye gaze direction had the desired effect. A significant effect of the visualization was found for under- and over-participants who, as a result of the feedback, changed their speaking behaviour to become less extreme. With respect to the perceived value of the service, it was found that the

visualization did not influence the participants' satisfaction with their team. Questionnaire and interview data showed that participants were slightly positive about the system, although several participants had concerns about the fact that the system distracted them from the discussion. About half of the participants indicated that they would like to use the system on a next occasion.

In sum, the results indicate that feedback on eye gaze and speaking behaviour may lead meeting participants to change their behaviour and thus may influence the social dynamics of meetings. Analysis of the micro-patterns of six under-participants and six over-participants indicated that the change in behaviour could not be attributed to the visualization of eye gaze direction so that we infer that the primary effect of the visualization is to make participants aware of their speaking activity.

One possible explanation for the finding that feedback about speaking behaviour is effective but feedback about gaze behaviour is not, is that this outcome is due to the concrete properties of the visualizations. As described in the Methods section, participants obtained an explanation by the experiment leader and then had 5 min to get to know the system using it in a warm-up discussion. Although this was enough for participants to understand the concept of the visualization, it may have been insufficient to really understand the meaning and the impact of the information shown. Indeed, some participants mentioned that they would need more extensive training with the system in order to fully grasp the intention of the circles and be able to use the information during the discussion before the system can be really useful. Also, several participants indicated that they did not really use the visualization, because thinking about what to do with the information would distract them too much from the actual discussion going on. In particular, some participants commented that they found the circles representing visual attention ('attention from listeners' and 'attention from speaker') difficult to interpret. It may take more than one meeting for participants to understand the meaning of the circles at a glance and use it effectively (i.e. to change one's behaviour) without being distracted too much.

Alternatively, the outcome that feedback about speaking behaviour is effective but feedback about gaze behaviour is not, may be related to the controllability of the behaviour. Although both speaking activity and visual attention may be consciously controlled, intuitively it appears much easier to control speaking activity than visual attention. Noticing that one has been speaking already for a long time, one may simply decide to stop speaking and hand over the turn (although personality may play a role as well). Similarly, participants who are little active may decide to become more active when becoming aware of their relative lack of participation. It seems plausible that

meeting participants feel inclined to do so especially when the evidence of their under- or over-participation is so clearly shown on the table. On the other hand, eye gaze in group meetings intuitively appears to be less under conscious control and to be ruled more by significant events in the environment and by engrained habits. In meetings, speaking is probably the most significant event, therewith pulling the listener's visual attention, and only when the listener gets bored, she/he will look away. With respect to engrained habits, it is common experience that speakers have to give effort learning to give their attention to all members of the audience instead of looking at papers or at a single member of the audience. These considerations lead us to believe that speaking time can be more easily changed on the basis of feedback than eye gaze. Finally, the results may not so much be related to specific effects of feedback on speaking behaviour but the feedback may have rather raised the general awareness of the social dynamics and potential unbalances in the social behaviour, most notably the speaking behaviour. This needs to be determined in future research.

6.1 Limitations

In interpreting the outcomes of the current study, the decisions that we made in setting up the study should be taken into consideration, as they impose limitations on the generality of the outcomes. Here, we discuss three main characteristics: the nature of the groups, the nature of the meetings and the nature of the visualization.

(1) It may be noted that the majority of the groups taking part in our experiment consisted of people that did not know each other or work together, so they did not have a history together and they would not be in any meeting together afterwards either. In such a situation, people are often rather polite, friendly and lenient; indeed, almost all participants indicated that they found the other group members kind and not irritating. This may have influenced the results since in such a situation, the social dynamics of the meeting usually are satisfactory and there may be less need for feedback. Moreover, if there would be problems concerning the social dynamics, for example if one person would take the lead and disregard some of the other participants, people might not feel the urge to change the situation, because it was just this one time that they had to deal with it. The situation may be rather different when people have to meet with the same group every week. Therefore, providing feedback about social dynamics may be more useful for groups that have just started and will continue to work together for some time, or for existing groups experiencing problems.

Another aspect concerning the nature of the groups is the hierarchical structure. In our study, equal participation

was considered an optimal strategy. In real life, in particular in working situations, many meetings involve a clear hierarchical structure with a chairman and/or one or more experts. In such situations, feedback on eye gaze might still be valuable, but feedback on information and its visualization would have to be reconsidered.

(2) In order to be able to give feedback on eye gaze, we set up the study such that participants would not have to consult papers or look at a joint display. Of course, this is different from many real-life situations. In such situations, the regulatory function of eye gaze operates differently (see e.g., [31]). In particular, the turn-giving role of eye gaze is concentrated near the end of the utterance. Obviously, this imposes additional challenges for the technology.

(3) As outlined previously, the visualizations might not have conveyed their intended meaning easily. Different visualizations, especially of the eye gaze information, might have given different results. Alternatively, a more longitudinal approach might have enabled participants to become familiar with the visualizations and to make more effective use of the information to adjust their behaviour.

6.2 Future work

Future work will need to validate and enrich the conclusions in several ways. In the first place, it needs to be investigated whether feedback on speaking behaviour is indeed much more effective than feedback on eye gaze behaviour; also, as was mentioned elsewhere, it needs to be determined to what extent it's a specific effect of feedback or a more general effect of creating awareness of certain aspects of the situation which slip our minds under nonaugmented circumstances. A second question that we would like to answer is how such systems might be deployed in meetings. In the current set-up, the information was available all the time and it was up to the participants to use the information as they felt meaningful, either by changing their behaviour or bringing it up for discussion. Instead, the system might note deviations from the optimal pattern and intervene to bring the social dynamics up for discussion during or after the meeting. For example, Pianesi et al. [23] provide information about the social and functional roles of the participant in the form of an individual report after the meeting. Other researchers have evaluated a system with which information about participation level could be reviewed offline in between two tasks [7]. Also, the system might be used as a support system for human facilitators or coaches, providing them with objective data concerning the social dynamics of the meeting as a basis for intervention. A third question for future research concerns the persistence of the effects. We would like people to learn to modify and control their behaviour, but it remains to be established to what extent such acquired behaviour will persist.

Acknowledgments The work reported in this paper is part of the FP6 CHIL project (FP6-506909), partially funded by the European Commission under the Information Society Technologies Programme (IST). We gratefully acknowledge J. DiMicco for making available the hidden profile tasks, A. Eyck, K. Geerlings, O. van Herwijnen, O. Kulyk and J. Wang for their assistance in setting up and conducting the experiment and the data analysis, and Fabio Pianesi, Massimo Zancanaro and Ilenia Graziola for fruitful discussions and their contribution in preparing the questionnaire. Also, we thank the anonymous reviewers for suggestions for improvement.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Abowd GD, Mynath ED, Rodden T (2002) The human experience. *IEEE Pervasive Comput* 1:48–57
2. Bakx I, van Turnhout K, Terken J (2003) Facial orientation during multi party interaction with information kiosks. In: Proceedings of interact
3. Basu S, Choudhury T, Clarkson B, Pentland A (2001) Towards measuring human interactions in conversational settings. In: Proceedings of the IEEE international workshop on cues in communication CUES
4. Brdiczka O, Maisonnasse J, Reignier P (2005) Automatic detection of interaction groups. In: Proceedings of the 7th international conference on multimodal interface, Trento, Italy, October 2005, pp 32–36
5. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Measur* 20:37–46
6. DiMicco J, Pandolfo A, Bender W (2004) Influencing group participation with a shared display. In: Proceedings of the 2004 ACM conference on computer supported cooperative work CSCW, pp 614–623
7. DiMicco J, Hollenbach K, Pandolfo A, Bender W (2007) The impact of increased awareness while face-to-face. *Human-Computer Interaction, Special Issue on Awareness Systems Design*, vol 22, pp 47–96
8. Dourish P, Bly S (1992) Portholes: supporting awareness in a distributed work group. In: Proceedings of the ACM conference on human factors in computer systems CHI'92
9. Doyle M, Straus D (1993) How to make meetings work. The Berkley Publishing Group, New York
10. Erickson T, Halverson C, Kellogg WA, Laff M, Wolf T (2002) Social translucence: designing social infrastructures that make collective activity visible. In: Preece J (ed) *Communications of the ACM—special issue on community*, vol 45, no 4, pp 40–44
11. Graziola E, Eyck A, Sturm J (2007) Modelling intention to use services supporting human-human communication and collaboration. Internal Report CHIL project. <http://chil.server.de/?12532>
12. Jayagopi DB, Hung H, Yeo C, Gatica-Perez D (2008) Predicting the dominant clique in meetings through fusion of nonverbal cues. In: Proceeding of the 16th ACM international conference on multimedia, ACM, New York, pp 809–812
13. Janis I (1972) Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes. Houghton Mifflin, Boston
14. Kalma AP (1992) Gazing in triads: a powerful signal in floor apportionment. *Br J Soc Psychol* 31:21–39
15. Kendon A (1967) Some functions of gaze direction in social interaction. *Acta Psychol* 26:22–63
16. Kipp M (2001) Anvil—a generic annotation tool for multimodal dialogue. In: Proceedings of the 7th European conference on speech communication and technology eurospeech, pp 1367–1370
17. Kraemer K, King J (1988) Computer-based systems for cooperative work and group decision making. *ACM Comput Surv* 20:115–146
18. Kulyk O, Wang C, Terken J (2006) Real-time feedback based on nonverbal behaviour to enhance social dynamics in small group meetings. In: Joint workshop on multimodal interaction and related machine learning algorithms (MLMI) (Lecture Notes in Computer Science), vol 3869, pp 150–161
19. Lightle JP, Kagel JH, Arkes HR (2009) Information exchange in group decision making: the hidden profile problem reconsidered. *Manage Sci* 55:568–581
20. Madan A, Caneel R, Pentland A (2004) GroupMedia: distributed multimodal interfaces. In: Proceedings of sixth international conference on multimodal interfaces ICMI04, pp 309–316
21. Matthews T, Rattenbury T, Carter S (2007) Defining, designing, and evaluating peripheral displays: an analysis using Activity Theory. *Human-Comput Interact* 22:221–261
22. McCowan I, Gatica-Perez D, Bengio S, Moore D, Boulard H (2004) Towards computer understanding of human interactions. In: Aarts E, Collier R, van Loenen E, de Ruyter B (eds) *Ambient intelligence, lecture notes in computer science*. Springer, Heidelberg, pp 235–251
23. Pianesi F, Zancanaro M, Falcon V, Not E (2006) Toward supporting group dynamics. In: Proceedings of artificial intelligence applications and innovations, pp 302–311
24. Rienks RJ, Zhang D, Gatica-Perez D, Post W (2006) Detection and application of influence rankings in small group meetings. In: Proceedings of the eighth international conference on multimodal interfaces ICMI'06, pp 257–264
25. Sacks H, Schegloff E, Jefferson G (1974) A simplest systematics for the organisation of turn-taking for conversation. *Lang J Linguist Soc Am* 50:696–735
26. Schulz-Hardt S, Frey D, Luthgens C, Moscovici S (2000) Biased information search in group decision making. *J Pers Soc Psychol* 78:655–669
27. Stasser G, Stewart D (1992) Discovery of hidden profiles by decision making groups: solving a problem versus making a judgment. *J Pers Soc Psychol* 63:426–434
28. Stasser G, Titus W (2003) Hidden profiles: a brief history. *Psychol Inq* 14:304–313
29. Stiefelhagen R, Zhu J (2002) Head orientation and gaze direction in meetings. In: Proceedings of human factors in computing systems CHI2002, pp 858–859
30. Takemae Y, Otsuka K, Mukawa N (2003) Video cut editing rule based on participants' gaze in multiparty conversation. In: Proceedings of the 11th ACM international conference on multimedia, pp 303–306
31. Terken J, Joris I, de Valk L (2007) Multimodal cues for addressee-hood in triadic communication with a human information retrieval agent. In: ACM international conference on multimodal interaction (ICMI), Nagoya, Japan, 12–14 November 2007
32. Tang JC (1991) Findings from observational studies of collaborative work. *Int J Man-Mach Stud* 34:143–160
33. Vertegaal R, Slagter R, van der Veer G, Nijholt A (2001) Eye gaze patterns in conversations: there is more to conversational agents than meets the eye. In: Proceedings of human factors in computing systems CHI2001, pp 301–308
34. Waibel A, Stiefelhagen R (2009) Computers in the human interaction loop. Springer, Heidelberg
35. Weisband SP, Schneider SK, Connolly T (1995) Computer-mediated communication and social information: status salience and status awareness. *Acad Manag J* 38:1124–1151