Pers Ubiquit Comput (2008) 12:181–195 DOI 10.1007/s00779-007-0144-5

Authors'post-print

Multimodal support to group dynamics

Fabio Pianesi · Massimo Zancanaro · Elena Not · Chiara Leonardi · Vera Falcon · Bruno Lepri

Abstract The complexity of group dynamics occurring in small group interactions often hinders the performance of teams. The availability of rich multimodal information about what is going on during the meeting makes it possible to explore the possibility of providing support to dysfunctional teams from facilitation to training sessions addressing both the individuals and the group as a whole. A necessary step in this direction is that of capturing and understanding group dynamics. In this paper, we discuss a particular scenario, in which meeting participants receive multimedia feedback on their relational behaviour, as a first step towards increasing self-awareness. We describe the background and the motivation for a coding scheme for annotating meeting recordings partially inspired by the Bales' Interaction Process Analysis. This coding scheme was aimed at identifying suitable observable behavioural sequences. The study is complemented with an experimental investigation on the acceptability of such a service.

1 Introduction

Meetings are more and more important in structuring daily work in organizations. Executives on average spend 40–50% of their working hours in meetings [1]. However, the success of a meeting is often hindered by the participants' behaviour: professionals agree that as much as 50% of meeting time is unproductive and that up to 25% of meeting time is spent discussing irrelevant issues [1]. In order to improve performance of meetings, 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 maintaining a fair and focused behaviour as well as directing and setting the pace of the discussion. Training experiences aim at increasing the relational skills of individual participants by providing an offline (with respect to meetings) guidance—or coaching—so that the team eventually will be able to overcome or to cope with its disfunctionalities.

In discussing the role of collaboration for teachers and in particular peer coaching, Andersen [2] suggests that coaching sessions provide a scheduled opportunity to think reflectively, and that the coaching process allows the externalisation of both though contents and processes that are normally internal, making them available to examination. By bringing a different perspective to the relationship, the coach can see circumstances and possibilities that the coachee cannot [3]. There are three stages in the reflective process [4]: (i) the return to experience (what happened?); (ii) attending to feelings (how did I feel, why did I (re)act this way?); and (iii) the re-evaluation of the experience (what does it mean?). In the present work, we mainly focused on the first stage.

In this paper, we present a multimodal system, called 'the relational report (RR)' that monitors groups, and generates individual reports about the participants' behaviour. The system observes the meeting as a coach would do, and not as a recorder. This means that the system does not keep verbatim trace of what people said and/or what they did during the meeting. The generated reports are not minutes; they do not address content, but present a more qualitative, meta-level interpretation of what happened in the social dynamics of the group. They do not contain information like ''in the first part of the meeting you have talked for ten minutes about machine learning techniques useful to solve the problem'' but rather ''in the first part of the meeting you have provided the group with background information'' or ''you have prevented others from intervening in the discussion''. The reports are delivered privately to each participant after the meeting, and their purpose is that of informing them about their behaviour rather than evaluating it. Hence, the system acts as a coach for the individual group participants.

In presenting this work, we aim at contributing to the definition of a new class of systems based on intelligent monitoring of behaviour as opposed to classical multimodal ones where perceptual components are exploited as smart input devices. These kind of systems are very challenging and, although multimodal components are becoming more robust, full implementations that can operate in a non-controlled environments are still not available. Employing a user-centred approach to their design is complicated by the fact that the users cannot actually experiment with robust versions thereof, or can do so only to a limited extent. Another big challenge posed by these systems is that they are likely to have a strong impact on acceptability since they are often based on some form or another of monitoring.

In the next two sections, we will discuss initial steps we took towards the relational report: focus groups conducted with potential users, and a Wizard-of-Oz experiment to assess the acceptability of an automatic coaching service. Then, we will introduce a coding scheme that, drawing on social psychology literature, describes the interactions within a group in terms of the roles played by the participants. This coding scheme was used to annotate a multimodal corpus of meetings that, in turn, was exploited to train a system to automatically detect those roles. Finally, we will discuss the system component that generates the reports from the automatically extracted roles time series.

2 Previous and related works

In the field of CSCW the focus is often on distributed meetings, and the social relationships among meeting participants have been recognized as a fundamental aspect of the meetings' efficacy since the seminal work of Tang [5]. Many different attempts have been made to bring the social dynamics at a ''visible'' level. For example, Dourish and Bly [6] 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 a more effective communication, and improved interactions, and can contribute to a shared sense of community. Another example in this respect is the work of Erikson and colleagues [7], which proposed the idea of ''social translucence'', that is graphical widgets that signal cues that are socially salient. The claim is that such a functionality—by supporting mutual awareness and accountability—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 communication; therefore awareness and visibility of the context are not problems; the impact of participants' perception of their own activity on the others could play an important role, though. An example of a work closer to ours in this respect is Di Micco and colleagues' [8], which investigates the effects of providing the team members with a feedback about their own speaking activity during a face-to-face meeting. Our approach, though similar in spirit, is different, especially because we address a larger set of basic information (beyond speech activity) to bear on the automatic understanding of relational behaviour.

Another work closer to our approach is Maloney-Krichmar and Preece [9] research on the dynamics of an online group community. They used a coding scheme similar to the one we will discuss later one, inspired to the same source as ours, and investigated inter-rater agreement by considering agreement rate (proportions). The schema was basically meant to serve for analytical and theory-building purposes while the one we will propose was devised to serve the automatic annotation of meetings, so that functionalities such as the relational report can be built.

Our work has deep roots in the field of multimodality. Most of the current research in this area is aimed at providing easy access to computerized services for the group to efficiently accomplish its tasks. For example, in the CHIL project, most of the services provided are aimed at offering better ways of connecting people (the Connector service) and supporting human memory (the Memory Jog) [10]. The research in the AMI project mostly focuses on off-line multimedia retrieval and multimedia browsing of

information obtained from meetings [11]. The DARPA-funded project CALO supports a group in creating a project schedule by automatically interpreting gestures and speech, including the learning of new words [12]. Our approach takes a different perspective aiming at improving team cohesion, and individual relational skills.

Recently, there has been some interest in the automatic analysis of group interaction. For example, McCowan and colleagues [11] 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 and colleagues [13] proposed a fusion algorithm that detects subgroup activities in a meeting.

3 Initial study: focus groups

Three focus groups were conducted in order to provide for a broad view of the attitudes of potential addressees of the RR towards the service. Two of the three focus groups were composed by (five) researchers and technicians from ITC-irst (the groups of experts). The third group consisted of (four) people from ITC-irst administrative staff (the group of non-experts). The distinction was motivated by the attempt to investigate whether and how different professional profiles—having different relationship with technology (technology developers/researchers vs. technology users)—corresponded to different attitudes towards the service. The facilitator was the same in all the focus groups. Each focus group was structured in three phases. During the first, the facilitator introduced the general topic (the relational report) and the rules of the discussion. The relational report was explained by first presenting two videos drawn from one of our recorded meetings, followed by a (mock up) multimodal relational report addressing one of the participants in that meeting, and constructed according to the principles briefly introduced above and to be better discussed below. It consisted of a videoclip in which a talking head described the behaviour of the relevant participant, providing suggestions as to how to improve his/her relational skills.

The second phase was devoted to discussing four specific issues, one at a time. The facilitator introduced each issue by asking a question that the focus group would then discuss. The allowed time was approx. 10 min per issue. During the third phase the facilitator presented a short summary of the discussion for the group to briefly discuss.

The issues investigated were the *perceived usefulness* (''what do you think about the usefulness of a report such as the one you have seen?—would you prefer a descriptive or a normative report?), the *reliability* of the report (''what do you think about the reliability of the report?'''), its *intrusiveness* (''What are your opinions about the possible intrusiveness of the report and of the equipment it needs?''), and its *acceptability* (''what do you think about the acceptability of the report?—does it change according to whether the feedback is positive or negative?''). The facilitator never intervened during the discussion, except when needed to keep the discussion to its topic, or to explain the questions. All the focus groups were video recorded. The facilitator used these recordings to compile a summary after the end of each focus group.

Concerning the *perceived usefulness* of the report, the consensus in all the focus groups was that it could be useful, though the utility was seen as dependent on the disposition of each addressee to consider criticisms. Two 'expert' participants rejected the usefulness of the feedback, on the grounds that people are already aware of their own behaviour, and that behaviour is not an important aspect for meeting success, respectively. In both cases, this negative attitude was at least partially determined by the lack of trust towards the computer reliability. One participant in the non-expert focus group, on the other hand, considered the possibility of showing the report to a supervisor, in order to obtain a kind of formative counsel.

As to the *reliability*, it was widely agreed by all groups that the report was reliable. Almost everyone, however, pointed out the need for more audio–video evidence for the statements of the RR. In details, the expert participants suggested adding both quantitative (e.g. statistics on turn taking, time spent on talking, overlapping speech, etc.) and qualitative information, also suggesting that the report should take into account the official (organizational) role of the addressee. The non-expert focus group emphasized that the report must be ''an objective synthesis of the behaviour exhibited during the meeting'', while fearing that this goal can be hindered by the incapability of the system to contextualise people's behaviour, this way leading to inaccurate or wrong reports. Here contextualisation does not refer to the immediate context of the interaction, but to the long-term relational history of the group and of the individuals composing it, including their official roles and positions within the organization.

Regarding *intrusiveness*, being video recorded was acknowledged to be more intrusive and annoying than the individual delivery of the behavioural feedback. There are differences between the expert and non-expert focus groups, though. The latter could not explain precisely why they were annoyed by the video recordings, maintaining that the embarrassment is automatic and not controllable (''it is the very idea of an ''eye watching you'' that is annoying; this is intrusive by definition''). The experts, on the other hand, explicitly linked their attitudes to privacy problems, fearing that the video recording could be used in

unfair ways. Interestingly, these negative feelings seemed to be triggered more by the visual part of the recordings than by the audio one.

The *acceptability* of the report turned out to depend mainly on the trust in the system, and on the subjective disposition and motivations to accept external feedbacks. Thus, the expert groups agreed that the option of choosing whether to receive or not the report improves acceptability (if you ask for the report, you trust in it). As with usefulness, the acceptability of the report is expected to depend on the quantity and quality of the evidence (quantitative or descriptive) the report comes with; as many put it, this information allows addressees to control the factual basis of the report. The motivations and the cost–benefit balance turned out to be important aspects determining acceptability (and usefulness). Thus many mentioned, again, the greater acceptability (and utility) the relational report can have for people involved in a formative path. Finally, many lamented the lack of the interaction that a human coach makes available: the possibility of explaining the reasons of one's behaviour and discussing them would improve the acceptability of the relational feedback.

The two kinds of subjects we used in our focus groups did indeed differ along the computer trust dimension. Whereas the experts confirmed their skepticism about the possibility for a computer to do the job we illustrated by means of our mock-up, this did not appear to be a major issue for the non-experts. On the contrary, the latter were much more interested in finding the right place to the relational reports in their own working experience and environments. Non-experts admitted that reports could be used to improve individual skills, but they also discussed the possibility of making it available to office managers and heads, to allow them to better monitor the relational skills of their people. In a way, our non-experts seemed to be keen to assimilating the system to one of the official authorities, or official sources of information, they are used to deal with. This attitude has a number of possible consequences that emerged in the course of the discussion: in the first place, the acceptability of the system as an authority is bound to its being objective. Secondly, the level of concern for privacy issues, and the felt intrusiveness are lower, given that the cost of being monitored is balanced by the willingness to find a proper place in their environment for the relational report. Finally, it motivates a more constructive approach, which manifests itself in a strong interest about the more appropriate means to convey the report. Hence, some of the non-experts observed that the talking head with its facial expressions, introduces an evaluative aspect that can have an important impact on acceptability. This was widely agreed upon by all the other members of the focus group; the discussion continued addressing the best ways to present the report, and much consensus gained the idea of using only text, in a way, a very objective and aseptic mean.

As anticipated, the (dis)trust factors coloured many of the statements of the expert people. They did not believe that the machine will (ever) be able to monitor human behaviour and meaningfully report about it; and they were clearly worried that this might be the case, appearing much concerned with privacy issues and with the potential intrusion in very delicate issues. Very few were the attempts at finding a place to the system in their environments. Finally, let us observe that besides being motivated by the higher awareness on the limits, defects and advantages of the technology, this general attitude could be related to a more individualistic conception of the work (most of the experts were researchers) and lower familiarity with external control.

4 The acceptability of the relational report

The second step of our research was to more precisely assess users' acceptance of a system that keeps track of the participants' behaviour and analyses their relational roles to produce reports about their relational behaviour. The assessment was conducted by comparing the acceptance of automatically produced relational reports with those produced by a human expert.

The experiment was organized as a Wizard-Of-Oz [14]: all the relational reports were produced by a human coach but half of the participants were told that an automatic system produced them, and the other half were told the truth. The experiment addressed the same four dimensions informally examined in the focus groups: (i) the perceived usefulness of the RR; (ii) its reliability (whether people think that an automatic system can reliably provide a report on such a delicate matter as individual behaviour in group situations); (iii) its intrusiveness (the perceived degree of intrusiveness of a service that monitors group and individual behaviour to provide reports on their relational behaviour); and (iv) its acceptability (what affects the acceptance of the report by addressees?).

4.1 The experiment

Eleven groups of four people engaged in a structured discussion about half an hour long, according to the Survival Task paradigm. This task is frequently used in experimental and social psychology to elicit decision-making processes in small groups. Originally designed by National Aeronautics and Space Administration (NASA) to train astronauts before the first Moon landing, the Survival Task proved to be a good indicator of group decision-making processes [15]. The exercise consists in promoting group

discussion by asking participants to reach consensus on how to survive in a disaster scenario, like moon landing or a plane crashing in Canada. The group has to rank a number (usually 15) of items according to their importance for crew members to survive.

A consensus decision-making scenario was chosen, because the intensive engagement requested to reach mutual agreement offers the possibility to observe a larger set of social dynamics and attitudes. In consensus decision-making processes, each participant is asked to express her/his opinion and the group is encouraged to discuss each individual proposal by weighing and evaluating their quality. Consensus was enforced by establishing that any participant's proposal would become part of the common sorted list only if she managed to convince the others of the validity of her proposal. An element of competition was also added by awarding a prize to the individual who proposed the greatest number of correct and consensually accepted items.

The participants (40% males and 60% females) involved in the study were all clerks from ITC-irst administrative services. In all cases, they knew each other, and had often been involved in common group activities in the past. The average age was 35 years. All the groups were mixed gender.

The groups were video-recorded using four fixed omni-directional cameras, close-talk microphones and seven T-shaped microphone arrays, each consisting of four omni directional microphones (Fig. 1). There was no attempt to hide the recording devices since one of the purposes of the experiment was to evaluate the acceptability of being recorded.

Few days after their session, participants received an individual report elaborated by a social psychologist (the coacher), describing her behaviour in terms of the functional roles played during the meeting. In writing the reports, the psychologist considered only non-verbal aspects of participants' behaviour, such as the posture and the tone of voice, and not aspects related to content. Each subject was convened individually so that she could not discuss the content of her report with the other participants; during that session, they were also asked to fill out a questionnaire designed to investigate the four dimensions mentioned above (see below).

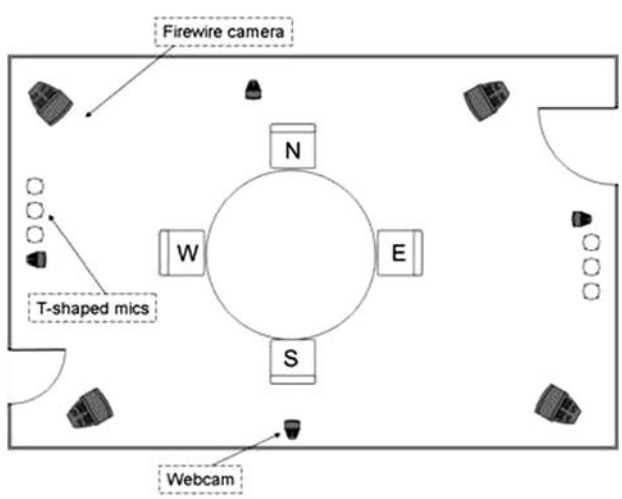

Fig. 1 The experimental setting

Half of the participants were told that their report was automatically produced by an intelligent system able to monitor the groups' behaviour, while the other half (the control group) were told that the report was written by a psychologist. The selection was randomised and balanced with respect to gender.

The attitude toward the report was tested by a seven-item questionnaire aimed at assessing the perceived usefulness, reliability, perceived degree of intrusiveness and acceptability of the RR. A semantic differential targeting the appropriateness, completeness and clarity of the report was also used (the semantic differential was part of the 6-scale questionnaire proposed by Garrison [16], with a Cronbach alpha of 0.9482). The questionnaire and the semantic differential scale are reported in the Appendix.

4.1.1 Results

The answers to the questionnaire were analysed by means of a multivariate ANOVA ($p = 0.05$), applied to the data from 41 questionnaires (three subjects did not fill theirs properly). The factor was the source of the report: 'human' for the control condition, and 'system' for the experimental one. Generally, there were no statistically significant differences among the responses to the questionnaire in the two groups. Regarding the subscales of the semantic differential, they were also analysed by means of a multivariate ANOVA with $p = 0.05$. The only difference we found concerned the appropriateness sub-scale ($F(1,39) = 4.883$, $p < 0.05$), where the 'system' group rated the appropriateness of their report higher than the 'human' one (estimated means and standard errors: $M_{expert} = 28.38$, $SE_{expert} = 2.03$; $M_{system} = 34.82$, $SE_{system} = 1.98$).

In the end, this study did not reveal any significant difference between the two groups concerning usefulness, reliability, degree of intrusiveness, acceptability, completeness and clarity of the report. The only significant difference concerned the appropriateness, with a slight advantage of the system over the human, a result that echoes comments of the clerks of the focus group who maintained that the system could be more objective than the human in its assessment of the interaction behaviour. As far as these results are concerned, there is no substantive evidence that an automatically produced report about own

relational behaviour in meetings would be accepted any differently than one produced by a human expert. Though to be confirmed by further study, this is an encouraging result for it supports the idea that meeting participants could indeed consider automatically produced reports to improve their own relational skills.

5 Observing group behaviour

The previous sections have provided both insights about potential users' attitudes and beliefs that can be used to extract initial requirements for the RR, and evidence that such a service could be acceptable and valuable for them. The next step consists in the development of a coding scheme capable of capturing relevant behavioural sequences, which can be used to train a system that automatically recognizes and classifies relational behaviour.

A coding scheme for group behaviour should be usable by human annotators to provide corpora for supervised learning, and its categories should be mappable onto patterns of low-level observations that can be automatically detected by means of acoustical and visual scene analysis.

In our search of suitable categories for the coding scheme, the goal of presenting individual profiles to participants suggested that we carefully consider those approaches to social dynamics that focus on the roles that members play inside the group. Eventually, we based our coding scheme on Benne and Sheats's functional roles [17], and on Bales [18] two-dimensional approach, adjusting them according to observations we performed on a number of face-to-face meetings (see [19] for more details about the coding scheme).

The Functional Role Coding Scheme (FRCS) consists of five labels for the task area and five labels for the socio emotional area. The task area includes functional roles related to the facilitation and coordination of the tasks the group is involved in, as well as to the technical skills of the members as they are deployed in the course of the meeting. The Socio Emotional Area involves roles oriented toward the functioning of the team as a group. Below we give a synthetic description of FRCS.

5.1 The task area functional roles

Orienteer (o). She orients the group by introducing the agenda, defining goals and procedures, keeping the group focused and on track and summarizing the most important arguments and the group decisions.

Giver (g). She provides factual information and answers to questions. She states her beliefs and attitudes about an idea, expresses personal values and factual information.

Seeker (s). She requests suggestions and information, as well as clarifications, to promote effective group decisions.

Procedural technician (pt). She uses the resources available to the group, managing them for the sake of the group.

Follower (f). She only listens, without actively participating in the interaction.

5.2 The socio-emotional functional roles

Attacker (a). She deflates the status of others, expresses disapproval, attacks the group or the problem.

Gate-keeper (gk). She is the group moderator, who mediates the communicative relations; she encourages and facilitates the participation and regulates the flow of communication.

Protagonist (p). She takes the floor, driving the conversation, assuming a personal perspective and asserting her authority.

Supporter (su). She shows a cooperative attitude demonstrating understanding, attention and acceptance as well as providing technical and relational support.

Neutral (n). She passively accepts the idea of others, serving as an audience in group discussion.

5.3 Studies on the reliability of the coding scheme

The coding scheme was applied to a corpus consisting of the video and audio recordings of nine group meetings (selected from real meetings held at our place), for a total of 12.5 h. Its reliability was assessed on a subset of the corpus consisting of 130 min of meetings for the socio-emotional area and 126 min for the Task Area (from three group interactions). Five participants were coded on the socio-emotional area and five in the task area by two trained annotators. Two confusion matrices were built, one for the task area and one for the socio-emotional one, to measure cross-judge consistency of class membership by means of Cohen's κ [20].

In the task area, Cohen's statistics was $\kappa = 0.70$ ($N = 758$, SE = 0.02, $p < 0.001$; confidence interval with $\alpha = 0.05$: 0.67–0.75). According to Landis and Koch's [21] criteria, the agreement on the task area is good ($0.6 < \kappa < 0.8$). Table 1 shows the confusion matrix (the table shows the occurrences of the different roles at sampling of 10 s).

Figure 2 shows the percentage of the different roles as they occurred in our corpus. The Orienteer is the most common role reflecting the nature of the interactions observed, which were mostly project meetings where teams had to report to their project managers about the status of the work.

Table 1 Confusion matrix for the roles in the task area (758 × 10 s = 126 min)

Judge 1 × Judge 2 Crosstabulation						
Count	Judge 2					Total
	g	n	o	r	s	
Judge 1						
g	115	55	13	3	0	186
n	3	140	15	18	1	177
o	2	18	231	0	16	267
r	1	7	0	81	0	89
s	0	8	3	0	28	39
Total	121	228	262	102	45	758

g Giver, *n* Follower, *o* Orienteer; *r* Procedural Technician, *s* Seeker

JUDGE1 * JUDGE2 Crosstabulation

Count		JUDGE2					Total
		g	n	o	r	s	
JUDGE1	g	115	55	13	3	0	186
	n	3	140	15	18	1	177
	o	2	18	231	0	16	267
	r	1	7	0	81	0	89
	s	0	8	3	0	28	39
Total		121	228	262	102	45	758

Fig. 2 Percentage of the different roles in the task area as annotated by the two judges

Regarding the socio-emotional area, the inter-annotator agreement was $\kappa = 0.60$ ($N = 783$, SE = 0.02, $p < 0.001$; confidence interval with $\alpha = 0.05$: 0.56–0.65). According to Landis and Koch's [21] criteria, the agreement on the Socio-Emotional roles is at the borderline between good ($0.6 < \kappa < 0.8$) and moderate ($0.4 < \kappa < 0.6$). Table 2 shows the confusion matrix (the table shows the occurrences of the different roles at sampling of 10 s).

Figure 3 shows the relative percentage of the different roles in the socio-emotional area in our corpus. It can be noted that the Gate-Keeper role was never observed by either annotator, this being probably due to the actual absence of a (either professional or de facto) facilitator in our meetings. The Attacker too is not well represented. Again, this reflects the nature of our meetings, which do not favour the emergence of strong contrasts among participants.

Table 2 Confusion matrix for the roles of the socio-emotional area (783 × 10 s = 130 min)

Judge 1 × Judge 2 Crosstabulation					
Count	Judge 2				Total
	a	n	p	s	
Judge 1					
a	26	1	5	0	32
n	3	241	29	105	378
p	0	32	233	12	277
s	0	14	7	75	96
Total	29	288	274	192	783

a Attacker; *n* Neutral, *p* Protagonist, *s* Supporter, *g* Gate-Keeper (not present)

5.3.1 *Analysis of the disagreements*

The class-wise analysis of the κ's for the task area shows that the most reliable classes are the *Orienteer* and the *Procedural Technician* (see also the values of the *z*-scores in Table 3). The least reliable class is the *Seeker*, mostly because of its high standard error. The *Giver* and the *Follower* fall in between. Considering the absolute values of the κ's and the lower bounds of the confidence intervals ($\alpha = 0.05$), the classes that deserve consideration in view of improvements are the *Seeker* and the *Follower*.

The use of the Pearson's standardized residuals from the independence model, enables us to pin point the disagreements that more closely follow a uniform pattern, hence those on which the judges diverge most. In these cases, the standardized residuals come close to zero. The data, reproduced in Table 4, confirm the results based on the κ statistics: almost all off-diagonal residuals are strongly negative, and often below the value of -3 that can be taken as a cut-off threshold for significance.

The most interesting disagreements between the two annotators concern two cases: in the first, judge1 classifies roles as *Giver* and judge2 classifies them as Follower; in the second, judge1 sees a *Procedural Technician* role whereas judge2 classifies it as a *Seeker*. Putting together these results with the discussion of Table 3 above, it can be concluded that in order to improve inter-annotators agreement in the task area, we must address, in the first place, the *Seeker* and the *Follower*, in particular reducing the '*giver–follower*' and the '*orienteer–seeker*' disagreement

Turning to the socio-emotional area, the class-wise analysis of the κ's, in Table 5, confirms that the social area is slightly less reliable than the task one (see also Table 6). The most reliable class is the *Protagonist*, and the by far less reliable one is the *Supporter*; the *Attacker*, despite its high κ value, needs some consideration, given its high standard error.

The analysis of Pearson's standardized residuals from the independence model shows the importance of the disagreement on *Neutral* and *Supporter* between judge1 and judge2, see Table 6.

In conclusion, the weakest class in the social area is the *Supporter*, which is involved in a strong disagreement with the *Neutral*.

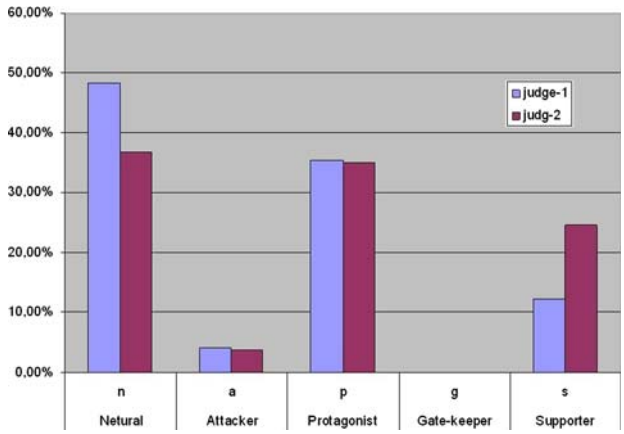

Fig. 3 Percentage of the different roles in the socio-emotional area as annotated by the two judges

Table 3 Class-wise κ values for the task area

	κ	SE	z-score	Conf. interval	
				Up. bound	Low. bound
g	0.69	0.032	21.53	0.75	0.63
f	0.58	0.033	17.61	0.65	0.52
o	0.81	0.023	35	0.85	0.76
r	0.83	0.031	26.65	0.89	0.77
s	0.65	0.062	10.44	0.77	0.53

g Giver, *f* Follower, *o* Orienteer, *r* Procedural Technician, *s* Seeker

Finally, an important feature of coding schemes is the symmetry of their confusion matrices. In a perfectly symmetric confusion matrix, for labels *a* and *b*, any *a* versus *b* disagreements between judge1 and judge2 correspond to a *b* versus *a* disagreements between judge2 and judge1. Symmetry can be assessed through the Bowker test [22], which yields a statistics that has asymptotic χ^2 distribution. In our case, the value of the Bowker statistics is 75.14 and 69.59 for the task and the social area, respectively, with 10 and 6 degree of freedom. In both cases, the null hypothesis

Table 4 Standardized residual for the roles in the task area. Model: independence

Judge 1	Judge 2				
	g	f	o	r	s
g	19.7	–0.2	–9.1	–5.4	–3.9
f	–5.9	16.2	–8.3	–1.5	–3.5
o	–8.4	–10.3	22.2	–8.0	0
r	–4.1	–4.9	–7.3	22.8	–2.5
s	–2.8	–1.3	–3.6	–2.5	17.9

Table 5 Class-wise κ values for the social area

	κ	SE	z-score	Conf. interval	
				Up. bound	Low. bound
a	0.85	0.05	16.92	0.94	0.75
p	0.53	0.03	17.53	0.59	0.47
n	0.76	0.024	31.75	0.81	0.72
s	0.43	0.039	10.95	0.50	0.35

a Attacker, *n* Neutral, *p* Protagonist, *s* Supporter

Table 6 Standardized residual for the social area

Judge 1	Judge 2			
	a	n	p	S
a	23.7	–4.0	–2.3	–3.3
n	–4.2	15.1	–15.5	2
p	–4.1	–10.8	21.3	–9.7
s	–2.1	–4.8	–6.1	13

Model: independence—*a* Attacker, *n* Neutral, *p* Protagonist, *s* Supporter

that the matrices are symmetric can be rejected with $p < 0.0001$. Tables 7 and 8 report the standardized residuals under the symmetry hypothesis.

The analysis of Pearson's standardized residuals under the symmetry hypothesis confirms that the offending cases are the same as those analysed above in connection with the independence hypothesis. In detail, the '*giver–follower*' and the '*orienteer–seeker*' disagreements are the main culprits of the lack of symmetry in the task area, whereas the '*neutral–supporter*' disagreement is the main responsible for the lack of symmetry in the social area.

To improve agreement, efforts had to be focused on the *giver–follower* and the *procedural technician* disagreements in the Task Area, with the goal of improving the κ values for the *follower* and the *seeker*, respectively, and the balance/symmetry of the annotation schema. In the socio-emotional area, the validity of the annotation schema could be ameliorated by reducing the *neutral-supporter* disagreements, improving the κ value of *supporter* and the overall schema balance.

A new set of guidelines for annotators were compiled, resulting in a uniform improvement of the κ statistics.

6 The survival task corpus

As a first step toward the actual implementation of a system component able to automatically classify the roles according to the FRCS, we collected a multimodal corpus

Table 7 Standardized residuals for the task area

	g	f	o	r	s
g	0	6.83	2.84	1	0
f	–6.83	0	–0.52	2.2	–2.33
o	2.84	0.52	0	0	2.98
r	–1	–2.2		0	0
s	0	2.33	–2.98	0	0

Model: symmetry. *g* Giver, *f* Follower, *o* Orienteer, *r* Procedural Technician, *s* Seeker

Table 8 Standardized residuals for the social area

	a	n	p	s
a	0	–1	2.24	0
n	1	0	–0.38	8.34
p	–2.24	0.38	0	1.15
s	0	–8.34	–1.15	0

Model: symmetry. *a* Attacker, *n* Neutral, *p* Protagonist, *s* Supporter

that includes manually annotated functional roles and audio and visual features. As described below, the first attempts at automatic classification are satisfactory, with accuracy ranging from 0.65 to 0.68 and the macro *F*-score from 0.52 to 0.55; all these figures are above the chosen baselines. In our view, these results show that the task of using multimodal information to support the production of automatic personalized reports about one's relational behaviour in group environment is feasible, even if more work needs to be done before attaining performances which are adequate for realistic applications.

The multimodal annotated corpus is based on the audio and video material recorded during the 11 meetings of the Survival Task experiment described above. The average duration of the meetings was 25 min, the range being 13′.08″–30′.06‴ the total length of the corpus is 3.44.55″ h.

6.1 Role annotation

Functional role annotations consists of tuples ⟨*role-area*; *participant-code*; *role-label*; start: *start-time*; end: *end-time*; duration: *duration*⟩ where role-area and role-label have been described above. Figure 4 presents an example: the tuple ⟨task: w; o; start:1.796; end:170.066; duration:168:26999″ or the duration:168.269⟩ refers to the role of orienteer ('o') belonging to the 'task' area, as played by participant *w* from time 1.796 till time 170.066, for a duration of 168.26999 s.

The corpus was annotated by a single judge trained according to the new guidelines.

6.2 Audio annotation: voice activity

The audio recordings were annotated with respect to the voice activity of the participants—that is, the presence/absence of human voice, without distinguishing between verbal and non-verbal activity.

Each session was segmented by first automatically labelling voice activity (as recorded by means of the close-talk microphones) by means of the voice activity detector (VAD), which uses the time energy of the signal [23]. For each speaker, VAD identifies the amount of speech activity, and produces an output such as ⟨*participant-code*, *start time*, *end time*, *label*⟩, where *label* takes on the value 'speech' and 'no-speech'.

Voice activity detector's output was then manually checked and improved. In the first place, errors of the automatic annotation were removed, and in particular, the wrong assignments due the fact that the voice activity of a subject could often be captured by the close-talk microphone of her neighbour. Secondly, VAD is based on time energy, and it is not able to distinguish between voice activity and other acoustic events. Manual annotation purified the VAD annotation from breaths, yawns, coughing, and noises caused by the subjects when touching the microphones.

6.3 Video annotation: 3D tracking of body activity

Visual cues were employed to derive head position and orientation as well as body activity. The subjects' position in the room was tracked through head position identification. 3D positions have an absolute timestamp and are referenced to an origin, which is on the floor under the centre of the table. The 3D co-ordinate system for the room is oriented in the following way: the *X*-axis represents a Westerly direction, the *Z*-axis a Northerly direction, and the *Y*-axis the height from the floor. For each participant the 3D tracking produces a tuple ⟨*timestamp*; *x-axis*; *z-axis*; *y-axis*⟩, where an absolute timestamp is followed by the cardinal point, which identifies head position in the room.

Starting from head position detection, colour and hedge features were used to track head orientation and to estimate focus of attention. The output from the 3D tracking consists of tuples such as ⟨*timestamp*; *head orientation*⟩ for each subject. Head orientation can take on one of the following values: ''down'', when subject head is oriented toward the table, ''S'', ''N'', ''W'', ''E'', when the head is oriented toward South, North, West or East (see Fig. 5), each of them referring to one of the other participants.

Fidgeting refers to localized repetitive motions such as when the hand remains stationary while the fingers are tapping the table, or playing with glasses, etc. Fidgeting was tracked by means of skin region features; temporal

motion is used as trigger; see [24] for details. For each subject, the output of the analysis consisted in the tuples ⟨*timestamp, fidgeting energy; hand/arm activity*⟩. For example, in the tuple ⟨1124358961419507; 16; 1⟩, an absolute timestamp is followed by two normalized fidgeting values. The first ('16') represents the fidgeting energy of the person's body and the second ('1') represents his hand/arm activity. The normalized values are referenced to that person's most vigorous fidgeting during the entire recorded sequence, hence they are person specific.

6.3.1 Automatic detection of functional roles

As an initial step toward an automatic detection of functional roles, we modelled role assignment as a multiclass-classification problem using the speech activity and the fidgeting features only. We used time windows of varying size, from 0 to 14 s in two conditions: (a) the whole window is on the left hand side of the time point to classify (which is preferable for online classification), and (b) half of the window is before and half is after the classification point (in this case, the classification for time t takes place with a delay of *width*/2 s, where *width* is the length of the window).

Support vector machine[1] was used were used as classifier by employing a bound-constrained classification algorithm with an RBF kernel $K(x,y) = \exp(-\gamma\|x - y\|^2)$ [25]. The cost parameter C and the kernel parameter γ were estimated with the grid technique by cross-fold validation using a factor of 5. Furthermore, the cost parameter C was weighted for each class with a factor inversely proportional to the class size. SVM were originally designed for binary classification but several methods have been proposed to construct multi-class classifier [26]. The ''one-against-one'' method [27] was used whereby each training vector is compared against two different classes by minimizing the error between the separating hyperplane margins. Classification is then accomplished through a voting strategy whereby the class that most frequently won is selected.

The results are encouraging, though more work is needed to find a solution that can be smoothly integrated in a running system. The performance for the Task area is rather good with a macro precision of 0.55 and a macro recall of 0.49 (left-only window of 14 s); yet the differences among the classes are not negligible (ranging from an F-score of 0.73 for the Giver role to a mere 0.30 for the Seeker). The results are a little bit worse for the Socio area roles: though the best performance are quite high—macro precision and macro recall respectively, 0.75 and 0.43—the differences among the classes are also high (varying from an F-score of 0.78 for the Neutral role to a very low 0.02 for the Attacker). As expected, for both task and socio-emotional roles, the worse classification results were on low-represented classes (the Seeker and the Attacker). Even if it is likely that roles distribution will always be unbalanced, larger corpora, possibly from different task may provide for more balance. For more details about the automatic detection of functional roles, see [28].

```
task;w; o; start:621.466; end:645.965; duration:24.499023
task;w; g; start:647.769; end:806.186; duration:158.41699
task;w; g; start:831.091; end:835.619; duration:4.528015
task;w; o; start:855.022; end:1783.348; duration:928.32605
task;w; g; start:1784.843; end:1878.873; duration:94.03003
task;w; s; start:1952.966; end:1957.62; duration:4.6540527
task;w; g; start:2016.882; end:2067.334; duration:50.452026
task;w; g; start:2087.179; end:2147.026; duration:59.846924
……..
```

Fig. 4 Functional role annotations of a meeting

7 From coding scheme to relational reports

Starting from the information described above, a relational report can be automatically generated for each participant, which highlights some major aspects of his behaviour during the meeting. The report is built according to the task and socio-emotional roles played by the subject during the interaction, and taking into account also the roles played by the other participants. Each final relational report has the form of a multimedia presentation where different modalities are used to present the relevant information, and provide evidence in support of it. The actual text of the report is intended to describe the behaviour in an informative rather than normative way, helping the user accomplishing the first step of the reflective process—namely, the return to experience (what happened during the meeting?). To improve effectiveness and emotional involvement, a virtual character is used which reads the report with emotional facial expressions appropriate to the content (e.g. a sad expression is used when something unpleasant, e.g. a serious contrast with a colleague, is being recalled). When appropriate, the presentation is enriched with short audio–video clips from the actual meeting, which exemplify the information presented. A graphical representation of the participant's behaviour is also provided, yielding a more explicit feedback about the system's internal interpretation of what was monitored during the meeting. The generated report is automatically composed as a SMIL[2] presentation. Figure 6 shows a snapshot of a sample relational report.

[1] The BSVM tool available at http://www.csie.ntu.edu.tw/~cjlin/bsvm/

[2] SMIL—Synchronized Multimedia Integration Language—is the standard language for multimodal presentations developed within the W3C. http://www.w3.org/AudioVideo/

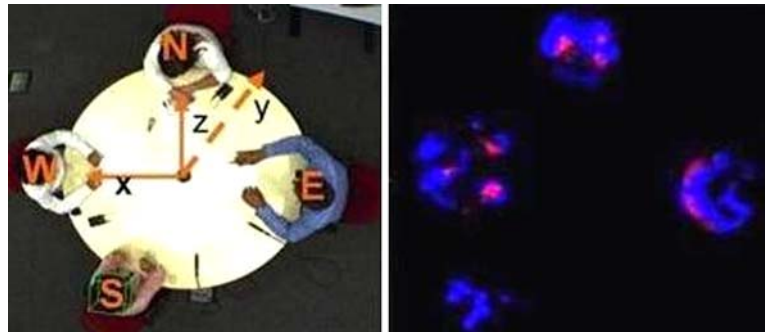

Fig. 5 Head orientation and head position detection

7.1 Content selection and phrasing

The approach to content selection and organization we adopted is based on the assumption that in behavioural reports, the conventions on how information is presented by human experts play a major role. These ''patterns of appropriate ordering'' (more widely known in the NLG community as schemas [29] or the Generic Structure Potential [30] of a text) have been exploited by many NLG systems to guide the text planner in organizing the text structure.

For our report generation we have re-used a general-purpose schema-based text planner [31], which accesses a repository of declarative discourse schemata. Each schema contains applicability conditions, which look in the FRCS annotation for specific patterns of roles played by the participants. Once the applicability conditions are satisfied, the text planner extracts from the body of the schema the instructions concerning what to say, what to show and in which order, and the different media synchronization. The schemata have been derived from the analysis of the actual reports written by the social psychologist involved in the Survival Task experiment described above, and from additional expert knowledge elicited by means of interviews. To exemplify, one of our schemas says that if the total amount of active task roles played by the participant during the meeting (orienteer + giver + seeker + recorder) is greater than 75%, then a text should be produced (i) expressing this active contribution (''You have very actively contributed to the discussion, with many verbal contributions..''), (ii) possibly indicating whether there has been dominance (''You have maintained a highly dominant role, keeping the attention of the other participants, as highlighted in the chart by the bar in red.''), (iii) including sentences emphasizing the amount of team guidance (''You've maintained a pivotal role in defining how to proceed with the discussion and in summarizing the results, as highlighted in the chart by the bar in blue.'') and (iv) information contribution (''You have significantly contributed with your ideas and opinions.'').

More elaborate discourse strategies involve reasoning about the behaviour of various participants at a given time. For example, should the target participant have often played the ''orienteer'' role at the beginning and at the end of the meeting, with the others being silent or neutral, the report could include a statement like ''*at the beginning of the*

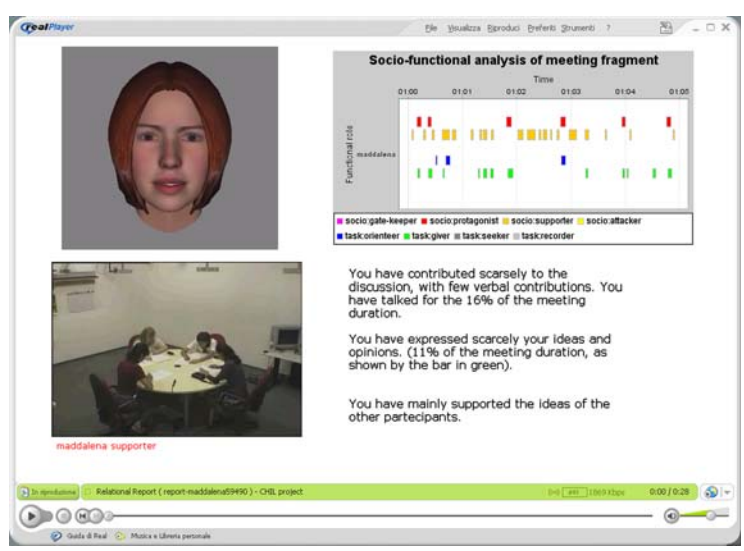

Fig. 6 An example of a multimedia relational report for one meeting participant

meeting you have helped define the agenda and initiate the discussion, summing up the outcome of the meeting at the end''. In case sequences of seeker-(recorder)-attacker are observed, the report could be complemented with: ''in some cases, however, you have displayed a critical and aggressive behaviour as a response to your colleagues' contribution''.

The linguistic realization of sentences is currently template-based. As a preliminary solution, the wording of sentences has been based on the typical one found in reports authored by human experts; more work is needed, however, to study the impact of improper lexical choices on users.

Schemas also tell when graphics should be included next to the text for clarification; they also introduce appropriate information about the emotions the virtual character should display. All this information is modelled as a SMIL script for a multimedia presentation.

7.2 Graphical feedback

Graphics can play a crucial role in intuitively conveying the dynamics of relational behaviour during a meeting, either as an aid for counsellors' work (e.g. [32, 33]), and as a visual feedback to meeting participants for self-reflection [34, 35].

Visual displays have been used for various purposes in group contexts. For instance, DiMicco [34] exploited them to provide quantitative feedback on the level of participation of each meeting member. In our case, a more qualitative interpretation of social dynamics may be provided by plotting the time each participant spent playing the different task- and socio-emotional roles during the meeting. For instance, the graphic in the top right part of Fig. 6 shows the temporal distribution of the various roles played by a participant during the meeting. Specific phrases are inserted in the text, which comment on, and explain the meaning of the various coloured bars of the graphic, (''You have only seldom expressed your ideas and opinions as shown by the green bar.'').

An alternative type of visualization is inspired to the work of Bales [32], where quantitative information about various dimensions of interaction are used to provide a qualitative view, called the 'interaction profile', of a participant's activity during the meeting (see Fig. 7). In his work on Interaction Process Analysis, Bales tested the efficacy of this type of displays, concluding that their qualitative characteristics, e.g. the position of the peak(s), may be used by counsellors to explain participants' behaviour in small-group interaction.

Starting from Bales' work, and adapting it to match our FRCS, we developed similar graphical views that express at a glance the interaction behaviour of a participant, by profiling the amount of time spent on playing task- and socio-emotional roles. Figure 8 reports examples, with the left-hand graph (a) being the display of a very talkative, orienteering participant and the right-hand (b) one that of passive, supportive, participant.

These graphics, automatically generated from the FRCS annotation of the meeting, are then made available for their

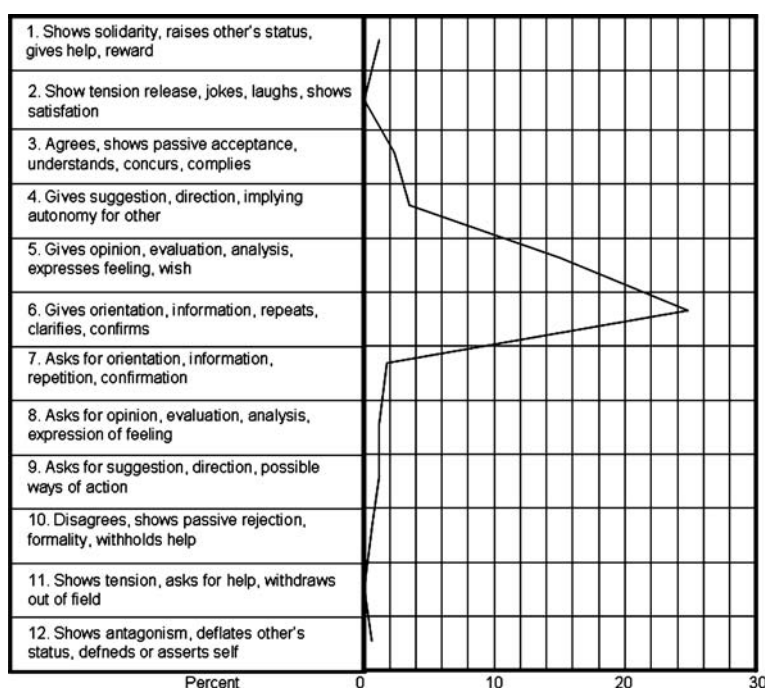

Fig. 7 Interaction profile of leader in democratic-directive role, according to Bales' Interaction Process Analysis

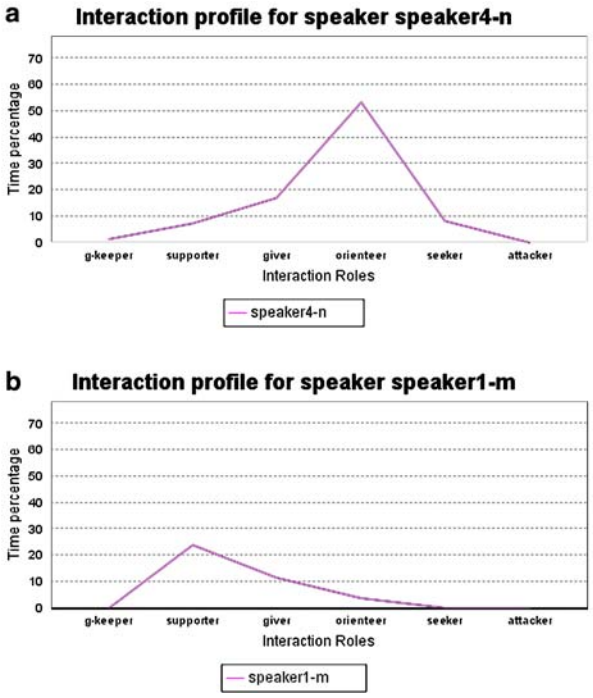

Fig. 8 Sample interaction profiles automatically generated based on socio-emotional roles: **a** Very talkative, orienteering meeting participant, **b** Passive, supportive meeting participant

integration in the multimedia relational report, as illustrated in Fig. 9.

Other sophisticated and intuitive views used by small group counsellors, e.g. those based on the SYMLOG framework [33], are also under investigation, though their automatic generation is not straightforward, given that they rely on richer information about participants behaviour than we have available at present.

8 Conclusions

The aim of this work was to contribute to an emerging new class of multimodal systems in which multimodality is not used to improve human–machine interaction, but as a core component of devices that by ''observing'', and staying in the loop of human–human interaction [10] provide various types of functionalities. Besides being technically challenging, these systems raise so many user-related (intrusiveness, acceptability, privacy violation, etc.) and ethical issues that a user-centred approach to their design is virtually necessary.

Building on the availability of rich multimodal information in meeting rooms equipped with technology for audio–visual scene analysis, we have explored the prospects of a functionality inspired to *coaching*. It consists of a report about the social behaviour of individual participants, which is generated from multimodal information, and privately delivered to them. The underlining idea is that the individual, the group(s) they are part of, and the whole organization might benefit from an increased awareness of participants about their own behaviour during meetings.

Fig. 9 Sample relational report integrating different graphical views

In our case, we used a mix of laboratory-based testing and attitudinal study to deal with the initial phases of the user centred design cycle.

User-related concerns were investigated by means of a mix of laboratory-based testing and attitudinal study. We first elicited attitudes and beliefs about the service through focus group; then we empirically investigated the acceptance of such a system by simulating it through a WOZ experiment comparing the acceptance of automatically generated reports with that of reports produced by a professional coach. These studies provided enough evidence that the service can actually be valuable for engaging people in the first stage of the reflective process [4]—that is, the return to experience (what happened?)—and allowed us to derive initial requirement specification.

With this much of a background and empirical support, we turned to illustrate steps taken towards the effective construction of such a service: (a) the development and validation of a reliable coding scheme to annotate group behaviour; (b) the production of an annotated multimodal corpus that was used to (c) train a component for the automatic extraction of functional roles from audio–visual observations; a generation component that, exploiting role information, assembles the multimedia report.

In the future, we plan to refine the automatic classification of functional roles by exploiting more multimodal features, and by experimenting with different techniques such as Hidden Markov Models. We also plan to investigate the impact of the different communicative strategies used in the relational reports (the talking head, the different graphic displays and so on) on the acceptability and effectiveness of the relational report.

9 Appendix

Here we present the questionnaire and the semantic differential scale used in the acceptability study. They both were in Italian. Table 9

Questionnaire

1. This kind of reports may help me improving my relational behavior
   I agree ° ° ° ° ° ° ° I disagree
2. This type of reports may help reflecting on some aspects of my behavior that I have never considered before
   I agree ° ° ° ° ° ° ° I disagree
3. This kind of reports may help me improving my interactions in the meeting I participate
   I agree ° ° ° ° ° ° ° I disagree
4. I will take into consideration the observation of this report at the next meeting
   I agree ° ° ° ° ° ° ° I disagree
5. The report cougth some relevant aspects of my behavior
   I agree ° ° ° ° ° ° ° I disagree
6. The report missed some relevant aspects of my behavior
   I agree ° ° ° ° ° ° ° I disagree
7. The expert/system that compiled the report had a good ability in analysing behavior
   I agree ° ° ° ° ° ° ° I disagree

Table 9 Semantic differential

	1	2	3	4	5	6	7	
Appropriateness scale								
Incorrect								Correct
Inaccurate								Accurate
Unusual								Usual
Appropriate								Inappropriate
Unwise								Wise
Incompetent								Competent
Trustworthy								Untrustworthy
Reputable								Disreputable
Qualified								Unqualified
Credible								Not credible
Reliable								Unreliable
Dishonest								Honest
Dangerous								Safe
Believable								not believable
Completeness scale								
Incomplete								Complete
Objective								Subjective
Deep								Shallow
Sufficient								Insufficient
Clarity scale								
Readable								Unreadable
Understandable								Not understandable
Unclear								Clear
Unimportant								Important

References

1. Doyle M, Straus D (1993) How to make meetings work. The Berkley Publishing Group, New York
2. Andersen C (2000) A theoretical framework for examining peer collaboration in preservice teacher education. In: Proceedings of the 2000 annual international conference of the association for the education of teachers in science
3. Bloom G, Castagna C, Warren B (2006) More than mentors: principal coaching. Leadership. May/June [cited 05/05/2006]. Available at: http://www.acsa.org/
4. Boud D, Keogh R, Walker D (1985) Reflection: turning experience into learning. Kogan Page, London
5. Tang JC (1991) Findings from observational studies of collaborative work. Int J Man-Mach Stud 34:143–160
6. 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
7. Erickson T, Halverson C, Kellogg W.A, Laff M, Wolf T (2002) Social translucence: designing social infrastructures that make collective activity visible. Commun ACM 45(4):40–44
8. DiMicco JM, Pandolfo A, Bender W (2004) Influencing group participation with a shared display. CSCW 2004:614–623
9. Maloney-Krichmar D, Preece J (2005) A multilevel analysis of sociability, usability, and community dynamics in an online health community. ACM Trans Comput-Hum Interact 12(2):201–232

10. Waibel A, Steusloff H, Stiefelhagen R (2004) CHIL: computer in the human interaction loop. In: NIST ICASSP meeting recognition workshop, Montreal, Canada
11. McCowan I, Gatica-Perez D, Bengio S, Moore D, Bourlard H (2004) Towards computer understanding of human interactions. In: Aarts E, Collier R, van Loenen E, de Ruyter B (eds) Ambient intelligence. Springer, Heidelberg, pp 235–251
12. Kaiser E, Demirdjian D, Gruenstein A, Li X, Niekrasz J, Wesson M, Kumar S (2004) A multimodal learning interface for sketch, speak and point creation of a schedule Chart. In: Proceedings of ICMI 2004, pp 329–330
13. Brdiczka O, Maisonnasse J, Reignier P (2005) Automatic detection of interaction groups. In: Proceedings of the 7th international conference on multimodal interface, Trento, Italy
14. Wilson J, Rosenberg D (1988) Rapid prototyping for user interface design. In Helander M (eds) Handbook of human–computer interaction, New York, pp 859–875
15. Hall JW, Watson WH (1970) The effects of a normative intervention on group decision-making performance. Hum Relat 23(4):299–317
16. Garrison B (2003) The perceived credibility of electronic mail in newspaper newsgathering. In: Proceedings of communication technology and policy division, Association for Educational in Journalism and Mass Communication Midwinter Conference, Boulder, Colorado
17. Benne K.D, Sheats P (1948) Functional roles of group members. J Soc Issues 4:41–49
18. Bales R.F (1970) Personality and interpersonal behaviour. Holt Rinehart and Winston, New York
19. Falcon V, Leonardi C, Not E, Pianesi F, Zancanaro M (2005) Observing multimodal behaviour to support group dynamics. In: Workshop on user-centred design and evaluation of services for human–human communication and collaboration held in conjunction with ICMI'05, Trento, Italy
20. Cohen JA (1960) Coefficient of agreement for nominal scales. Educ Psychol Measure 20(1):37–46
21. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
22. Agresti A (2002) Categorical data analysis. Wiley, New York
23. Carli G, Gretter G (1992) A start-end point detection algorithm for a real-time acoustic front-end based on DSP32C VME board. In: Proceedings of ICSPAT, Boston, USA
24. Chippendale P (2006) Towards automatic body language annotation. International conference on automatic face and gesture recognition—FG2006 (IEEE) Southampton, UK, pp 487–492
25. Cristianini N, Shawe-Taylor J (2000) Support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge
26. Hsu C-W, Lin C-JA (2002) Comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13:415–425
27. Kressel U (1999) Pairwise classification and support vector machines. In: Scholkopf B, Burges CJC, Smola AJ (eds) Advances in Kernel methods—support vector learning. MIT Press, Cambridge
28. Zancanaro M, Lepri B, Pianesi F (2006) Automatic detection of group functional roles in face to face interactions. In: Proceedings of international conference of multimodal interfaces ICMI-06
29. McKeown KR (1985) Text generation: using discourse strategies and focus constraints to generate natural language text. Cambridge University Press, Cambridge
30. Halliday MAK, Hasan R (1985) Language, context and text: aspects of language in a social-semiotic perspective. Deakin University Press, Geelong
31. Callaway C, Not E, Stock O (2006) Report generation for post-visit summaries in museum environments. In: Stock O, Zancanaro M (eds) PEACH: intelligent interfaces for museum visits. Cognitive technologies series. Springer, Heidelberg, pp 71–92
32. Bales RF (1976) Interaction process analysis: a method for the study of small groups. University of Chicago press, Chicago
33. Bales RF, Cohen SP (1979) SYMLOG: a system for the multiple level observation of groups. Collier Macmillan Publishers, London
34. DiMicco JM (2005) Changing small group interaction through visual reflections of social behaviour. PhD Thesis, MIT Media Lab
35. Losada M, Markovitch S (1990) GroupAnalyzer: a system for dynamic analysis of group interaction. In: Proceedings of 23rd Hawaii international conference for system sciences, Kailua-Kona, Hawaii, pp 101–110