Socially-Aware Interfaces
for Supporting Collocated Interaction

Gianluca Schiavo[1]
[1] CIMeC, Center for Mind/Brain Sciences,
University of Trento, and FBK – Fondazione Bruno Kessler, Trento, Italy
gianluca.schiavo@unitn.it

Abstract. One of the important challenges in ubiquitous computing is to improve the computer's access to information available in the social context. The goal of my PhD project is to investigate how to design interfaces that support collocated multi-users interactions taking into account users' non-verbal behaviour (specifically, gaze, postures, and body movements). In particular, the research activities are twofold: to understand which non-verbal cues and social signals reflect joint attention, cooperation and synchronization in collocated group activities and to design systems that can handle and utilise this information. To this end, I present an integrated research approach for designing multi-user interactions based on the social signal processing approach. I also discuss the progress to date targeting the development of systems able to sense and react to social context. Finally I propose some relevant points that I would like to explore during the doctoral consortium.

1   Introduction

Multi-user interfaces are collaborative computing interfaces that support simultaneous participation of multiple users who are collocated in the same place. Examples of technologies that integrate multi-user interfaces are interactive tabletops, public displays (e.g. large wall displays) and peripheral displays. These technologies differ from standard personal computers or non-located technologies because they provide a shared workspace and a platform for the cooperation and collaboration of several users engaged in the same task. Multi-user interfaces are in fact assumed to be better for group activities since they can support equal and flexible forms of collaboration [16].

Designing for multi-user interfaces can raise different issues to single-user interfaces. There are, in fact, different social dynamics characterizing the interaction of a group with a multi-user interface that should be taken into account in the design process. Many researchers [13, 16] have examined the social interactions that occurred while users interact with multi-user interfaces including territoriality, awareness of others' activity, group formation, parallel or collaborative interaction; these are all relevant social dynamics that should be addressed when dealing with multi-user interfaces. In fact, multi-user interfaces, more than single-user ones, should be designed with the social context in mind, leveraging the benefits of group interactions.

My research project aims to expand systems' capabilities for sensing the social context, based on the assumption that multi-user interfaces should provide information in a way that best suits the users' needs both as individuals and as group members. From this perspective, automatic machine sensing converges with psychological and social research, addressing how multi-user technology is enabled to sense and properly react to social dynamics.

1.1 Exploring Social Signals

The project can surely benefit from the extensive research that has been conducted in different areas of psychology and computer science, notably from the research fields of non-verbal behaviour, affective computing and social signal processing.

Research in psychology and in social cognition regarding the mechanisms of non-verbal behaviour has suggested that some social cues are the result of automatic processes. Experimental studies have proved that many social constructs and actions are determined by the display and the interpretation of specific non-verbal cues like gaze, postures, and body movements [2].

Affective computing is a domain within computer science pioneered by the research of Rosalind Picard [9] that studies the development of machines able to recognize and respond appropriately to emotions. Advances in this field have provided systems with the ability to detect and recognize different emotional states from various verbal and non-verbal behavioural cues like facial expressions, body postures, and physiological signals. This is a solid basis for further explorations and applications of these notions when moving from individual to group perspective on emotions.

In line with this perspective is the social signal analysis approach. According to related literature, social signals are "verbal and non-verbal behaviours that, either directly or indirectly, conveys information about social actions, social interactions and social emotions" [10]. Social signal processing (SSP) is indeed the area of research that investigates the automatic identification and analysis of social signals that humans display during their social interactions, adopting systems that sense multimodal behaviours and recognise the underlying socio-emotional patterns [15]. This research domain is mainly focused on the computational analysis of verbal and non-verbal behaviour, and in the modelling of these signals in significant social dynamics. More specifically, SSP research explores method for recognition and identification of human behaviour in naturalistic data, using algorithms and computing systems able to infer social signals from simple features extracted from the acoustical and visual scene analysis. This is made possible by the use of context-aware systems that can automatically sense and interpret cues to detect and analyse human behaviour. This domain draws on methods and theoretical approaches from different disciplines: on the one hand, cognitive science and ethnography can help to define and understand behavioural cues [10]; on the other hand, computer science, artificial intelligence and engineering allow the development of context-aware systems. Therefore, research on SSP can provide the basic components for recognising and modelling multimodal behaviour of users. In this research domain,

investigations have been conducted to explore computational models of many social interactions in small groups such as joint attention and interest, social relations and group cohesion [4], but applications and studies in the HCI domain are still scarce.

My research project involves social signal processing (SSP) as a method for capturing multi-modal human behaviours and recognizing the underlining socio-emotional patterns. The purpose is to investigate and design multi-user systems that adapt their interface, taking into account the users' social contexts.

In the research field of HCI, considerable research has been conducted in intelligent and context-aware interfaces but only a few works have tried to automatically measure social aspects of human-human interaction, using this information to inform multi-user interfaces and support group activities.

The research of Pentland et al. [8, 7] have used sociometric badges to monitor communication patterns and other social signals during team activity, reporting a graphical representation of group dynamics to the members themselves.

Similarly, DiMicco et al. [3] have investigated peripheral displays that visualize the amount of participation (in terms of speaking activity) of the member of a small group during a meeting, with the purpose of stimulating individual reflections on the on-going activity, thus harnessing social collaboration.

Considering physiological signals, Slovak et al. [14] have investigated how communicating signals as the heartbeat rate can implicitly change the experience of a social situation, and the behaviour displayed toward other people interacting.

Notably, Balaam et al. [1] showed how a multi-user public display can enhance interactional synchrony displaying subtle feedback about users' behaviour. The findings from this study suggest that social dynamics, like rapport, can be leveraged by machine to support group behaviour, without requiring a direct and exclusive interaction with the users. In fact, in this research, and similarly in other studies that investigated persuasive and empathic technologies [6], the focus is specifically on human-human interaction rather than on explicit human-machine interaction.

My research thus considers both psychological and computational models of group behaviour with an emphasis on the application of multi-user interfaces that integrate these models in order to sense and react to their surrounding social environment.

In the next sections, the research challenges, the approach and the progress to date are presented.

2 Research Challenges

Among the research challenges, the following three are of particular relevance for this project:

- *Defining the information that certain social cues can provide to multi-user systems*. While there are many social cues that have been described in the psychological literature, it is important to identify a basic subset of signals that promise to be relevant for the characterization of individual and group states, and for which automatic extraction seems feasible.
- *Exploring novel interaction techniques for multi-user systems that leverage social signal processing*. Exploring aspects related to indirect interaction,

including the mechanisms for communicate with the members of the group in a non-invasive but effective way.

- *Designing guidelines for applications of socially aware multi-user interfaces.* The project should provide some best practices and guidelines for the design and the evaluation of socially-aware interfaces.

3 Research Approach

The project adopts an integrated approach for designing and implementing socially-aware multi-user interfaces.

A. Sensing Individual and Group Behaviour

The first step of the research approach is to define which verbal and non-verbal behaviours can be reliably detected by the system. This first step comprises both theoretical and technical challenges. The theoretical challenge is to understand and define the relevant behaviours and social signals that should be addressed. To this end, research from non-verbal behaviour and communication can provide insight on which behavioural cues should be considered and on which information can be extracted from them. The technical challenges are identifiable in how to accurately detect the intended behaviour in real world scenarios, which might be characterized by high levels of noise and extended inter-subjects variability.

B. Modelling Behaviour and Social Signals

Once intended non-verbal cues are extracted from the social context, models of the raw data are required to highlight significant patterns in the data and to fuse cues that carry different amounts of information with respect to the social signals of interest. In this specific challenge, affective computing and social signal processing are relevant research fields dealing with models of multimodal behaviour of single users and groups.

C. Integrating Social Signals in Multi-User Interfaces

The last step of the research approach is to integrate the information obtained from the previous steps ("sensing" and "modelling") in order to provide the interface with meaningful information about the social context.

The interface should in fact actively support both human-computer and human-human interaction, with the goal to augment the interaction process. To this end, the interface should enact feedback mechanisms in the background to support the activity. Such mechanisms should be investigated to explore what feedback modalities and temporal characteristics are most effective.

Moreover, factors determining acceptability and effectiveness of the system should be assessed [5] using different methodologies: from questionnaires and interviews to observational methods using behavioural coding of audio-video recordings.

3   Progress To Date

During the first year of my PhD I have made some progress and presented my work in conferences related to HCI (e.g. CHI 2013 [11], C&T 2013 [12]). In the following paragraphs I will discuss the progress to date targeting the development of systems that sense and react to social context. Three studies will be thus presented as well as their design and evaluation processes.

The Engagement Study

In this study, focused more on the first two steps of the research approach, we investigated a novel approach in order to detect three affective states that characterize the activity of video game playing: engagement, stress and boredom. We developed a system that estimates the player's affective state through recognizing non-verbal behavioural cues using off-the-shelf hardware, such as a webcam, a keyboard and a mouse.

The system was tested in an empirical study with the aim to gather data and model the player's affective states. Facial expressions, head movements, keyboard and mouse activities were recorded while the participants played a single shooter video game. We used an adapted version of the experience sampling methodology to gather the ground truth and trained an SVM (Support Vector Machine) that recognized the affective states with an accuracy of 73%. The results showed that affective states as engagement, stress and boredom may be estimated by taking non-verbal behaviours as head movements and facial expressions into consideration.

The Mediaplayer: Sensing and Reacting to Users' Interest

In this second study the concept of engagement was considered in a broader perspective, taking into account the behaviour of groups of people and exploring socially-aware interfaces for public displays.

Public displays are encouraging multi-user systems for public and semi-public spaces because of their:

(i) *ubiquitous potential*, they can provide ubiquitous access to information; (ii) *social-aware potential*, they are media that can support both individual and group interactions in public and social contexts; (iii) *context-aware potential*, they are situated artefacts deeply embedded in their specific physical and social environment. In this work [11], we mostly focused on the two latter points, proposing a social-aware public display that provides different level of information accordingly to the perceived interest of the user(s) and the social context

We designed and developed a public display, named the Mediaplayer, which detects the audience interest and adapts the on-screen content accordingly. The Mediaplayer is a public display that can track the surrounding area by means of a 3D depth sensor (Microsoft's Kinect sensor). The depth information is used to detect audience's non-verbal cues, including users' spatial position, gaze orientation and

group configuration (Figure 1). This information is then used to estimate the level of attention and interest of the audience and to automatically adapt the interface providing more information on the screen if the users show more interest to the display.

In order to explore in an ecological setting how users interact with a context-aware public display, a field study was carried on. The Mediaplayer was deployed during a cultural event in a public space. We compared the adaptive system with a control condition, consisting of a non-adaptive system where the same content is offered without any adaptation. In the control condition, the information about the visual scene was collected as described above, but not used by the system. In order to minimize the influence of time, people affluence and light conditions on the results, the two conditions were counterbalanced during the study, switching automatically every 60 minutes. Roughly 350 people interacted with the Mediaplayer and ecological data about behaviour of individuals and groups in front of the screen was collected.

The field study revealed that the context-aware public display is more appealing for the audience than the control condition: more people approached the display while in the adaptive mode, people showed higher levels of interest and considered the experience more engaging as reported in a post-interaction questionnaire.

This study showed that behavioural-based measures are valuable data to inform and adapt a public display in a socially-aware way, improving users' engagement with the technology.


Fig. 1. User in front of the Mediaplayer (left image), the related depth image from the Kinect sensor (middle) and the final elaborated image with information about the user's distance and head orientation (right).

Agora2.0: A public display for Civic Participation

Agora2.0 is a platform for civic participation composed of two equally relevant features: an online system for voting ideas and an interactive public display deployed in a public space relevant to the community [12]. The system allows the public administration staff and the citizenry to post polls and gather opinions about local issues through questions that are answered online or onsite. Agora2.0 was evaluated in a realistic deployment in a public setting, where the system was used by actual citizens and their public administration (Figure 2).

The aim of this project was to encourage civic engagement bridging a virtual space for public deliberation with a physical space typically used by the community to discuss local issues. Agora2.0 is still a research-in-progress: nevertheless, its deployment showed some relevant points that can be considered when designing and deploying a big display in a public space. The next step in this direction would be to provide a system like Agora2.0 with the adaptability of the Mediaplayer in order to combine the potentialities of the two systems. Socially-aware public displays could have the potential to attract and maintain the audience attention to the system, promoting users' engagement and participation to the platform. This scenario opens also opportunities for exploring new interaction techniques for public systems. These techniques can be either implicit or explicit, and they should support collaboration and cooperation among users.
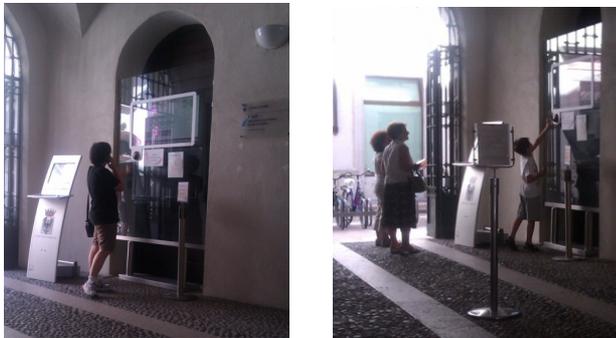
Fig. 2. User (left) and a group of users (right) interacting with Agora2.0

Future Steps: the Conversation Table

Future steps in this project will involve the exploration of a socially-aware system for collocated group activities taking place around a table. In an on-going study, the focus shifts from a public situation to a more semi-public context, targeting team activities. This scenario provides some advantages on the technical side and an optimal setting for controlled studies.

The conversation table is a socially aware system enhanced with sensors and peripheral displays that support the group communication activity. The system will be equipped with Kinect sensors to get information about non-verbal behaviour of group's members. The conversation table will continuously monitors the group social dynamics and uses this knowledge to plan and deploy strategies to support and influence the group behaviour through peripheral displays. The system should follow the three main steps of the research approach:

- Recording natural individual and group interaction data detecting multimodal behavioural cues;
- Learning models from these features to detect relevant social signals (e.g. dominance, equitable communication, subgroups);
- Generating appropriate feedbacks in real-time using the outputs of the models to influence the group activity.

Each of these steps presents challenges:

- The multimodal features extracted in real-time from the recording signals (including vocal activity, gaze orientation, and head movements) need to be identified and modelled in meaningful social signals such as group attention, mood and cohesion.
- By combining these multimodal features and interpretations, the system needs to reason about, plan and realize the actions to perform in response to a series of constraints (e.g. correct timing, deliver the appropriate stimuli to the specific target). To this end, continuous perception and interpretation of the group dynamics is required to keep the system effective.
- The development and evaluation of such real-time continuous systems also requires careful attention to the research methodology in terms of experimental design and evaluation methods.

4 Expected Contributions

I believe that attending the CHItaly 2013 Doctoral Consortium will give me a chance to share my work and vision on socially-aware systems with other students as well as with senior researchers in the consortium. The participation to the DC will surely provide me with an opportunity to get fresh perspectives on my project from researchers with different backgrounds, hopefully getting guidance on future directions of my project.

Short Bio

I am a psychologist, currently a second year Phd candidate in cognitive science at the University of Trento (Italy) and at FBK - Fondazione Bruno Kessler, working under the supervision of Dr. Massimo Zancanaro.

I completed my MSc degree in experimental psychology and cognitive sciences at the University of Padova (Italy) in 2011. During my studies, I worked at the Human Technology Lab (HTLab) at the University of Padova (Italy) and I completed an internship at the Helsinki Institute of Information Technology (HIIT), Finland.

My current research interests are oriented towards human-computer interaction (HCI) and co-located multi-user systems (e.g. interactive tabletops, public displays and proximity-based technologies). The key point of my PhD project is to explore how to design socially-aware multi-user interfaces for supporting co-located interaction, applying theoretical frameworks and tools from cognitive science and machine learning, such as social signal processing. As part of my research, I am investigating the role of non-verbal behaviour in human-computer as well in human-human interactions to investigate how multi-user systems can leverage this information.

References

1. Balaam, M., Fitzpatrick, G., Good, J., & Harris, E. Enhancing interactional synchrony with an ambient display. In *Proc. CHI 2011*, ACM Press (2011), 867-876.

2. Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. Automaticity in social-cognitive processes. *Trends in cognitive sciences 16*, 12 (2012), 593–605.

3. DiMicco, J. M., Hollenbach, K. J., Pandolfo, A., & Bender, W. The impact of increased awareness while face-to-face. *Human–Computer Interaction 22,* 1-2 (2007), 47-96.

4. Gatica-Perez, D. Automatic non-verbal analysis of social interaction in small groups: A review. *Image and Vision Computing 27,* 12 (2009), 1775-1787.

5. Höök, K. Steps to take before intelligent user interfaces become real. *Interacting with computers 12*, 4 (2000), 409-426.

6. Janssen, J. H. A three-component framework for empathic technologies to augment human interaction. *Journal on Multimodal User Interfaces 6*, 3-4 (2012), 143-161.

7. Kim, T., Hinds, P., & Pentland, A. S. Awareness as an antidote to distance: Making distributed groups cooperative and consistent. In *Proc. of CSCW 2012*, ACM Press (2012), 1237-1246

8. Pentland, A., with T. Heibeck. Honest Signals: How They Shape Our World, MIT Press, 2008.

9. Picard, R. W. Affective computing. MIT press, 2000.

10. Poggi, I., & D'Errico, F. Social signals: A psychological perspective. In Salah, A.A., Gevers, T. (eds.) *Computer Analysis of Human Behavior*. Springer (2011), 185-225.

11. Schiavo, G., Mencarini, E., Vovard, K., & Zancanaro, M. Sensing and reacting to users' interest: an adaptive public display. *In Proc. EA CHI 2013*, ACM Press (2013), 1545-1550.

12. Schiavo, G., Milano, M., Saldivar, J., Nasir., T., Zancanaro, M., & Convertino, G. Agora 2.0: Enhancing Civic Participation through a Public Display. *In Proc. of C&T'13*, ACM Press (2013).

13. Scott, S.D., Grant, K.D., & Mandryk, R.L. System Guidelines for Co-located, Collaborative Work on a Tabletop Display. In *Proc. of ECSCW'03*. Springer (2003).

14. Slovák, P., Janssen, J., & Fitzpatrick, G. Understanding heart rate sharing: towards unpacking physiosocial space (2012). In *Proc. CHI 2012*, ACM Press (2012), 859-868.

15. Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schröder, M. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing 3*, 1 (2012), 69-87.

16. Yuill, N., & Rogers, Y. Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces (2012). *ACM Transactions on Computer-Human Interaction* (TOCHI), 19(1), 1-25.