Mobile Social Signal Processing: Vision and Research Issues

Alessandro Vinciarelli
Dept. of Computing Science
University of Glasgow
G12 8QQ Glasgow (UK)
Idiap Research Institute
CP592 - Martigny (CH)
vincia@dcs.gla.ac.uk

Roderick Murray-Smith
Dept. of Computing Science
University of Glasgow
G12 8QQ Glasgow (UK)
rod@dcs.gla.ac.uk

Hervé Bourlard
Idiap Research Institute
CP592 - 1920 Martigny (CH)
EPFL - 1015 Lausanne (CH)
bourlard@idiap.ch

ABSTRACT

This paper introduces the First International Workshop on Mobile Social Signal Processing (SSP). The Workshop aims at bringing together the Mobile HCI and Social Signal Processing research communities. The former investigates approaches for effective interaction with mobile and wearable devices, while the latter focuses on modeling, analysis and synthesis of nonverbal behavior in human–human and human–machine interactions. While dealing with similar problems, the two domains have different goals and methodologies. However, mutual exchange of expertise is likely to raise new research questions as well as to improve approaches in both domains. After providing a brief survey of Mobile HCI and SSP, the paper introduces general aspects of the workshop (including topics, keynote speakers and dissemination means).

Categories and Subject Descriptors

A.0 [**General**]: [Conference Proceedings]; J.4 [**Social and Behavioral Sciences**]: [Psychology, Sociology]

General Terms

Human Factors

Keywords

Mobile HCI, Social Signal Processing

1. INTRODUCTION

Conversation is the "primordial site of human sociality and social life" [24]. Thus, it is not surprising to observe that mobile phones, allowing one to talk with virtually anybody at virtually any moment, have pervaded our everyday life more quickly and deeply than any previous technology. However, while becoming a preeminent form of social interaction, mobile phone conversations have been the subject of limited investigation from both a psychological and technological points of view [3, 12]. The reason is not only that the diffusion of mobile phones is a relatively recent phenomenon, but also that phone conversations have traditionally been considered nothing more than particular cases of face-to-face conversations, characterized by speech being the only information at disposition, in contrast with actual face-to-face conversations where humans are known to exchange not only words, but also a wide spectrum of nonverbal behavioral cues accounting for social, affective and relational phenomena [10, 19, 22].

This leaves open a major gap in the moment where two important phenomena take place in the scientific and technological landscape. The first is that nowadays standard mobile phones contain a large number of sensors (e.g., GPS, accelerometers, magnetometers, capacitive touch and, in the near future, pressure sensing). Also the increasing processing power and the potential to use server-side processing allows the use of algorithms previously considered only possible on powerful PCs, capturing, with unprecedented depth and precision, context and behavior of their users (e.g., position, movement, hand grip behavior, proximity to social network members, gait type, auditory context). This behavior can also potentially be compared with large numbers of other users, to categorise the style of interaction [20]. The second is that automatic analysis, synthesis and understanding of verbal and nonverbal communication, typically captured with multiple sensors, is one of the hottest topics in the computing community. This applies in particular to Social Signal Processing (SSP), the new, emerging domain aimed at bringing social intelligence in machines) [28, 29].

The rest of the paper is organized as follows: Sections 2 and 3 provide brief surveys of SSP and Mobile HCI, respectively, Section 4 introduces goals and research questions addressed by the workshop, Section 5 describes the dissemination means that will be used to ensure proper diffusion of the workshop content, and Section 6 briefly introduces the keynote speakers.

2. SOCIAL SIGNAL PROCESSING

SSP is the new, emerging domain aimed at bringing social intelligence in computers through analysis and synthesis of nonverbal behavior in social interaction [28, 29]. The rationale is that nonverbal behavioral cues (e.g., facial expressions, postures, gestures, vocalizations, etc.) in human-human communication convey social signals, i.e. the mutual relational attitudes exchanged between people involved in interactions (e.g., empathy, hostility, agreement, dominance, etc.) [2, 1]. In other words, nonverbal behavioral cues are used as a physical, machine detectable evidence of social

MobileHCI'10, September 7–10, 2010, Lisbon, Portugal.
ACM 978-1-60558-835-3/10/09.

phenomena non-otherwise accessible to human perception and machine sensing [27].

The SSP approach is supported by several decades of research in social psychology showing that nonverbal communication is the channel through which we perceive social aspects of our interactions [10, 19, 22]. While no dictionary of nonverbal behavioral cues can be built, i.e. it is not possible to associate each cue with a precise social signal (e.g. a smile with sympathy or a frown with hostility), still the attitude people infer from a given constellation of cues tends to be consistent and coherent at least for a given culture or social context. The same applies, symmetrically, to the cues people display to others when they want to communicate their attitude [21, 17].

The above means that nonverbal cues, when detected with cameras, microphones and any other suitable sensor, can be used to infer social signals via probabilistic approaches. Furthermore, it means that nonverbal cues, when synthesized via artificial faces, voices or other forms of embodied agents can elicit the same kind of social perceptions they elicit when they are displayed by humans.

The main technologies involved in SSP are as follows:

- *Data capture.* The sensors used for capturing interactions determine the kind of analysis that it is possible to perform (if only microphones are used, no facial expression analysis can be performed) and influence the ecological validity of the data (if sensors are intrusive, people might be less spontaneous). Possible experimental apparatuses range from the webcam available on an ordinary laptop to a fully equipped smart meeting room [7], depending on the particular application targeted.

- *Person detection.* Technologies like speaker diarization [26] and face detection [30] allow the identification of those parts of the data, typically audio and video recordings, that correspond to a person. This step is important because it helps to associate nonverbal behavioral cues extracted from a given part of the data with a given person.

- *Nonverbal behavioral cues detection and synthesis.* Technologies like facial expression analysis [31], gesture recognition [14], prosody analysis and synthesis [25], etc., allow one to automatically detect nonverbal behavioral cues in signals. In synthesis scenarios, nonverbal behavioral cues are artificially generated via expressive speech synthesis, artificial faces, Embodied Conversational Agents, etc [8].

- *Social Behavior Understanding* In analysis scenarios, nonverbal behavioral cues as detected in data are used to infer social signals, i.e. relational attitudes, using probabilistic approaches such as Graphical Models, Neural Networks, Support Vector Machines, etc [6, 11].

The social phenomena most commonly addressed so far in SSP are roles, personality, dominance, agreement and disagreement, emotions, politeness, etc.

3. MOBILE HCI

Mobile Human–Computer Interaction (HCI) is the study of how humans interact with mobile devices, and how that empowers them and allows them to interact with their environment and social network [13]. Modern mobile devices have long since moved beyond basic voice and text communications, and are now equipped with significant sensing and processing ability [4, 5]. The small size of these devices, the reduced input capability and the diverse usage situations are key challenges in the design of mobile systems. Evaluation of such systems in realistic settings is difficult, especially when the impact of the design on social aspects is an important factor [18, 9].

This mobility and diverse usage does, however, provide interesting new research opportunities to measure and influence social interactions in ways that would have been extremely difficult only a few years ago. Modern mobile devices can sense movement, muscle tremor, location, proximity of other devices, can sample audio and video signals and magnetic field disturbances. This gives us opportunities to record in greater detail than ever before human activity, including that of social interactions [16, 15]. It also allows us to design experiments which can stimulate users in specific contexts, allowing a trade-off between realistic conditions and experimental control [23]. Potential benefits for Mobile HCI research from the SSP community include the use of techniques to help infer emotional consequences for users of different mobile interaction designs.

4. GOALS AND RESEARCH QUESTIONS

This workshop gathers, for the first time, researchers active in the communities that have dealt, so far separately, with the two phenomena described earlier:

- Social Signal Processing, including analysis, synthesis and modeling of verbal and nonverbal behavior in human-human communication.

- Mobile Human Computer Interaction.

This is expected to extend the investigation area of the two domains and identify a number of research questions that not only promise to bring significant novelty in both SSP and Mobile HCI, but also require the application of knowledge from both domains to be effectively investigated:

- Is it possible to integrate the input of mobile phone sensors in current approaches for automatic analysis of social phenomena in conversations? If yes, does this allow a better understanding of human-human communication in general and mobile phone conversations in particular? Are there differences between face-to-face and phone mediated conversations? If yes, what are the differences? Do the methodologies applied for analyzing conversations need to be changed?

- Does context influence the communication behavior of people talking on the phone? If yes, what are the most important contextual factors? Does the context misalignment between people talking on the phone influences negatively the quality of conversation? If yes, is it possible to smooth the effect? Is it possible to detect the context? Are there contextual factors that are more important than others? If yes, which ones?

- Does the transmission of nonverbal behavioral cues, so important in face-to-face communication, improve phone conversation experience? Is it possible to make

speakers aware of the context of their interlocutors? If yes, does this improve overall conversation experience?

- Does a better understanding of communication behavior influence the design of mobile phones? Can mobile phones enhance and support human-human communication? Are mobile phones sufficiently human-centered? If not, what do they miss?
- Can we evaluate how the use of a mobile phone affects the key social interaction variables of "trust" and "competence" evaluation?. What effects do the reduced frequency range, delays during interaction and effect of external disturbances have on perception of voice qualities? Can we deliberately alter the channel properties to transform perception of trust and competence?
- Can we create metrics which help us evaluate the effect on social interaction of augmenting the voice channel with other feedback channels?
- Can we create non-vocal, but embodied interaction techniques which are appropriate for mobile use (e.g., mapping gestures to sounds or haptic feedback)?
- What would be the ethical issues related to the everyday use of in-hand, automated social signal analysis?

5. DISSEMINATION

The workshop will be funded by the Social Signal Processing Network (`http://www.sspnet.eu`), a European network of excellence aimed at fostering SSP research in Europe. This will help to increase the impact of the workshop and will ensure that awareness about Mobile HCI issues will be diffused throughout a large community of researchers working on the analysis of human behavior. Correspondingly, the presence of the SSPNet at Mobile HCI will ensure that information about SSP, and more in general human behavior analysis, will be disseminated in the Mobile HCI community. The workshop presentations will be recorded and they will be posted on the SSPNet Virtual Learning Center[1], a repository of lectures and presentations visited, on average, by 30 people per day. This will guarantee the maximization of the workshop impact in the months following the event. Furthermore, the proceedings of the workshop will be published as a volume of the Springer *Lecture Notes in Computer Science*.[2]

6. KEYNOTE SPEAKERS AND PANEL

The workshop includes two keynote talks aimed at covering some of the most important aspects of the workshop, the analysis of human-human communication in conversation and the analysis of large-scale group behavior via mobile phones. The first aspect will be presented by Jens Allwood[3], from the University of Göteborg. Prof. Allwood will illustrate his view on human-human communication as it has been developed during 4 decades of investigations dedicated to linguistics, pragmatics, semantics, intercultural communication, etc. Furthermore, Professor Allwood will assess the role of Mobile HCI technologies as an instrument for further investigation of conversation, possibly under a perspective not explored so far.

The second aspect will be presented by Juha Laurila from the Nokia Research Centre in Lausanne.[4] The main mission of the center is rich context modeling: "*Rich context is characterized by the use of a wide range of sensed and historic information, aggregated into a coherent model of a user's state and surroundings; including things like their location, motion, weather, connectivity options, and proximity to others. This data and its analysis form the backbone for a new class of services in areas like weather, traffic, wellness, or entertainment.*"[5] In particular, Dr Laurila will show how data captured with sensors embedded in mobile phones help to model the behavior of large groups of individuals (their acquaintances, their habits, etc.).

The workshop will include as well a panel where the integration between SSP and Mobile HCI will be discussed not only by attendees and keynote speakers, but also by Scott Jenson (director of mobile search at Google) and by the organizers of the workshop, Rod Murray-Smith (University of Glasgow), and Alessandro Vinciarelli (University of Glasgow and Idiap Research Institute), coordinator of the Social Signal Processing Network (`http://www.sspnet.eu`).

7. CONCLUSION

As mobile phones are one of the most important instruments of our social life, the cross-pollination between Mobile HCI and Social Signal Processing is likely to foster on one hand a better understanding of the way people interact via phone and, on the other hand, of how to make mobile phones more social interaction centered. This workshop is a first step towards better integration between the two domains and identification of new, promising research directions.

8. ACKNOWLEDGMENTS

This work has been supported in part by the European Community 7^{th} Framework Program via the Network of Excellence Social Signal Processing Network (SSPNet, Grant Agreement 231287), in part by the Swiss National Science Foundation via the National Centre of Competence in Research IM2 (Interactive Multimodal Information Management), and in part by the ESRC/MRC *Social Interaction* grant RES-060-25-0010, and EPSRC grant EP/E042740/1.

9. REFERENCES

[1] N. Ambady, F. Bernieri, and J. Richeson. Towards a histology of social behavior: judgmental accuracy from thin slices of behavior. In M. Zanna, editor, *Advances in Experimental Social Psychology*, pages 201–272. 2000.

[2] N. Ambady and R. Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychological Bulletin*, 111(2):256–274, 1992.

[1]See `http://sspnet.eu/virtual-learning-centre/`

[2]See `http://www.springer.com/computer/lncs?SGWID=0-164-0-0-0`

[3]See `http://www.ling.gu.se/~jens/`

[4]See `http://research.nokia.com/research/rich_context_modeling`

[5]This text has been quoted from the home page of the Nokia Research Centre in Lausanne.

[3] I. Arminen and A. Weilenmann. Mobile presence and intimacy–Reshaping social actions in mobile contextual configuration. *Journal of Pragmatics*, 41(10):1905–1923, 2009.

[4] V. Bellotti, M. Back, W. K. Edwards, R. E. Grinter, A. Henderson, and C. Lopes. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, CHI '02, pages 415–422, New York, NY, USA, 2002. ACM.

[5] S. Benford, H. Schnädelbach, B. K., R. Anastasi, C. Greenhalgh, T. Rodden, J. Green, A. Ghali, T. Pridmore, B. Gaver, A. Boucher, B. Walker, S. Pennington, A. Schmidt, H. Gellersen, and A. Steed. Expected, sensed, and desired: A framework for designing sensing-based interaction. *ACM Trans. Comput.-Hum. Interact.*, 12:3–30, March 2005.

[6] C. Bishop et al. *Pattern recognition and machine learning*. Springer, 2006.

[7] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, et al. The AMI meeting corpus: A pre-announcement. In *Lecture notes in computer science*, volume 3869, pages 28–39, 2005.

[8] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, page 527. ACM, 1999.

[9] A. Crossan, R. Murray-Smith, S. Brewster, and B. Musizza. Instrumented Usability Analysis for Mobile Devices. *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, 2007.

[10] M. Knapp and J. Hall. *Nonverbal Communication in Human Interaction*. Harcourt Brace College Publishers, 1972.

[11] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[12] C. Licoppe. Recognizing mutual "proximity" at a distance: Weaving together mobility, sociality and technology. *Journal of Pragmatics*, 41(10):1924–1937, 2009.

[13] J. Lumsden. *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*. IGI Global, 2008.

[14] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3):311–324, 2007.

[15] A. Morrison, M. Bell, and M. Chalmers. Visualisation of spectator activity at stadium events. In *Information Visualisation, 2009 13th International Conference*, pages 219–226, 2009.

[16] R. Murray-Smith, A. Ramsay, S. Garrod, M. Jackson, and B. Musizza. Gait alignment in mobile phone conversations. In *Proc. of Mobile HCI 07*, pages 214–221, 2007.

[17] C. Nass and S. Brave. *Wired for speech: How voice activates and advances the Human-Computer relationship*. The MIT Press, 2005.

[18] A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 919–928, New York, NY, USA, 2005. ACM.

[19] I. Poggi. *Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication*. Weidler Buchverlag Berlin, 2007.

[20] M. Raento, A. Oulasvirta, and N. Eagle. Smartphones: an emerging tool for social scientists. *Sociological Methods & Research*, 37(3):426, 2009.

[21] B. Reeves and C. Nass. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press New York, NY, USA, New York (USA), 1996.

[22] V. Richmond and J. McCroskey. *Nonverbal Behaviors in interpersonal relations*. Allyn and Bacon, 1995.

[23] V. Roto, A. Oulasvirta, T. Haikarainen, J. Kuorelahti, H. Lehmuskallio, and T. Nyyssonen. Examining mobile phone use in the wild with quasi-experimentation. *Helsinki Institute for Information Technology (HIIT), Technical Report*, 1, 2004.

[24] E. Schegloff. Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, 50(2):101–114, 1987.

[25] M. Schröder. Expressive speech synthesis: Past, present, and possible futures. In *Affective Information Processing*, pages 111–126. Springer, 2008.

[26] S. Tranter and D. Reynolds. An overview of automatic speaker diarization systems. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1557–1565, 2006.

[27] A. Vinciarelli. Capturing order in social interactions. *IEEE Signal Processing Magazine*, 26(5):133–137, 2009.

[28] A. Vinciarelli, M. Pantic, and H. Bourlard. Social Signal Processing: Survey of an emerging domain. *Image and Vision Computing Journal*, 27(12):1743–1759, 2009.

[29] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland. Social Signal Processing: State-of-the-art and future perspectives of an emerging domain. In *Proceedings of the ACM International Conference on Multimedia*, pages 1061–1070, 2008.

[30] M. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.

[31] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: audio, visual and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.