Mobile Phones and Social Signal Processing for Analysis and Understanding of Dyadic Conversations

Alessandro Vinciarelli

University of Glasgow
Sir A.Williams Building, G12 8QQ Glasgow, UK
Idiap Research Institute
CP592, 1920 Martigny, Switzerland

Abstract. Social Signal Processing is the domain aimed at bridging the social intelligence gap betweeen humans and machines via modeling, analysis and synthesis of nonverbal behavior in social interactions. One of the main challenges of the domain is to sense unobtrusively the behavior of social interaction participants, one of the key conditions to preserve the spontaneity and naturalness of the interactions under exam. In this respect, mobile devices offer a major opportunity because they are equipped with a wide array of sensors that, while capturing the behavior of their users with an unprecedented depth, are still invisible. This is particularly important because mobile devices are part of the everyday life of a large number of individuals and, hence, they can be used to investigate and sense natural and spontaneous scenarios.

1 Introduction

The number of mobile phone users in the world has been recently estimated to be around 3.5 billions, more than 50% of the current world population [13]. The diffusion changes significantly depending on the country: while in Papua New Guinea only 0.44 percent of the population subscribes to a mobile telephony service, the same figure is 154 percent in the case of Luxemburg (more than one phone per person). In the developed countries (in particular Europe and the Americas) virtually everybody holds a mobile subscription, but the penetration is high and growing in the developing world as well (300 millions new users are expected in India in the next few years) [12]. The same variability across countries can be observed for what concens the amount of time spent on the phone, ranging between 22 and 800 minutes per month [13].

A mere 15 years ago it was hard to predict the impressive figures above. Even in a country like Italy, where the density of mobile phones is today among the highest in the world, sociologists used to observe prevailing negative feelings in surveys about the acceptance of mobile technologies [11]. The main change since then is that mobile phones are no longer an instrument for professional or emergency calls only (as it used to be at the beginning of their diffusion), but one of the main channels through which we get involved in social interactions. Mobile phones provide the possibility of starting a conversation, the "primordial site of human sociality and social life" [27], at virtually every moment of the day, almost independently of where we are and what we do. Furthermore, mobile phones extend our opportunities for social contacts well beyond conversations to include the exchange of text messages (roughly 2×10^5 SMS per second have been exchanged worldwide in 2010 [12]) as well as the access to popular social media (e.g., Facebook, LinkedIn, etc.). In this respect, mobile phones seem to be a key support for our social life and an ideal response to the needs of the "social animal" [35].

Thus, it is not surprising to observe that both social scientists and computing researchers have identified mobile phones or, more generally, mobile and wearable devices as an instrument to access social life with at an unprecedented depth and scale [23]. This applies in particular to naturalistic settings difficult to observe in the laboratory, whether this means to identify daily routines in the life of social groups [9], to look for personality traces in everyday speaking behavior [20], or to sense the overall behavior of an organization [22], just to name a few examples. In all cases above, mobile devices have been used as an unobtrusive, but ubiquitous and pervasive sensor that can be carried without effort and, to a certain extent, without awareness in the most natural settings of our everyday life (see [21] for an example of how unobtrusiveness is assessed).

In such a perspective, mobile phones have a major advantage with respect to other wearable devices because they are an everyday object and are carried spontaneously, in contrast with any other device designed for sensing and collecting data. Furthermore, standard mobile phones are now equipped with an increasingly wider range of sensors (magnetometers, GPS, accelerometers, etc.) that reduce the sensing capability gap with respect to devices explicitly designed for scientific experiments.

For the reasons above, mobile phones appear to be particularly suitable for research in Social Signal Processing (SSP), the domain aimed at automatic understanding of social interactions via modeling, analysis and synthesis of nonverbal behavior (see Section 2 for more details) [34]. In fact, the sensors of a standard mobile phone allow one to capture not only nonverbal speech aspects (prosody, vocalizations, pauses, etc.), but also non verbal cues related to body movement (via accelerometers, gyroscopes and magnetometers) that are typically difficult to capture otherwise in ecologically valid settings, but still carry socially relevant information [15].

In particular, SSP appears to be one of the most suitable paradigms to develop approaches for automatic analysis and understanding of dyadic conversations, an interaction scenario that, despite its primacy and frequency (phones are used most of the times to call even though the younger generations tend to favor the use of SMS), has been so far neglected from both a technological and psychological points of view. As a result, mobile phones could reduce the social intelligence gap with respect to their users [35], support the effectiveness of task oriented calls (e.g., moderating people talking too much or deflating conflicts),

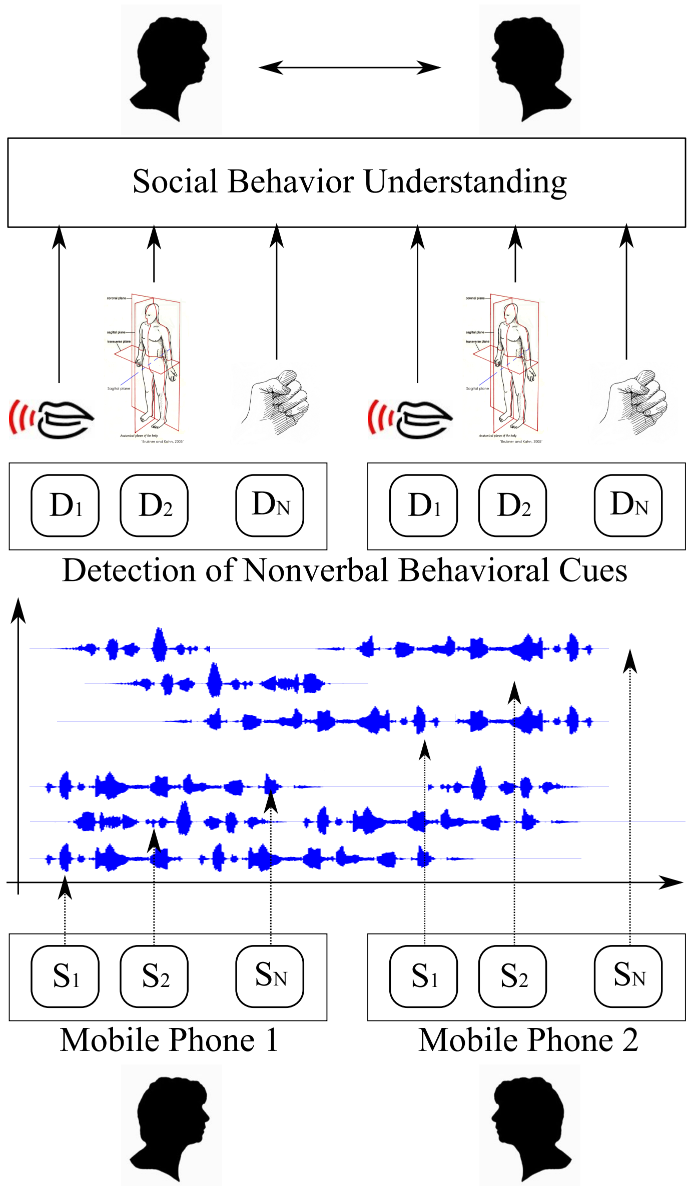

Fig. 1. Overall scheme of an SSP approach applied to mobile phone conversations. The signals captured with the sensors of the two phones $(S_1, \ldots, S_N)$ are fed to non-verbal cues detectors $(D_1, \ldots, D_N)$. The output of these latter is then automatically interpreted to identify the social signals being exchanged between the speakers.

activate services appropriate to the social context (e.g., by canceling background noise in case of formal conversations), etc.

The rest of this paper shows how mobile phones can be used to perform SSP research (Section 2) and what are the main challenges facing the application of SSP in mobile environments (Section 3). The final Section 4 draws some conclusions.

2 Social Signal Processing and Mobile Phones

Social interactions are accompanied by a wide spectrum of nonverbal behavioral cues (facial expression, vocalizations, gestures, postures, etc.) [15, 24] that add layers of meaning, typically related to social and affective aspects of an interaction, to the words being said [36]. While our attention tends to focus on what people say, a number of cognitve processes (typically taking place outside conscious awareness) interpret nonverbal behavior of others in terms of socially relevant cues, including values, beliefs, emotions, goals, intentions, etc. [30]. These processes take place independently of any actual need or will for them taking place, but they influence to a large, sometimes dominant extent our social behavior, especially in the earliest stages of an interaction [31].

Social Signal Processing (SSP) relies on the phenomenon above and proposes to use nonverbal communication as a physical, machine detectable evidence of social signals, the perceivable stimuli (including nonverbal behavioral cues) that are produced during social interactions and "[...] *play a part in the formation and adjustment of relationships and interactions* [...] *or provide information about the agents; and that can be addressed by technologies of signal processing and synthesis*"[1] (see [34, 35] for an extensive survey of the domain). The choice of nonverbal behavior as a privileged cue for understanding social phenomena results from several decades of investigations in psychology, anthropology and other human sciences (see [15, 24] for extenisve monographies about the subject): "*thin slices of behavior*" [2], short samples of nonverbal behavior collected during a social interaction, appear to be sufficient to provide accurate social judgments in a large number of situations [1, 6].

Figure 1 shows how the SSP paradigm can be applied in the case of mobile phone conversations. When two people are involved in a phone conversation, they naturally make use of a number of sensors embedded nowadays in a large number of standard phones available on the market. Besides microphones, without which phone calls would be obviously impossible, the most common sensors available on a phone include accelerometers, magnetometers, Global Positioning Systems, gyroscopes, etc. Thus, each of the phones can be thought of as an array of sensors $(S_1, \dots, S_N)$ capturing signals that, potentially, carry information about the nonverbal behavior of their users.

[1] The quote comes from the "Belfast Declaration", the document issued by the Social Signal Processing Network (European Network of Excellence on SSP). The document is available for download at the following link: `http://sspnet.eu/about/`.

The main difference with respect to sensing approaches commonly applied in SSP is the lack of cameras, essential to capture fundamental nonverbal cues such as facial expressions and gaze behavior. However, this should not represent a major problem for two main reasons: the first is that approaches based on vocal behavior (accessible via the phone microphones) tend to achieve, at least in the SSP works presented so far in the literature, satisfactory performances [35]. The second is that the lack of visual information about interlocutors corresponds to the actual condition of people talking on the phone. Hence, the lack of cameras portraying the interactants simply reflects the condition of the users. Furthermore, many phones allow one to perform video-calls and such an opportunity, not particulary exploited today, might extend the analysis of face and gaze behavior to mobile phone based interaction scenarios.

Once the signals have been captured, it is possible to detect nonverbal behavioral cues using different approaches (identified as D_i in Figure 1) depending on the particular sensor. The extraction of vocal cues from speech signals is the subject of a large number of works in the literature, in particular when it comes to emotion recognition (see [4, 26, 28] for psychological research and [32] for technological approaches), inference of social information from turn-organization (see [33] for an introduction to the problem and [35] for an extensive survey), and analysis of traits (see [29] for an exhaustive description of cues currently extracted from speech).

The other sensors available on the phone (accelerometers, magnetometers, etc.) have not been used extensively in SSP, at least for what concerns face-to-face scenarios. SSP works aimed at the analysis of large social networks (see, e.g., [9, 22]) generally make use of proximity detectors (e.g., bluetooth and RFID) to identify direct interactions between people, but do not consider accelerometers. In contrast, accelerometers have been used extensively in the ubiquitous computing community, especially to recognize the "context" (see [7] for a definition of what it is meant by this) and the actions being performed by users (see [10, 16] for extensive surveys). Furthermore, accelerometers have been used to improve interaction with machines (e.g., in a gesture based design system [14]), or computer mediated communication (e.g., in a system aimed at sharing information about travels [25]).

3 Main Challenges

From a technological point of view, Mobile SSP faces the same challenges as any other SSP investigation (see [35] for an extensive survey), including fusion of multiple modalities where behavioral cues take place at different time-scales, modeling of annotation variability in judgmental studies involving multiple raters, definition of continuous rather than categorical descriptors of social and psychological phenomena, etc. However, two main challeges are specific of the application of SSP in mobile conversations, namely the modeling of principles and laws underlying phone mediated conversations and the redefinition of the concept of privacy. The rest of this section will focus on these.

Phone conversations tend to be considered as a specific case of face-to-face interaction where visual cues are not available. However, such a view does not consider that talking through a phone does not simply eliminate the visual channel, but it constrains the array of cues that people can use to convey social meaning. Therefore, communication practices must undergo significant changes to accomplish simple social goals like, e.g., the communication of immediacy [3] and proximity [18]. Furthermore, people participating in mobile phone conversations are often immersed in contexts where they are interacting with other, co-located individuals and this induces further changes in the social needs to be addressed [8, 19]. Taking into account this type of issues is a crucial step towards the improvement of Mobile SSP technologies.

In a context where personal data is considered "*the new oil of the internet and the new currency of the digital world*" [37], mobile SSP can attract significant interest. On one hand, the analysis of nonverbal communication respects the privacy because it does not take into account what people say. On the other hand, recent work on social media shows that privacy protected information can be effectively inferred from publicly available cues [17]. In other words, the very concept of privacy should be redesigned in light of mobile SSP progresses. This is a major issue that can make the difference between SSP technologies being accepted or not by the users.

4 Conclusions

This article has outlined research opportunities and challenges that can emerge from the cross-pollination between Social Signal Processing - the domain aimed at modelling, analysis and synthesis of nonverbal behavior in social interactions - and mobile Human-Computer Interaction. The increasingly wide array of sensors embedded on standard mobile devices is transforming these latter in a laboratory for human behavior analysis [23]. However, technologies capable of analyzing social and psychological phenomena at the level of one-to-one conversations might become a significant threat for the privacy of people.

The identification of a correct tradeoff between the two conflicting phenomena above is beyond the scope of this article and, in any case, it requires a large societal debate [5]. From a strictly scientific point of view, the analysis of mobile phone conversations in a laboratory context, where subjects are aware of being recorded, promises to bring significant progress in domains like understanding of human behavior, development of new sensors, and improvement of automatic behavior analysis techniques. In other words, SSP can contribute to make mobile phones, one of the main infrastructures of our social life, more socially intelligent.

References

1. N Ambady, F Bernieri, and J Richeson. Towards a histology of social behavior: judgmental accuracy from thin slices of behavior. In M P 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.
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. Jo-Anne Bachorowski. Vocal Expression and Perception of Emotion. *Current Directions in Psychological Science*, 8(2):53–57, 1999.
5. Z. Bauman and D. Lyon. *Liquid Surveillance*. Polity Press, 2013.
6. J R Curhan and A Pentland. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92(3):802–811, 2007.
7. P. Dourish. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1):19–30, 2004.
8. P. Dourish and G. Bell. *Divining a digital future: mess and mythology in ubiquitous computing*. MIT Press, 2011.
9. N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2005.
10. D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso. Preprocessing techniques for context recognition from accelerometers data. *Personal and Ubiquitous Computing*, 14(7):645–662, 2010.
11. L. Fortunati and A.M. Manganelli. The social representation of communications. *Personal and Ubiquitous Computing*, 12(6):421–431, 2008.
12. ITU. The World in 2010: ICT Facts and Figures. Technical report, International Telecommunication Union.
13. K. Kalba. The Global Adoption and Diffusion of Mobile Phones. Technical Report December, Center for Information Policy Research Harvard University, 2008.
14. J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, and S. Di Marca. Accelerometer-based gesture control for a design environment. *Personal and Ubiquitous Computing*, 10(5):285–299, 2006.
15. M.L. Knapp and J.A. Hall. *Nonverbal Communication in Human Interaction*. Harcourt Brace College Publishers, 1972.
16. J.F. Knight, H. W. Bristow, S. Anastopoulou, C. Baber, A. Schwirtz, and T. N. Arvanitis. Uses of accelerometer data collected from a wearable system. *Personal and Ubiquitous Computing*, 11(2):117–132, 2007.
17. M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceeedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.
18. C. Licoppe. Recognizing mutual 'proximity' at a distance: Weaving together mobility, sociality and technology. *Journal of Pragmatics*, 41(10):1924–1937, 2009.
19. R. Ling. *New tech, new ties*. MIT Press, 2008.
20. M. R. Mehl, J. W. Pennebaker, D. M. Crow, J. Dabbs, and J. H. Price. The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments and Computers*, 33(4):517–523, 2001.
21. M.R. Mehl and S.E. Holleran. An Empirical Analysis of the Obtrusiveness of and Participants' Compliance with the Electronically Activated Recorder (EAR). *European Journal of Psychological Assessment*, 23(4):248–257, 2007.
22. D. Olguin Olguin, B. N Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: technology and methodology for automatically measuring

organizational behavior. *IEEE Transactions on Systems, Man and Cybernetics Part B*, 39(1):43–55, 2009.
23. M. Raento, A. Oulasvirta, and N. Eagle. Smartphones: An emerging tool for social scientists. *Sociological Methods & Research*, 37(3):426–454, 2009.
24. V.P. Richmond and J.C. McCroskey. *Nonverbal Behavior in Interpersonal Relations*. Allyn and Bacon, 2000.
25. S. Robinson, P. Eslambolchilar, and M. Jones. Exploring casual point-and-tilt interactions for mobile geo-blogging. *Personal and Ubiquitous Computing*, 14(4):363–379, 2010.
26. J A Russell, J A Bachorowski, and J M Fernandez-Dols. Facial and Vocal Expressions of Emotion. *Annual Reviews in Psychology*, 54(1):329–349, 2003.
27. E. A. Schegloff. Analyzing Single Episodes of Interaction: An Exercise in Conversation Analysis. *Social Psychology Quarterly*, 50(2):101–114, 1987.
28. K R Scherer. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2):227–256, 2003.
29. B. Schuller. *Voice and Speech Analysis in Search of States and Traits*, pages 233–258. Springer, 2011.
30. J S Uleman, L S Newman, and G B Moskowitz. *People as flexible interpreters: Evidence and issues from spontaneous trait inference*, volume 28, pages 211–279. Elsevier, 1996.
31. J S Uleman, S A Saribay, and C M Gonzalez. Spontaneous inferences, implicit impressions, and implicit theories. *Annual Reviews of Psychology*, 59:329–360, 2008.
32. D Ververidis and C Kotropoulos. Emotional speech recognition : Resources, features, and methods. 48(9):1162–1181, 2006.
33. A Vinciarelli. Capturing Order in Social Interactions. *IEEE Signal Processing Magazine*, 26(5):133–137, 2009.
34. 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.
35. A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder. Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Transactions on Affective Computing*, 3(1):69–87, 2012.
36. T. Wharton. *The Pragmatics of Non-Verbal Communication*. Cambridge University Press, 2009.
37. World Economic Forum. Personal data: the emergence of a new asset class. Technical report, World Economic Forum, 2011.