Social Signals, their Function, and Automatic Analysis: A Survey

Alessandro Vinciarelli[1,2], Maja Pantic[3,4], Hervé Bourlard[1,2] and Alex Pentland[5]
[1]Idiap research institute, CP592 - 1920 Martigny (CH)
[2]Ecole Polytechnique Federale de Lausanne - 1015 Lausanne (CH)
[3]Computing, Imperial College London, 180 Queen's Gate - London SW7 2AZ (U.K.)
[4]EEMCS, University of Twente, Drienerlolaan 5 - 7522 NB Enschede (NL)
[5]The MIT Media Laboratory, 20 Ames St. - Cambridge, MA 01239 (USA)
vincia@idiap.ch, m.pantic@imperial.ac.uk, bourlard@idiap.ch, pentland@media.mit.edu

ABSTRACT

Social Signal Processing (SSP) aims at the analysis of social behaviour in both Human-Human and Human-Computer interactions. SSP revolves around automatic sensing and interpretation of social signals, complex aggregates of nonverbal behaviours through which individuals express their attitudes towards other human (and virtual) participants in the current social context. As such, SSP integrates both engineering (speech analysis, computer vision, etc.) and human sciences (social psychology, anthropology, etc.) as it requires multimodal and multidisciplinary approaches. As of today, SSP is still in its early infancy, but the domain is quickly developing, and a growing number of works is appearing in the literature. This paper provides an introduction to nonverbal behaviour involved in social signals and a survey of the main results obtained so far in SSP. It also outlines possibilities and challenges that SSP is expected to face in the next years if it is to reach its full maturity.

Categories and Subject Descriptors: H.3.1 [Content Analysis and Indexing].

General Terms: Experimentation.

Keywords: Social Signal Processing, Social Behaviour Analysis, Computer Vision, Speech Analysis.

1. INTRODUCTION

Social Signal Processing (SSP) is the new pioneering domain aimed at bringing *social intelligence* to computers [52]. This is one of the multiple facets of human intelligence and can be thought of as the ability of dealing effectively with social interactions, whether this means to be accepted as leader in a working environment, to be an understanding parent, to be respected in a community, or to capture the attention of the audience. Since humans spend most of their life being involved in social interactions, social intelligence is definitely a key ability that can make the difference between success and failure in life [1].

How can a computer, or a machine in general, develop social intelligence when this seems to pertain to aspects of human psychology that are far from being detectable and accessible by sensors? The answer is in *nonverbal communication*, the phenomenon that psychologists have been studying for more than one century, and that consists of the wide spectrum of nonverbal behaviours through which humans communicate what cannot be said with words including feelings and attitudes [38][55]. Nonverbal communication can be considered as one of the physical, detectable, and measurable evidences of our inner life, the other being the content of our verbal messages. But unlike the latter, nonverbal communication is typically *honest* [19][53] and reliable because it is mostly out of the reach of conscious control, thus it leaks information about our actual state, and not about what we want to show as such.

In the case of social interactions, nonverbal communication takes the form of *social signals* [2][3], complex aggregates of behavioural cues accounting for our attitudes towards other human (and virtual) participants in the current social context. Social signals include phenomena such as attention, empathy, politeness, flirting, and (dis)agreement, and are conveyed through multiple behavioural cues including posture, facial expression, voice quality, gestures, etc. Social signals are what we experience when we watch a television program in a language we do not understand and still we are able to capture most of the social landscape, i.e., what kind of relationships people have (are there hierarchic relationships, Are people happy with one another, and similar).

As an example, consider Figure 1. The information at dispositon is limited to two silhouettes, but still people observing the picture assess correctlyin 50% of cases the interaction that takes place: the man and the woman form a couple and they are fighting. In the other 50% of the cases, assessors guess correctly at least some elements of the actual situation; they understand that the two persons are members of the same family or close friends (people do not allow mere acquaintnances to come so close), and they clearly see that the interaction is intense (the backward posture and the tension of the hand leave few doubts about).

When it comes to automatic sensing and processing of so-

ICMI'08 October 20-22, 2008, Chania, Greece.

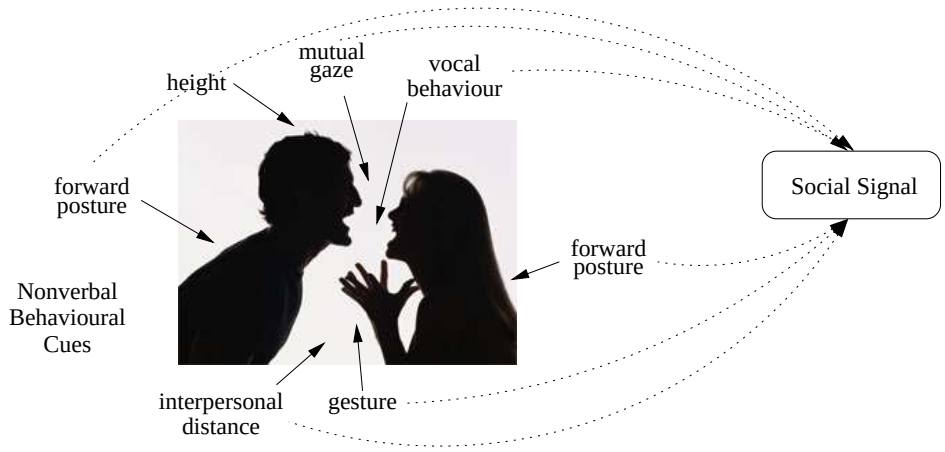

Figure 1: Behavioural cues and social signals. Combinations of multiple behavioural cues (vocal behaviour, posture, mutual gaze, interpersonal distance, etc.) produce social signals (in this case aggressivity or disagreement) that are evident even from static images showing only the silhouettes of the individuals involved in the interaction.

cial signals, the problem is tractable because social signals are based on behavioural cues that are detectable through sensors as simple as cameras and microphones. However, although the importance of social signals in everyday life situations is evident, and in spite of recent advances in machine analysis and synthesis of relevant behavioural cues like gaze exchange, blinks, smiles, head nods, crossed arms, laughter, and similar, the research efforts in machine analysis and synthesis of human social signals are still tentative and pioneering efforts. Yet, the importance of studying social interactions and developing automated assessing of human social behaviour from audiovisual recordings is undisputable. It will result in valuable multimodal tools that could revolutionise basic research in cognitive and social sciences by raising the quality and shortening the time to conduct research that is now lengthy, laborious, and often imprecise. At the same time, such tools form a large step ahead in realising naturalistic, socially-aware computing and interfaces, built for humans, based on models of human behaviour.

This paper provides a survey of the main approaches proposed so far for the analysis social behaviour. Section 2 describes in detail nonverbal communication and its role in social behaviour, Section 3 shows the main results obtained so far in SSP, and Section 4 outlines the main challenges that the researchers in the domain face.

2. NONVERBAL BEHAVIOUR IN SOCIAL INTERACTIONS

While interacting with other people, our attention tends to focus only on verbal messages. Yet, social interactions are rich in nonverbal aspects which influence heavily not only the meaning of the words, but also our perception of social contexts [33]. The correct interpretation of multiple nonverbal behavioral cues we exchange with others, consciously and unconsciously, is the key ability that makes the difference between dealing appropriately and inappropriately with social interactions [1]. Figure 2 shows the three main components of nonverbal communication: *nonverbal behavioural cues*, i.e., observable changes in facial and body gesture that accompany our communication, *codes*, i.e., classes of nonverbal behavioural cues related to specific communication means/ modality, and *functions*, i.e., the goals that nonverbal communication is aimed at.

Behavioural cues accompany any social interaction and, more in general, any moment of life. *In this sense, humans cannot "not communicate".* Even when sleeping, humans communicate through their movement or their position [55]. When talking on the phone, people still use gestures to punctuate the discourse even though the listener cannot see them [38]. When they are alone, people still display their emotions. Nonverbal behaviour is one of the most pervasive aspects of human life [31].

2.1 Codes and Behavioural Cues

Codes are classes of nonverbal behaviours sharing a common function or, more often, represent a common communication mean/ modality (e.g., the voice or the face). Five codes (and the related behavioural cues) can be identified:

Physical appearance.

This code includes bodily characteristics such as height, body shape, physiognomy, skin and hair color, as well as artificial characteristics such as clothes, ornaments, make up, and other manufacts used to modify/ accentuate the facial/ body aspects.

Gestures and postures.

Gestures are often used consciously, e.g., when waiving a hand in a greeting gesture. However, postures are typically assumed unconsciously and they are one of the most reliable cues about the rapport between people [55].

Face and eyes behaviour.

Facial expressions and gaze behaviour (who looks at whom and how much) are arguably the most reliable cues when interpreting social signals. Psychological experiments show

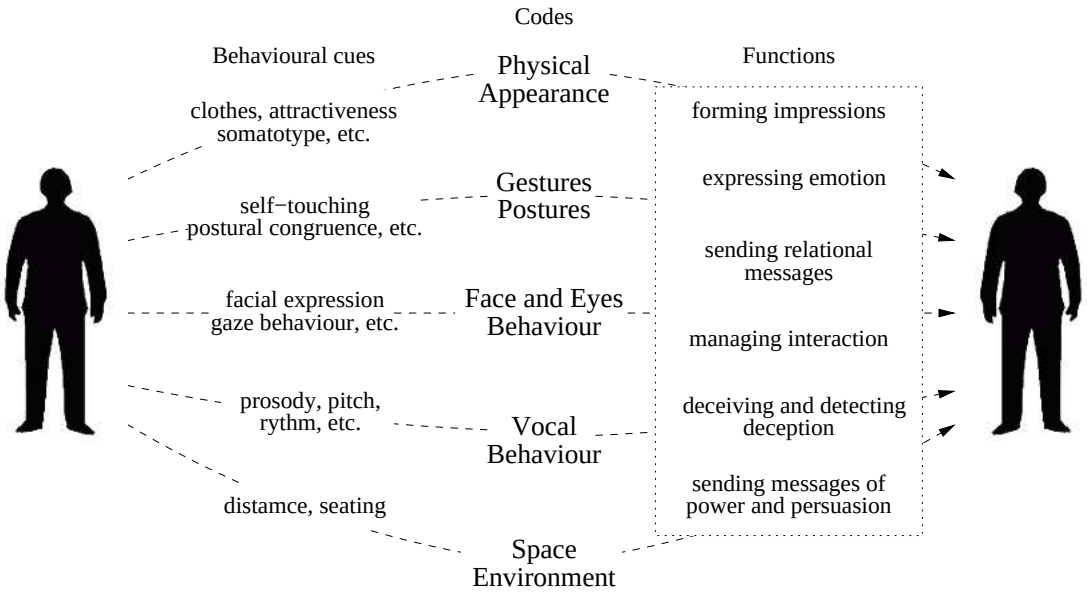

Figure 2: Behavioural cues, codes and functions. *Nonverbal behavioural cues* are organized into *codes* and fulfill *functions* aimed at affecting the perception of others.

that rapport judgments based on facial expression only are far more accurate than the judgments based on any other behavioural cue used alone [3].

Vocal behaviour.

All spoken cues that accompany a verbal message such as *voice quality* (*how* something is said), *linguistic* and *non-linguistic vocalizations* (the former are expressions like "*ehm*" and "*ah-ah*", that are used as words even though they are not real words; the latter are laughter, cries, etc.), *silences*, and *turn-taking patterns* (who talks to whom and how smooth is the transition) influence its actual meaning [12].

Space and environment behaviour.

Physical distances between individuals often correspond to their social distances. Anthropologists have shown that people tend to split the space around them into concetric regions where others are allowed depending on social criteria [32]: only family members come closer than 0.5 meters, friends and people we meet frequently (e.g., colleagues) stay between 0.5 and 1.2 meters, formal relationships (e.g., with professionals or hierarchical superoris) take place between 1.2 and 2.0 meters. The figures apply to western cultures, but can change in other contexts.

2.2 The Functions

Combined together, different codes (classes of nonverbal behaviours) send powerful messages that influence specific aspects of social interplay. Table 1 shows that different behavioural cues affect some of the most important social signals, including personality, status, dominance, persuasion, interaction regulation, and rapport. Overall, psychologists have identified six main functions of nonverbal communication.

Forming impressions.

People make judgments about others even before starting an interaction. Behavioural cues like movement and appearance contribute significantly to the impression we form in others because they are noticed first. Overall, the first impression we have about others is completely dominated by nonverbal communication [3].

Although common wisdom suggests that the appearance is not important, psychological observations suggest the reverse. For example, attractiveness elicits desirable social perceptions like high status or good personality even though no objective basis for such an assumption exists (this phenomenon if referred to as "*what is beautiful is good*" [16]). Tall people are attributed, on average higher social status [26], and the body shapes (round and soft, bony and muscular, or thin and fragile) tend to elicit the attribution of certain personality traits [11].

Managing interaction.

Appropriate nonverbal messages regulate the interaction flow during conversations, showing when it is the right moment to intervene, when a turn is going to its end, etc. [54][72]. Frequent interactions between specific individuals typically lead to coordination and synchronicity of nonverbal messages, so that the resulting interaction is fluent and smooth. The regulation in conversations relies on behaviours (including voice quality and gaze) aimed at maintaining, yielding, denying, or requesting the turn [72]. When the interaction is satisfying, the speaker transitions tend to be smooth and no interruptions or long latency times are observed. When the interactions are not positive, interruptions and other behaviours related to aggressivity and dominance appear more frequently [66]. Note, however, that the amount of overlapping speech accounts for up to 10% of the total time even in normal conversations [63].

Expressing emotion.

Nonverbal communication is the primary mean for expressing emotions [20]. Affective arousal modulates all human communicative signals, but facial expressions and body gestures appear to be the most important in the human judgment of affective behaviour [3][20].

Sending relational messages.

Arguably, this is the most important function in social interactions. Nonverbal messages communicate how people feel about others, what kind of relationship they want to establish, whether they like or dislike others, etc.

In general, nonverbal messages communicate attitudes towards the others without the need of verbalizing them. This is evident in psychological experiments where human observers judge the rapport between people using a single behavioural cue. Usually, independently of the target judged behaviour, the results obtained using the facial expressions alone lead to the best accuracy [3]. Facial expressions, typically represented with the *Facial Action Coding System* (FACS) [21], express cognitive states like interest and puzzlement [13], social behaviours like accord and rapport [3][13], personality traits like extraversion and temperament [23], and social signals like status, trustworthiness, emblems (i.e., culture-specific interactive signals like wink), regulators (i.e., conversational mediators like nod and gaze exchange), and illustrators (i.e., cues accompanying speech like raised eyebrows) [20].

Relational messages are also conveyed through gestures and postures. Gestures like self-touching and manipulation of small objects, called *adaptors*, occur typically due to boredom or negative attitudes towards others [38]. Postures are typically assumed unconsciously and they are one of the most reliable cues about the rapport between people [55]. Three main criteria define the social meaning of a posture [59]: *inclusion vs. exclusion* (facing in the direction opposite to others shows a negative attitude), *parallel vs. face-to-face* (the choice of face-to-face postures in absence of constraints shows engagement in the interaction), and *congruence vs. non-congruence* (people having satisfying interactions tend to assume the same posture).

Relational messages are also conveyed by vocal behaviour, which communicates emotions [61], and influences the perception of dominance, extroversion, competence and persuasiveness [60]. *Linguistic vocalizations* include all non-words that are used as if they were actual words, e.g., "*ehm*","*ah-ah*", "*uhm*", etc. They typically account for embarassment or difficulty with respect to a social interaction [27], but they are also used when someone else speaks (the *back-channel*) to show attention, agreement, wonder or contradiction [64]. The *non-linguistic vocalizations* include nonverbal sounds like laughing, sobbing, crying, whispering, groaning, and similar. These may or may not accompany words, and can be used to reward desirable social behaviour (e.g., through laughter [36]), or to show strong social bonds (e.g., when crying in empathy).

Deceiving and detecting deception.

Nonverbal behaviour plays a major role in the deception process. Deception-related behavioural cues are typically split into *strategic* and *non-strategic*, the former are shown consciously to make lies more credible, the latter are out of control and typically leak negative feelings which can then be used to catch liars. Research in spychology actually suggests that the visual channel carrying facial expressions and body gestures is the most important in the human judgment od deceptive behaviour [3][19]. While facial expressions are often strategic behavioural cues (especially the easily activated smiles), body gestures are usually non-strategic cues as people consider them unimportant and do not pay an effort to control.

	Example Social Signals						
Social Cues	emotion	personality	status	dominance	persuasion	regulation	rapport
Physical appearance							
height			√	√			
attractiveness		√	√	√	√		√
body shape		√		√			
Gesture and posture							
hand gestures	√				√	√	√
posture	√	√	√	√	√	√	√
walking		√	√	√			
Face and eyes behaviour							
facial expressions	√	√	√	√	√	√	√
gaze behaviour	√	√	√	√	√	√	√
focus of attention	√	√			√	√	√
Vocal behaviour							
prosody	√	√			√		
turn taking			√	√		√	√
vocalizations	√	√		√	√	√	√
silence							√
Space and Environment							
distance		√	√		√		√
seating arrangement				√			√

Table 1: The table shows the behavioural cues associated to some of the most important social behaviours.

Sending Messages of Power and Persuasion.

Nonverbal behavioural cues are also signs of social control, power, and dominance. Powerful people touch more than they are touched, look at others less than they are looked at, and have control over time and space [4].

3. STATE-OF-THE-ART

The expression *Social Signal Processing* has been used for the first time in [52] to group under a collective definition several pioneering works of Alex Pentland and his group at MIT. These works aimed at two main applications: the prediction of behavioural outcomes like the result of salary negotiations, hiring interviews, or speed-dating conversations [14], and the analysis of large groups of individuals (around 100 people) through smart cellular phones equipped with proximity detectors and vocal activity analyzers [18][51] (an application called *reality mining*).

In approximately the same period, few other groups worked on the analysis of social interactions in multimedia recordings targeting three main areas: the analysis of interactions in small groups, the recognition of roles, and the sensing

Ref.	Data	Time	Source	Performance
Dominance Detection				
[34]	Meetings from AMI Corpus (34 segments)	3h.00m	simulated	Most dominant person correctly detected in 85% of segments
[56]	Meetings (8 meetings)	1h.35m	simulated	Most dominant person correctly detected in 75% of meetings
[57]	Meetings (40 recordings)	20h.00m	simulated	Most dominant person correctly detected in 60% of meetings
Collective Action Recognition				
[15]	Meetings (30 recordings, publicly available)	2h.30m	simulated	Action Error Rate of 12.5%
[41]	Meetings (60 recordings, publicly available)	5h.00m	simulated	Action Error Rate of 8.9%
Role Recognition				
[7]	Meetings (2 recordings, 3 roles)	0h.45m	simulated	50.0% of segments (up to 60 seconds long) correctly classified
[8]	NIST TREC SDR Corpus (35 recordings, publicly available 3 roles)	17h.00m	real	80.0% of the news stories correctly labeled in terms of role
[17]	The Survival Corpus (11 recordings, publicly available, 5 roles)	4h.30m	simulated	90% of precision in role assignment
[24]	AMI Meeting Corpus (138 recordings, publicly available, 4 roles)	45h.00m	simulated	67.9% of the data time correctly labeled in terms of role
[69]	Radio news bulletins (96 recordings, 6 roles)	25h.00m	real	80% of the data time correctly labeled in terms of role
[71]	Movies (3 recordings , 4 roles)	5h.46m	real	95% of roles correctly assigned
[73]	The Survival Corpus (11 recordings, publicly available, 5 roles)	4h.30m	real	Up to 65% of analysis windows (around 10 seconds long) correctly classfied in terms of role
Interest Level Detection				
[25]	Meetings (50 recordings, 3 interest levels)	unknown	simulated	75% Precision
[43]	Children playing with video games (10 recordings, 3 interest levels)	3h.20m	real	82% recognition rate

Table 2: Results obtained by several Social Signal Processing works. For each work, information about the data (kind of interaction, availability, size), the total duration of the recordings, the distinction between real-world and simulated data, and the reported performance, are summarized.

of users attitudes towards computer interfaces. Results for problems that have been addressed by more than one group are reported in Table 2.

The research on interactions in small groups has focused on the detection of dominant persons and on the recognition of collective actions. The problem of dominance is addressed in [34][56][57], where multimodal approaches combine several nonverbal features, mainly speaking energy and body movement, to identify at each moment who is the dominant individual. The same kind of features has been applied in [15][41] to recognize the actions performed in meetings like discussions, presentations, etc. The combination of the information extracted from different modalities is performed with different algorithms including Dynamic Bayesian Networks [44] and layered Hidden Markov Models [46].

The recognition of roles has been addressed in two main contexts: broadcast material [8][69][71] and small scale meetings [7][17][73]. The works in [69][71] apply Social Network Analysis [70] to detect the role of people in broadcast news and movies, respectively. The approach in [8] recognizes the roles of speakers in broadcast news using vocal behaviour and lexical features. The roles in meetings are recognized using nonverbal behaviour in the case of [7], while a multimodal approach including both audio and visual features is applied in [17][73].

The interest level has been investigated in [25][43]. The first work applies features extracted from video and audio, while the second is mainly based on pressure sensors detecting the posture.

The reaction of users to social signals exhibited by computers has been investigated in several works showing that computers are *social actors*, i.e., they elicit the same reactions and perceptions as humans [45]. This happens, e.g., when children tend to imitate the voice quality of cartoon characters appearing on the interface of didactic applications [47], or when beta testers provide higher appreciation scores for interfaces exhibiting some form of *mimicry*, i.e. of the behaviour imitation typically displayed by humans to mean affiliation and liking [5].

4. CONCLUSIONS AND FUTURE CHALLENGES

Social Signal Processing has the ambitious goal of bringing social intelligence [1][28] to computers. The first results in this research domain have been sufficiently impressive to attract broader scientific [29] and businness [9] communities. More importantly, the pioneering efforts in the field have

established a viable interface between human sciences and engineering, which is necessary if socially aware computing is to become our reality - social interactions and behaviours, although complex and rooted in yet to be explored aspects of human psychology, can be analyzed automatically with the help of computers. This "cultural" breakthrough is, in our opinion, the most important result of research in SSP so far. In fact, the pioneering contributions in SSP - [49][50]- have shown that the social signals, typically described as so elusive and subtle that only trained psychologists can recognize them [26], are actually evident and detectable enough to be captured through sensors like microphones and cameras, and interpreted through analysis techniques like machine learning and statistics.

However, although fundamental, these are only the first steps and the journey towards *artificial social intelligence* and *socially-aware computing* is still long. In the rest of this section we discuss four challenges facing the researchers in the field, for which we believe are the crucial turnover issues that need to be addressed before the research in the field can enter its next phase - the deployment phase.

The first issue relates to *tightening of the collaboration between social scientists and engineers*. The analysis of human behaviour in general, and social behaviour in particular, is an inherently multidisciplinary problem [48]. More specifically no automatic analysis of social interactions is possible without taking into account the basic mechanisms governing social behaviours that the psychologists have investigated for decades, such as the *chameleon effect* (mutual imitation of people aimed at showing liking or affiliation) [10][39], the interpersonal adaptation (mutual accommodation of behavioural patterns between interacting individuals) [30], the interactional synchrony (degree of coordination during interactions) [37], the presence or roles in groups [6][67], the dynamics of conversations [54][72], etc. The collaboration between technology and social sciences demands a mutual effort of the two disciplines. On one hand, engineers must include social sciences in their reflection, while on the other hand, social scientists must formulate their findings in a form useful for engineers and their work on SSP.

The second issue relates to the need of implementing *multi-cue, multi-modal, approaches* to SSP. Nonverbal behaviours cannot be read like words in a book [38][55]; they are not unequivocally associated to a specific meaning and their appearance can depend on factors that have nothing to do with social behaviour. Postures correspond in general to social attitudes, but sometimes they are simply comfortable [59]; physical distances typically account for social distances, but sometimes they are simply the effect of physical constraints [32]. Moreover, the same signal can correspond to different social behaviour interpretations depending on context and culture [68] (although many advocate that social signals are natural rather than cultural [62]). In other words, social signals are intrinsically ambiguous and the best way to deal with such problem is to use multiple behavioural cues extracted from multiple modalities. Numerous studies have theoretically and empirically demonstrated the advantage of integration of multiple modalities (at least audio and visual) in human behaviour analysis over single modalities (e.g., [58]). This corresponds, from a technological point of view, to the combination of different classifiers that has extensivley been shown to be more effective than single classifiers, as long as they are sufficiently *diverse*, i.e., account for different aspects of the same problem. It is therefore not surprising that some of the most successful works in SSP so far use features extracted from multiple modalities (e.g., [15][34][41]. Note, however, that the relative contributions of different modalities and the related behavioural cues to affect judgment of displayed behaviour depend on the targeted behavioural category and the context in which the behaviour occurs [22][58].

The third issue relates to *the use of real-world data*. Both psychologists and engineers tend to produce their data in laboratories and artificial settings (see e.g., [14][41]), in order to limit parasitic effects and elicit the specific phenomena they want to observe. However, this is likely to simplify excessively the situation and to improve artificially the performance of the automatic approaches. Social interaction is one of the most ubiquitous phenomena in the world - the media (radio and television) show almost exclusively social interactions (debates, movies, talk-shows) [42]. Also other, less common kinds of data are centered on social interactions, e.g., meeting recordings [40], surveillance material [35], and similar. The use of real-world data will allow analysis of interactions that have an actual impact on the life of the participants, thus, will show the actual effects of goals and motivations that typically drive human behaviour. This includes also the analysis of *group interactions*, a task difficult from both technological and social point of view because it involves the need of observing multiple people involved in a large number of one-to-one interactions.

The last, but not least, challenging issue relates to the *the identification of applications likely to benefit from SSP*. Applications have the important advantage of linking the effectiveness of detecting social signals to the reality. For example, one of the earliest applications is the prediction of the outcome in transactions recorded at a call center and the results show that the number of successful calls can be increased by around 20% by stopping early the calls that are not promising [9]. This can have not only a positive impact on the marketplace, but also provide *benchmarking procedures* for the SSP research, one of the best means to improve the overall quality of a research domain as extensively shown in fields where international evaluations take place every year (e.g. video analysis in TrecVid [65]).

Acknowledgements. The work of Dr. Vinciarelli is supported by the Swiss National Science Foundation through the National Center of Competence in Research on Interactive Multimodal Information Management (IM2). The work of Dr. Pantic is supported in part by the EU IST Programme project FP6-0027787 (AMIDA) and the EC's 7^{th} Framework Programme (FP7/2007-2013) under grant agreement no. 211486 (SEMAINE).

5. REFERENCES

[1] K. Albrecht. *Social Intelligence: The new science of success*. John Wiley & Sons Ltd, 2005.

[2] 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.

[3] 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.

[4] M. Argyle. *The Psychology of Interpersonal Behaviour*. Penguin, 1967.

[5] J. Bailenson, N. Yee, K. Patel, and A. Beall. Detecting digital chameleons. *Computers in Human Behavior*, 24(1):66–87, 2008.

[6] R. Bales. *Interaction Process Analysis: A Method for the Study of Small Groups*. Addison-Wesley, 1950.

[7] S. Banerjee and A. Rudnicky. Using simple speech based features to detect the state of a meeting and the roles of the meeting participants. In *Proceedings of International Conference on Spoken Language Processing*, pages 2189–2192, 2004.

[8] R. Barzilay, M. Collins, J. Hirschberg, and S. Whittaker. The rules behind the roles: identifying speaker roles in radio broadcasts. In *Proceedings of American Association of Artificial Intelligence Symposium*, pages 679–684, 2000.

[9] M. Buchanan. The science of subtle signals. *Strategy+Business*, 48:68–77, 2007.

[10] T. Chartrand and J. Bargh. The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893–910, 1999.

[11] J. Cortes and F. Gatti. Physique and self-description of temperament. *Journal of Consulting Psychology*, 29(5):432–439, 1965.

[12] D. Crystal. *Prosodic Systems and Intonation in English*. Cambridge University Press, 1969.

[13] D. Cunningham, M. Kleiner, H. Bülthoff, and C. Wallraven. The components of conversational facial expressions. *Proceedings of the Symposium on Applied Perception in Graphics and Visualization*, pages 143–150, 2004.

[14] J. 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.

[15] A. Dielmann and S. Renals. Automatic meeting segmentation using dynamic bayesian networks. *IEEE Transactions on Multimedia*, 9(1):25, 2007.

[16] K. Dion, E. Berscheid, and E. Walster. What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3):285–290, 1972.

[17] W. Dong, B. Lepri, A. Cappelletti, A. Pentland, F. Pianesi, and M. Zancanaro. Using the influence model to recognize functional roles in meetings. In *Proceedings of the International Conference on Multimodal Interfaces*, pages 271–278, 2007.

[18] N. Eagle and A. Pentland. Reality mining: sensing complex social signals. *Journal of Personal and Ubiquitous Computing*, 10(4):255–268, 2006.

[19] P. Ekman. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1000(1):205–221, 2003.

[20] P. Ekman and W. Friesen. The repertoire of nonverbal behavior. *Semiotica*, 1:49–98, 1969.

[21] P. Ekman and W. Friesen. Facial action coding system (FACS): Manual. 2002.

[22] P. Ekman, T. Huang, T. Sejnowski, and J. Hager, editors. *Final Report to NSF of the Planning Workshop on Facial Expression Understanding*. Human Interaction Laboratory, University of California, San Francisco, 1993.

[23] P. Ekman and E. Rosenberg. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press, 2005.

[24] N. Garg, S. Favre, H. Salamin, D. Hakkani-Tür, and A. Vinciarelli. Role recognition for meeting participants: an approach based on lexical information and social network analysis. In *Proceedings of the ACM International Conference on Multimedia*, 2008.

[25] D. Gatica-Perez, I. McCowan, D. Zhang, and S. Bengio. Detecting group interest-level in meetings. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 489–492, 2005.

[26] M. Gladwell. *Blink: The Power of Thinking without Thinking*. Little Brown & Company, 2005.

[27] C. Glass, T. Merluzzi, J. Biever, and K. Larsen. Cognitive assessment of social anxiety: Development and validation of a self-statement questionnaire. *Cognitive Therapy and Research*, 6(1):37–55, 1982.

[28] D. Goleman. *Social intelligence*. Hutchinson, 2006.

[29] K. Greene. 10 emerging technologies 2008. *MIT Technology Review*, february 2008.

[30] S. Gregory, K. Dagan, and S. Webster. Evaluating the relation of vocal accommodation in conversation partners fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, 21(1):23–43, 1997.

[31] L. Guerrero, J. DeVito, and M. Hecht. *The Nonverbal Communication Reader. Classic and Contemporary Readings*. Waveland Press, 1999.

[32] E. Hall. *The silent language*. Doubleday, 1959.

[33] M. Hecht, J. De Vito, and L. Guerrero. Perspectives on nonverbal communication. codes, functions and contexts. In L. Guerrero, J. De Vito, and M. Hecht, editors, *The Nonverbal Communication Reader*, pages 3–18. 1999.

[34] H. Hung, D. Jayagopi, C. Yeo, G. Friedland, S. Ba, J. Odobez, K. Ramchandran, N. Mirghafori, and D. Gatica-Perez. Using audio and video features to classify the most dominant person in a group meeting. In *Proceedings of the ACM International Conference on Multimedia*, pages 835–838, 2007.

[35] Y. Ivanov, C. Stauffer, A. Bobick, and W. Grimson. Video surveillance of interactions. *Proceeding of the Workshop on Visual Surveillance at Computer Vision and Pattern Recognition*, 1999.

[36] D. Keltner and J. Haidt. Social functions of emotions at four levels of analysis. *Cognition and Emotion*, 13(5):505–521, 1999.

[37] M. Kimura and I. Daibo. Interactional synchrony in conversations about emotional episodes: A measurement by "the between-participants pseudosynchrony experimental paradigm". *Journal of Nonverbal Behavior*, 30(3):115–126, 2006.

[38] M. Knapp and J. Hall. *Nonverbal Communication in Human Interaction*. Harcourt Brace College

Publishers, 1972.

[39] J. Lakin, V. Jefferis, C. Cheng, and T. Chartrand. The Chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27(3):145–162, 2003.

[40] I. McCowan, S. Bengio, D. Gatica-Perez, G. Lathoud, F. Monay, D. Moore, P. Wellner, and H. Bourlard. Modeling human interaction in meetings. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 748–751, 2003.

[41] I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, and D. Zhang. Automatic analysis of multimodal group actions in meetings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):305–317, 2005.

[42] D. Morris. *Peoplewatching*. Vintage, 2007.

[43] S. Mota and R. Picard. Automated posture analysis for detecting learner's interest level. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, pages 49–56, 2003.

[44] K. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California Berkeley, 2002.

[45] C. Nass and K. Lee. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, 7(3):171–181, 2001.

[46] N. Oliver, E. Horvitz, and A. Garg. Layered representations for human activity recognition. In *Proceedings of the International Conference on Multimodal Interfaces*, pages 3–8, 2002.

[47] S. Oviatt, C. Darves, and R. Coulston. Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction*, 11(3):300–328, 2004.

[48] M. Pantic, A. Pentland, A. Nijholt, and T. Huang. Human-centred intelligent human-computer interaction (HCI2): How far are we from attaining it? *International Journal of Autonomous and Adaptive Communications Systems, to appear*, 2008.

[49] A. Pentland. Social dynamics: Signals and behavior. In *International Conference on Developmental Learning*, 2004.

[50] A. Pentland. Socially aware computation and communication. *IEEE Computer*, 38(3):33–40, 2005.

[51] A. Pentland. Automatic mapping and modeling of human networks. *Physica A*, 378:59–67, 2007.

[52] A. Pentland. Social signal processing. *IEEE Signal Processing Magazine*, 24(4):108–111, 2007.

[53] A. Pentland. *Honest signals: how they shape our world*. MIT Press, 2008.

[54] G. Psathas. *Conversation Analysis - The study of talk-in-interaction*. Sage Publications, 1995.

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

[56] R. Rienks and D. Heylen. Dominance Detection in Meetings Using Easily Obtainable Features. *Proceedings of the 2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Lecture Notes in Computer Science 3869:76–86, 2006.

[57] R. Rienks, D. Zhang, and D. Gatica-Perez. Detection and application of influence rankings in small group meetings. In *Proceedings of the International Conference on Multimodal Interfaces*, pages 257–264, 2006.

[58] J. Russell, J. Bachorowski, and J. Fernandez-Dols. Facial and vocal expressions of emotion. *Annual Reviews in Psychology*, 54(1):329–349, 2003.

[59] A. Scheflen. The significance of posture in communication systems. *Psychiatry*, 27:316–331, 1964.

[60] K. Scherer. *Personality markers in speech*. Cambridge University Press, 1979.

[61] K. Scherer. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2):227–256, 2003.

[62] U. Segerstrale and P. Molnar, editors. *Nonverbal communication: where nature meets culture*. Lawrence Erlbaum Associates, 1997.

[63] E. Shriberg, A. Stolcke, and D. Baron. Observations of overlap: findings and implications for automatic processing of multiparty conversation. In *Proceedings of Eurospeech*, pages 1359–1362, 2001.

[64] P. Shrout and D. Fiske. Nonverbal behaviors and social evaluation. *Journal of Personality*, 49(2):115–128, 1981.

[65] A. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVid. In *Proceedings of the ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, 2006.

[66] L. Smith-Lovin and C. Brody. Interruptions in group discussions: the effects of gender and group composition. *American Sociological Review*, 54(3):424–435, 1989.

[67] H. Tischler. *Introduction to Sociology*. Harcourt Brace College Publishers, 1990.

[68] H. Triandis. *Culture and social behavior*. McGraw-Hill, 1994.

[69] A. Vinciarelli. Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Transactions on Multimedia*, 9(9):1215–1226, 2007.

[70] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[71] C. Weng, W. Chu, and J. Wu. Movie analysis based on roles social network. In *proceedings of IEEE International Conference on Multimedia and Expo*, pages 1403–1406, 2007.

[72] G. Yule. *Pragmatics*. Oxford University Press, 1996.

[73] M. Zancanaro, B. Lepri, and F. Pianesi. Automatic detection of group functional roles in face to face interactions. In *Proceedings of the International Conference on Multimodal Interfaces*, pages 28–34, 2006.