Social Signal Processing

Maja Pantic and Alessandro Vinciarelli

Abstract

Social Signal Processing (SSP) is the new cross-disciplinary research domain that aims at understanding and modelling social interactions (research in human-sciences) and at providing computers with similar abilities (research in computer science). SSP is still in its formative phase, and the journey towards artificial social intelligence and socially-aware computing is still long. This chapter surveys the current state of the art and summarizes issues that the researchers in this field face.

Keywords/keyphrases

Social signal processing, artificial social intelligence, socially-aware computing

Glossary terms

- *Social Signals*: communicative or informative signals which provide information about social facts (social interactions, social emotions, social evaluations, social attitudes and social relations)
- *Social Interactions*: Social interactions are events in which actually or virtually present agents exchange an array of *social actions*, i.e. communicative and informative signals performed by one agent in relation to one or more other agents
- *Social Emotions*: Emotions like admiration, envy, and compassion are that can be felt only toward another person
- *Social Evaluations*: Social evaluations relate to assessing whether and how much the characteristics of a person comply with our standards of beauty, intelligence, strength, justice, altruism, etc.
- *Social Attitudes*: positive or negative evaluation of a person or a group of people. Social attitudes include cognitive elements like beliefs, opinions, and social emotions.
- *Social Relations*: A social relation is a relation between two (or more) persons in which these persons have related goals.

1. Social intelligence in men and machines

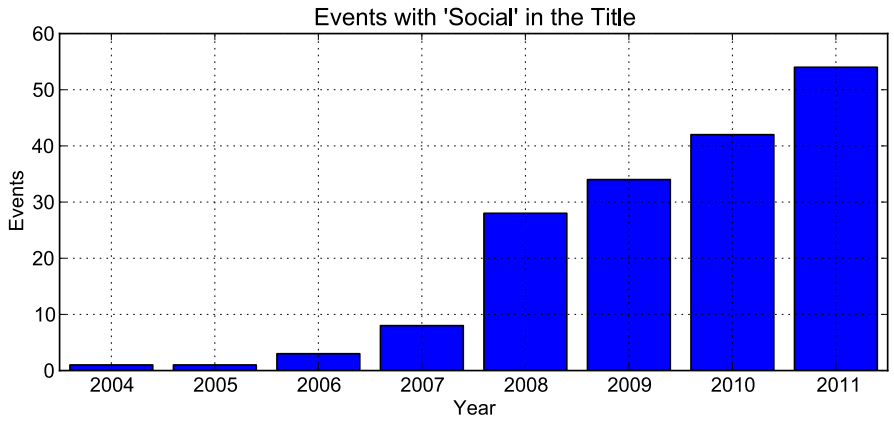
The need of dealing effectively with social interactions has driven the evolution of brain structures and cognitive abilities in all species characterized by complex social exchanges including, in particular, humans (Gallese, 2006). The relationship between degree of expansion of the neocortex and size of the groups in primates is one of the most conclusive and important evidences of such a process (Dunbar, 1992). Therefore, it is not surprising to observe that the computing community considers the development of socially intelligent machines an important priority (Vinciarelli et al., 2012), especially since computers left their traditional role of enhanced versions of old tools (e.g., word processors replacing typewriters) and became full social actors expected to seamlessly integrate into our everyday life (Nass et al., 1994; Vinciarelli 2009b).

Social Signal Processing (SSP) is one of the domains that contribute to the efforts aimed at endowing machines with social intelligence (see Section 2) and, in

particular, it focuses on modelling, analysis and synthesis of nonverbal behaviour in social interactions (Vinciarelli et al., 2009). The key idea of SSP is that computers can participate in social interactions by automatically understanding and/or synthesising the many nonverbal behavioural cues (facial expressions, vocalisations, gestures, postures, etc.) that people use to express, or leak, socially relevant information (attitudes, beliefs, intentions, stances, etc.).

Overall, SSP stems from three major research areas, namely human behaviour understanding, social psychology, and computer science. The former provides methodologies for dealing with non-verbal behaviour as a physical (machine detectable) phenomenon. Social psychology provides quantitative analyses of the relationship between nonverbal behaviour and social/psychological phenomena. Computer science provides technologies for machine detection and synthesis of these relevant phenomena within both human-human and human-computer interaction context. The result is an interdisciplinary domain where the target is machine modelling and understanding the social meaning of human behaviour in interactive contexts.

While in its early stages – the expression "*Social Signal Processing*" was coined only a few years ago (Pentland, 2007) – the SSP research area have witnessed an impressive development in the past years, in terms of both knowledge accumulation and increase of interest from the research community. The domain has made significant progress in terms of social phenomena made accessible to technological investigation (roles, personality, conflict, leadership, mimicry, attraction, stances, etc.), methodologies adopted (regression and prediction approaches for dimensional assessments, probabilistic inference for modelling and recognition of multimodal sequences of human behaviour, combinations of multiple ratings and crowdsourcing for attaining a more reliable ground truth, etc.) and benchmarking campaigns carried out (facial expression recognition, automatic personality perception, vocalisations detection, etc.). Furthermore, major efforts have been done towards the definition of social signals (Mehu & Scherer, 2012; Poggi & D'Errico, 2012), the delimitation of the domain's scope (Brunet et al., 2012), and setting a research agenda for the progress in the field (Pantic et al., 2011). Figure 1 shows the number of technology oriented events (workshops, conferences, and symposia) and publications revolving around social interactions. The trend speaks for itself and is still growing at the moment this article is being written.



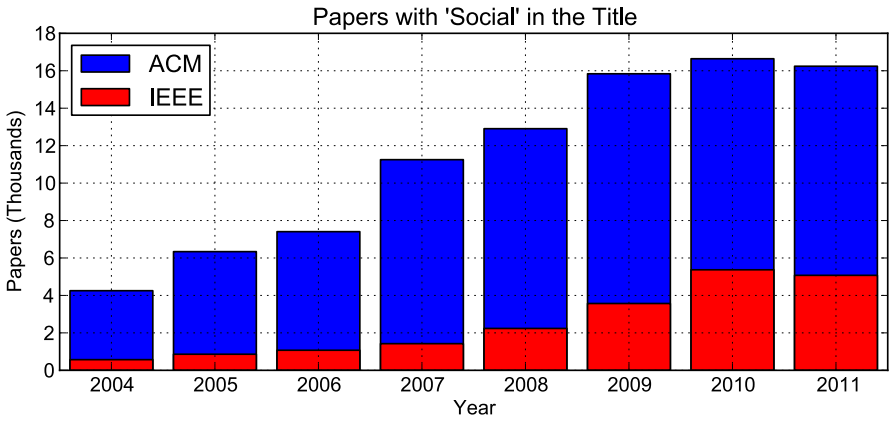


Figure 1. The upper plot shows the number of technology-oriented events (workshops, summer schools, symposia, etc.) with the word "social" in their title, as advertised in the "dbworld" mailing list. The lower plot shows the same information for the papers available via the IEEE-Xplore and the ACM Digital Library.

The rest of this chapter provides an account of the main results achieved so far as well as an indication of remaining challenges and most promising applications.

2. Social Signal Processing: Definition and Context

In 2007, Alex Pentland coined the expression "Social Signal Processing" (Pentland, 2007), to describe pioneering efforts in inferring socially relevant information from non-verbal behavioral cues (e.g., at predicting the outcome of a salary negotiation based on the way the participants talk, but not on what the participants say). Since then, the domain has continued to grow and addresses an increasingly wider spectrum of scenarios. The scope of the field, according to a widely accepted definition (Brunet et al., 2012), is to study signals (in a broad everyday sense of the word) that:

- are produced during social interactions,
- that either play a part in the formation and adjustment of relationships and interactions between agents (human and artificial),
- or provide information about the agents,
- and that can be addressed by technologies of signal processing and synthesis.

The relationship between SSP and the other socially aware technologies can be analyzed in terms of two main dimensions, namely the scale of the interactions under consideration and the processing level. The first dimension ranges between dyads and online communities including millions of individuals, the second between high level, easily detectable electronic evidences (e.g., the exchange of an e-mail or a "connection" in social media like "LinkedIn") and low-level, subtle behavioral cues that need complex signal processing and machine learning techniques to be detected (e.g., individual action units in facial expressions or short-term changes in speech prosody).

In such a framework of reference, SSP considers only small-scale scenarios (rarely more than four individuals) where it applies low-level processing techniques, mostly to audio and video data. SSP approaches typically rely on subtle behavioral cues and address social phenomena as complex as role-playing, personality, conflict, emotions, etc. At the opposite side of the spectrum, Social Network Analysis approaches can take into account millions of people, but typically depend on electronic traces left during the usage of web-based technologies (see above).

In between these extremes, it is possible to find areas that target middle scale groups (50-150 individuals), often corresponding to actual communities such as the members of an organization (e.g., a company or a school) analyzed during its operations. One example is *reality mining* (Eagle & Pentland, 2005; Raento et al., 2009), the domain using the smartphones as sensors for social and other activities. Interaction evidences used in this case include both high level cues (e.g., phone calls or text messages) and low-level behavioral signals such as fidgeting (captured via accelerometers) or proximity to others (captured via Bluetooth). Another example is the design of sociotechnical systems (de Bruijn and Herder, 2009). In this case the goal is to analyze and optimize the impact of technologies on groups of people sharing a particular setting (e.g., the employees of a company or the inhabitants of a building). This area considers high-level evidence such as usage logs, field observations and questionnaires.

A recent trend is to apply SSP inspired approaches to data collected in social media like, e.g., blogs, YouTube videos, etc. The number of involved subjects is typically high (100-500 people), but these tend to be considered individually and not as a community. The main difference with respect to "standard" SSP approaches is the adoption of social network inspired features (e.g., number of times a video has been watched, on-line ratings, etc.) typically available in social media (Biel and Gatica-Perez, 2012; Salvagnini et al., 2012). Last, but not least, is the research on socially aware approaches aimed at computer supported communication and collaboration. In this field, the goal is not to understand or synthesize social interactions, but to support - and possibly enhance - social contacts between individuals expected to accomplish common tasks or communicate via computer systems (Grudin and Poltrock, 2012). In this case, the focus is typically on building infrastructures (virtual spaces, interfaces, etc.) that facilitate basic social mechanisms such as eye contact, information sharing, turn-organization, focus of attention, etc. Such technologies typically address small groups (2-10 people) of non co-located individuals.

3. Social Signals

Social signals are the key concept of SSP and their definition is still subject of research in the human sciences community (Mehu et al., 2012; Poggi et al., 2012). In an evolutionary-ethological perspective, social signals are behaviours that have co-evolved across multiple subjects to make social interaction possible (Mehu and Scherer, 2012). From a social psychological point of view, social signals include any behaviour aimed at engaging others in a joint activity, often communication (Brunet and Cowie, 2012). This work adopts the cognitive perspective proposed by Poggi and D'Errico (2012), where social signals are defined as "*communicative or informative signals which [...] provide information about social facts*", i.e. about social (inter)actions, social emotions, social evaluations, social attitudes and social relations.

Social (Inter)actions -- Social interactions are events in which actually or virtually present agents exchange an array of *social actions*, i.e. communicative and informative signals performed by one agent in relation to one or more other agents. Typical communicative signals in social interactions are backchannel signals such as head nods, gaze exchanges, and rapport, which inform the recipient that her interaction partner is following and understanding her (Miles et al, 2009).

Social Emotions -- A clear distinction can be made between *individual* and *social* emotions. Happiness and sadness are typical examples of individual emotions – we can be happy or sad on our own; our feelings are not directed to any other person. On the other hand, admiration, envy, and compassion are typical examples of social emotions – we have these feelings toward another person. Signals revealing individual emotions of a person and those communicating social emotions both include facial expressions, vocal intonations and outbursts, and body gestures and postures (Mayne & Bonanno, 2001).

Social Evaluations -- Social evaluation of a person relates to assessing whether and how much the characteristics of this person comply with our standards of beauty, intelligence, strength, justice, altruism, etc. We judge other people because based on our evaluation we decide whether to engage in a social interaction with them, what types of social actions to perform, and what relations to establish with them (Gladwell, 2005). Typical signals shown in social evaluation are approval and disapproval, at least when it comes to the evaluator. As far as the evaluated person is concerned, typical signals involve those conveying desired characteristics such as pride, self-confidence, and mental strength, which include raised chin, erected posture, easy and relaxed movements, etc. (Manusov & Patterson, 2006)

Social Attitudes -- Social attitude can be defined as a positive or negative evaluation of a person or a group of people (Gilbert et al., 1998). Social attitudes include cognitive elements like beliefs, opinions, and social emotions. All these elements determine (and are determined by) preferences and intentions (Fishbein & Ajzen, 1975). Agreement and disagreement can be seen as being related to social attitudes. If two persons agree then this usually entails an alliance and a mutually positive attitude. This is in contrast to disagreement, which typically implies conflict and mutually negative attitude. Typical signals of agreement and disagreement are head nods and head shakes, smiles, crossed arms, etc. (Bousmalis et al., 2012).

Social Relations -- A social relation is a relation between two (or more) persons in which these persons have related goals (Kelley & Thibaut, 1978). Hence, not every relation is a social relation. Two persons sitting next to each other in a bus have a physical proximity relation, but this is not a social relation, although one can rise from it. We can have many different kinds of social relations with other people: dependency, competition, cooperation, love, exploitation, etc. Typical signals revealing social relations include the manner of greeting (saying 'hello' signals the wish for a positive social relation, saluting signals belonging to a specific group like the army), the manner of conversing (e.g., using the word 'professor' signals submission), mirroring (signalling wish to have a positive social relation), spatial

positioning (e.g., making a circle around a certain person distinguishes that person as the leader), etc.

4. Machine Analysis of Social Signals

The core idea behind machine analysis of social signals is that these are physical, machine detectable traces of social and psychological phenomena that may not be observed directly (Vinciarelli et al., 2012). For this reason, typical SSP technologies include two main components (Vinciarelli et al., 2009). The first aims at detecting the morphology (or the very simple existence) of social signals in data captured with a wide array of sensors, most commonly microphones and cameras. The second aims at interpreting detected social signals in terms of social facts (see above), according to rules/ principles proposed in the large body of literature in human sciences.

Social (Inter)actions – In the past decade, a significant progress in automatic audio and/or visual recognition of communicative signals like head nods, smiles, laughter and hesitation has been reported (De la Torre & Cohn, 2011; Schuller et al., 2013). Reviews of such technologies are included in Section 2 of this volume. However, a multitude of social signals underlying the manifestation of various social facts involve explicit representation of the context, time, and interplay between different modalities. For example, in order to model gaze exchanges or mimicry (Delaherche et al., 2012), which are crucial for inferring rapport, empathy, and dominance, all interacting parties and their mutual multimodal interplay in time should be modelled. Yet, most of the present approaches to machine analysis of social signals and human behaviours are neither multimodal, nor context-sensitive, nor suitable for handling multiple interacting parties and longer time scales (Pantic, 2009; De la Torre & Cohn, 2011, Delaherche et al., 2012). Hence, proper machine modelling of social interactions and the related phenomena like rapport and interaction cohesion is yet to be attempted.

Social Emotions -- Whilst the state of the art in machine analysis of basic emotions such as happiness, anger, fear and disgust, is fairly advanced, especially when it comes to analysis of acted displays recorded in constrained lab settings (Zeng et al., 2009), machine analysis of social emotions such as empathy, envy, admiration, etc., is yet to be attempted. Although some of social emotions could be arguably represented in terms of affect dimensions -- valence, arousal, expectation, power, and intensity -- and pioneering efforts towards automatic dimensional and continuous emotion recognition have been recently proposed (Gunes & Pantic, 2010; Nicolaou et al., 2012), a number of crucial issues need to be addressed first if these approaches to automatic dimensional and continuous emotion recognition are to be used with freely moving subjects in real-world multi-party scenarios like patient-doctor discussions, talk-shows, job interviews, etc. In particular, published techniques revolve around the emotional expressions of a single subject rather than around the dynamics of the emotional feedback exchange between two subjects, which is the crux in the analysis of any social emotions. Moreover, the state of the art techniques are still unable to handle natural scenarios such as incomplete information due to occlusions, large and sudden changes in head pose, and other

temporal dynamics typical of natural facial expressions (Zeng et al. 2009), which must be expected in real-world scenarios in which social emotions occur.

Social Evaluations -- Only recently, efforts have been reported towards automatic prediction of social evaluations including personality and beauty estimation. Automatic attribution of personality traits, in terms of the "Big Five" personality model, attracts increasingly more attention in the last years (Pianesi et al., 2008; Olguin-Olguin et al., 2009; Zen et al., 2010; Mairesse et al., 2012; Polzehl et al., 2010; Pianesi, 2013). Most of the works rely on speech, especially after personality perception benchmarking campaigns organized in the speech processing community (Schuller et al., 2013; Lee et. al. this volume). The cues most commonly adopted include prosody (pitch, speaking rate, energy and their statistics across time), voice quality (statistical spectral measurements) and, whenever the subject is involved in interactions, turn-organization features (see above). Other cues that appear to have an influence, especially from a perception point of view, are facial expressions, focus of attention, fidgeting, interpersonal distances, etc. However, automated approaches using such visual cues are yet to be attempted. The results change depending on the setting, but it is frequent to observe that the best predicted traits are extraversion and conscientiousness, in line with psychology findings showing that such personality dimensions are the most reliably perceived in humans as well (Judd et al., 2005).

Automatic facial attractiveness estimation have been attempted based on the facial shape (e.g., Gunes & Piccardi, 2006, Schmid et al., 2008, Zhang et al., 2011) as well as based on facial appearance information encoded in terms of Gabor filters responses (Whithehill & Movellan, 2008) or eigenfaces (Sutic et al., 2010). A survey of the efforts on the topic is reported by Bottino & Laurentini (2010). However, the research in this domain is still in its very first stage and many basic research questions remain unanswered including exactly which features (and modalities) are the most informative for the target problem.

Social Attitudes – Similarly to social emotions and social evaluations, automatic assessment of social attitudes has been attempted only recently and there are just a few studies on the topic. Conflict and disagreement have been detected and measured, in both dimensional and categorical terms using prosody, overlapping speech, facial expressions, gestures and head movements (Kim et al., 2012; Bousmalis et al., 2012, 2013). Dominance has been studied in particular in meetings, where turn-organization features and received visual attention were shown to be the best predictors (Hung et al., 2009; Gatica-Perez, 2009).

Social Relations – One of the most common problems addressed in SSP is the recognition of roles, whether this means to identify people fulfilling specific functions in well-defined settings like, e.g., anchorman in a talk show or chairman in a meeting (Barzilay et al., 2000; Liu, 2006; Laskowski et al., 2008; Gatica-Perez, 2009; Salamin and Vinciarelli, 2012), or to address the very structure of social interactions in small groups by tackling roles observable in every social situation like, e.g., attacker, neutral, supporter, etc. (Banerjee and Rudnicky, 2004; Dong et al., 2007, Valente and Vinciarelli, 2011). The social signals that appear to be most effective in this problem are those related to turn-organization -*who talks when, how much and with whom* - in line with the indications of conversation analysis, the

domain that studies the social meaning behind the way interaction is organized (Sacks et al., 1974). Speaking time distribution across different interaction participants, adjacency pair statistics between different individuals, average length of turns, number of turns per individual, number of turns between consecutive turns of the same individual, and variants of these measurements lead to high role recognition performances in almost every setting considered in the literature. The analysis of turn-organization is typically performed by applying speaker diarization approaches to audio data, i.e. technologies that segment audio data into time intervals expected to correspond to an individual voice. After such a step, it is possible to measure turn-organization features and apply pattern recognition to assign each person a role. Limited, but statistically significant improvements come from a variety of other cues, including lexical choices, fidgeting, focus of attention, prosody, etc. None of these cues produces, individually, satisfactory results. Therefore they appear only in multimodal approaches where they improve the performance achieved with turn-organization cues.

5. Machine Synthesis of Social Signals

Most of the efforts in machine synthesis of social signals aim at artificially generating social actions, informative and communicative signals displayed by an artificial agent in relation to another, typically human (Poggi and D'Errico, 2012). However, the latest efforts target the synthesis of more complex constructs, in particular emotions and attitudes, that typically require the coordinated synthesis of several social actions at the same time (Vinciarelli et al., 2012).

Social (Inter)actions – One of the most challenging goals for an artificial agent is to get involved in conversations with humans. Therefore, social actions typical of such a setting are those that have received most attention. Since an agent is expected to actively participate, the ability of appropriately grabbing and releasing the floor is a priority and it is typically modeled via action-perception loops (Bonaiuto and Thorisson, 2008) or imitation (Prepin and Revel, 2007). However, in order to appear natural, agents must be active not only when they intervene and talk, but also when they listen. Such a goal is achieved by simulating back-channel cues like head-nodding, laughter, vocalizations (e.g., "yeah", "ah-ah", etc.) and other behaviors people display to show attention. The main issue is to identify the moments when such cues are appropriate. The most common approaches consist of reacting when the speaker shows certain cues (Maatman et al., 2005), using probabilistic models that predict the best back-channel "spots" (Morency et al., 2007), or analyzing what the interlocutors say (Kopp et al., 2008). When the agent is a robot, or any other machine that can move, listening behavior includes proxemics as well, i.e. the use of space and distances as a social cue. Two approaches are commonly adopted for this purpose, the social force model (Jan and Traum, 2007) and the simulation of human territoriality (Pedica and Vilhjalmsson, 2009).

Social Emotions – In many scenarios, the expression of social emotions like empathy through a virtual humans face (Niewiadomski et al. 2008, Ochs et al. 2010) and voice (Schroeder, 2009) or any other form of nonverbal behavior is very important. Besides expression synthesis, the research community has devoted much

energy in defining and implementing computational models of behaviors that underlie the decisions on the choice of emotional expression. For an overview see Marsella et al. (2010).

Social Evaluations – The computational models of emotions, based on appraisal models typically contain variables that deal with the evaluation of the human interlocutor and the situation the agent is in. On the other hand, many studies dealing with the evaluation of virtual humans (Ruttkay & Pelachaud 2004) consider the other side of the coin: the question of how the agent is perceived by the human. This can pertain to any of the behaviors exhibited by the agent and any dimension. For instance, Ter Maat and Heylen (2009) consider how different turn-taking strategies evoke different impressions, while De Melo & Gratch (2009) consider the effect of wrinkles, just to give two extreme examples of behaviors and dimensions of expression that have been related to social evaluation.

Social Attitudes – The synthesis of attitudes requires the artificial generation of several cues in a coordinated fashion as well as coherence in the behavior displayed by the agent. Since artificial agents are used in scenarios where they are expected to provide a service (museum guiding, tutoring, help-desk dialogues, etc.), the attitude most commonly addressed is politeness. In the simplest approaches, politeness does not arise from an analysis of the interlocutor's behavior, but from predefined settings that account for power distance (Gupta et al., 2007; Porayska-Pomsta and Mellish, 2004). Such a problem is overcome in (de Jong et al., 2008), where the politeness degree of the interlocutor is matched by the agent in a museum guide scenario.

A crucial channel through which any attitude can be conveyed is speech and major efforts have been spent towards the synthesis of "expressive" voices, i.e. voices capable of conveying something more than just the words being uttered (Schroeder, 2009). Initial approaches were based on the collection of short speech snippets extracted from natural speech expressing different attitudes. The snippets were then played back to reproduce the same attitude. Such an approach has been used to make agents capable of reporting differently on good rather than bad news (Pitrelli et al., 2006), of giving orders (Johnson et al., 2002) or of playing characters (Gebhard et al., 2008). The main drawback of such approaches is that it is necessary to collect examples for each and every attitude to be synthesized. Thus, current techniques try to represent expressiveness in terms of parameters that can be manipulated to allow agents to express desired attitudes (Schroeder, 2007; Zovato et al., 2004).

Social Relations – The Laura agent was one of the first agents that was extensively studied in a longitudinal study (Bickmore & Picard, 2005). One of the major research interests in developing the agent for this study was modeling the long-term relations that might develop between the agent and the user over the course of repeated interactions. This involved modeling many social psychological theories on relationships formation and friendship. Currently, there is a surge of work on companion agents and robots (Leite et al. 2010, Robins et al, 2012).

6. Conclusions

Social Signal Processing (SSP) is the new research and technological domain that aims at providing computers with the ability to sense and understand human social signals. SSP is in its initial phase and the researchers in the field face many challenges (Pantic et al., 2011). Given the current state of the art in automatic analysis of social signals, the focus of future research efforts in the field should be on tackling the problem of context-constrained and multi-party analysis of multimodal behavioural signals shown in temporal intervals of various length. As suggested by Pantic (2009), this should be treated as one complex problem rather than a number of detached problems in human sensing, context sensing, and human behaviour understanding. Given the current state of the art in automatic analysis of social signals, it may take decades to fully understand and be able to synthesize various combinations of social signals that are appropriate for different contexts and different conversational agents. There are many issues involved and one of those is that it is not self-evident that synthetic agents should behave in the same way as humans do, or that they should exhibit faithful copy of human social behaviours. On the contrary, evidence from the cartoon industry suggests that, in order to be believable, cartoon characters need to show strongly exaggerated behaviour. This suggests further that a trade-off between the degree of naturalness and the type of (exaggerated) gestural and vocal expression may be necessary for modeling believable conversational agents' behavior. All in all, the journey towards artificial social intelligence and socially-aware computing is still long and many aspects of it are yet to be attempted.

Acknowledgements

The research that has led to this work has been supported in part by the European Community's Seventh Framework Programme (FP7/2007-2013), under grant agreement no. 231287 (SSPNet).

References

K. Albrecht (2005). *Social Intelligence: The new science of success*. John Wiley & Sons.

S. Banerjee, A. Rudnicky (2004). Using simple speech based features to detect the state of a meeting and the roles of the meeting participants. In *Proc. Int'l Conf. Spoken Language Processing* (pp. 221-231).

R. Barzilay, M. Collins, J. Hirschberg, S. Whittaker (2000). The rules behind the roles: identifying speaker roles in radio broadcasts. In *Proc. Conf. Artificial Intelligence* (pp. 679–684).

T. Bickmore, R. Picard (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Trans. Computer Human Interaction*, 59(1), 21-30.

J. Biel and D. Gatica-Perez (2012). The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs. *IEEE Trans. Multimedia*, to appear.

J. Bonaiuto, K. R. Thorisson (2008). Towards a neurocognitive model of realtime turntaking in face-to-face dialogue. In G. K. I. Wachsmuth, M. Lenzen (Eds.), *Embodied Communication in Humans And Machines* (pp. 451–484). Oxford University Press, .

A. Bottino, A. Laurentini (2010). The Analysis of Facial Beauty: An Emerging Area of Research in Pattern Analysis. In *Lecture Notes in Computer Science* (Vol. 6111, pp. 425-435).

K. Bousmalis, M. Mehu, M. Pantic (2012). Spotting agreement and disagreement based on noverbal audiovisual cues: A Survey. *Image and Vision Computing Journal*.

K. Bousmalis, S. Zafeiriou, L.P. Morencey, M. Pantic (2013). Infinite Hidden Conditional Random Fields for human behavior analysis. *IEEE Trans. Neural Networks and Learning Systems*, 24(1), 170-177.

H. de Bruijn and P.M. Herder (2009). System and actor perspectives on sociotechnical systems. *IEEE Trans. Systems, Man and Cybernetics*, 39(5), 981-992.

P. Brunet, R. Cowie (2012). Towards a conceptual framework of research on social signal processing. *Journal of Multimodal User Interfaces*, 6(3–4), 101–115.

P.M. Brunet, R. Cowie, D. Heylen, A. Nijholt, M. Schroeder (2012). Conceptual frameworks for multimodal social signal processing. *Journal of Multimodal User Interfaces*.

M. Buchanan (2007). The science of subtle signals. *Strategy+Business*, 48, 68-77.

M. de Jong, M. Theune, D. Hofs (2008). Politeness and alignment in dialogues with a virtual guide. In *Proc. Int'l Conf. Autonomous Agents and Multiagent Systems* (pp. 207–214).

E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, D. Cohen (2012). Interpersonal Synchrony: A Survey Of Evaluation Methods Across Disciplines. *IEEE Trans. Affective Computing*, 3(3),349-365.

F. De La Torre, J.F. Cohn. Facial expression analysis. In T.B. Moeslund, A. Hilton, V. Kruger, L. Sigal (Eds.), *Visual Analysis of Humans* pp. 377-409). Springer Verlag.

C. de Melo, J. Gratch (2009). Expression of Emotions using Wrinkles, Blushing, Sweating and Tears. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 188-200).

W. Dong, B. Lepri, A. Cappelletti, A. Pentland, F. Pianesi, M. Zancanaro (2007). Using the influence model to recognize functional roles in meetings. In *Proc. Int'l Conf. Multimodal Interfaces* (pp. 271–278).

R.Dunbar (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 20, 469-493.

N. Eagle and A. Pentland (2005). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, (10)4, 255–268.

M. Fishbein and I. Ajzen (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley.

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

V.Gallese (2006). Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism. *Brain Research*, 1079(1), 15-24.

D.T. Gilbert, S.T. Fiske, and G. Lindzey (Eds.). (1998). *Handbook of Social Psychology*. McGraw-Hill.

M. Gladwell (2005). *Blink: The Power of Thinking without Thinking*. Little Brown & Company.

H. Gunes, M. Pantic (2010). Automatic, Dimensional and Continuous Emotion Recognition (A Survey). *Int'l Journal of Synthetic Emotions*, 1(1), 68-99.

H. Gunes, M. Piccardi (2006). Assessing facial beauty through proportion analysis by image processing and supervised learning. *Int'l Journal of Human-Computer Studies*, 64, 1184-1199.

D.B. Jayagopi, H. Hung, C. Yeo, D. Gatica-Perez (2009). Modeling dominance in group conversations using nonverbal activity cues. *IEEE Trans. Audio, Speech, and Language Processing*, 17(3), 501-513.

J. Grudin, S. Poltrock (2012). Taxonomy and Theory in Computer Supported Cooperative Work. In S.W.J. Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology*. Oxford University Press.

S. Gupta, M.A. Walker, D.M. Romano (2007). Generating politeness in task based interaction: An evaluation of the effect of linguistic form and culture. In *Proc. European Workshop on Natural Language Generation* (pp. 57-64).

D. Jan and D. R. Traum (2007). Dynamic movement and positioning of embodied agents in multiparty conversations. In *Proc. Int'l Joint Conf. Autonomous Agents and Multiagent Systems*.

C. Judd, L. James-Hawkins, V. Yzerbyt, and Y. Kashima (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6), 899–913.

H.H. Kelley and J. Thibaut (1978). *Interpersonal relations: A theory of interdependence*. Wiley.

S.Kim, M.Filippone, F.Valente A.Vinciarelli (2012). Predicting the Conflict Level in Television Political Debates: an Approach Based on Crowdsourcing, Nonverbal Communication and Gaussian Processes. In *Proc. ACM Int'l Conf. Multimedia* (pp. 793-796).

S. Kopp, T. Stocksmeier, D. Gibbon (2007). Incremental multimodal feedback for conversational agents. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 139-146).

K. Laskowski, M. Ostendorf, T. Schultz (2008). Modeling vocal interaction for text-independent participant characterization in multi-party conversation. In *Proc. ISCA/ACL SIGdial Workshop on Discourse and Dialogue* (pp. 148-155).

I. Leite, S. Mascarenhas, A. Pereira, C. Martinho, R. Prada, A. Paiva (2010). Why Can't We Be Friends? -- An Empathic Game Companion for Long-Term Interaction. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 315-321).

Y. Liu (2006). Initial study on automatic identification of speaker role in broadcast news speech. In *Proc. Human Language Technology Conf. of the NAACL* (pp. 81–84).

R.M. Maatman, J. Gratch, and S. Marsella (2005). Natural behavior of a listening agent. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 25–36).

F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30, 457-500.

V. Manusov and M.L. Patterson (Eds.). (2006). *The SAGE Handbook of Nonverbal Communication*. Sage Publishers.

S. Marsella, J. Gratch, P. Petta (2010). Computational Models of Emotions. In K.R. Scherer, T. Banzinger, E. Roesch (Eds.), *A blueprint for an affectively competent Agent*. Oxford University Press.

T.J. Mayne and G.A. Bonanno (2001). *Emotions: Current Issues and Future Directions*. Guildford Press.

M. Mehu and K. Scherer (2012). A psycho-ethological approach to social signal processing. *Cognitive Processing*, 13(2), 397–414.

M. Mehu, F. D'Errico and D.Heylen (2012). Conceptual analysis of social signals: the importance of clarifying terminology. *Journal on Multimodal User Interfaces*, 6(3-4), 179-189.

L.K. Miles, L.K. Nind, and C.N. Macrae (2009). The rhythm of rapport: Interpersonal synchrony and social perception. *J. Experimental Social Psychology*, 45, 585-589.

L. Morency, I. de Kok, and J. Gratch (2008). Predicting listener backchannels: A probabilistic multimodal approach. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 176–190).

C. Nass, J. Steuer and E.R. Tauber (1994). Computers are social actors. In *Proc. SIGCHI Conf. Human factors in computing systems: celebrating interdependence* (pp. 72-78).

M. Nicolaou, V. Pavlovic, M. Pantic (2012). Dynamic Probabilistic CCA for analysis of affective behaviour. In *Proc. European Conf. Computer Vision*.

R. Niewiadomski, M. Ochs, C. Pelachaud (2008). Expressions of empathy in ECAs. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 37-44).

M. Ochs, R. Niewiadomski, C. Pelachaud (2010). How a virtual agent should smile? Morphological and dynamic characteristics of virtual agent's smiles. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 427-440).

D. Olguin Olguin, P.A. Gloor, A. Pentland (2009). Capturing individual and group behavior with wearable sensors. In *Proc. AAAI Spring Symposium*.

M. Pantic (2009). Machine analysis of facial Behaviour: Naturalistic and dynamic behaviour. *Phyl. Trans. Royal Society B*, 364, 3505-3513.

M. Pantic, R. Cowie, F. D'Errico, D, Heylen, M. Mehu, C. Pelachaud, I. Poggi, M. Schroeder, and A. Vinciarelli (2011). Social Signal Processing: The Research Agenda. In T.B. Moeslund, A. Hilton, V. Kruger and L. Sigal (Eds.), *Visual Analysis of Humans* (pp. 511-538). Springer Verlag.

C. Pedica, H.H. Vilhjalmsson (2009). Spontaneous avatar behavior for human territoriality. In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 344–357).

A. Pentland (2005). Socially aware computation and communication. *IEEE Computer*, 38(3), 33-40.

A. Pentland (2007). Social Signal Processing. *IEEE Signal Processing Magazine*, 24(4), 108–111.

F. Pianesi, N. Mana, and A. Cappelletti (2008). Multimodal recognition of personality traits in social interactions. In *Proc. Int'l Conf. Multimodal Interfaces* (pp. 53-60).

F.Pianesi. Searching for personality (2013). *IEEE Signal Processing Magazine*, to appear.

I. Poggi, F. D'Errico (2012). Social Signals: a framework in terms of goals and beliefs. *Cognitive Processing*, 13(2), 427–445.

I. Poggi, F. D'Errico, A.Vinciarelli (2012). Social Signals: from theory to application. *Cognitive Processing*, 13(2), 189-196.

T. Polzehl, S. Moller, and F. Metze (2010). Automatically assessing personality from speech. In *Proc. IEEE Int'l Conf. Semantic Computing* (pp. 134–140).

K. Porayska-Pomsta and C. Mellish (2004). Modelling politeness in natural language generation. In *Proc. Int'l Conf. Natural Language Generation, LNAI 3123* (pp. 141–150).

K. Prepin, A. Revel (2007). Human-machine interaction as a model of machine-machine interaction: how to make machines interact as humans do. *Advanced Robotics*, 21(15), 1709–1723.

M. Raento, A. Oulasvirta, N. Eagle (2009). Smartphones: An Emerging Tool for Social Scientists. *Sociological Methods & Research*, 37(3), 426–454.

B. Robins, K. Dautenhahn, E. Ferrari, G. Kronreif, B. Prazak-Aram, P. Marti, I. Iacono, G.J. Gelderblom, T. Bernd, F. Caprino, E. Laudanna (2012). Scenarios of robot-assisted play for children with cognitive and physical disabilities. *Interaction Studies*, 13(2), 189-234.

Z. Ruttkay, C. Pelachaud (Eds.). (2004). *From brows to trust: Evaluating Embodied Conversational Agents*. Kluwer Academic Publishing.

H. Sacks, E. Schegloff, G. Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 696-735.

H.Salamin, A.Vinciarelli (2012). Automatic Role Recognition in Multiparty Conversations: an Approach Based on Turn Organization, Prosody and Conditional Random Fields. *IEEE Trans. Multimedia*, 14(2), 338-345.

P. Salvagnini, H. Salamin, M. Cristani, A. Vinciarelli, V. Murino (2012). Learning to teach from Videolectures: Predicting lecture ratings based on lecturer's nonverbal behaviour. In *Proc. IEEE Int'l Conf. Cognitive InfoCommunications* (pp. 415-419).

K. Schmid, D. Marx, A. Samal (2008). Computation of face attractiveness index based on neoclassic canons, symmetry and golden ratio. *Pattern Recognition*, 41, 2710-2717.

M. Schroeder (2007). Interpolating expressions in unit selection. In *Proc. Int'l Conf. Affective Computing and Intelligent Interaction* (pp. 718-720).

M. Schroeder. Expressive speech synthesis: Past, present, and possible futures (2009). In J. Tao and T. Tan (Eds.), *Affective Information Processing* (pp. 111-126). Springer Verlag.

B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Muller, S. Narayanan (2013). Paralinguistics in speech and language – State of the art and the challenge. *Computer Speech and Language*, 27, 4-39.

D. Sutic, I. Breskovic, R. Huic, I. Jukic (2010). Automatic evaluation of facial attractiveness. In *Proc. Int'l MIRO Convention* (pp. 1339-1342).

M. ter Maat, D. Heylen (2009). Turn Management or Impressions Management? In *Proc. Int'l Conf. Intelligent Virtual Agents* (pp. 467-473).

F.Valente and A.Vinciarelli (2011). Language-Independent Socio-Emotional Role Recognition in the AMI Meetings Corpus. In *Proc. Interspeech* (pp. 3077-3080).

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

A.Vinciarelli (2009). Capturing Order in Social Interactions. *IEEE Signal Processing Magazine*, 26(5), 133-137.

A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, M. Schroeder (2012). Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Trans. Affective Computing*, 3(1), 69–87.

J. Whithehill, J. Movellan (2008). Personalized facial attractiveness prediction. In *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*.

G. Zen, B. Lepri, E. Ricci, O. Lanz (2010). Space speaks: towards socially and personality aware visual surveillance. In *Proc. ACM Int'l Workshop on Multimodal Pervasive Video Analysis* (pp. 37–42).

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

D. Zhang, Q. Zhao, F. Chen (2011). Quantitative analysis of human facial beauty using geometric features. *Pattern Recognition*, 44(4), 940-950.

E. Zovato, A. Pacchiotti, S. Quazza, and S. Sandri (2004). Towards emotional speech synthesis: A rule based approach. In *Proc. ISCA Speech Synthesis Workshop* (pp. 219–220).