New Social Signals in a New Interaction World: The Next Frontier for Social Signal Processing

Alessandro Vinciarelli *Member, IEEE*, and Alex Pentland *Member, IEEE*

Abstract—Social Signal Processing is the domain aimed at modeling, analysis and synthesis of social behaviour, in particular when it comes to nonverbal aspects. So far, the field has focused on face-to-face interactions where it is possible to use the whole range of nonverbal cues that people have at disposition to communicate (gestures, facial expressions, vocalizations, etc.). However, increasingly more interactions take place through communication technologies that limit the use of nonverbal cues (e.g., phones do not allow one to display facial expressions) or require the adoption of artificial cues that do not belong to the natural repertoire of human beings (e.g., "likes" on social media). This opens a new frontier for Social Signal Processing where the main questions are whether people still exchange social signals and, if yes, how.

Index Terms—Social Signal Processing; Nonverbal Behaviour; Social Physics; Technology Mediated Communication; Reality Mining.

1 Introduction

There is no social interaction without *social signals*, i.e. without observable cues allowing people to understand and reasonably predict the behaviour of others. The literature defines social signals in different ways: "*acts or structures that influence the behavior or internal state of other individuals*" [1], "*actions whose function is to bring about some reaction or to engage in some process*" [2], "*communicative or informative signals which [...] provide information about social facts*" [3], etc. While emphasizing different aspects of social interaction, all definitions agree on one point, namely that social signals are observable behaviours that produce, intentionally or not, tangible changes in others, whether this means to modify their inner state (e.g., to stimulate the emotions they experience), to modify their observable behaviour (e.g., to make them laugh in response to a joke) or to change their beliefs about the social setting (e.g., to make them aware of conflict or disagreement).

Social Signal Processing (SSP) is the computing domain aimed at modeling, analysis and synthesis of social signals in human-human and human-machine interactions [4], [5], [6]. Modeling refers to the study of principles and laws that govern the use of social signals; Analysis is the development of automatic approaches capable to detect and interpret social signals (e.g., to understand whether a smile communicates sympathy or sarcasm); Synthesis is the generation of artificial social signals that produce the same effects as those displayed by humans (e.g., to synthesize voices that sound extravert or competent). Overall, the key-idea of SSP - inspired by several decades of research in social psychology - is that social signals are the physical, machine detectable traces of psychological and social phenomena (e.g., conflict, empathy, interest, emotions, etc.). These cannot be observed and sensed directly, but only inferred from the behaviour of others.

SSP is still a young domain - the expression "*Social Signal Processing*" was coined less than ten years ago [4] - and so far it has focused mainly on one particular type of social signals, i.e. the nonverbal behavioural cues (facial expressions, gestures, vocalizations, etc.) that people exchange in face-to-face, co-located interactions. This is not surprising given that such a setting is the "*primary site of human sociality*" [7] and the very ability of exchanging social signals is the result of an evolutionary process during which face-to-face interactions were the only possible ones [8]. Furthermore, nonverbal cues tend to be displayed outside conscious awareness and, therefore, tend to be *honest*, i.e. to leak reliable information independently of the intention to do so [9].

SSP extends the scope of other computing domains revolving around human behaviour like, e.g., Affective Computing (focusing mainly on emotions), Action Recognition (focusing on what people do rather than on social and psychological motivations underlying actions), Automatic Speech Recognition and Natural Language Processing (focusing on verbal rather than nonverbal aspects of communication), etc. In other words, SSP fills a gap left open by other machine intelligence domains dealing with people involved in co-located social interactions.

Still, one of the most notable phenomena of the last decade is that mobile technologies and social networking platforms have multiplied the chances for social interaction to a large extent. Nowadays, people can interact with virtually anybody at virtually every moment [10]. As a result, social interactions take place increasingly more frequently through technology and not necessarily

- *A.Vinciarelli is with the University of Glasgow (School of Computing Science and Institute of Neuroscience and Psychology). E-mail: vincia@dcs.gla.ac.uk*
- *Alex Pentland is with Massachussets Institute of Technology (Media Lab). E-mail: pentland@mit.edu*

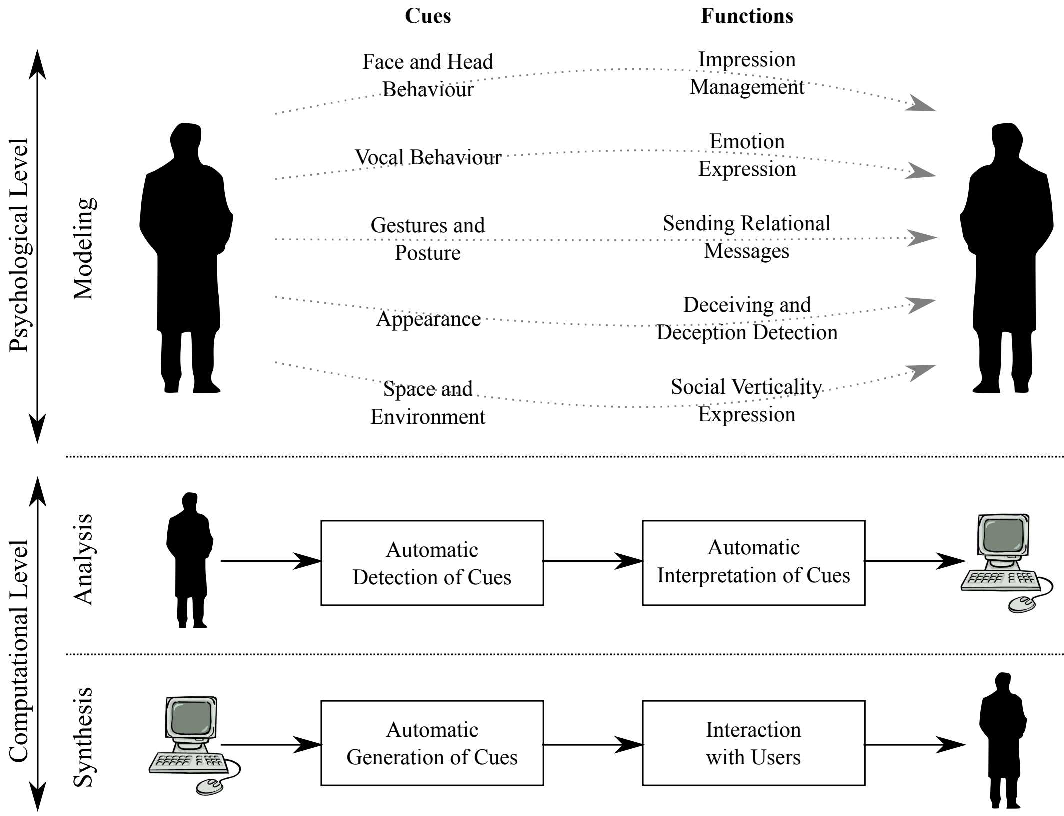

Fig. 1. Social Signal Processing. The upper part of the figure (psychological level) shows the two main elements of face-to-face interactions, namely nonverbal cues and functions that these contribute to perform (the subject of modeling in SSP). The lower part (computational level) shows the main technological components of analysis and synthesis approaches.

face-to-face. According to the latest statistics[1], Internet users worldwide are now three billions (roughly 40% of the population), for a total of over two billions of active social media accounts (29% of the world population).

The initial reaction to the major change described above was an attempt to transfer the social signals typical of face-to-face interactions to the virtual world. This is the case, e.g., of *emoticons*, a popular way to communicate emotional states, and *Second Life*, an interaction platform where people act through avatars reproducing human behaviour. However, the lack of similar functionalities does not appear to be an obstacle towards the success of social media and mobile apps allowing people to interact. Therefore, the question is not only how to transfer face-to-face social signals to the online world, but also what are the social signals in these new, technology mediated settings and how they are exchanged.

From this point of view, there are two main situations. The first is the use of technologies that limit the range of social signals that people can display like, e.g., phones that allow one to use only vocal behaviour. The second is the use of technologies that require people to use social signals fully alternative to those available in face-to-face interactions. This is the case of social networking platforms where people interact extensively, but cannot use any of the social signals that have been studied so far in SSP.

The rest of this article is organized as follows: Section 2 surveys the most important SSP contributions so far, Section 3 outlines early results on the exchange of social signals in new settings, and the Section 4 will draw some conclusions.

2 State-of-the-Art

So far, Social Signal Processing has focused on settings like the one depicted in the upper part of Figure 1: One individual displays observable behaviour - in particular nonverbal cues - to perform all functions necessary to interact with others [11], [12].

The repertoire of available nonverbal cues is wide and, according to a taxonomy commonly applied in the literature [13], it includes five major classes (see column "*Cues*" in Figure 1), namely *face and head behaviour* (facial expressions, gaze contact management, head nods and shakes, etc.), *vocal behaviour* (everything in speech except words, including vocalizations, prosody, pauses, paralinguistics, etc.), *gestures and postures* (self-touching, body orientation, postural congruence, folding arms, etc.), *appearance* (somatotype, clothes, attractiveness, ornaments,

1. http://wearesocial.net/tag/sdmw/

uniforms, etc.), and *space and environment* (mutual distances, territoriality, proxemics, etc.).

Similarly, the number of functions that nonverbal communication fulfills in social interaction is large. However, the psychologists consider that the most common ones are those listed on the right hand side of Figure 1 [13], namely *impression management* (conscious or subconscious processes aimed at influencing others' perception), *emotions expression* (communication about one's inner affective and emotional state), *sending relational messages* (communication of relationship and attitudes towards others), *social verticality* (communication of hierarchical and status differences), *deception management* (ability to deceive and detect others' attempts to deceive), etc.

The setting of Figure 1 has shaped the scope of SSP and, in particular, has determined the three main problems addressed in the field (see [5], [6] for extensive surveys), i.e. modeling, analysis and synthesis.

Modeling (see psychological level in Figure 1): *identification of quantitative relationships between nonverbal cues and functions in terms suitable for computer processing*. The modeling problem concerns mainly human sciences (psychology, anthropology, etc.) and its main goal is to understand how people use social signals (in particular nonverbal cues) to interact with one another. Typical modeling questions are, e.g., what are the nonverbal cues that people use to convey a certain type of impression? Is there a relation between the frequency of a given nonverbal cue and the perception of a certain social phenomenon? Etc.

Analysis (see computational level in Figure 1): *Inference of social and psychological phenomena (i.e., functions being performed in an interaction) from automatically detected nonverbal cues*. The analysis problem concerns mainly computing science and consists of replacing the person on the right hand side of Figure 1 with a machine. The main goal is to build devices capable to sense and understand the social landscape like humans do. Typical analysis questions are, e.g., is it possible to automatically detect a given nonverbal cue? Is it possible to automatically intepret detected cues in terms of the functions mentioned in Figure 1? What are the best sensors to detect the cue related to a given social phenomenon? Etc.

Synthesis (see computational level in Figure 1): *Generation of artificial cues aimed at allowing machines to perform one or more of the functions mentioned in Figure 1*. The synthesis problem concerns mainly computing science and consists of replacing the person on the left-hand side of Figure 1 with a machine capable to exhibit human-like behavior (e.g., social robots or speech synthesizers). The main synthesis goal is to build machines that interact with their users by tapping the same psychological and cognitive processes as humans. Typical synthesis questions are, e.g., is it possible to generate artificial cues capable to perform the functions on the right-hand side of Figure 1? What is the best form of embodiment to perform a given interaction function? Is truly social interaction possible between humans and machines?

So far, the most important SSP efforts have mirrored the taxonomy outlined above (see [5] for an extensive survey) and have targeted facial expressions [14], [15], paralanguage [16], vocalizations (e.g., laughter [17]), bodily movements [18], [19], attractiveness [20], proxemics [21], etc. When it comes to functions, the literature proposes works that address a wide spectrum of social and psychological phenomena, including personality impressions [22], emotions [23], relational messages (e.g., conflict and disagreement [24], mimicry [25], etc.), social verticality (e.g., dominance [26] and roles [27]), etc.

The common point of the approaches mentioned above, and the many others that the literature proposes, is that they operate in settings where interacting people are co-located and can use the full spectrum of cues that people have at disposition by nature (like in the case of nonverbal behaviour) and culture (like in the case of clothes, status symbols, etc.). However, new interaction settings acquire increasingly more importance where the exchange of social signals takes place in different, still largely to be explored forms.

3 Towards New Interaction Settings

Section 2 has shown that the state-of-the-art in SSP revolves around face-to-face interactions, in part because such a setting is of primary importance, in part because it reflects the investigations of social psychologists in the last eight decades. However, nowadays technologies allow and, sometimes, require people to interact in a much wider spectrum of settings. In some cases, this means to lose some of the cues at disposition in co-located interactions (e.g., people talking on the phone can use only speech and vocal cues). In other cases, this means to adopt a set of cues fully alternative to those available in direct interactions (e.g., social media users can *like* and *connect*, but cannot smile or establish eye contact).

To the best of our knowledge, it is still an open question how the signals that people exchange in these new communication settings correspond to the traditional social signals. In other words, it is still unclear whether the model underlying the scheme of Figure 1, where people display observable cues to perform functions necessary to social interaction, can be transferred to new comunication technologies.

The rest of this section tries to address, at least to an initial extent, the question above. In particular, the section addresses three scenarios where the setting is increasingly different from those traditionally considered in SSP and social psychology. The results seem to suggest that people keep behaving according to the model of Figure 1. However, the way this happens includes unexpected effects.

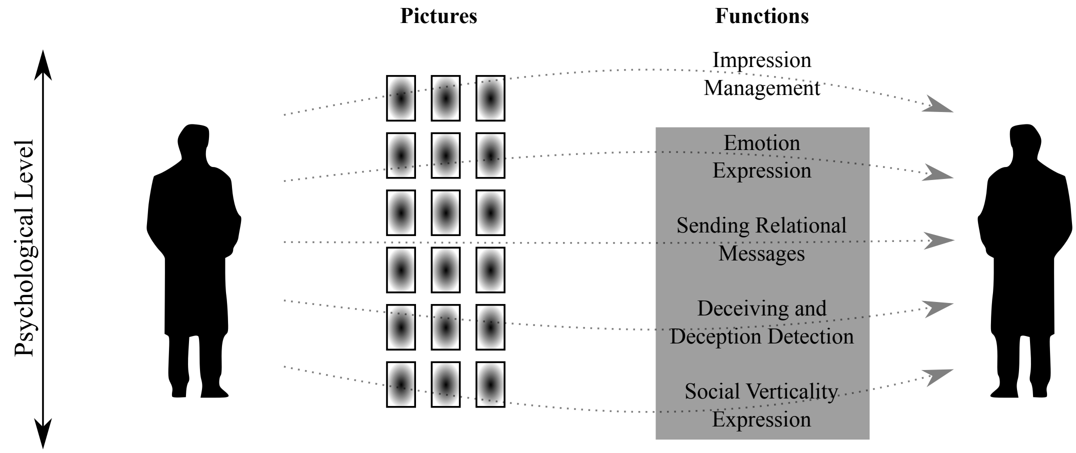

Fig. 2. Pictures as social signals. In the case of social media, pictures might act as social signals, i.e. might allow people to perform the functions necessary to social interaction (at least impression management according to the experiments presented in [28]). The functions that have been investigated only in face-to-face settings are shadowed.

3.1 Less Might Be More

What does it happen when most of the cues available in face-to-face interactions are no longer at disposition? In the case of phone calls, where the only source of social signals is speech, everyday experience suggests that there are no major problems in interacting with others. This is probably why phone calls tend to be considered substantially equivalent to co-located conversations [29]. However, as the time spent in calls increases as a result of the diffusion of mobile phones, evidence shows that these latter do not merely transmit speech signals from one place to the other, but change the way people interact [30].

Negotiation, one of the most common interaction modes of everyday life,provides an interesting example. The experiments presented in [31] analyze negotiations carried over the phone by 60 pairs of unacquainted individuals (120 subjects in total). Experimental protocol and scenario are designed to put the two subjects involved in a negotiation on an even foot. The only, inevitable difference between the participants of a call is that one makes the call while the other receives it (the assignment to one of the two conditions is random). In principle, this should not influence the outcome of the negotiations, but the results show that the subjects receiving a call win 70% of the times while the subjects making it win only in the remaining 30% of the cases ($p < 0.01$ according to a two-tailed binomial test). In other words, the very simple fact of being the receiver - a role assigned randomly in the experiment - literally doubles the chance of winning the negotiation with an unacquainted individual.

Further analysis shows that receivers and callers use social signals differently. In particular, the receivers tend to show significantly more frequently social signals associated to dominance and higher status (e.g., tendency to initiate overlapping speech) while the callers tend to display significantly more frequently cues associated to lower status and submissive attitude (e.g., hesitations and filled pauses like "*ehm*" or "*uhm*") [32], thus explaining the different chances of success in the negotiations.

While it is not a surprise that people manage to accomplish a task as complex as a negotiation over the phone (even if they lack most of the cues available in face-to-face interactions), the effect of the setting turns out to be unexpected. The difference between calling and receiveing results into a difference of social verticality that produces effects in terms of both observable behaviour and outcomes of the interaction.

3.2 Pictures as Social Signals

How does social interaction work when none of the cues available in face-to-face settings can be used? This might sound like an artificial and unrealistic situation, but it is exactly what happens on social media where millions of people interact without displaying any of the social signals they typically use in co-located interactions. Some surrogates are available - emoticons and thumbs-up icons are probably an attempt to transfer facial expressions and gestures to the online world - but most of the interactions involve cues that have no equivalent in ordinary, face-to-face social contacts.

Pictures provide a possible example. According to the surveys of the Pew Research Center, 46% of the American Internet users post original pictures or share online images posted by others [33]. Furthermore, one of the main motivations behind the use of photo-sharing platforms is to maintain contact with others [34]. In this respect, pictures play the role of an online "*social currency*" [33]. However, it is still an open question whether the pictures work as social signals according to the definitions provided at the beginning of this article [1], [2], [3]. In other words, it is still an open question whether the pictures work like nonverbal cues in face-to-face encounters and fulfill at least some of the functions illustrated in Figure 1.

The experiments presented in [28] address the question above and the results suggest that pictures play a role in impression management among the users of Flickr, one of the most popular photo-sharing platforms of the web. In the work, a pool of raters have been asked to assess the Big-Five personality traits of 300 Flickr users based on the images these latter tag as "*favourite*", i.e. as particularly appealing from an aesthetic and/or affective point of view. The setting reflects the situation depicted in Figure 2, where Flickr users interact with others through the photo-sharing platform and, therefore, the pictures are the only social signals they have at disposition.

The results show that the personality scores attributed to a given user can be inferred with significant accuracy - correlation up to 0.6 between actual and predicted scores - from low-level, content independent features extracted automatically from the favourite images (e.g., statistics about the distribution of colours or textural properties). Therefore, favourite pictures convey an impression and the statistical relationship between low-level, visual characteristics and attributed personality traits is solid enough to allow one to perform automatic prediction. In other words, visual characteristics of favourite pictures co-variate with personality traits (e.g., people that tend to use brighter colours tend to be perceived as more extravert than the others [28]) in the same way as nonverbal cues do in face-to-face interactions (e.g., people that speak louder tend to be perceived as more extravert [35]).

3.3 Social Physics

Are there signals generated by groups of people as well as individuals? During the last decade the SSP research community has developed the ability to analyze the signals of entire social organisms - groups, companies, and whole communities - on a millisecond-by-millisecond basis for up to years at a time, an implementation of what is known as a *living lab* [36], [37]. The method is simple: Measurements are made by collecting digital breadcrumbs such as the sensors from cell phones, postings on social media, purchases with credit cards, and more. By combining these fine-grain, objective measurements with traditional surveys and other social science tools, it is possible to provide both richer context and greater detail to the traditional measurement techniques. In many cases this allows the accurate prediction of effect sizes even in complex, natural circumstances. It is also possible to estimate subjective biases, and compare socially-constructed reality to objective reality. In other words, we can track how even the most apparently insignificant signals help shape social phenomena at the scale of communities as large as entire cities.

Using this methodology it was possible to find, for instance, that there are behavioral markers that signal the impending onset of flu [38], a fact that is now being used commercially by `http://ginger.io`. It was also found that certain aggregate changes in behavior, for instance, elders abandoning a favorite town square, predict future high crime rates [39]. And it was discovered that changes in whom people associate with accurately predict changes in voting behavior [40]. But what is perhaps just as remarkable is that these studies take only days from conception to launch, and their cost can be limited to the cost of encouraging people to download the appropriate mobile phone app.

An example of such a living lab is the open data city launched last year with the city of Trento in Italy, by the MIT Human Dynamics Lab along with Telecom Italia, Telefónica, the research university Fondazione Bruno Kessler, the Institute for Data Driven Design, and local companies. This living lab has the approval and informed consent of all of its participants - they know that they are part of a large experiment whose goal is to better understand human behavior and to use that knowledge to invent a better way of living. More detail on these living labs can be found at `http://realitycommons.media.mit.edu` and `http://www.mobileterritoriallab.eu/`.

4 Conclusions

The core-idea of SSP is that social signals are the physical, machine detectable evidence of social and psychological phenomena that mediate the interactions of different people. The technological implication of such a statement is that social signals can be, on the one hand, automatically detected and interpreted to understand what goes on between interacting people and, on the other hand, automatically generated to build machines that interact with people like people do. On the long-term, this is expected to bring socially intelligent machines, i.e. machines that sense and understand social interactions in the same way as people do in everyday life while seamlessly interacting with their users.

This article has shown that the same principle applies not only to face-to-face interactions, for which it has helped to develop a large number of socially intelligent technologies [5], [6], but also to new human-human communication settings resulting from the widespread diffusion of technologies like mobile phones and social networking platforms. The main difference is that the social signals at disposition can be limited (like in the case of the phones) or fully alternative (like in the case of social networking platforms) with respect to those available by nature in face-to-face social exchanges (see Figure 1).

The effect known as "*Media Equation*" [8] shows that people are sensitive to social signals not only when they meet others in person, but also when they see them, e.g., in videos or other media. This is why social interactions are possible through the phone, when people hear others only through the speakers, even if unexpected effects are still possible (see Section 3.1). However, the results illustrated in this article go one step further and show that digital traces can work as social signals even when they do not correspond to any of the cues that people display in co-located interactions. This is the case of the pictures that convey personality impressions (see Section 3.2) or of the digital breadcrumbs that people leak in everyday life (see Section 3.3).

It is maybe not a surprise that people interpret most of the signals they receive as social [41]. However, the way this happens in nowadays technological landscape is still largely to be explored. This appears to be a likely new frontier for Social Signal Processing, the natural continuation of the extensive work done in the last decade to model, analyze and synthesize social signals in face-to-face settings.

REFERENCES

[1] M. Mehu and K. Scherer, "A psycho-ethological approach to Social Signal Processing," *Cognitive Processing*, vol. 13, no. 2, pp. 397–414, 2012.

[2] P. Brunet and R. Cowie, "Towards a conceptual framework of research on Social Signal Processing," *Journal of Multimodal User Interfaces*, vol. 6, no. 3-4, pp. 101–115, 2012.

[3] I. Poggi and F. D'Errico, "Social Signals: a framework in terms of goals and beliefs," *Cognitive Processing*, vol. 13, no. 2, pp. 427–445, 2012.

[4] A. Pentland, "Social Signal Processing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 108–111, 2007.

[5] 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*, vol. 3, no. 1, pp. 69–87, 2012.

[6] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social Signal Processing: Survey of an emerging domain," *Image and Vision Computing Journal*, vol. 27, no. 12, pp. 1743–1759, 2009.

[7] E. Schegloff, "Analyzing single episodes of interaction: An exercise in conversation analysis," *Social Psychology Quarterly*, vol. 50, no. 2, pp. 101–114, 1987.

[8] B. Reeves and C. Nass, *How people treat computers, television, and new media like real people and places*. Cambridge University Press, 1996.

[9] A. Pentland, *Honest signals*. MIT Press, 2008.

[10] R. Harper, *Texture: Human expression in the age of communications overload*. MIT Press, 2010.

[11] V. Richmond, J. McCroskey, and S. Payne, *Nonverbal behavior in interpersonal relations*. Prentice Hall, 1991.

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

[13] M. Hecht, J. De Vito, and L. Guerrero, "Perspectives on nonverbal communication: Codes, functions, and contexts," in *The nonverbal communication reader*, L. Guerrero, J. De Vito, and M. Hecht, Eds. Waveland Press, 1999, pp. 3–18.

[14] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[15] Z. Deng and J. Noh, "Computer facial animation: A survey," in *Data-Driven 3D Facial Animation*, Z. Deng and U. Neumann, Eds. Springer Verlag, 2008, pp. 1–28.

[16] B. Schuller and A. Batliner, *Computational paralinguistics: emotion, affect and personality in speech and language processing*. John Wiley & Sons, 2013.

[17] R. Niewiadomski, J. Hofmann, J. Urbain, T. Platt, J. Wagner, B. Piot, H. Cakmak, S. Pammi, T. Baur, S. Dupont, M. Geist, F. Lingenfelser, G. McKeown, O. Pietquin, and W. Ruch, "Laugh-aware virtual agent and its impact on user amusement," in *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems*, 2013, pp. 619–626.

[18] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, 2010.

[19] B. Hartmann, M. Mancini, and C. Pelachaud, "Implementing expressive gesture synthesis for embodied conversational agents," in *Gesture in Human-Computer Interaction and Simulation*, S. Gibet, N. Courty, and J.-F. Kamp, Eds. Springer, 2006, pp. 188–199.

[20] H. Gunes, "A survey of perception and computation of human beauty," in *Proceedings of the joint ACM Workshop on Human Gesture and Behavior Understanding*, 2011, pp. 19–24.

[21] M. Cristani, G. Paggetti, A. Vinciarelli, L. Bazzani, G. Menegaz, and V. Murino, "Towards computational proxemics: Inferring social relations from interpersonal distances," in *Proceedings of IEEE International Conference on Social Computing*, 2011, pp. 290–297.

[22] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[23] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[24] K. Bousmalis, M. Mehu, and M. Pantic, "Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools," *Image and Vision Computing*, vol. 31, no. 2, pp. 203–221, 2013.

[25] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 349–365, 2012.

[26] D. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez, "Modeling dominance in group conversations using nonverbal activity cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 3, pp. 501–513, 2009.

[27] H. Salamin and A. Vinciarelli, "Automatic role recognition in multiparty conversations: An approach based on turn organization, prosody, and conditional random fields," *IEEE Transactions on Multimedia*, vol. 14, no. 2, pp. 338–345, 2012.

[28] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis," in *Proceedings of the ACM International Conference on Multimedia*, 2013, pp. 213–222.

[29] I. Arminen and A. Weilenmann, "Mobile presence and intimacy - reshaping social actions in mobile contextual configuration," *Journal of Pragmatics*, vol. 41, no. 10, pp. 1905–1923, 2009.

[30] R. Ling, *New tech, new ties*. MIT Press, 2008.

[31] A. Vinciarelli, H. Salamin, and A. Polychroniou, "Negotiating over mobile phones: calling or being called can make the difference," *Cognitive Computation*, vol. 6, no. 4, pp. 677–688, 2014.

[32] A. Vinciarelli, P. Chatziioannou, and A. Esposito, "When the words are not everything: the use of laughter, fillers, back-channel, silence and overlapping speech in phone calls," *Frontiers in ICT*, vol. 2, no. 4, 2015.

[33] L. Rainie, J. Brenner, and K. Purcell, "Photos and videos as social currency online," Pew Research Center, Tech. Rep., 2012.

[34] N. Van House, "Personal photography, digital technologies and the uses of the visual," *Visual Studies*, vol. 26, no. 2, pp. 125–134, 2011.

[35] K. Scherer, "Personality inference from voice quality: The loud voice of extroversion," *European Journal of Social Psychology*, vol. 8, no. 4, pp. 467–487, 1978.

[36] A. Pentland, *Social Physics: How Good Ideas Spread - The Lessons from a New Science*. Penguin, 2014.

[37] N. Eagle and K. Greene, *Reality mining: Using big data to engineer a better world*. MIT Press, 2014.

[38] A. Madan, M. Cebrian, D. Lazer, and A. Pentland, "Social sensing for epidemiological behavior change," in *Proceedings of the ACM International Conference on Ubiquitous Computing*, 2010, pp. 291–300.

[39] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pentland, "Once upon a crime: Towards crime prediction from demographics and mobile data," in *Proceedings of the ACM International Conference on Multimodal Interaction*, 2014, pp. 427–434.

[40] A. Madan, K. Farrahi, D. Gatica-Perez, and A. Pentland, "Pervasive sensing to model political opinions in face-to-face networks," in *Pervasive Computing*. Springer, 2011, pp. 214–231.

[41] C. Nass and S. Brave, *Wired for speech: How voice activates and advances the human-computer relationship*. MIT Press, 2005.